

	Data de Depósito:
	Assinatura:
Técnicas de classifica	ação hierárquica multirrótulo1
R	Ricardo Cerri
Orientador: Prof. Dr. André	Carlos Ponce de Leon Ferreira de Carvalho
	rtação apresentada ao Instituto de Ciências Mate-
	as e de Computação - ICMC-USP, como parte dos sitos necessários à obtenção do título de Mestre em
Ciênc	ias da Computação e Matemática Computacional.
	P - São Carlos aneiro/2010
	,

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

¹Trabalho Realizado com Auxílio da FAPESP

Esta dissertação foi preparada com o formatador de textos LATEX. A bibliografia é gerada automaticamente pelo BibTeX, utilizando o estilo <i>apalike</i> com modificações para o português. O estilo utilizado no documento e as modificações no estilo <i>apalike</i> foram desenvolvidas por Ronaldo Cristiano Prati.
©Copyright 2010 - Ricardo Cerri Todos os direitos reservados

Agradecimentos

Primeiramente agradeço a Deus, que me deu saúde para realizar minhas atividades.

Agradeço a meus pais, Irineu e Cleide, que sempre me apoiaram e foram meus grandes amigos durante todo o período de meu mestrado.

Ao meu orientador Prof. Dr. André de Carvalho, pela constante presença e orientação durante toda a pesquisa, e também por sua confiança. Sua ajuda foi muito importante para a realização deste trabalho.

Ao meu amigo Eduardo Costa, que sempre me ajudou desde o período de graduação, trocando idéias e informações. Sua participação foi muito importante, principalmente no período inicial deste mestrado.

Ao meu irmão Rodrigo e a toda a minha família, que sempre torceu por mim em todos os momentos.

À minha namorada Talita, pela paciência e por muitas vezes compreender minha ausência, por me ajudar nas correções do texto, e por seu companheirismo que me faz tão bem.

Aos meus amigos do Laboratório de Computação Bioinspirada do ICMC/USP: Thiago, Danilo, Jonathan, Débora, Pablo, Márcio, Murilo, Luiz, André, Lucas e todos os outros. Todos eles sempre estiveram dispostos a trocar idéias sobre a pesquisa. Também os agradeço pelos vários momentos agradáveis que passamos juntos.

A todos os amigos que fiz no ICMC/USP: Daniel, Francisco, Renato, Paulo e todos os outros amigos que, com certeza, tornam nossa vida muito melhor.

Aos meus amigos de graduação da UNESP de Rio Claro, Márcio, Felipe, Kleber, Fernando, Luiz, Tiago, Eduardo, Roberto, Rafaela e muitos outros cuja amizade continua sendo muito importante.

Aos amigos da Companhia de Dança de Salão Júnior & Ellen: Vinícius, Rodrigo, Júnior, Ellen, Carol, Milene, Anne, Luiza, Elisandra, Niuro, Cainá, Natassia e Deco. Com eles passei muitos momentos agradáveis, com muita dança e amizade.

Aos amigos do Ministério de Música Dom de Deus, da Paróquia Nossa Senhora Aparecida de Rio Claro. Agradeço a companhia, as risadas e a música que me faz muito bem.

A todos os amigos do Madrigal Cor da Voz de Rio Claro. Com certeza o convívio com o canto tornou meus dias muito mais alegres.

A todos os professores e funcionários do ICMC/USP - São Carlos, que contribuíram diretamente ou indiretamente para a realização deste trabalho.

Aos amigos com quem morei durante esse período, Denis, Vinícius, Matheus, João Paulo e Thiago, pela amizade nesses dois anos.

À Amanda Clare, pelo código do algoritmo C4.5H, utilizado nos experimentos.

À Fapesp e ao CNPq, pelo auxílio financeiro para a realização desta pesquisa.

Resumo

Muitos dos problemas de classificação descritos na literatura de Aprendizado de Máquina e Mineração de Dados dizem respeito à classificação de dados em que cada exemplo a ser classificado pertence a um conjunto finito, e geralmente pequeno, de classes que estão em um mesmo nível. Vários problemas de classificação, entretanto, são de natureza hierárquica, em que classes podem ser subclasses ou superclasses de outras classes. Em muitos problemas hierárquicos, principalmente no campo da Bioinformática, um ou mais exemplos podem ser associados a mais de uma classe simultaneamente. Esses problemas são conhecidos como problemas de classificação hierárquica multirrótulo. Nesta pesquisa, foram investigadas diferentes técnicas para lidar com esses tipos de problemas. Essas técnicas são baseadas em duas abordagens: local ou Top-Down e global ou One-Shot. Três técnicas descritas na literatura foram utilizadas. A primeira delas, chamada HMC-BR, é baseada na abordagem Top-Down, e utiliza uma estratégia de classificação binária chamada Um-Contra-Todos. As outras duas técnicas, baseadas na abordagem One-Shot, são chamadas C4.5H (uma extensão do algoritmo de indução de árvores de decisão C4.5), e de Clus-HMC (baseada na noção de *Predictive Clustering Trees*, em que árvores de decisão são estruturadas como uma hierarquia de grupos (clusters)). Além das técnicas descritas na literatura, duas novas técnicas foram propostas e implementadas nesta pesquisa, chamadas de HMC-LP e HMC-CT. Essas técnicas são variações hierárquicas de técnicas de classificação multirrótulo não hierárquicas. A técnica HMC-LP utiliza uma estratégia de combinação de classes e a técnica HMC-CT utiliza uma estratégia de decomposição de classes. Para a avaliação das técnicas, foram utilizadas medidas específicas para esse tipo de classificação. Os resultados experimentais mostraram que as técnicas propostas obtiveram desempenhos superiores ou semelhantes aos das técnicas descritas na literatura, dependendo da medida de avaliação utilizada e das características dos conjuntos de dados.

Abstract

Many of the classification problems described in the literature of Machine Learning and Data Mining are related to data classification where each example to be classified belongs to a finite, and usually small, set of classes located at the same level. There are many classification problems, however, that are of hierarchical nature, where classes can be subclasses or superclasses of other classes. In many hierarchical problems, mainly in the Bioinformatics field, one or more examples can be associated to more than one class simultaneously. These problems are known as hierarchical multilabel classification problems. In this research, different techniques to deal with these kinds of problems were investigated, based on two approaches, named local or Top-Down and global or One-Shot. Three techniques described in the literature were used. The first one, named HMC-BR, is based on the Top-Down approach, and uses a binary classification strategy named One-Against-All. The other two techniques, based on the One-Shot approach, are named C4.5H (an extension of the decision tree induction algorithm C4.5), and Clus-HMC (based on the notion of Predictive Clustering Trees, where decision trees are structured as a hierarchy of clusters). In addition to the techniques described in the literature, two new techniques were proposed, named HMC-LP and HMC-CT. These techniques are hierarchical variations of non-hierarchical multilabel classification techniques. The HMC-LP technique uses a label combination strategy and the HMC-CT technique uses a label decomposition strategy. The evaluation of the techniques was performed using specific metrics for this kind of classification. The experimental results showed that the proposed techniques achieved better or similar performances than the techniques described in the literature, depending on the evaluation metric used and on the characteristics of the datasets.

Sumário

Li	sta d	le Abreviaturas	xix
Li	sta d	le Figuras	xxi
Li	sta d	le Tabelas	xxiii
Li	sta d	le Algoritmos	xxxi
	List	a de Publicações	. xxxiii
1	Intr	rodução	1
	1.1	Objetivos	. 2
	1.2	Motivação	. 3
	1.3	Técnicas Investigadas	. 4
	1.4	Resumo dos Resultados	. 6
	1.5	Organização do Documento	. 6
2	Cor	nceitos Fundamentais de Classificação e Classificação Hierárquica	7
	2.1	Conceitos Fundamentais de Classificação	. 7
		2.1.1 Paradigmas de Classificação de Dados	. 9
		2.1.2 Avaliação de Classificadores	. 17
	2.2	Conceitos Fundamentais de Classificação Hierárquica	. 21
		2.2.1 Abordagens Utilizadas em Problemas Hierárquicos	. 22
		2.2.2 Avaliando Modelos de Classificação Hierárquica	. 25
	2.3	Considerações Finais	. 26
3	Cla	ssificação Multirrótulo	29
	3.1	Abordagens para Tratar Problemas Multirrótulo	. 30
		3.1.1 Abordagem Independente de Algoritmo	. 31
		3.1.2 Abordagem Dependente de Algoritmo	. 37
	3.2	Densidade e Cardinalidade de Rótulo	. 39
	3.3	Medidas de Avaliação	. 40

	3.4	Consid	derações Finais	41			
4	Cla	ssificaç	ção Hierárquica Multirrótulo	43			
	4.1	Conce	eitos Fundamentais	43			
	4.2	Técnie	cas Investigadas na Literatura	45			
	4.3	Medid	las de Avaliação	52			
		4.3.1	Medidas Baseadas nas Relações de Ancestralidade e Descendência .	52			
		4.3.2	Medidas Baseadas em Distância	55			
		4.3.3	Medidas Baseadas em Similaridade	58			
		4.3.4	Outras Medidas	60			
	4.4	Consid	derações Finais	62			
5	Téc	nicas 1	Propostas, Investigadas e Comparadas	63			
	5.1	Utiliza	ação de PCT - Clus-HMC	64			
	5.2	Exten	são do Algoritmo C4.5 - C4.5H	68			
		5.2.1	Leitura e Armazenamento da Hierarquia de Classes	69			
		5.2.2	Testes para Verificar a Pertinência de uma Dada Classe	70			
		5.2.3	Encontrar a Melhor Classe ou Classes para Representar um Nó da				
			AD	70			
		5.2.4	Cálculo da Entropia	70			
	5.3	Classi	ficação Binária - HMC-BR	73			
	5.4	Comb	inação de Classes - HMC-LP	75			
	5.5	Decon	mposição de Classes - HMC-CT	76			
	5.6	Considerações Finais					
6	Exp	oerime	ntos	7 9			
	6.1	Conju	intos de Dados	79			
	6.2	Metoo	dologia	82			
		6.2.1	Preparação dos Dados	82			
		6.2.2	Ferramentas de Software	85			
		6.2.3	Avaliação	85			
		6.2.4	Testes Estatísticos	87			
	6.3	Result	tados Experimentais	89			
		6.3.1	Resultados no Primeiro Nível das Hierarquias	90			
		6.3.2	Resultados no Segundo Nível das Hierarquias	98			
		6.3.3	Resultados no Terceiro Nível das Hierarquias	105			
		6.3.4	Resultados no Quarto Nível das Hierarquias	112			
	6.4	Consid	derações Gerais sobre os Desempenhos das Técnicas e sobre as Medi-				
		das de	e Avaliação	119			
	6.5	Consid	derações Finais	124			

7	Con	onclusão					
	7.1	Principais Resultados					
	7.2	Contribuições do Trabalho					
	7.3	Trabalhos Futuros					
Re	eferê	ncias Bibliográficas 131					
A	Exp	perimentos Preliminares 141					
	A.1	Experimentos com Classificação Hierárquica					
		A.1.1 Bases de Dados Utilizadas					
		A.1.2 Teste Estatístico Aplicado					
		A.1.3 Experimento 1					
		A.1.4 Experimento 2					
	A.2 Experimentos com Classificação Multirrótulo						
		A.2.1 Bases de Dados Utilizadas					
		A.2.2 Teste Estatístico Aplicado					
	A.2.3 Experimento						
	A.3 Considerações Finais						
В	Valo	ores de Precisão e Revocação Hierárquicas 153					
\mathbf{C}	Test	tes Estatísticos 195					
	C.1	Análise dos Resultados da Medida <i>Hierarchical Loss Function</i> 196					
	C.2	Análise dos Resultados da Medida F1 Hierárquica					
	C.3	Análise dos Resultados da Medida Micro F1 Hierárquica 200					
	C.4	Análise dos Resultados da Medida Micro F1 Hierárquica Ponderada 202					
	C.5	Análise dos Resultados da Medida Macro F1 Hierárquica					
	C.6	Análise dos Resultados da Medida Macro F1 Hierárquica Ponderada 206					

Lista de Abreviaturas

AM Aprendizado de Máquina

MD Mineração de Dados

AD Árvores de Decisão

EC Enzyme Commission

GO Gene Ontology

SVMs Máquinas de Vetores de Suporte

DAG Grafo Acíclico Direcionado

KNN k-Nearest Neighbor

VP Verdadeiros Positivos

FP Falsos Positivos

VN Verdadeiros Negativos

FN Falsos Negativos

TA Taxa de Acerto

E Especificidade

S Sensibilidade

TFP Taxa de Falsos Positivos (FP)

P Precisão

ROC Receiver Operating Characteristics

AUC Area Under the ROC Curve

ADT Alternating Decision Tree

MMAC multi-class multi-label associative classification

TDIDT Top-Down Induction of Decision Trees

AIS Sistema Imunológico Artificial

MHC-AIS Multi-label Hierarchical Classification with an Artificial Immune System

SC Sequential Covering

RE Rule Evolution

H-Loss Hierarchical Loss Function

PR Precisão-Revocação

AUPRC área abaixo da curva PR

ORF Open Reading Frames

MIPS Munich Information Center for Protein Sequences

BioCom Laboratório de Computação Bioinspirada

GPCR G-Protein-Coupled Receptor

MIPS Munich Information Center for Protein Sequences

Ripper Repeated Incremental Pruning to Produce Error Reduction

RNA Redes Neurais Artificiais

PCT Predictive Clustering Trees

EUROFAN European Functional Analysis Projects

TRIPLES Transposon-Insertion Phenotypes, Localization, and Expression in Saccharomyces

HMC-BR Hierarchical Multilabel Classification with Binary-Relevance

HMC-LP Hierarchical Multilabel Classification with Label-Powerset

HMC-CT Hierarchical Multilabel Classification with Cross-Training

Lista de Figuras

1.1	Exemplo de Hierarquia do Esquema FunCat	5
2.1	Processo de Classificação	8
2.2	Exemplo de Função de Classificação	9
2.3	AD para a Classificação de Vertebrados	11
2.4	Exemplos de Problema de Separação de Dados usando SVMs	14
2.5	Exemplo de Hiperplano de Separação das SVMs	14
2.6	Exemplo de Rede Bayesiana para Detectar Doenças Cardíacas e Azia	15
2.7	Exemplo de Classificação Utilizando a técnica KNN para K igual a 1, 2 e 3	17
2.8	Exemplo de curva ROC	20
2.9	Hierarquia de classes estruturada como uma árvore	21
2.10	Hierarquia de classes estruturada como um DAG	22
3.1	(a) Típico problema de classificação. (b) Problema de classificação multir-	
	rótulo	30
3.2	Técnicas para Classificação Multirrótulo	31
3.3	Transformação Baseada em Rótulos	32
3.4	Eliminação de exemplos com mais de um rótulo	33
3.5	Criação de novas classes	34
3.6	Conversão por simplificação de rótulos	35
3.7	Processo de decomposição utilizando método aditivo	36
3.8	Processo de decomposição utilizando método multiplicativo	37
4.1	Problema hierárquico multirrótulo estruturado como uma árvore	44
4.2	Exemplo de uma hierarquia de classes	47
4.3	Representação gráfica das medidas precisão e revocação hierárquicas	53
4.4	Representação gráfica da função H-Loss	55
5.1	Exemplo Simples de Hierarquia de Classes	66
5.2	Exemplo de PCT	67

5.3	Exemplo de Arquivo de Configuração do Clus-HMC	68
5.4	Exemplo de Arquivo com Hierarquia de Classes para o C4.5H	69
5.5	Processo de Combinação de Classes da Técnica HMC-LP	75
5.6	Processo de Decomposição de Classes da Técnica HMC-CT	77
6.1	Exemplo de Vetor de Classes do Conjunto de Dados	83
6.2	Exemplo de Vetor de Classes Obtido pelas Técnicas da Abordagem ${\it One-Shot}$	86
6.3	Hierarquia de Classes	123

Lista de Tabelas

2.1	Conjunto de Dados de Vertebrados	10
2.2	Conjunto de Dados de Vertebrados	12
2.3	Exemplo de Matriz de Confusão	18
3.1	Comparação das técnicas independentes de algoritmo	37
4.1	Trabalhos que tratam problemas hierárquicos multirrótulo	51
4.2	Abordagens para tratar problemas hierárquicos multirrótulo	52
6.1	Categorias Principais do Esquema FunCat	80
6.2	Características dos Conjuntos de Dados Originais	82
6.3	Características Multirrótulo dos Conjuntos de Dados	84
6.4	Número de Classes por Nível	85
6.5	Valores Críticos para o Teste de Nemenyi Bicaudal	89
6.6	Comparação das Abordagens <i>Top-Down</i> e <i>One-Shot</i> Utilizando a Medida	
	H-Loss no Primeiro Nível da Hierarquia	92
6.7	Comparação das Abordagens <i>Top-Down</i> e <i>One-Shot</i> Utilizando a Medida	
	F1 Hierárquica no Primeiro Nível da Hierarquia	93
6.8	Comparação das Abordagens <i>Top-Down</i> e <i>One-Shot</i> Utilizando a Medida	
	Micro F1 Hierárquica no Primeiro Nível da Hierarquia	94
6.9	Comparação das Abordagens <i>Top-Down</i> e <i>One-Shot</i> Utilizando a Medida	
	Macro F1 Hierárquica no Primeiro Nível da Hierarquia	95
6.10	Comparação das Abordagens <i>Top-Down</i> e <i>One-Shot</i> Utilizando a Medida	
	Micro F1 Hierárquica Ponderada no Primeiro Nível da Hierarquia	96
6.11	Comparação das Abordagens <i>Top-Down</i> e <i>One-Shot</i> Utilizando a Medida	
	Macro F1 Hierárquica Ponderada no Primeiro Nível da Hierarquia	97
6.12	Comparação das Abordagens <i>Top-Down</i> e <i>One-Shot</i> Utilizando a Medida	
	H-Loss no Segundo Nível da Hierarquia	99
6.13	Comparação das Abordagens <i>Top-Down</i> e <i>One-Shot</i> Utilizando a Medida	
	F1 Hierárquica no Segundo Nível da Hierarquia	100

6.14	Comparação das Abordagens <i>Top-Down</i> e <i>One-Shot</i> Utilizando a Medida	
	Micro F1 Hierárquica no Segundo Nível da Hierarquia	101
6.15	Comparação das Abordagens Top-Down e One-Shot Utilizando a Medida	
	Macro F1 Hierárquica no Segundo Nível da Hierarquia	102
6.16	Comparação das Abordagens $Top\text{-}Down$ e $One\text{-}Shot$ Utilizando a Medida	
	Micro F1 Hierárquica Ponderada no Segundo Nível da Hierarquia $\ \ .$	103
6.17	Comparação das Abordagens <i>Top-Down</i> e <i>One-Shot</i> Utilizando a Medida	
	Macro F1 Hierárquica Ponderada no Segundo Nível da Hierarquia	104
6.18	Comparação das Abordagens <i>Top-Down</i> e <i>One-Shot</i> Utilizando a Medida	
	H-Loss no Terceiro Nível da Hierarquia	106
6.19	Comparação das Abordagens <i>Top-Down</i> e <i>One-Shot</i> Utilizando a Medida	
	F1 Hierárquica no Terceiro Nível da Hierarquia	107
6.20	Comparação das Abordagens <i>Top-Down</i> e <i>One-Shot</i> Utilizando a Medida	
	Micro F1 Hierárquica no Terceiro Nível da Hierarquia	108
6.21	Comparação das Abordagens <i>Top-Down</i> e <i>One-Shot</i> Utilizando a Medida	
	Macro F1 Hierárquica no Terceiro Nível da Hierarquia	109
6.22	Comparação das Abordagens <i>Top-Down</i> e <i>One-Shot</i> Utilizando a Medida	
	Micro F1 Hierárquica Ponderada no Terceiro Nível da Hierarquia	110
6.23	Comparação das Abordagens <i>Top-Down</i> e <i>One-Shot</i> Utilizando a Medida	
	Macro F1 Hierárquica Ponderada no Terceiro Nível da Hierarquia	111
6.24	Comparação das Abordagens Top-Down e One-Shot Utilizando a Medida	
	H-Loss no Quarto Nível da Hierarquia	113
6.25	Comparação das Abordagens Top-Down e One-Shot Utilizando a Medida	
	F1 Hierárquica no Quarto Nível da Hierarquia	114
6.26	Comparação das Abordagens <i>Top-Down</i> e <i>One-Shot</i> Utilizando a Medida	
	Micro F1 Hierárquica no Quarto Nível da Hierarquia	115
6.27	Comparação das Abordagens <i>Top-Down</i> e <i>One-Shot</i> Utilizando a Medida	110
C O O	Macro F1 Hierárquica no Quarto Nível da Hierarquia	116
6.28	Comparação das Abordagens <i>Top-Down</i> e <i>One-Shot</i> Utilizando a Medida	117
<i>c</i> 00	Micro F1 Hierárquica Ponderada no Quarto Nível da Hierarquia	117
0.29	Comparação das Abordagens <i>Top-Down</i> e <i>One-Shot</i> Utilizando a Medida	110
6.20	Macro F1 Hierárquica Ponderada no Quarto Nível da Hierarquia	
	Ranking Médio dos Algoritmos Utilizando a Medida H-Loss	
	Ranking Médio dos Algoritmos Utilizando a Medida F1 Hierárquica	
	Ranking Médio dos Algoritmos Utilizando a Medida Micro F1 Hierárquica	119
0.33	Ranking Médio dos Algoritmos Utilizando a Medida Micro F1 Hierárquica Ponderada	190
6 24	Ranking Médio dos Algoritmos Utilizando a Medida Macro F1 Hierárquica	
	Ranking Médio dos Algoritmos Utilizando a Medida Macro F1 Hierárquica Ranking Médio dos Algoritmos Utilizando a Medida Macro F1 Hierárquica	140
0.00	Ponderada	190
	1 \(\tau \)	140

A.1	Resultados na base de dados de GPCRs utilizando abordagem <i>Top-Down</i>	
	e medida baseada em custo uniforme	143
A.2	Resultados na base de dados de GPCRs utilizando abordagem <i>Top-Down</i> e medida baseada em distância independente de profundidade	143
A.3	Resultados na base de dados de GPCRs utilizando abordagem <i>Top-Down</i> e medida baseada em distância dependente de profundidade	
A.4	Resultados na base de dados de GPCRs utilizando abordagem <i>Top-Down</i>	
A.5	e medida baseada nas relações de ancestralidade e descendência	144
A . C	e medida baseada em custo uniforme	144
A.6	Resultados na base de dados de enzimas utilizando abordagem <i>Top-Down</i> e medida baseada em distância independente de profundidade	144
A.7	Resultados na base de dados de enzimas utilizando abordagem <i>Top-Down</i> e medida baseada em distância dependente de profundidade	144
A.8	Resultados na base de dados de enzimas utilizando abordagem $Top\text{-}Down$	
A.9	e medida baseada nas relações de ancestralidade e descendência Resultados na base de dados de GPCRs utilizando abordagens <i>Top-Down</i>	145
Δ 10	e <i>One-Shot</i> com C4.5 e medida baseada em custo uniforme	145
A.10	e <i>One-Shot</i> com C4.5 e medida baseada em distância independente de pro- fundidade	1.45
A.11	Resultados na base de dados de GPCRs utilizando abordagens $Top\text{-}Down$ e $One\text{-}Shot$ com C4.5 e medida baseada em distância dependente de pro-	
A.12	fundidade	146
A 13	descendência	146
	e <i>One-Shot</i> com C4.5 e medida baseada em custo uniforme	146
	fundidade	146
A.15	Resultados na base de dados de enzimas utilizando abordagens <i>Top-Down</i> e <i>One-Shot</i> com C4.5 e medida baseada em distância dependente de profundidade	146
A.16	Resultados na base de dados de enzimas utilizando abordagens <i>Top-Down</i> e <i>One-Shot</i> com C4.5 e medida baseada nas relações de ancestralidade e	110
Δ 17	descendência	146
11.11	Yeasts	148

A.19	Resultados Utilizando a Técnica <i>Label-Powerset</i> na base de dados de <i>Yeasts</i> Resultados Utilizando a Técnica <i>Cross-Training</i> na base de dados de <i>Yeasts</i> Resultados Utilizando a Técnica RAKEL na base de dados de <i>Yeasts</i>	149
A.21	Resultados Utilizando a Técnica ML-kNN na base de dados de <i>Yeasts</i> Resultados Utilizando a Técnica Um-Contra-Todos na base de dados de	149
A.23	famílias de proteínas	
A.24	lias de proteínas	
A.25	lias de proteínas	
A.26	proteínas	
B.1	Comparação dos Abordagons Ton Down o One Shot Utilizando a Modida	150
В.1	Comparação das Abordagens <i>Top-Down</i> e <i>One-Shot</i> Utilizando a Medida Precisão Hierárquica no Primeiro Nível da Hierarquia	155
B.3	Precisão Hierárquica no Segundo Nível da Hierarquia	156
B.4	Precisão Hierárquica no Terceiro Nível da Hierarquia Comparação das Abordagens $Top\text{-}Down$ e $One\text{-}Shot$ Utilizando a Medida	157
B.5	Precisão Hierárquica no Quarto Nível da Hierarquia	158
B.6	Revocação Hierárquica no Primeiro Nível da Hierarquia	
B.7	Revocação Hierárquica no Segundo Nível da Hierarquia	
B.8	Revocação Hierárquica no Terceiro Nível da Hierarquia	161
B.9	Revocação Hierárquica no Quarto Nível da Hierarquia	162
B.10	Micro Precisão Hierárquica no Primeiro Nível da Hierarquia	163
B.11	Micro Precisão Hierárquica no Segundo Nível da Hierarquia	164
B.12	Micro Precisão Hierárquica no Terceiro Nível da Hierarquia	165
B.13	Micro Precisão Hierárquica no Quarto Nível da Hierarquia Comparação das Abordagens $Top\text{-}Down$ e $One\text{-}Shot$ Utilizando a Medida	166
	Micro Revocação Hierárquica no Primeiro Nível da Hierarquia	167

B.14	Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida	
	Micro Revocação Hierárquica no Segundo Nível da Hierarquia $\ \ldots \ \ldots \ \ldots$	168
B.15	Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida	
	Micro Revocação Hierárquica no Terceiro Nível da Hierarquia $\ \ldots \ \ldots \ \ldots$	169
B.16	Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida	
	Micro Revocação Hierárquica no Quarto Nível da Hierarquia	170
B.17	Comparação das Abordagens Top-Down e One-Shot Utilizando a Medida	
	Macro Precisão Hierárquica no Primeiro Nível da Hierarquia	171
B.18	Comparação das Abordagens Top-Down e One-Shot Utilizando a Medida	
	Macro Precisão Hierárquica no Segundo Nível da Hierarquia $\ \ldots \ \ldots \ \ldots$	172
B.19	Comparação das Abordagens Top-Down e One-Shot Utilizando a Medida	
	Macro Precisão Hierárquica no Terceiro Nível da Hierarquia	173
B.20	Comparação das Abordagens Top-Down e One-Shot Utilizando a Medida	
	Macro Precisão Hierárquica no Quarto Nível da Hierarquia	174
B.21	Comparação das Abordagens Top-Down e One-Shot Utilizando a Medida	
	Macro Revocação Hierárquica no Primeiro Nível da Hierarquia	175
B.22	Comparação das Abordagens Top-Down e One-Shot Utilizando a Medida	
	Macro Revocação Hierárquica no Segundo Nível da Hierarquia	176
B.23	Comparação das Abordagens Top-Down e One-Shot Utilizando a Medida	
	Macro Revocação Hierárquica no Terceiro Nível da Hierarquia $\ \ldots \ \ldots \ \ldots$	177
B.24	Comparação das Abordagens Top-Down e One-Shot Utilizando a Medida	
	Macro Revocação Hierárquica no Quarto Nível da Hierarquia	178
B.25	Comparação das Abordagens $Top\text{-}Down$ e $One\text{-}Shot$ Utilizando a Medida	
	Micro Precisão Hierárquica Ponderada no Primeiro Nível da Hierarquia $\ .$.	179
B.26	Comparação das Abordagens $Top\text{-}Down$ e $One\text{-}Shot$ Utilizando a Medida	
	Micro Precisão Hierárquica Ponderada no Segundo Nível da Hierarquia $$	180
B.27	Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida	
	Micro Precisão Hierárquica Ponderada no Terceiro Nível da Hierarquia $$	181
B.28	Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida	
	Micro Precisão Hierárquica Ponderada no Quarto Nível da Hierarquia	182
B.29	Comparação das Abordagens $Top\text{-}Down$ e $One\text{-}Shot$ Utilizando a Medida	
	Micro Revocação Hierárquica Ponderada no Primeiro Nível da Hierarquia $$.	183
B.30	Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida	
	Micro Revocação Hierárquica Ponderada no Segundo Nível da Hierarquia $% \left(1\right) =\left(1\right) +\left(1\right$	184
B.31	Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida	
	Micro Revocação Hierárquica Ponderada no Terceiro Nível da Hierarquia $$.	185
B.32	Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida	
	Micro Revocação Hierárquica Ponderada no Quarto Nível da Hierarquia . $$.	186
B.33	Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida	
	Macro Precisão Hierárquica Ponderada no Primeiro Nível da Hierarquia	187

B.34	Comparação das Abordagens <i>Top-Down</i> e <i>One-Shot</i> Utilizando a Medida Macro Precisão Hierárquica Ponderada no Segundo Nível da Hierarquia	. 188
B.35	Comparação das Abordagens <i>Top-Down</i> e <i>One-Shot</i> Utilizando a Medida Macro Precisão Hierárquica Ponderada no Terceiro Nível da Hierarquia	
B.36	Comparação das Abordagens <i>Top-Down</i> e <i>One-Shot</i> Utilizando a Medida Macro Precisão Hierárquica Ponderada no Quarto Nível da Hierarquia	
B.37	Comparação das Abordagens <i>Top-Down</i> e <i>One-Shot</i> Utilizando a Medida Macro Revocação Hierárquica Ponderada no Primeiro Nível da Hierarquia	191
B.38	Comparação das Abordagens <i>Top-Down</i> e <i>One-Shot</i> Utilizando a Medida Macro Revocação Hierárquica Ponderada no Segundo Nível da Hierarquia .	
B.39	Comparação das Abordagens <i>Top-Down</i> e <i>One-Shot</i> Utilizando a Medida Macro Revocação Hierárquica Ponderada no Terceiro Nível da Hierarquia .	
B.40	Comparação das Abordagens <i>Top-Down</i> e <i>One-Shot</i> Utilizando a Medida Macro Revocação Hierárquica Ponderada no Quarto Nível da Hierarquia .	
C.1	Legenda para os Resultados dos Testes Estatísticos	
C.2	Análise dos Resultados da Medida <i>Hierarchical Loss Function</i> no Primeiro Nível das Hierarquias	. 197
C.3	Análise dos Resultados da Medida <i>Hierarchical Loss Function</i> no Segundo Nível das Hierarquias	. 197
C.4	Análise dos Resultados da Medida <i>Hierarchical Loss Function</i> no Terceiro Nível das Hierarquias	. 197
C.5	Análise dos Resultados da Medida <i>Hierarchical Loss Function</i> no Quarto Nível das Hierarquias	. 198
C.6	Análise dos Resultados da Medida F1 Hierárquica no Primeiro Nível das Hierarquias	. 198
C.7	Análise dos Resultados da Medida F1 Hierárquica no Segundo Nível das Hierarquias	. 199
C.8	Análise dos Resultados da Medida F1 Hierárquica no Terceiro Nível das Hierarquias	. 199
C.9	Análise dos Resultados da Medida F1 Hierárquica no Quarto Nível das Hierarquias	
C.10	Análise dos Resultados da Medida Micro F1 Hierárquica no Primeiro Nível das Hierarquias	
C.11	Análise dos Resultados da Medida Micro F1 Hierárquica no Segundo Nível das Hierarquias	
C.12	Análise dos Resultados da Medida Micro F1 Hierárquica no Terceiro Nível das Hierarquias	
C.13	Análise dos Resultados da Medida Micro F1 Hierárquica no Quarto Nível	
	das Hierarquias	. ZUI

C.14	Análise dos Resultados da Medida Micro F1 Hierárquica Ponderada no	
	Primeiro Nível das Hierarquias	202
C.15	Análise dos Resultados da Medida Micro F1 Hierárquica Ponderada no	
	Segundo Nível das Hierarquias	202
C.16	Análise dos Resultados da Medida Micro F1 Hierárquica Ponderada no	
	Terceiro Nível das Hierarquias	203
C.17	Análise dos Resultados da Medida Micro F1 Hierárquica Ponderada no	
	Quarto Nível das Hierarquias	203
C.18	Análise dos Resultados da Medida Macro F1 Hierárquica no Primeiro Nível	
	das Hierarquias	204
C.19	Análise dos Resultados da Medida Macro F1 Hierárquica no Segundo Nível	
	das Hierarquias	204
C.20	Análise dos Resultados da Medida Macro F1 Hierárquica no Terceiro Nível	
	das Hierarquias	205
C.21	Análise dos Resultados da Medida Macro F1 Hierárquica no Quarto Nível	
	das Hierarquias	205
C.22	Análise dos Resultados da Medida Macro F1 Hierárquica Ponderada no	
	Primeiro Nível das Hierarquias	206
C.23	Análise dos Resultados da Medida Macro F1 Hierárquica Ponderada no	
	Segundo Nível das Hierarquias	207
C.24	Análise dos Resultados da Medida Macro F1 Hierárquica Ponderada no	
	Terceiro Nível das Hierarquias	207
C.25	Análise dos Resultados da Medida Macro F1 Hierárquica Ponderada no	
	Quarto Nível das Hierarquias	207

Lista de Algoritmos

1	Procedimento para a Construção <i>Predictive Clustering Trees</i>	64
2	Procedimento para Encontrar o Melhor Atributo em uma PCT	65
3	Algoritmo Básico C4.5	69
4	Processo de Classificação da Técnica HMC-BR	74
5	Processo de Combinação de Classes da Técnica HMC-LP	76
6	Processo de Classificação da Técnica HMC-CT	77

Lista de Publicações

Publicações que foram geradas a partir desta pesquisa:

- 1. Cerri, R.; Carvalho, A.; Costa, E. P.; Freitas, A. Classificação Hierárquica de Proteínas Utilizando Abordagens Top-Down e Big-Bang. In: IV Workshop em Algoritmos e Aplicações de Mineração de Dados, Campinas. Anais do IV Workshop em Algoritmos e Aplicações de Mineração de Dados. Porto Alegre: SBC, 2008. p. 46-54.
- Cerri, R.; Costa, E. P.; Carvalho, A. Classificação Hierárquica de Proteínas Utilizando Técnicas de Aprendizado de Máquina. In: II Workshop on Computational Intelligence, Salvador. Anais do II Workshop on Computational Intelligence. Porto Alegre: SBC, 2008. p. 1-6.
- 3. Cerri, R.; Silva, R.; Carvalho, A. Comparing methods for multilabel classification of proteins using machine learning techniques. In: 4th Brazilian Symposium on Bioinformatics, BSB'09, Porto Alegre. Lecture Notes in Bioinformatics. Heidelberg: Springer, 2009. v. 5676. p. 109-120.
- 4. Cerri, R.; Carvalho, A. C. P. L. F. Comparing Local and Global Hierarchical Multilabel Classification Methods Using Decision Trees. In: V Workshop em Algoritmos e Aplicações de Mineração de Dados, Fortaleza. Anais do V Workshop em Algoritmos e Aplicações de Mineração de Dados. Porto Alegre: SBC, 2009. p. 75-82.

Capítulo

Introdução

Classificação é uma das mais importantes tarefas de Aprendizado de Máquina (AM) e Mineração de Dados (MD). Nessas áreas de pesquisa, um problema de classificação pode ser definido como: dado um conjunto de exemplos de treinamento compostos por pares (T_i, y_i) , no qual T_i representa uma tupla de atributos de entrada que descrevem um exemplo e y_i sua classe associada, encontrar uma função que mapeie cada T_i para sua classe associada y_i , tal que i = 1, 2, ..., m, em que m é o número de exemplos de treinamento.

A grande maioria dos problemas de classificação descritos na literatura diz respeito a problemas de classificação não hierárquicos (*Flat Classification*), em que cada exemplo é associado a uma classe pertencente a um conjunto finito de classes, todas em um mesmo nível. No entanto, existe um grande número de problemas em que uma ou mais classes podem ser divididas em subclasses ou agrupadas em superclasses. Esses problemas são conhecidos na literatura de AM como problemas de classificação hierárquica.

Muitos dos problemas hierárquicos estudados dizem respeito a problemas de classificação hierárquica multiclasse simples-rótulo. Nesses tipos de problemas, um exemplo pode pertencer a apenas uma classe dentre as várias classes que pertencem ao problema de classificação. Existem, porém, problemas hierárquicos em que duas ou mais classes podem ser atribuídas a um mesmo exemplo. Esses problemas são muito importantes na prática, principalmente nas áreas de classificação de textos e Bioinformática. Esses problemas, que combinam classificação hierárquica com classificação multirrótulo, são chamados de problemas de classificação hierárquica multirrótulo.

No campo da Bioinformática, um dos problemas que podem ser tratados utilizando classificação hierárquica multirrótulo é a predição de funções de proteínas, que são macromoléculas formadas por longas sequências de aminoácidos e que executam quase todas as funções celulares nos seres vivos. Outro problema muito investigado é o problema de

2 1 Introdução

classificação de textos. Muitos textos podem ser classificados como pertencentes a mais de uma categoria. A correta classificação de textos e artigos científicos torna mais fácil e rápida a organização e recuperação de informações.

Este capítulo está organizado da seguinte maneira: a Seção 1.1 apresenta os principais objetivos desta pesquisa; na Seção 1.2 são destacadas as motivações para a realização desta pesquisa; a Seção 1.3 apresenta as técnicas investigadas; um resumo dos principais resultados é apresentado na Seção 1.4; e na Seção 1.5 é apresentada a organização geral deste documento.

1.1 Objetivos

Os objetivos desta pesquisa, em linhas gerais, são descritos a seguir:

- Investigação de técnicas de classificação hierárquica multirrótulo: esse objetivo consistiu em realizar uma pesquisa bibliográfica das principais abordagens e técnicas propostas para resolução de problemas hierárquicos multirrótulo, a fim de verificar suas características, vantagens e deficiências.
- Implementação de novas técnicas: neste trabalho foram propostas duas novas técnicas de classificação e comparadas com técnicas existentes na literatura, de modo a realizar um estudo comparativo entre elas. As técnicas propostas são variações hierárquicas de técnicas de classificação multirrótulo não hierárquicas.
- Aplicação das técnicas de classificação hierárquica multirrótulo a um problema atual
 e relevante da área de Bioinformática: vários problemas relevantes para a área de Bioinformática podem ser investigados utilizando técnicas de classificação hierárquica
 multirrótulo. Nesta pesquisa, foram utilizados dez conjuntos de dados biológicos
 relativos à predição de funções de proteínas.
- Avaliação dos resultados por meio de diferentes medidas específicas para classificação hierárquica multirrótulo: levando em consideração as características peculiares de problemas de classificação hierárquica multirrótulo, medidas de avaliação específicas devem ser utilizadas para avaliar classificadores desenvolvidos para tais problemas. Assim, nesta pesquisa foram avaliados, por meio dessas medidas específicas, as técnicas de classificação implementadas.
- Análise de diferentes medidas de avaliação: Também foi um dos objetivos desta pesquisa o levantamento das principais medidas de avaliação existentes na literatura para problemas hierárquicos multirrótulo.

1.2 Motivação 3

1.2 Motivação

Várias motivações podem ser citadas para investigar técnicas de classificação hierárquica multirrótulo, principalmente na área de Bioinformática.

Inicialmente, problemas de classificação hierárquica multirrótulo são relativamente pouco explorados, havendo ainda muitas oportunidades para a realização de pesquisas, melhoria das técnicas atuais e contribuição com novas propostas que permitam o avanço da área. Apesar de técnicas para tratar problemas hierárquicos multirrótulo terem sido propostas na literatura, ainda há muitas contribuições a serem feitas, principalmente quanto ao desenvolvimento de novas técnicas de classificação e medidas de avaliação específicas. Uma contribuição que também pode ser dada para a área é a comparação de diferentes técnicas de classificação e diferentes medidas de avaliação, de maneira a medir o desempenho das técnicas e analisar o comportamento das medidas em diferentes bases de dados.

No campo da Bioinformática, com o avanço das pesquisas em biologia molecular e finalização do sequenciamento do genoma humano, deu-se início a pesquisas na área de proteômica, visando a identificação de proteínas expressas pelo genoma e a predição de suas funções. Essa predição pode ser feita por meio de homologia, comparando uma nova sequência com inúmeras sequências de proteínas, com funções previamente identificadas, presentes em um conjunto de dados. Apesar desse método ser muito utilizado, ele apresenta limitações. Duas proteínas podem apresentar sequências similares, mas desempenhar funções diferentes ou apresentar sequências diferentes e desempenhar funções iguais ou similares. Proteínas sendo comparadas podem ser similares em regiões de suas sequências não determinantes para suas funções. Adicionalmente, a predição por meio de homologia ignora as muitas propriedades bioquímicas das proteínas. Muitas bases de dados são organizadas de maneira hierárquica, quando são distribuídas em classes e subclasses, e alguns problemas hierárquicos são multirrótulo, quando podem ser atribuídas duas ou mais classes a um mesmo exemplo. Proteínas, por exemplo, podem desempenhar mais de uma função, dificultando ainda mais a tarefa de classificação. Sendo assim, técnicas de predição baseadas em AM têm sido exploradas na classificação de funções de proteínas.

Um exemplo de uma hierarquia de classes de proteínas é a hierarquia EC, provida pela Enzyme Commission (ECC, 1992), que organiza a classificação funcional de enzimas (proteínas que funcionam como catalisadores de reações químicas). Nessa classificação, as classes são identificadas por um esquema numérico composto por quatro números, em que o primeiro específica a classe mais geral e o último a classe mais específica. Por exemplo, a enzima tripéptido aminopeptidase tem o código EC 3.4.11.4, em que EC 3 representa uma hidrolase (enzima que usa a água para catalisar algumas moléculas), EC 3.4 representa hidrolase que atua sobre as ligações peptídicas, EC 3.4.11 representa aquelas que atuam sobre o amino-terminal de aminoácidos de um polipeptídeo, e EC 3.4.11.4 são aquelas que

4 1 Introdução

atuam sobre o amino-terminal final de um tripeptídeo.

A Gene Ontology (GO) (Ashburner et al., 2000) é outro exemplo de hierarquia de classes multirrótulo. Ela está organizada em três ontologias: ontologia de componentes celulares, ontologia de processos biológicos e ontologia de funções moleculares. Nessas ontologias, genes e proteínas, por exemplo, estão organizados de maneira hierárquica, e podem apresentar mais de uma função ou característica. Desenvolver ferramentas para organizar e recuperar essas informações é muito importante para a correta predição da função de novas proteínas.

No campo da classificação de textos, mas ainda relativo à área das ciências biológicas e saúde, documentos podem ser simultaneamente classificados em mais de uma classe. A base de dados MedLine (MedLine database, 2008) é um exemplo de repositório hierárquico multirrótulo que contém aproximadamente 11 milhões de registros científicos. Apesar de repositórios como o MedLine terem uma grande quantidade de informação, eles ainda carecem de bons mecanismos para recuperar essas informações. Sendo assim, técnicas de AM podem desempenhar um importante papel na classificação de documentos.

Existem também conjuntos de dados biológicos estruturados de acordo com o esquema FunCat (http://mips.gsf.de/projects/funcat) desenvolvido pelo MIPS (Mewes et al., 2002). O Funcat é um esquema de anotação para a descrição funcional de proteínas de procariontes, eucariontes unicelulares, plantas e animais. Sua hierarquia tem até seis níveis de profundidade, e consiste de 28 principais categorias funcionais que cobrem áreas como transporte celular, metabolismo e comunicação celular. A Figura 1.1 apresenta um exemplo de uma hierarquia estruturada de acordo com o Funcat. Como no esquema EC, as classes também são identificadas por um esquema numérico.

Por fim, do ponto de vista biológico, o estudo de técnicas computacionais para a classificação funcional de proteínas e outros dados biológicos é uma grande motivação, uma vez que o conhecimento adquirido é essencial para o desenvolvimento de novos medicamentos, tratamento de doenças e métodos de diagnóstico, por exemplo.

1.3 Técnicas Investigadas

Diversas técnicas foram propostas na literatura para lidar com problemas hierárquicos multirrótulo. Nesta pesquisa foram investigadas técnicas baseadas em duas abordagens, chamadas local ou *Top-Down* e global ou *One-Shot*.

Baseadas na abordagem *One-Shot*, foram utilizadas duas técnicas propostas na literatura. A primeira delas, chamada C4.5H, é uma variação do algoritmo de indução de Árvores de Decisão (AD) C4.5 e foi proposta por Clare (2003). Essa variação consistiu em modificações na maneira como é calculada a entropia das classes. A segunda técnica é chamada de Clus-HMC, baseada na noção de *Predictive Clustering Trees* (Blockeel et al., 2002). Nessa técnica, AD são estruturadas como uma hierarquia de grupos. A idéia geral é particionar o conjunto de classes em grupos, de maneira que a distância intra-grupos

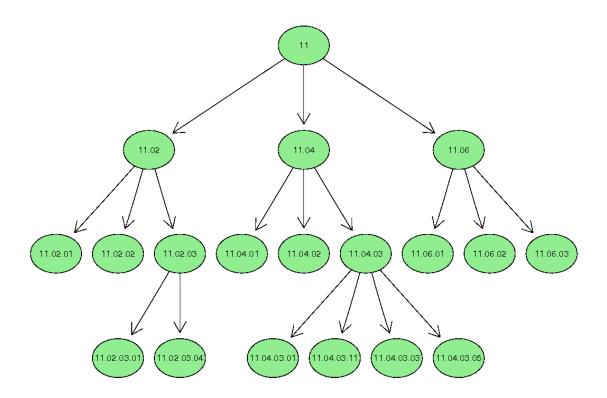


Figura 1.1: Exemplo de Hierarquia do Esquema FunCat

seja minimizada.

Na abordagem *Top-Down* foram utilizadas três diferentes técnicas. A primeira delas é chamada de *Hierarchical Multilabel Classification with Binary Relevance* (HMC-BR) e utiliza classificação binária (um-contra-todos). Nessa técnica, é associado um classificador a cada classe da hierarquia, e cada classificador é treinado para resolver um problema de classificação binário. Duas outras técnicas foram propostas neste trabalho, chamadas de *Hierarchical Multilabel Classification with Label Powerset* (HMC-LP) e *Hierarchical Multilabel Classification with Cross-Training* (HMC-CT). A técnica HMC-LP é uma variação hierárquica de uma técnica multirrótulo não hierárquica chamada *Label Powerset*, utilizada nos trabalhos de Tsoumakas e Vlahavas (2007) e Boutell et al. (2004), que utiliza uma estratégia de combinação de rótulos. A técnica HMC-CT utiliza um processo de decomposição de rótulos, e é uma extensão hierárquica de um algoritmo multirrótulo não hierárquico chamado *Cross-Training*, proposto no trabalho de Shen et al. (2004).

Ao contrário da abordagem *One-Shot*, nas técnicas baseadas na abordagem *Top-Down* podem ser utilizados algoritmos de classificação convencionais. Dessa forma, foram utilizados cinco algoritmos de classificação: *k-Nearest Neighbor* (KNN) (Aha et al., 1991), *Repeated Incremental Pruning to Produce Error Reduction* (Ripper) (Cohen, 1995), Redes Bayesianas (Friedman et al., 1997), Máquinas de Vetores de Suporte (SVMs) (Vapnik, 1999) e C4.5 (Quinlan, 1993).

6 1 Introdução

1.4 Resumo dos Resultados

Para a realização dos experimentos, foram selecionados dez conjuntos de dados biológicos, relativos à classificação funcional de proteínas do organismo Saccharomyces cerevisiae, também conhecido como Yeast. Esses conjuntos de dados foram utilizados no trabalho de Vens et al. (2008), e estão disponíveis no endereço http://www.cs.kuleuven.be/~dtai/clus/hmcdatasets/.

Os experimentos consistiram na comparação das abordagens *Top-Down* e *One-Shot* em quatro níveis da hierarquia de classes do esquema FunCat. As comparações e os resultados foram feitos e reportados nível a nível da hierarquia, utilizando diferentes medidas de avaliação específicas para problemas de classificação hierárquica multirrótulo.

Como foram utilizadas diferentes medidas de avaliação, os desempenhos das técnicas foram diferentes em cada medida, pois cada uma delas faz considerações diferentes para avaliar os classificadores. Mesmo com as variações na avaliação, os melhores resultados no primeiro nível das hierarquias foram obtidos pela técnica HMC-CT, utilizando os algoritmos Ripper e SVM. No segundo nível, os desempenhos das técnicas da abordagem *Top-Down* não se diferenciaram tanto uns dos outros. Podem ser destacados os desempenhos do algoritmo BayesNet na técnica HMC-BR, do algoritmo SVM na técnica HMC-LP e do algoritmo Ripper com a técnica HMC-CT.

No terceiro e quarto níveis da hierarquia, pode-se destacar os resultados da técnica HMC-LP. No terceiro nível, os melhores resultados foram obtidos pela técnica HMC-LP utilizando o algoritmo SVM. A técnica HMC-BR com o algoritmo BayesNet também apresentou alguns bons resultados. No quarto nível, a técnica HMC-LP, utilizando os algoritmo SVM, C4.5 e Ripper, obteve o melhor desempenho. Também pode-se observar alguns bons resultados da técnica HMC-BR com a utilização do algoritmo BayesNet.

1.5 Organização do Documento

Este documento está organizado da seguinte maneira: o Capítulo 2 apresenta os conceitos fundamentais de classificação e classificação hierárquica de dados; o Capítulo 3 trata do problema de classificação multirrótulo; o Capítulo 4 apresenta os conceitos de classificação hierárquica multirrótulo e algumas medidas de avaliação, bem como uma revisão bibliográfica de alguns trabalhos que vem sendo feitos envolvendo esse tipo de classificação; o Capítulo 5 descreve as técnicas que foram utilizadas neste trabalho; os experimentos são apresentados no Capítulo 6, e, finalmente, as conclusões da pesquisa são apresentadas no Capítulo 7; o Apêndice A apresenta alguns experimentos preliminares que foram realizados com classificação hierárquica simples-rótulo e classificação multirrótulo não hierárquica; resultados adicionais sobre os desempenhos das técnicas são mostrados no Apêndice B; e, no Apêndice C, são apresentados os resultados dos testes estatísticos realizados nos resultados experimentais.

Capítulo

Conceitos Fundamentais de Classificação e Classificação Hierárquica

2.1 Conceitos Fundamentais de Classificação

Classificação de dados é o processo de encontrar um modelo que descreva as diferentes classes presentes em um conjunto de dados, ou seja, extrair informações a partir de um conjunto de dados por meio de sua categorização. Por exemplo, em uma aplicação bancária, clientes que possuam um cartão de crédito podem ser classificados como "risco baixo", "risco normal" ou "risco alto". De maneira geral, o processo de classificação consiste em atribuir rótulos aos dados, de maneira que esses rótulos representem de alguma forma os dados categorizados sob o mesmo rótulo.

A classificação faz parte de um tipo de aprendizado de máquina chamado de aprendizado supervisionado, em que são desenvolvidos algoritmos que realizam induções de classificadores a partir de exemplos previamente classificados. Assim, é obtido um classificador utilizando exemplos que contêm a informação da sua saída esperada. Esse classificador é obtido por meio de um algoritmo de indução (indutor), que tem como objetivo fazer com que o classificador seja capaz de classificar corretamente novos exemplos.

Inicialmente, os dados pertencentes ao domínio sobre o qual será aplicado o algoritmo de classificação devem ser preparados para serem representados de forma adequada para processamento. Eles devem ser organizados em um conjunto de exemplos de maneira que cada exemplo seja representado por uma tupla de atributos. Os atributos de entrada representam as características de cada exemplo (variáveis independentes), e são utilizados para induzir o classificador. O atributo de saída representa as classes que são associadas aos exemplos (variável dependente). A Figura 2.1 apresenta um esquema de um processo

de classificação.

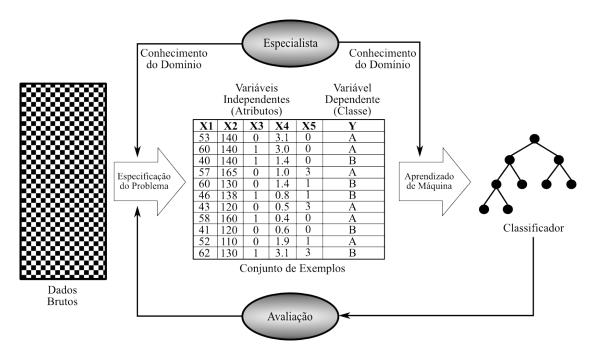


Figura 2.1: Processo de Classificação (Monard e Baranauskas, 2003)

Após a preparação, os dados são utilizados por um algoritmo de treinamento (indutor). Durante o treinamento, cada exemplo é representado por um par (T_i, y_i) , tal que i = 1, 2, 3, ..., m, em que m é o número de exemplos de treinamento, T_i é uma tupla de atributos de entrada que descrevem cada exemplo e y_i é o atributo de saída que possui o rótulo que classifica o exemplo. Segundo Tan et al. (2005), o processo de classificação consiste do aprendizado de uma função objetivo f que mapeia cada conjunto de atributos T_i em uma das classes predefinidas y_i . Essa função objetivo é encontrada por meio de ajustes dos parâmetros livres do algoritmo de indução. O conhecimento de um especialista do domínio sendo investigado pode ser utilizado para auxiliar na preparação dos dados e no processo de obtenção do classificador.

Depois de terminada a fase de treinamento, o modelo resultante consiste de um classificador que deve ser capaz de predizer corretamente a classe à qual pertence um exemplo de entrada. Dependendo de como foi feito o treinamento, pode acontecer de o modelo estar sub-ajustado ou super-ajustado aos dados. Essas situações são indesejadas. O sub-ajustamento (underfitting) ocorre quando o modelo ajusta-se muito pouco aos exemplos de treinamento, resultando em um classificador que não consegue predizer corretamente a classe de um grande número de exemplos de treinamento. Essa situação ocorre, por exemplo, quando a amostra dos dados utilizados no treinamento é pouco representativa. O super-ajustamento (overfitting) ocorre quando o modelo resultante é muito específico, ajustando-se demais aos exemplos de treinamento e sendo ineficaz na predição da classe de novos exemplos.

Após a obtenção do classificador, o mesmo é avaliado por meio da apresentação de novos exemplos, que não foram utilizados durante o treinamento. Essa etapa é conhecida

como fase de teste ou validação. Alguns aspectos importantes em relação ao classificador e ao algoritmo que o induz devem ser levados em consideração, como a habilidade do classificador de prever a classe correta dos novos exemplos, o custo computacional associado ao algoritmo e sua escalabilidade. A qualidade do classificador pode ser medida comparando as classes atribuídas por ele com as classes verdadeiras dos exemplos. A Figura 2.2 apresenta uma função de classificação obtida para o exemplo de classificação de clientes que possuam cartão de crédito. A função ajustada é capaz de distinguir entre exemplos pertencentes às classes "Risco Alto", "Risco Baixo" e "Risco Normal", e é representada pelas retas que separam as três classes. Na figura, X_1 e X_2 representam atributos, como por exemplo o salário e idade de um cliente.

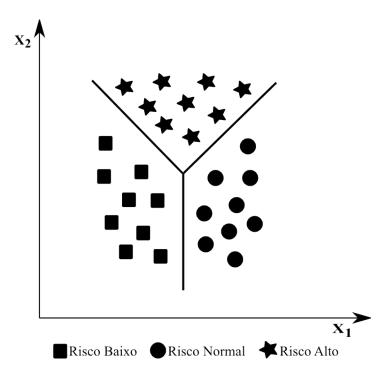


Figura 2.2: Exemplo de Função de Classificação

Na literatura podem ser encontrados muitos algoritmos de classificação. Esses algoritmos são divididos de acordo com o paradigma de classificação a que pertencem. A próxima seção descreve brevemente alguns desses algoritmos e seus respectivos paradigmas.

2.1.1 Paradigmas de Classificação de Dados

No aprendizado supervisionado, podem-se destacar diversos paradigmas de aprendizado, que determinam a abordagem utilizada por um algoritmo de AM durante o processo de indução de um classificador. Esta seção apresenta brevemente três paradigmas muito utilizados: simbólico, estatístico e baseado em exemplos. Algoritmos de classificação baseados nesses paradigmas foram utilizados nos experimentos desta pesquisa.

Para ilustrar o funcionamento de alguns algoritmos de classificação apresentados nesta seção, será utilizado como exemplo um conjunto de dados formado por animais vertebrados e suas características. Dependendo de suas características, cada vertebrado é classificado

em uma das cinco categorias: "Mamífero", "Réptil", "Peixe", "Anfíbio" ou "Pássaro". A Tabela 2.1 apresenta esse conjunto de dados.

	- ·	0.1		a	<u> </u>	ъ .		
Animal	Temperatura	Cobertura	Ovíparo	Criatura	Criatura	Possui	Hiborna	Classe
Allillai	Corporal	da pele	Oviparo	Aquática	Aérea	Pernas	IIIDeilia	Classe
Humano	Quente	Cabelos	Não	Não	Não	Sim	Não	Mamífero
Cobra	Fria	Escamas	Sim	Não	Não	Não	Sim	Réptil
Salmão	Fria	Escamas	Sim	Sim	Não	Não	Não	Peixe
Baleia	Quente	Cabelos	Não	Sim	Não	Não	Não	Mamífero
Sapo	Fria	Não possui	Sim	Semi	Não	Sim	Sim	Anfíbio
Morcego	Quente	Cabelo	Não	Não	Sim	Sim	Sim	Mamífero
Pomba	Quente	Penas	Sim	Não	Sim	Sim	Não	Pássaro
Gato	Quente	Pelos	Não	Não	Não	Sim	Não	Mamífero
Tartaruga	Fria	Escamas	Sim	Semi	Não	Sim	Não	Réptil
Pinguim	Quente	Penas	Sim	Semi	Não	Sim	Não	Pássaro
Porco Espinho	Quente	Espinhos	Não	Não	Não	Sim	Sim	Mamífero
Enguia	Fria	Escamas	Sim	Sim	Não	Não	Não	Peixe
Salamandra	Fria	Não possui	Sim	Semi	Não	Sim	Sim	Anfíbio

Tabela 2.1: Conjunto de Dados de Vertebrados (Tan et al., 2005)

Paradigma Simbólico

O paradigma simbólico fundamenta-se na construção de representações simbólicas para a generalização do conhecimento. Em classificação, essas representações são formadas por meio de relacionamentos lógicos entre os atributos de entrada do conjunto de dados e as classes envolvidas no problema. As representações simbólicas geradas geralmente podem ser interpretadas em linguagem natural, de maneira similar à representação do conhecimento utilizada por seres humanos, na forma de alguma expressão lógica, árvores de decisão, regras ou rede semântica (Monard e Baranauskas, 2003).

Sistemas baseados nesse paradigma são muito úteis quando o classificador gerado precisa ser interpretado por especialistas da área do problema ao qual o classificador está sendo aplicado. Duas técnicas muito utilizadas em aprendizado simbólico são técnicas para indução de AD e técnicas para extração de regras de decisão, das quais pode-se destacar os algoritmos C4.5 e Ripper, respectivamente, brevemente apresentados a seguir.

Árvores de Decisão

Uma maneira natural e intuitiva de se classificar um padrão é por meio de uma sequência de decisões, em que a próxima decisão depende da decisão atual (Larranaga et al., 2006). Essa sequência de decisões pode ser representada por uma estrutura de dados do tipo árvore, a qual é definida recursivamente como: um nó folha que corresponde a uma classe ou um nó interno, de decisão, que contém uma decisão sobre algum atributo. Para cada resultado da decisão, existe uma aresta para uma subárvore. Essa técnica de classificação está entre as mais populares e tem sido aplicada em tarefas como, por exemplo, diagnóstico médico (Mitchell, 1997).

Na Figura 2.3 é apresentado um exemplo de árvore de decisão para a classificação dos vertebrados da Tabela 2.1. A partir das características de cada animal, a árvore de decisão realiza inferências para classificá-lo. Supondo que uma nova espécie X de um animal seja descoberta e deva ser classificada como "Mamífero" ou "Não Mamífero", as características da tabela podem ser utilizadas para essa classificação. Através das características da tabela, se um animal tem a temperatura do corpo fria, ele definitivamente não é um mamífero. Se a temperatura de seu corpo é quente, ele pode ser um pássaro ou um mamífero. Assim, é necessário verificar se ele é um ovíparo. Aqueles animais que não são ovíparos são mamíferos, enquanto os outros animais não são mamíferos.

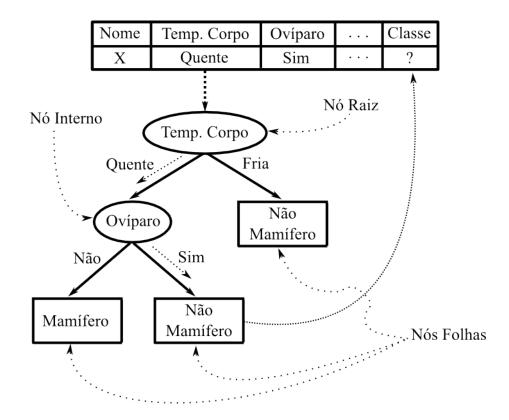


Figura 2.3: AD para a Classificação de Vertebrados (Tan et al., 2005)

A classificação de dados em uma AD se inicia na raiz em direção a um nó-folha, que define a classe. Nesse processo, ao testar um atributo em um determinado nó, move-se para baixo na árvore por meio do ramo relacionado ao valor do atributo presente no exemplo de teste. O processo é repetido para a subárvore enraizada no novo nó. O algoritmo C4.5 (Quinlan, 1993) é um exemplo de algoritmo de indução de AD muito utilizado em AM.

As ADs foram escolhidas como uma das técnicas de classificação utilizadas neste trabalho devido, principalmente, a sua alta interpretabilidade e facilidade de entendimento.

Regras de Decisão

Um classificador baseado em regras classifica exemplos utilizando uma coleção de regras da forma "Se... Então...", extraídas do conjunto de exemplos de treinamento (Tan et al., 2005). As regras são extraídas utilizando algoritmos de cobertura sequencial de

exemplos, que seguem um estratégia gulosa para a obtenção das regras. Um conjunto de regras, inicialmente vazio, é incrementado à medida que as regras são selecionadas e, a cada iteração, uma classe é analisada, sendo considerada positiva e todas as outras como negativas. A melhor regra é selecionada para compor o conjunto de regras, baseado em uma medida de avaliação. É desejável que uma regra cubra a maioria dos exemplos positivos e nenhum (ou poucos) exemplo(s) negativo(s). A seguir, é apresentado um conjunto de regras extraído do conjunto de dados de vertebrados da Tabela 2.1.

- R_1 : SE (Criatura Aquática = Sim) E (Temperatura Corporal = Fria) $ENT\tilde{A}O$ (Classe = Peixe)
- R_2 : SE (Cobertura da Pele = Cabelos) E (Criatura Aquática = Não) ENTÃO (Classe = Mamífero)
- R₃ : SE (Cobertura da Pele = Escamas) E (Criatura Aquática = Não) ENTÃO (Classe = Réptil)
- $R_4: SE \ (Oviparo = Sim) \ E \ (Criatura \ A\'erea = SIM) \ ENT\~AO \ (Classe = P\'assaro)$
- $R_5: SE \ (Cobertura Pele = N\~{a}o \ possui) \ ENT\~{A}O \ (Classe = Anfibio)$

Um exemplo de teste é classificado de acordo com a regra na qual suas características se enquadram. Por exemplo, dado um novo animal com as características apresentadas na tabela 2.2, suas características se enquadram na regra R_1 , e portanto sua classificação será "Peixe".

Tabela 2.2: Conjunto de Dados de Vertebrados (Tan et al., 2005)

Animal	Temperatura	Cobertura	0	Criatura	Criatura	Possui	TT:1	
	Corporal	da pele	Ovíparo	Aquática	Aérea	Pernas	Hiberna	
Τι	ubarão	Fria	Escamas	Não	Sim	Não	Não	Não

O algoritmo Ripper (Cohen, 1995) é um dos algoritmos mais utilizados para a extração de regras. Ele ordena, de maneira crescente, as classes envolvidas no problema de acordo com sua frequência no conjunto de treinamento, e é adequado para a construção de modelos em conjuntos de dados desbalanceados. Além disso, ele trabalha bem com dados com ruído, devido ao seu mecanismo de validação, que previne o super-ajuste do modelo ao conjunto de treinamento. Devido a essas características, esse algoritmo foi escolhido como um dos algoritmos utilizados nesta pesquisa.

Paradigma Estatístico

A idéia geral de técnicas que se baseiam nesse paradigma é explorar as dependências funcionais de um conjunto de dados por meio de modelos estatísticos, e assim encontrar um classificador que separe corretamente os dados em suas classes.

Várias dessas técnicas assumem alguma forma de modelo, e então encontram valores apropriados para os parâmetros do modelo a partir dos dados de treinamento. Um classificador linear, por exemplo, assume que as classes podem ser expressas como combinação linear dos valores dos atributos, e então procura uma combinação linear particular que fornece a melhor aproximação sobre o conjunto de dados (Monard e Baranauskas, 2003).

Inicialmente, assume-se que os dados são gerados de forma independente e identicamente distribuída, de acordo com o modelo de distribuição de probabilidade assumido para o problema. Duas técnicas muito utilizadas baseadas nesse paradigma são SVMs e Redes Bayesianas.

Máquinas de Vetores de Suporte

As SVMs, baseadas na teoria de aprendizado estatístico, utilizam funções de kernel para mapear os vetores de características dos exemplos para um espaço de dimensão mais elevada, geralmente muito maior que o espaço original (Cristianini e Shawe-Taylor, 2000). Com um mapeamento apropriado para uma dimensão suficientemente grande, é possível separar dados de duas classes por meio de um hiperplano.

Para problemas com mais de duas classes, são utilizadas duas estratégias. A primeira é conhecida como um-contra-todos, na qual o problema é decomposto em K problemas binários, sendo K o número de classes. Um classificador binário é então associado a cada classe e especializado em separar sua classe associada de todas as outras classes. Na segunda estratégia, chamada um-contra-um, são utilizados K(K-1)/2 classificadores binários, e cada classificador é utilizado para fazer a distinção entre um par de classes.

O objetivo no treinamento das SVMs é encontrar um hiperplano que separa os dados de diferentes classes com a maior margem possível. É esperado que quanto maior for essa margem, maior é a capacidade de generalização do classificador. A margem de separação entre as classes é um conceito fundamental no projeto de SVMs e está associada ao erro permitido na classificação. Os exemplos que se encontram dentro da margem de separação ou sobre ela são chamados de vetores de suporte e definem a superfície de separação. A Figura 2.4 ilustra o mapeamento dos dados do problema de classificação de vertebrados, considerando duas características, para um espaço de maior dimensão. O conjunto de dados bidimensional representado na Figura 2.4.a é mapeado para um espaço tridimensional representado na Figura 2.4.b. Nesse espaço, é determinado o hiperplano de separação entre as classes.

A Figura 2.5 ilustra um hiperplano de separação para o problema de classificação dos vertebrados nas classes "Mamífero" e "Não Mamífero".

A escolha das SVMs como uma das técnicas de classificação utilizadas nesta pesquisa foi motivada pela sua boa capacidade de generalização, mesmo para problemas com muitos atributos.

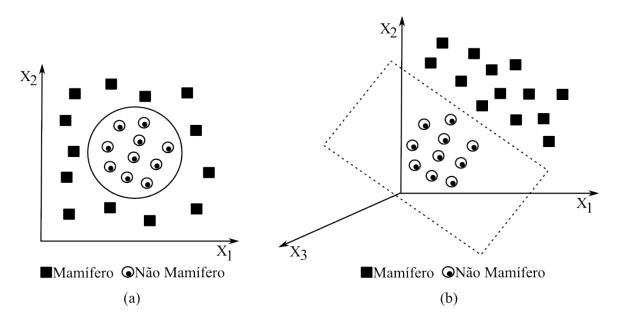


Figura 2.4: Exemplos de Problema de Separação de Dados usando SVMs

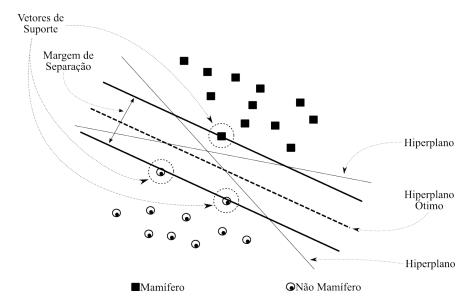


Figura 2.5: Exemplo de Hiperplano de Separação das SVMs

Redes Bayesianas

As Redes Bayesianas estão fundamentadas na manipulação de probabilidades explícitas para hipóteses. O cálculo das probabilidades é feito com base no teorema de Bayes (Bayes, 1763). Segundo esse teorema, um problema de classificação pode ser formalizado de maneira estatística. Sendo T um conjunto de atributos e y uma classe, se y tem uma relação não determinística com os atributos, T e y podem ser tratatos como variáveis aleatórias e ter sua relação probabilística capturada por P(y|T). Essa probabilidade condicional é conhecida como probabilidade a posteriori de y, em oposição à sua probabilidade a priori, P(y) (Tan et al., 2005).

Durante a fase de treinamento, as probabilidades a posteriori P(y|T) de todas as combinações de T e y são obtidas, baseado nas informações dos exemplos de treinamento. Através dessas probabilidades, um exemplo de teste T' pode ser classificado encontrando-

se a classe y' que maximiza a probabilidade a posteriori P(y', |T').

A abordagem de Redes Bayesianas utiliza uma representação gráfica que descreve os relacionamentos entre as variáveis do conjunto de dados (atributos e classes). Essa representação é constituída de um Grafo Acíclico Direcionado (DAG) (Directed Acyclic Graph) e de uma tabela de probabilidades para cada nó do DAG. Os nós representam variáveis e os arcos representam relações de dependência entre pares de variáveis.

A Figura 2.6 ilustra um exemplo de utilização de uma Rede Bayesiana para modelar um problema de pacientes com doenças cardíacas e azia. Cada variável do diagrama possui apenas dois valores. Os nós pais do nó que corresponde à variável doenças cardíacas (DC) representam fatores de risco que podem influenciar na doença, como prática de exercícios (PE) e dieta saudável (DS). Os nós filhos do nó (DC) correspondem aos sintomas da doença, como dor no peito (DP) e pressão sanguínea alta (PA). Os nós associados aos fatores de risco contêm as probabilidades a priori, enquanto os nós (DC), azia (A) e seus sintomas contêm as probabilidades a posteriori.

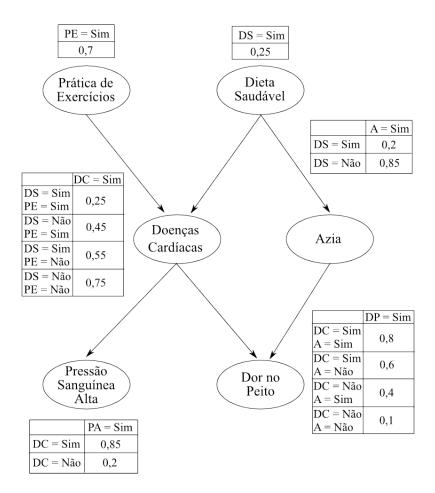


Figura 2.6: Exemplo de Rede Bayesiana para Detectar Doenças Cardíacas e Azia (adaptado de Tan et al. (2005))

Através do exemplo da Figura 2.6 pode-se, por exemplo, obter a probabilidade a posteriori de uma pessoa não ter uma doença cardíaca dado que ela não pratica exercícios,

porém tem uma dieta saudável. Essa probabilidade condicional é dada pela Equação 2.1.

$$P(DC = N\tilde{a}o|PE = N\tilde{a}o, DS = Sim) = 1 - P(DC = Sim|PE = N\tilde{a}o, DS = Sim)$$
$$= 1 - 0,55$$
$$= 0,45$$
(2.1)

Em uma Rede Bayesiana, uma variável A é dita condicionalmente dependente de B quando o nó que a representa tem como nó pai o nó que representa a variável B. A variável DP, por exemplo, é condicionalmente dependente das variáveis DC e A. Uma vez gerada a topologia da Rede Bayesiana, a classe predita é dada pelo rótulo mais provável, baseado no valor dos atributos de entrada do exemplo apresentado.

A técnica de Rede Bayesiana foi escolhida como uma das técnicas de classificação desta pesquisa devido ao fato de capturar o conhecimento a priori do conjunto de dados utilizando um modelo gráfico, fornecendo probabilidades que denotam o grau de certeza da classificação. Além disso, a técnica é robusta ao super-ajustamento (Tan et al., 2005).

Paradigma Baseado em Exemplos

Esse paradigma tem como pressuposto que se dois exemplos são similares, então eles pertencem à mesma classe. Desta forma, quando um novo exemplo é similar a um exemplo conhecido, a classe deste é atribuída ao novo exemplo. Ao contrário de outros paradigmas, que durante a fase de treinamento geram um modelo de classificação explícito e depois descartam os exemplos de treinamento, algoritmos desse paradigma precisam manter os exemplos de treinamento na memória para classificar novos exemplos, por isso são chamados de preguiçosos (lazy) (Monard e Baranauskas, 2003). Uma técnica muito utilizada é a técnica dos K-vizinhos mais próximos KNN, brevemente apresentada a seguir.

K-Vizinhos mais próximos

O algoritmo KNN relaciona cada um dos exemplos de treinamento a um ponto em um espaço n-dimensional, sendo n o número de atributos de entrada que descrevem o conjunto de dados (Fix e Hodges, 1951). Para classificar um novo exemplo, a similaridade com outros exemplos já conhecidos é calculada por meio do cálculo da distância de tais exemplos ao novo exemplo. Essa distância geralmente é calculada pela medida da distância Euclidiana entre os exemplos, considerando os valores de seus atributos de entrada. O número de exemplos a serem comparados com um novo exemplo é dado pelo parâmetro K do algoritmo.

A Figura 2.7 ilustra exemplos de classificação utilizando o algoritmo KNN para valores de K iguais a 1, 2 e 3. Um novo exemplo X é classificado baseado nas classes de seus K vizinhos mais próximos.

Ao contrário de algoritmos como ADs e baseados em regras, que buscam um modelo global para ajustar todo o espaço de busca, o algoritmo KNN faz suas predições baseado

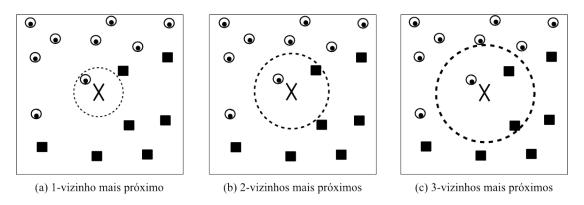


Figura 2.7: Exemplo de Classificação Utilizando a técnica KNN para K igual a 1, 2 e 3 (adaptado de Tan et al. (2005))

em informações locais. Devido a essa característica, o algoritmo KNN é mais suscetível a ruído nos dados. Apesar disso, esse algoritmo produz fronteiras de decisão mais arbitrárias, fornecendo um modelo mais flexível que algoritmos como ADs e baseados em regras. A variabilidade das fronteiras de decisão do algoritmo KNN também é maior, já que elas dependem da composição dos exemplos de treinamento (Tan et al., 2005). Devido a essas características, o algoritmo KNN foi um dos escolhidos para compor o conjunto de algoritmos de classificação utilizados nesta pesquisa.

2.1.2 Avaliação de Classificadores

Uma vez treinados, os classificadores precisam ser avaliados quanto à capacidade de predizer a classe de novos exemplos. Esta seção trata da tarefa de avaliação de classificadores.

Para a avaliação de algoritmos de classificação, deve-se tomar cuidado na partição dos exemplos que serão utilizados como treino e teste do classificador. Existe um método chamado resubstituição, que utiliza os mesmos exemplos utilizados no treinamento para o teste do classificador. Esse método faz com que a avaliação do desempenho desse classificador seja bem otimista. Esse desempenho não é necessariamente mantido quando o classificador é aplicado a novos exemplos.

Existem também métodos de reamostragem, que fazem distinção entre o conjunto de treinamento e teste do classificador. Alguns métodos de reamostragem são o *Holdoult*, amostragem aleatória, *K-fold cross-validation*, *leave-one-out* e *bootstrap* (Mitchell, 1997).

No método Holdout, os exemplos são divididos em uma porcentagem fixa (p) de exemplos para treinamento e (1-p) exemplos para o conjunto de teste.

A amostragem aleatória gera L conjuntos de treinamento e teste aleatoriamente e então induz L hipóteses, uma a partir de cada conjunto de treinamento. O erro final é obtido calculando-se a média entre os erros de cada hipótese para os conjuntos de teste.

No método *K-fold cross-validation*, o conjunto de exemplos é dividido em *k* subconjuntos ou partições. A cada iteração do algoritmo, uma partição é utilizada como teste e as demais para o treinamento. Esse processo é repetido tantas vezes quanto for o número de

partições. O stratified K-fold cross-validation utiliza a mesma proporção de exemplos de cada classe para formar as partições. Assim, cada partição reflete a mesma porcentagem de distribuição de exemplos por classe do conjunto original.

O método leave-one-out é um caso especial do cross-validation, em que cada partição tem tamanho 1. Assim, têm-se tantas partições quanto forem o número de exemplos. Esse método é computacionalmente custoso e, portanto, inviável para bases de dados grandes (Witten e Frank, 2005).

O bootstrap é um método que gera novos conjuntos de exemplos a partir dos exemplos originais utilizando reamostragem com reposição. Esse processo é repetido várias vezes de modo a gerar múltiplos conjuntos de treinamento e teste. Os dados que não foram selecionados para formar um conjunto de treinamento são utilizados para formar o conjunto de teste.

Geralmente, um problema de classificação pode ser avaliado com base em uma tabela chamada matriz de confusão. Essa matriz é composta do número de exemplos classificados correta e incorretamente para cada classe. A Tabela 2.3 apresenta uma matriz de confusão para um problema de classificação binário, que é composto apenas por duas classes (positiva e negativa).

Classe Predita Classe Verdadeira Positiva Negativa Positiva VP FN FP VN Negativa

Tabela 2.3: Exemplo de Matriz de Confusão

As siglas VP, FN, FP e VN são definidas como:

- VP: Verdadeiros positivos exemplos corretamente preditos como sendo da classe positiva;
- FP: Falsos positivos exemplos preditos como sendo da classe positiva, mas que pertencem à negativa;
- VN: Verdadeiros negativos exemplos preditos corretamente como sendo da classe negativa;
- FN: Falsos negativos exemplos preditos como sendo da classe negativa, mas que pertencem à positiva.

Para avaliar o desempenho de um classificador, são utilizadas algumas medidas de avaliação. A medida de avaliação mais utilizada é a taxa de acerto ou acurácia (Equação 2.2), que mostra a porcentagem de exemplos preditos corretamente pelo classificador.

$$TA = \frac{VN + VP}{FP + FN + VP + VN} \tag{2.2}$$

As Equações 2.3 e 2.4 apresentam outras medidas que também são utilizadas na avaliação do desempenho de classificadores. Essas medidas são chamadas de Sensibilidade (S), também conhecida como revocação (recall) ou taxa de VP, e Especificidade (E). A revocação estima a probabilidade de um exemplo pertencente à classe positiva ser predito como positivo, e a especificidade estima a probabilidade de um exemplo pertencente à classe negativa ser predito como negativo.

$$S = \frac{VP}{VP + FN} \tag{2.3}$$

$$E = \frac{VN}{VN + FP} \tag{2.4}$$

Outra medida utilizada é a Taxa de FP (TFP), que estima a probabilidade de um exemplo pertencente à classe negativa ser predito como positivo. Essa taxa é apresentada na Equação 2.5.

$$TFP = \frac{FP}{FP + FN} \tag{2.5}$$

Outra medida muito utilizada é a Precisão (P). Essa medida estima a probabilidade de uma predição positiva estar correta. A fórmula da precisão é dada pela Equação 2.6.

$$P = \frac{VP}{VP + FP} \tag{2.6}$$

A precisão pode ser combinada com a revocação, dando origem a outra medida, chamada de medida F (F-measure), que é uma combinação balanceada (média harmônica ponderada) das duas medidas. Para a realização do balanceamento, é utilizada uma constante β . A Equação 2.7 apresenta o cálculo da medida F. O valor de β define a importância dada aos valores de Precisão e Revocação. Aumentando o valor de β , aumenta-se o peso da Revocação, e diminuindo o valor de β , aumenta-se o peso da Precisão.

$$F - measure = \frac{(\beta^2 + 1) * P * S}{\beta^2 * P + S}$$
 (2.7)

A relação entre sensibilidade e especificidade de um classificador pode ser verificada por meio das curvas *Receiver Operating Characteristics* (ROC), como pode ser observado no exemplo apresentado na Figura 2.8. A curva é formada por um conjunto de pontos que correspondem a pares de valores da taxa de VP e da taxa de FP(1 - E). Por meio de uma curva ROC, pode ser definido quando um exemplo é classificado como positivo e quando ele é classificado como negativo.

Muitos algoritmos de classificação fornecem um valor contínuo para suas predições. As Redes Neurais Artificiais (RNA), por exemplo, geralmente fornecem como saída valores entre 0 e 1. Assim, é estabelecido um limiar para decidir se um exemplo pertence ou não a uma classe. Predições com valores acima do limiar determinado são atribuídas a uma classe e as predições abaixo do limiar são atribuídas à outra. Através da ordenação dos

exemplos, pode-se simular a escolha de vários valores de limiar para decisão de classificação, variando esse limiar desde o valor mais restritivo até o valor mais liberal, obtendo uma curva. A curva ROC retrata várias relações entre custos (taxa de FP) e benefícios (taxa de VP) da predição de um exemplo como positivo (Fawcett, 2004).

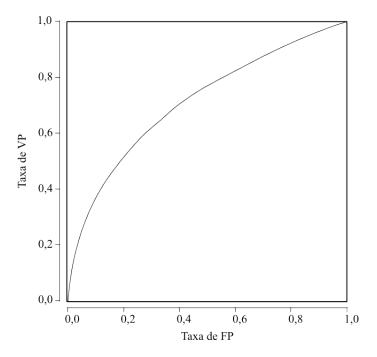


Figura 2.8: Exemplo de curva ROC

A área abaixo de uma curva ROC (AUC, do inglês Area Under the ROC Curve), é uma medida com valor entre 0 e 1, que pode ser utilizada para a comparação de dois algoritmos. A AUC de um classificador é a probabilidade de que, para um exemplo positivo e outro negativo, ambos escolhidos aleatoriamente, o exemplo positivo seja ordenado primeiro que o exemplo negativo (Egan, 1975). Quanto mais próximo de 1 é o valor da AUC, melhor é o desempenho do classificador.

2.2 Conceitos Fundamentais de Classificação Hierárquica

A grande maioria dos problemas de classificação descritos na literatura diz respeito a problemas de classificação não hierárquica (Flat Classification), em que cada exemplo é associado a uma classe pertencente a um conjunto finito de classes, não considerando assim relacionamentos hierárquicos. No entanto, existe um grande número de problemas em que uma ou mais classes podem ser divididas em subclasses ou agrupadas em superclasses. Nesse caso, as classes são dispostas em uma estrutura hierárquica, tal como uma árvore ou um Grafo Acíclico Direcionado (DAG). Esses problemas são conhecidos na literatura de Aprendizado de Máquina (AM) como problemas de classificação hierárquica. Tais problemas têm por objetivo a classificação de cada novo exemplo de entrada em um dos nós-folha. Pode ocorrer, no entanto, do classificador ter uma baixa confiança na classificação em uma das classes desse nível, sendo mais seguro classificá-lo em uma das classes dos níveis mais elevados da hierarquia.

As Figuras 2.9 e 2.10 ilustram o exemplo de um problema de classificação hierárquica de textos científicos, estruturado como uma árvore e como um DAG. A principal diferença entre a estrutura em árvore (Figura 2.9) e a estrutura DAG (Figura 2.10) é que, na estrutura em árvore, cada nó tem somente um nó pai, enquanto que na DAG cada nó pode ter um ou mais nós pais. Tanto na estrutura em árvore quanto na estrutura em DAG, os nós representam as classes envolvidas no problema.

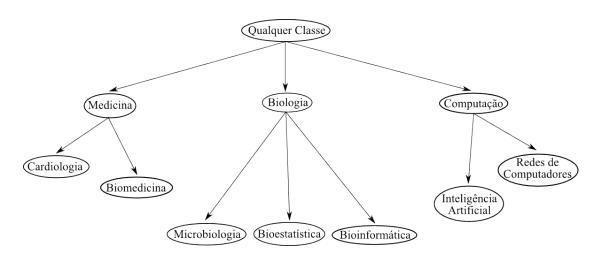


Figura 2.9: Hierarquia de classes estruturada como uma árvore

Na estrutura em árvore, quanto mais profundo é o nível, geralmente é mais difícil a predição da classe correta. Isso ocorre devido ao fato de que classes nos níveis mais profundos representam informações mais específicas e são produzidas por modelos induzidos a partir de um número menor de exemplos de treinamento. Para a estrutura DAG, essa análise é mais complexa. Nessa estrutura, um nó pode ter mais de um pai. Assim, modelos em níveis mais profundos podem ser induzidos com um número maior de exemplos de treinamento do que seus nós pais. Apesar disso, na prática, mesmo para DAGs, a precisão na predição decresce com o aumento da profundidade.

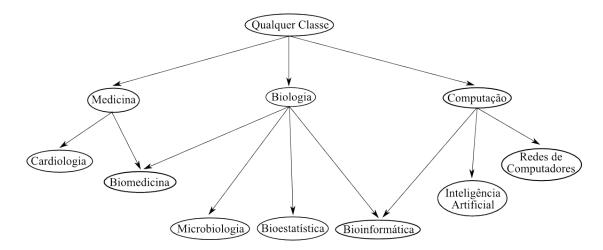


Figura 2.10: Hierarquia de classes estruturada como um DAG

Nas técnicas hierárquicas de classificação, o algoritmo de aprendizado induz um classificador que captura os relacionamentos mais relevantes entre as classes funcionais no conjunto de dados de treinamento, considerando os relacionamentos hierárquicos entre as classes. Dessa maneira, pode ser utilizada uma grande variedade de algoritmos de AM no processo de classificação (Freitas e Carvalho, 2007).

Em alguns problemas, todos os exemplos de entrada devem ser associados a classes representadas pelos nós-folha. Esses problemas são chamados de "problemas de predição obrigatória em nós-folha". Quando essa obrigação não ocorre, o problema de classificação é chamado de "problema de predição opcional em nós-folha".

Nos experimentos desta pesquisa, foram utilizadas hierarquias estruturadas como árvores e optou-se por fazer a predição obrigatoriamente nos nós-folha. Essa escolha foi feita porque dessa maneira as comparações entre as técnicas poderiam ser feitas de uma maneira mais controlada, e ainda porque as classes dos nós-folha da hierarquia representam informações mais específicas e importantes sobre os exemplos.

Definidos os conceitos fundamentais da classificação hierárquica, a próxima seção apresenta algumas abordagens propostas na literatura para tratar problemas hierárquicos.

2.2.1 Abordagens Utilizadas em Problemas Hierárquicos

Segundo Freitas e Carvalho (2007), quatro abordagens podem ser utilizadas para tratar problemas de classificação hierárquica:

- Transformação de um problema de classificação hierárquica em um problema de classificação não hierárquica;
- Predição hierárquica com algoritmos de classificação não hierárquicos;
- Classificação hierárquica local ou *Top-Down*;
- Classificação hierárquica global ou *One-Shot*.

Esta seção apresenta brevemente essas quatro abordagens.

Transformação de um Problema de Classificação Hierárquica em um Problema de Classificação não Hierárquica

Essa abordagem transforma o problema de classificação hierárquica original em um problema de classificação não hierárquica. A idéia baseia-se no fato de que um problema de classificação não hierárquica pode ser visto como um caso particular de problema de classificação hierárquica, no qual não existem subclasses e superclasses. Assim, técnicas tradicionais podem ser utilizadas nessa abordagem, sem a necessidade de alterações (Freitas e Carvalho, 2007).

Considerando a Figura 2.9, o problema de classificação hierárquica pode ser transformado no problema de predizer apenas as classes do primeiro nível. Nesse caso, informações úteis e mais específicas sobre o problema seriam perdidas, pois classes do segundo nível não seriam preditas. Por outro lado, o problema pode ser transformado no problema de predição das classes do segundo nível. Nesse caso, ao predizer as classes do segundo nível, as classes do primeiro nível são preditas implicitamente. Porém, a predição das classes do primeiro nível, que é mais confiável, deixaria de ser feita.

Predição Hierárquica com Algoritmos de Classificação não Hierárquicos

Essa abordagem divide um problema de classificação hierárquica em um conjunto de problemas de classificação não hierárquica. A principal diferença da abordagem anterior é a possibilidade de considerar os diferentes níveis da hierarquia simultaneamente. Diferente da abordagem anterior, na qual é escolhido apenas um nível para a classificação, nessa abordagem cada nível de classes é tratado como um problema de classificação independente, e para cada nível, são utilizados algoritmos de classificação não hierárquicos (Freitas e Carvalho, 2007).

No problema de classificação da Figura 2.9, um algoritmo de classificação não hierárquico seria associado a cada nível da hierarquia. Dessa maneira, o problema de classificação seria transformado em dois problemas: predizer as classes do primeiro nível e predizer as classes do segundo nível. Como esses dois problemas de classificação são distintos, os algoritmos de classificação são executados de maneira independente, tanto na fase de treinamento quanto na fase de teste. Um problema com essa abordagem é que não há garantia de que as execuções independentes dos algoritmos em níveis diferentes da hierarquia resultem na predição de classes compatíveis.

Classificação Hierárquica Top-Down

Em uma classificação *Top-Down*, durante a fase de treinamento do algoritmo, a hierarquia de classes é processada um nível de cada vez, produzindo um ou mais classificadores para cada nível da hierarquia. Esse processo produz uma árvore de classificadores. O classificador raiz é treinado com todos os exemplos de treinamento, e então, nos próximos níveis da hierarquia, um classificador é treinado usando apenas exemplos pertencentes

às classes preditas pelo classificador anterior na hierarquia. Considerando o exemplo da Figura 2.9, o classificador associado com a classe "Biologia" seria treinado apenas com os exemplos pertencentes às classes "Microbiologia", "Bioestatística" e "Bioinformática".

Na fase de teste, cada exemplo é classificado de maneira *Top-Down*. Quando atribuído a uma classe, um exemplo é então submetido a outro classificador de maneira a ser predita à qual subclasse dessa classe o exemplo pertence. Esse processo continua até uma classe em um nó-folha da árvore de classificação ser alcançada, ou nenhuma predição adicional puder ser feita para um exemplo. A abordagem *Top-Down* torna o processo de indução de um classificador simples e intuitivo, pois discriminar classes nível por nível na hierarquia é um processo parecido com o que um humano faria.

Apesar dessa abordagem produzir uma árvore de classificadores, cada classificador é construído utilizando um algoritmo de classificação não hierárquico. A desvantagem da abordagem *Top-Down* é que erros cometidos em níveis mais elevados da hierarquia de classes são propagados para os níveis mais específicos. Porém, a abordagem permite que sejam utilizados quaisquer algoritmos de classificação.

Nesta pesquisa, as técnicas de classificação hierárquica multirrótulo propostas foram baseadas nessa abordagem, pois a mesma apresentou bons resultados em experimentos considerando problemas de classificação hierárquica simples-rótulo (Costa et al., 2007, 2008) e também por permitir que vários algoritmos de classificação convencionais sejam utilizados. Além disso, a classificação ocorre de maneira mais natural, pois o processo de discriminação de classes nível a nível é mais parecido com a maneira que um humano faria a classificação.

Classificação Hierárquica One-Shot

Em uma classificação *One-Shot*, um modelo de classificação é criado considerando a hierarquia de classes como um todo, em uma única iteração do algoritmo de indução. Por isso, essa abordagem apresenta uma complexidade de implementação maior, mas evita o problema de propagação de erros da abordagem *Top-Down* (Freitas e Carvalho, 2007). Durante o processo de classificação, a predição de classes de novos exemplos é feita em apenas um passo. Por essa razão, a abordagem *One-Shot* não pode utilizar técnicas de classificação não hierárquica como a abordagem *Top-Down*. Se uma técnica de classificação não hierárquica for utilizada, ela tem que ser adaptada para considerar toda a hierarquia de classes. Uma desvantagem dessa abordagem é que como ela geralmente faz suas predições diretamente nos nós-folha da hierarquia, uma quantidade muito grande de classes e o desbalanceamento podem prejudicar seu desempenho.

Apesar da maior complexidade, técnicas baseadas em regras que seguem a abordagem One-Shot geralmente geram um conjunto de regras menor, facilitando o processo de interpretação dessas regras. Por isso, o desenvolvimento de técnicas baseadas nessa abordagem tem sido considerado na literatura.

Um exemplo dessa abordagem pode ser visto em Clare e King (2003). Nesse trabalho,

o algoritmo de indução de Árvores de Decisão (AD) C4.5 (Quinlan, 1993) foi modificado para a tarefa de classificação hierárquica. A idéia básica dessa modificação foi substituir a fórmula da entropia, utilizada para decidir qual atributo será selecionado para um dado nó da AD, por uma entropia ponderada. Essa entropia ponderada leva em consideração que classes em níveis mais elevados da hierarquia tendem a ter menores valores de entropia do que classes em níveis mais profundos, mas classes em níveis mais profundos são preferíveis, pois fornecem conhecimento mais específico.

A técnica desenvolvida por Clare e King (2003) e também outra técnica baseada na abordagem *One-Shot*, utilizada no trabalho de Vens et al. (2008), foram utilizadas nos experimentos desta pesquisa.

Uma vez treinados, os classificadores precisam ser avaliados quanto à capacidade de predizer a classe de novos exemplos. A próxima seção trata da tarefa de avaliação de classificadores hierárquicos.

2.2.2 Avaliando Modelos de Classificação Hierárquica

Antes de se avaliar um modelo de classificação hierárquica, deve-se decidir a forma com que os resultados da avaliação serão reportados. Basicamente, existem três diferentes formas de se fazer isso:

- Reportar uma taxa de desempenho para o modelo como um todo;
- Reportar uma taxa de desempenho para cada nível da hierarquia separadamente;
- Reportar uma taxa de desempenho para cada classe.

Reportar um valor de desempenho preditivo para todo o modelo é uma abordagem muito geral, fazendo com que muitas informações relevantes sejam perdidas, tornando a análise dos resultados mais difícil e imprecisa. Por outro lado, reportar valores de desempenho preditivo para cada classe é uma abordagem muito específica, pois problemas hierárquicos geralmente possuem um número elevado de classes, e reportar uma medida para cada uma delas é inviável, tornando difícil a análise dos resultados. A abordagem que reporta um desempenho preditivo para cada nível da hierarquia tem a vantagem de não ser tão geral quando se considera o processo como um todo, e nem tão específica quando se considera o desempenho em cada classe. Devido a essas características, essa foi a maneira escolhida para reportar os valores de desempenho das técnicas neste trabalho.

Outra consideração a ser feita na avaliação de modelos de classificação hierárquica é relativa à construção das matrizes de confusão dos resultados. Essas matrizes não podem ser utilizadas da mesma maneira que são utilizadas para a avaliação de classificadores não hierárquicos, pois diversas considerações devem ser feitas. Como existem relacionamentos de classes e subclasses (ou classes e superclasses), esses relacionamentos devem ser considerados para a construção da matriz de confusão. Isso é importante, pois quando um

classificador faz uma predição de um exemplo como pertencente a uma classe da hierarquia, seja essa classe representada por um nó-folha ou por um nó interno, essa predição é transmitida para os níveis superiores.

As medidas de avaliação também devem ser consideradas. Medidas de avaliação convencionais são inadequadas para modelos de classificação hierárquica, por não levarem em consideração a estrutura hierárquica desses problemas. Elas ignoram o fato de que a dificuldade na classificação aumenta conforme aumenta a profundidade da hierarquia de classes, pois estão baseadas na atribuição de um custo uniforme aos erros de classificação, independente das posições da classe verdadeira e da classe predita na hierarquia.

Como alternativas às medidas de avaliação de classificadores convencionais, medidas de desempenho específicas para classificadores hierárquicos foram propostas. Algumas delas são:

- Medidas baseadas em distância independentes de profundidade;
- Medidas baseadas em distância dependentes de profundidade;
- Medidas baseadas nas relações de ancestralidade e descendência.

Medidas baseadas em distância independentes de profundidade consideram que quanto mais perto umas das outras, mais semelhantes são as classes da hierarquia. Ao contrário, medidas baseadas em distância dependentes de profundidade consideram que conforme aumenta a profundidade da hierarquia, mais difícil fica a predição das classes e, portanto, pesos diferentes devem ser atribuídos às arestas que conectam as classes na hierarquia. Medidas baseadas nas relações de ancestralidade e descendência consideram as classes filhas das classes preditas e também as classes pais das classes preditas no momento de se medir o desempenho do classificador.

Algumas dessas medidas de avaliação foram utilizadas para a avaliação das técnicas de classificação hierárquica multirrótulo utilizadas nesta pesquisa, e serão explicadas em detalhes no Capítulo 4.

2.3 Considerações Finais

Neste capítulo foram apresentados os conceitos fundamentais de classificação e classificação hierárquica. Alguns paradigmas muito utilizados em tarefas de classificação foram discutidos de maneira breve, apresentando exemplos de técnicas desses paradigmas. Os principais métodos de particionamento do conjunto de dados utilizado para treinamento e teste de classificadores também foram descritos, além de algumas medidas utilizadas para medir a qualidade de um classificador. Abordagens para tratar problemas hierárquicos também foram apresentadas, além da questão de se considerar as características dos problemas hierárquicos para se avaliar um classificador obtido. Isso é necessário para a construção das matrizes de confusão utilizadas. Também foram introduzidos alguns tipos

de medidas de avaliação utilizadas, que serão explicadas detalhadamente no Capítulo 4, no contexto de classificação hierárquica multirrótulo. O próximo capítulo apresentará os conceitos de classificação multirrótulo, necessários para a compreensão de problemas de classificação hierárquica que também são multirrótulo.

Capítulo 3

Classificação Multirrótulo

Problemas de classificação que não são multirrótulo são chamados de problemas simplesrótulo. Nesses problemas, um classificador é treinado em um conjunto de exemplos que estão associados com uma única classe l de um conjunto de classes disjuntas L, onde |L| > 1. Se |L| = 2, então o problema é chamado de problema de classificação binária, e se |L| > 2, o problema é chamado de problema de classificação multiclasse (Tsoumakas e Katakis, 2007), como visto anteriormente.

Em uma classificação multirrótulo, cada exemplo pode ser associado a duas ou mais classes ao mesmo tempo. Um classificador multirrótulo pode ser definido como uma função $H: x \to 2^L$ que mapeia um exemplo x em um conjunto de classes $C \in 2^L$, em que 2^L é o conjunto potência de L, ou seja, o conjunto formado por todos os subconjuntos de L.

No passado, o estudo de métodos de classificação multirrótulo era motivado principalmente pelas tarefas de classificação de textos e diagnósticos médicos. Em um problema de classificação de textos, cada documento pode pertencer simultaneamente a mais de uma classe (ou tópico). Um documento, por exemplo, pode ser classificado como pertencente à área de Ciência da Computação e Física. Um artigo de jornal que aborda as reações da Igreja Católica com o lançamento do filme "O Código da Vinci" pode ser classificado ao mesmo tempo nas categorias Sociedade/Religião e Artes/Cinema. Na área de diagnósticos médicos, um paciente pode sofrer de diabetes e câncer de próstata ao mesmo tempo (Tsoumakas e Katakis, 2007). A área de classificação de textos é a que tem maior aplicação de técnicas de classificação multirrótulo (Gonçalves e Quaresma, 2003; Lauser e Hotho, 2003; Luo e Zincir-Heywood, 2005). Entretanto, muitos trabalhos podem ser encontrados nas áreas de Bioinformática (Clare e King, 2001; Zhang e Zhou, 2005; Elisseeff e Weston, 2001), diagnóstico médico (Karalic e Pirnat, 1991) e classificação de imagens (Boutell et al., 2004; Shen et al., 2004).

A Figura 3.1 apresenta uma comparação entre um problema de classificação conven-

cional, em que exemplos podem ser associados a apenas uma classe, e um problema de classificação multirrótulo. Na Figura 3.1.a é ilustrado um problema de classificação no qual um documento pode pertencer ou à classe "Física" ou à classe "Ciência da Computação", mas nunca às duas classes ao mesmo tempo. A Figura 3.1.b ilustra um exemplo de classificação multirrótulo, em que os documentos pertencentes simultaneamente às classes "Física" e "Ciência da Computação" são classificados na classe "Física Computacional".

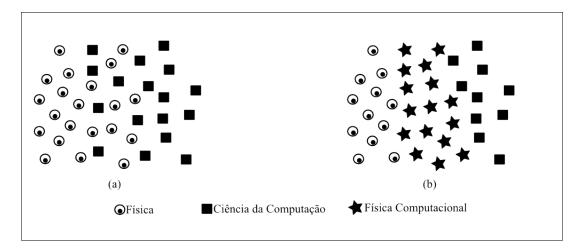


Figura 3.1: (a) Típico problema de classificação. (b) Problema de classificação multirrótulo

Diferentes técnicas têm sido propostas na literatura para tratar problemas de classificação multirrótulo. Em algumas dessas técnicas, classificadores simples-rótulo podem ser combinados para tratar problemas de classificação multirrótulo. Outras técnicas modificam classificadores simples-rótulo, através de adaptações em seus mecanismos internos, para permitir que sejam utilizados em problemas multirrótulo. Ainda, novos algoritmos podem ser desenvolvidos especificamente para tratar problemas de classificação multirrótulo (Carvalho e Freitas, 2009). A próxima seção apresenta algumas abordagens utilizadas para tratar problemas de classificação multirrótulo não hierárquicos.

3.1 Abordagens para Tratar Problemas Multirrótulo

Quando um classificador é treinado, uma probabilidade pode ser associada a cada uma das classes existentes no problema, e então ser utilizada para a classificação de um novo exemplo. Se o problema de classificação tem L classes, uma probabilidade p_i , com $1 \le i \le L$, no qual $0 \le p_i \le 1$, pode ser atribuída a cada classe. Quando o classificador é treinado em um problema de classificação simples-rótulo, há uma restrição que diz que $\sum_{p_i} = 1$. Em problemas de classificação multirrótulo, essa restrição não é adotada (Carvalho e Freitas, 2009). Problemas de classificação binária e multiclasse podem ser considerados como casos especiais de problemas de classificação multirrótulo, nos quais o número de classes atribuídas a cada exemplo é igual a 1 (Elisseeff e Weston, 2001).

A Figura 3.2 ilustra uma visão geral das diferentes técnicas propostas na literatura

para tratar problemas de classificação multirrótulo. De acordo com a figura, essas técnicas pertencem a duas grandes abordagems: abordagem independente de algoritmo e abordagem dependente de algoritmo. Como pode ser observado na figura, a abordagem independente de algoritmo utiliza algoritmos tradicionais de classificação para tratar problemas multirrótulo, transformando o problema multirrótulo original em um conjunto de problemas simples-rótulo. A abordagem dependente de algoritmo cria algoritmos específicos para tratar o problema multirrótulo. Esses algoritmos podem ser baseados em técnicas de classificação convencionais, como SVMs e AD.

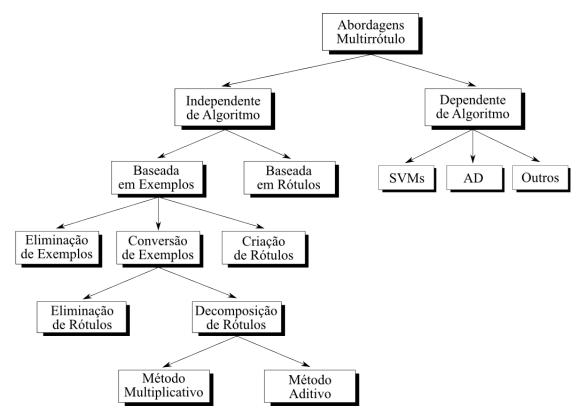


Figura 3.2: Técnicas para Classificação Multirrótulo (Carvalho e Freitas, 2009)

Para ilustrar o funcionamento das técnicas multirrótulo, será utilizado como exemplo um problema de classificação de textos científicos. Os textos podem ser classificados nas categorias "Computação", "Biologia" e "Medicina".

3.1.1 Abordagem Independente de Algoritmo

Nessa abordagem, qualquer algoritmo tradicional de classificação pode ser utilizado para tratar o problema multirrótulo. A idéia é transformar o problema original em um conjunto de problemas de classificação simples-rótulo. Essa transformação pode ser baseada nos rótulos de classes dos exemplos ou nos próprios exemplos de treinamento.

Transformação Baseada nos Rótulos das Classes

Nesse tipo de transformação, são utilizados L classificadores, sendo L o número de classes que estão envolvidas no problema. Cada classificador é então associado a uma

classe e treinado para resolver um problema de classificação binária, na qual é considerada a classe a qual ele está associado contra todas as outras classes envolvidas. Essa técnica é também chamada de técnica binária ou um-contra-todos ou *Binary-Relevance* (Tsoumakas e Vlahavas, 2007).

Um exemplo dessa técnica é ilustrado na Figura 3.3, na qual é considerado o problema multirrótulo de classificação de textos científicos nas classes "Computação", "Biologia" e "Medicina". Como cada classificador é associado a uma classe, três classificadores são treinados. O problema é então dividido em três problemas de classificação binários, sendo um para cada classe. O *i*-ésimo classificador é treinado para considerar os exemplos pertencentes à *i*-ésima classe como positivos e os outros exemplos como negativos. Assim, cada classificador torna-se especializado na classificação de uma classe particular. Quando um novo exemplo é apresentado, as classes para as quais os classificadores apresentarem uma saída positiva são atribuídas a ele.

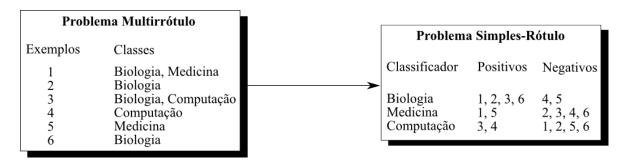


Figura 3.3: Transformação Baseada em Rótulos

Um ponto fraco dessa técnica é que ela assume que as classes atribuídas a um exemplo são independentes entre si. Isso nem sempre é verdade, e ignorar as possíveis correlações entre as classes pode fazer com que o método tenha pouca capacidade de generalização. Seu processo de transformação é reversível, ou seja, é possível recuperar as classes do problema original a partir do novo problema criado.

Nos experimentos desta pesquisa, uma variação hierárquica da técnica um-contra-todos foi implementada e utilizada para a tarefa de classificação hierárquica multirrótulo, por meio da utilização da abordagem *Top-Down*. Essa variação será explicada com detalhes na Seção 5.3.

Transformação Baseada nos Exemplos

Nesse tipo de transformação, o conjunto de classes associado a cada exemplo é redefinido, de maneira a converter o problema multirrótulo original em um ou mais problemas simples-rótulo. Ao contrário da técnica anterior, essa técnica não produz apenas problemas de classificação binária, podendo produzir tanto problemas binários quanto multiclasse.

Três diferentes estratégias são propostas para esse tipo de transformação (Carvalho e Freitas, 2009):

- Eliminação de exemplos multirrótulo;
- Criação de novos rótulos para os exemplos multirrótulo existentes;
- Conversão de exemplos multirrótulo em exemplos simples-rótulo.

Eliminação de exemplos multirrótulo

A estratégia mais simples que existe da transformação baseada em exemplos, mas também a mais ineficaz, é eliminar do conjunto de dados os exemplos que são multirrótulo. A eliminação dos exemplos com mais de uma classe não resolve o problema multirrótulo original. Ela apenas muda o problema, tranformando-o em outro mais simples e provavelmente não tão relevante quanto o original. A Figura 3.4 ilustra um exemplo dessa estratégia. Os exemplos multirrótulo 1 e 3 são eliminados do conjunto de dados original para transformar o problema em simples-rótulo. Os exemplos removidos, porém, podem representar informações importantes sobre o domínio do problema, e essas informações são perdidas.

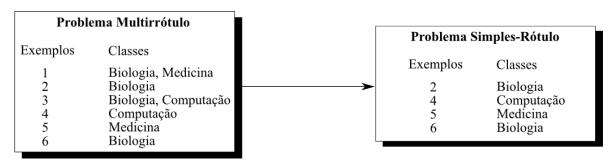


Figura 3.4: Eliminação de exemplos com mais de um rótulo

Um exemplo de perda de informação causada por essa estratégia pode ser dado através de um problema de classificação de proteínas. Proteínas podem desempenhar mais de uma função, e eliminar essas proteínas do conjunto de dados reduz a significância do classificador (Carvalho e Freitas, 2009), que não teria como predizer as outras funções de uma proteína. Essa transformação é irreversível, pois não é possível descobrir no novo problema criado, quais exemplos foram eliminados do problema original. O número de classificadores necessários também não é alterado.

Criação de novos rótulos para os exemplos multirrótulo existentes

Nessa estratégia, para cada exemplo, todas as classes atribuídas àquele exemplo são combinadas em uma nova e única classe. Com essa combinação, o número de classes envolvidas no problema pode aumentar consideravelmente, e algumas classes podem terminar com poucos exemplos. A Figura 3.5 ilustra um exemplo. Pode-se observar que as classes dos exemplos 1 e 3 foram combinadas, dando origem a uma nova classe "Biomedicina". Essa técnica de combinação de rótulos foi utilizada nos trabalhos de Tsoumakas e Vlahavas (2007) e Boutell et al. (2004), e chamada de Label-Powerset.

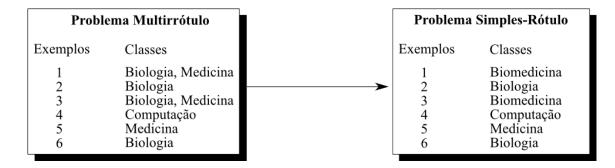


Figura 3.5: Criação de novas classes

Com a criação de novas classes, as classes do problema original não são perdidas. Se forem utilizados classificadores multiclasse no novo problema criado, a quantidade desses classificadores se mantém a mesma. Se forem utilizados classificadores binários, o número de classificadores necessários aumenta, devido ao aumento do número de classes.

Uma das técnicas propostas neste trabalho para a tarefa de classificação hierárquica multirrótulo consiste de uma variação hierárquica da técnica *Label-Powerset*, utilizando a abordagem *Top-Down*. Essa variação será explicada com detalhes na Seção 5.4.

Ainda no trabalho de Tsoumakas e Vlahavas (2007), foi criada uma técnica chamada $RAndom\ k\text{-}LabELsets$ (RAKEL), baseada na técnica Label-Powerset. Esta técnica iterativamente constrói uma combinação de m classificadores Label-Powerset. Sendo $L = \{\lambda_i\}$, i = 1..|L| o conjunto de rótulos do problema, um k-labelset é dado por um subconjunto $Y \subseteq L$, com k = |Y|. O termo L^k representa o conjunto de todos os k-labelsets de L. A cada iteração, 1..m, um k-labelset Y_i é selecionado randomicamente de L^k , sem reposição. Um classificador H_i é então treinado para Y_i .

Para a classificação de um novo exemplo x, cada classificador H_i toma uma decisão binária para cada rótulo λ_j do k-labelset Y_i correspondente. Uma decisão média é calculada para cada rótulo λ_j em L, e a decisão final é positiva para um dado rótulo se a decisão média for maior que um dado limiar t.

O propósito da técnica RAKEL é levar em consideração as correlações entre as classes e, ao mesmo tempo, evitar a desvantagem da técnica *Label-Powerset*, em que algumas classes podem terminar com poucos exemplos.

Conversão de exemplos multirrótulo em exemplos simples-rótulo

Existem duas variações para essa estratégia. Na primeira, todos os exemplos multirrótulo são convertidos para exemplos simples-rótulo, em um processo chamado de simplificação ou eliminação de rótulos. Uma segunda variação, chamada de decomposição de rótulos, decompõe todos os exemplos multirrótulo em um conjunto de exemplos simples-rótulo.

Na simplificação de rótulos, quando um exemplo possui mais de uma classe, uma de suas classes é escolhida e as outras são eliminadas. Essa escolha pode ser feita de maneira determinística, selecionando, dentre as classes as quais o exemplo pertence, aquela que

possui mais chance de ser a classe verdadeira, ou pode ser feita de maneira aleatória, selecionando uma classe randomicamente. A Figura 3.6 mostra um exemplo.

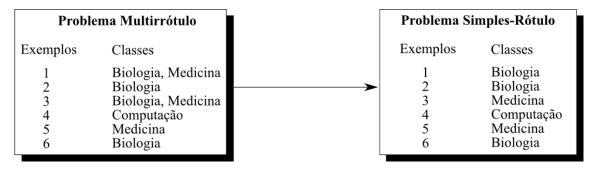


Figura 3.6: Transformação por simplificação de rótulos

É possível ver na figura que um processo de simplificação foi empregado, no qual foram escolhidas as classes "Biologia" e "Medicina" dos exemplos 1 e 3 respectivamente. A escolha de uma classe simplifica o problema, mas também apresenta desvantagens. Tomando novamente o exemplo de um problema de classificação de proteínas, se uma proteína, por exemplo, tem mais de uma função, ela seria classificada como tendo apenas uma função e, portanto, informação relevante seria perdida.

Se for adotado um critério determinístico na escolha das classes, é possível retornar ao problema multirrótulo original a partir do problema simples-rótulo criado. Se as classes forem escolhidas de maneira randômica, esse retorno não é possível. O número de classificadores utilizados no problema multirrótulo e no simples-rótulo geralmente é o mesmo.

No processo de decomposição de rótulos, um problema multirrótulo com L classes e M exemplos é dividido em K conjuntos de problemas simples-rótulo. O valor de K varia de 1, quando nenhum exemplo possui mais do que uma classe, a $(L-1)^M$, se todos os exemplos possuem L-1 classes. O processo de decomposição pode ser dividido em dois métodos: aditivo e multiplicativo (Carvalho e Freitas, 2009).

O método aditivo é ilustrado na Figura 3.7. Nesse método, para cada exemplo, cada uma das possíveis classes é considerada como sendo a classe positiva, em sequência. Para o exemplo de conjunto de dados de textos com rótulos "Biologia", "Medicina" e "Computação", quando um classificador para a classe "Biologia" for treinado, todos os exemplos multirrótulo que pertencerem à classe "Biologia" tornam-se exemplos simples-rótulo para a classe "Biologia", e o mesmo acontece para as outras classes. O método foi proposto por Shen et al. (2004) e é chamado de *Cross-Training*.

Nesta pesquisa, uma variação hierárquica da técnica *Cross-Training* foi proposta para a tarefa de classificação hierárquica multirrótulo, por meio da abordagem *Top-Down*. Essa nova variação será explicada com detalhes na Seção 5.5.

O número de classificadores utilizados é igual ao número de classes que rotulam pelo menos um exemplo multirrótulo. Esse método permite que o problema multirrótulo original seja recuperado a partir do problema simples-rótulo criado. No exemplo da Figura 3.7, o número de classificadores utilizados é igual a 2, pois as classes "Biologia" e "Medicina" rotulam exemplos multirrótulo.

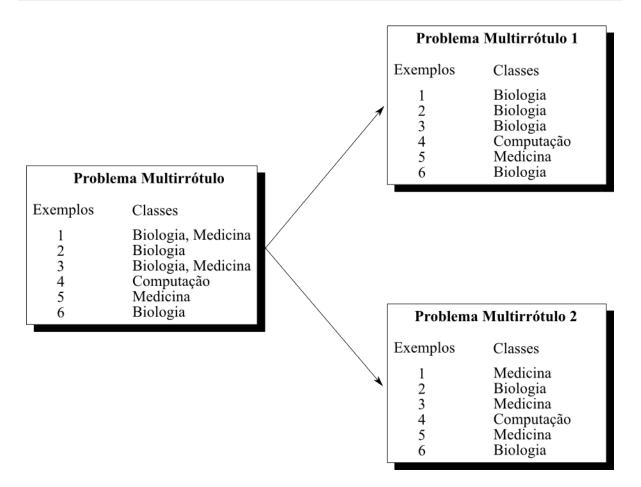


Figura 3.7: Processo de decomposição utilizando método aditivo

No método multiplicativo, por outro lado, uma combinação de todos os possíveis problemas simples-rótulo é utilizada. Esse método é similar a um método chamado um-contra-um (Carvalho e Freitas, 2009), utilizado para dividir um problema multiclasse em um conjunto de problemas binários.

O número de classificadores utilizados no método multiplicativo é igual ao $\prod l_i$, que é o produto do número de classes presentes em cada exemplo do conjunto de dados. A Figura 3.8 ilustra um exemplo, no qual o número de classificadores é dado pelo produto $2 \times 1 \times 2 \times 1 \times 1$. Nesse produto, cada número corresponde à quantidade de classes que rotulam cada exemplo do conjunto de dados.

O número de classificadores desse método cresce exponencialmente com o número de classes de cada exemplo do conjunto de dados. É possível notar também que o método aditivo produz um subconjunto de problemas simples-rótulo produzidos pelo método multiplicativo. O método multiplicativo também é reversível, permitindo a recuperação do problema multirrótulo original.

O método multiplicativo também minimiza a deficiência de métodos que eliminam ou combinam rótulos e perdem informação. Porém, ele não leva em consideração possíveis relações entre os rótulos de classes de um mesmo exemplo.

Esta seção abordou as principais técnicas utilizadas para tratar problemas de classificação multirrótulo, utilizando a abordagem independente de algoritmo. A tabela 3.1

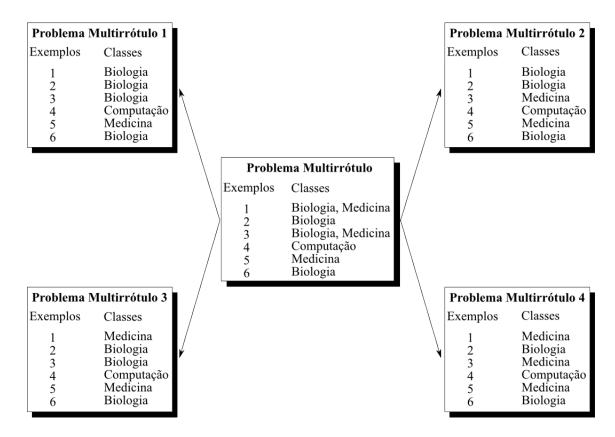


Figura 3.8: Processo de decomposição utilizando método multiplicativo

apresenta uma comparação dessas técnicas, em que L representa o número de classes, l_i representa o número de classes que rotulam o i-ésimo exemplo e K representa o número de classes do conjunto de dados que rotulam pelo menos um exemplo multirrótulo. Podese notar, pela tabela, que as técnicas diferem principalmente quanto à reversibilidade, ao número de classificadores utilizados, e ao tamanho do conjunto de dados após sua transformação.

Tabela 3.1: Comparação das técnicas independentes de algoritmo

Técnica de Transformação	Reversibilidade	N. Classificadores	N. Exemplos
Baseada nos rótulos	SIM	L	Se mantém
Eliminação de exemplos	NÃO	Se mantém	Reduz
Criação de rótulos	SIM	Se mantém	Se mantém
Simplificação de rótulos	Depende de critério	Se mantém	Se mantém
Decomp.: método aditivo	SIM	K	Aumenta
Decomp.: método multiplicativo	SIM	$\prod l_i$	Aumenta

3.1.2 Abordagem Dependente de Algoritmo

Na abordagem dependente de algoritmo, como o próprio nome sugere, novos algoritmos são propostos especificamente para tratar problemas de classificação multirrótulo. Um algoritmo específico, feito para um determinado problema de classificação, pode apresentar resultados melhores do que técnicas que seguem a abordagem independente de algoritmo. Diversas soluções baseadas nessa abordagem foram propostas para a resolução de problemas de classificação multirrótulo. Esta seção apresenta algumas delas.

Árvores de Decisão

Uma nova técnica de classificação baseada em árvores de decisão, chamada de "Árvore de Decisão Alternada" (Alternating Decision Tree (ADT)), foi proposta por Freund e Mason (1999). Essa técnica é uma generalização das árvores de decisão e seu princípio indutivo é baseado no método boosting (Freund e Schapire, 1999). Uma extensão da técnica ADT foi proposta por Comite et al. (2003), e é baseada nos métodos AdaBoost (Freund e Schapire, 1995) e ADTBoost (Freund e Mason, 1999). Esse algoritmo estende o ADT pela decomposição de problemas multiclasse usando a abordagem um-contra-todos (Carvalho e Freitas, 2009).

Outro trabalho que utiliza árvores de decisão foi proposto por Clare e King (2001). Nesse trabalho, os autores modificaram o algoritmo C4.5 (Quinlan, 1993) para a classificação de proteínas de acordo com suas funções. O algoritmo C4.5 define os nós da árvore de decisão através de uma medida chamada entropia. Os autores modificaram a fórmula dessa medida, originalmente elaborada para problemas simples-rótulo, de maneira a permitir seu uso em problemas multirrótulo. Outra modificação feita pelos autores foi a utilização dos nós-folha da árvore para representar conjuntos de rótulos de classes. Quando um nó-folha, alcançado na classificação de um exemplo, contém um conjunto de classes, uma regra separada é produzida para cada classe (Carvalho e Freitas, 2009).

Máquinas de Vetores de Suporte

Muitas das técnicas propostas para tratar problemas multirrótulo fazem uso de SVMs (Cristianini e Shawe-Taylor, 2000). No trabalho de Pavlidis e Grundy (1999), foram utilizadas SVMs para a classificação multirrótulo de funções de genes, utilizando decomposição binária. Para os experimentos, os autores utilizaram uma base de dados heterogênea, consistindo de dados de expressão gênica e perfis filogenéticos. De acordo com os autores, esses dois tipos de dados proporcionam uma visão mais exata de subconjuntos sobrepostos de categorias de funções gênicas presentes em uma célula, resultando em um melhor desempenho na classificação. Foi observado também que essa melhoria não é uniformemente distribuída entre as classes presentes no problema, portanto essa combinação só deve ser utilizada se há evidências de seu benefício.

Em Su et al. (2005), um problema multirrótulo foi decomposto em um conjunto de problemas binários utilizando a estratégia um-contra-todos. Os experimentos foram feitos utilizando um conjunto de dados de predição da localização subcelular de proteínas.

Outros métodos

Em Zhang e Zhou (2005) é proposta uma nova técnica para classificação multirrótulo baseada no algoritmo KNN, chamada ML-kNN. Nessa técnica, para cada exemplo, as classes associadas com os K exemplos vizinhos mais próximos são recuperadas, e é feita uma contagem dos vizinhos associados a cada classe. Então, o princípio $maximum\ a$

posteriori (Saridis, 1983) é utilizado para definir o conjunto de classes de um novo exemplo.

Em Schapire e Singer (1999) e Schapire e Singer (2000) são propostas duas extensões para o algoritmo Adaboost (Freund e Schapire, 1995), de maneira a permitir seu uso em problemas multirrótulo. Na primeira, é feita uma modificação na maneira de se avaliar o desempenho preditivo do modelo induzido, verificando sua capacidade de predizer um conjunto correto de classes para um dado exemplo. Na segunda, uma mudança no algoritmo faz com que ele passe a predizer um ranking de classes para cada exemplo de entrada.

Em Thabtah et al. (2004), é proposto um algoritmo multirrótulo baseado em regras de associação de classes. O algoritmo foi chamado de classificação associativa multiclasse multirrótulo (multi-class multi-label associative classification (MMAC)).

Um algoritmo de classificação baseado em entropia foi utilizado por Zhu et al. (2005) para a tarefa de recuperação de informação. Os autores utilizaram o modelo para explorar correlações entre as classes de documentos multirrótulo.

O número de exemplos multirrótulo e de classes de uma base de dados pode influenciar o desempenho das técnicas de classificação. A próxima seção faz algumas considerações a respeito disso.

3.2 Densidade e Cardinalidade de Rótulo

Os conjuntos de dados não são todos igualmente multirrótulo. Em alguns casos, o número de classes |L| de cada exemplo é pequeno se comparado com o número total de exemplos, enquanto em outros, esse número é grande. Esse número pode ser um parâmetro que influencia o desempenho das diferentes técnicas de classificação multirrótulo (Tsoumakas e Katakis, 2007). Nesta seção são apresentados os conceitos de cardinalidade de rótulo e densidade de rótulo em um conjunto de dados. Sendo D um conjunto de dados multirrótulo consistindo de |D| exemplos multirrótulo (x_i, Y_i) , com i = 1...|D| e $Y_i \subseteq L$, pode-se definir cardinalidade e densidade como:

 A cardinalidade de rótulo de D é dada pelo número médio de rótulos dos exemplos de D:

$$CR(D) = \frac{1}{|D|} \sum_{i=1}^{|D|} |Y_i|$$
 (3.1)

• A densidade de rótulo de D é dada pelo número médio de rótulos dos exemplos de D dividido por |L|:

$$DR(D) = \frac{1}{|D|} \sum_{i=1}^{|D|} \frac{|Y_i|}{|L|}$$
(3.2)

A cardinalidade de rótulo é independente do número de rótulos |L| presentes no problema de classificação, e é utilizada para quantificar o número de rótulos alternativos que caracterizam os exemplos de um conjunto de dados multirrótulo. A densidade de rótulo

leva em consideração o número de rótulos do problema de classificação. Dois conjuntos de dados com a mesma cardinalidade de rótulo, mas com uma grande diferença no número de rótulos (diferentes densidades de rótulo), podem não ter as mesmas propriedades e fazer com que os métodos de classificação multirrótulo apresentem comportamentos diferentes. Essas duas medidas são relacionadas uma com a outra: $CR(D) = |L| \times DR(D)$ (Tsoumakas e Katakis, 2007).

Para avaliar o desempenho dos classificadores multirrótulo, algumas medidas foram propostas na literatura. A próxima seção apresenta algumas delas.

3.3 Medidas de Avaliação

A avaliação de classificadores multirrótulo requer medidas diferentes das utilizadas em problemas de classificação simples-rótulo. Diferente da classificação simples-rótulo, na qual um exemplo é classificado de maneira errada ou correta, na classificação multirrótulo, um exemplo pode ser classificado de maneira parcialmente errada ou parcialmente correta. Esses casos acontecem quando um classificador atribui corretamente a um exemplo pelo menos uma das classes a que ele pertence, mas também não atribui ao exemplo uma ou mais classes as quais ele pertence. Pode acontecer também de o classificador atribuir a um exemplo uma ou mais classes as quais ele não pertence. Nesta seção serão apresentadas algumas das medidas propostas na literatura para avaliar classificadores multirrótulo.

Basicamente, o critério de avaliação utilizado pode ser baseado na classificação multirrótulo feita pelo classificador, que utiliza os rótulos atribuídos por um classificador a um dado exemplo, ou baseado em uma função de *ranking*, na qual, para cada exemplo, o classificador produz um *ranking* de rótulos.

Seja H um classificador multirrótulo com $Z_i = H(x_i)$ o conjunto de classes preditas por H para um dado exemplo x_i .

Para a avaliação baseada na classificação, uma medida muito comum é o *Hamming Loss*. No trabalho de Schapire e Singer (2000), essa medida foi utilizada. Ela é definida na Equação 3.3.

Hamming Loss
$$(H,D) = \frac{1}{|D|} \sum_{i=1}^{|D|} \frac{|Y_i \Delta Z_i|}{|L|}$$
 (3.3)

Nessa medida, o Δ representa a diferença simétrica entre dois conjuntos, e corresponde à operação XOR da lógica booleana (Tsoumakas e Katakis, 2007). Quanto menor for o valor do $Hamming\ Loss$, melhor é a classificação. A situação perfeita ocorre quando o seu valor é igual a 0.

No mesmo trabalho, o autor utilizou outras medidas de avaliação baseadas em ranking. Essas medidas são chamadas na literatura de one-error, Coverage e Average Precision. A medida one-error calcula o número de vezes que um rótulo com a melhor posição no ranking de rótulos não está presente no conjunto de rótulos corretos de um dado

exemplo de entrada (Shen et al., 2004). Na medida *Coverage*, é calculado o quão longe, em média, se deve percorrer o *ranking* de rótulos ordenados de maneira a atribuir a um exemplo de entrada, todos os rótulos que deveriam ser atribuídos a ele. A terceira medida, originalmente proposta para a tarefa de recuperação de informação, avalia a proporção média de rótulos que ocupam, no *ranking*, uma posição superior a um dado rótulo particular, e que pertencem ao conjunto de rótulos desejados (Carvalho e Freitas, 2009).

Outras medidas (Acurácia, Precisão e Revocação), foram utilizadas no trabalho de Godbole e Sarawagi (2004). Elas são apresentadas nas Equações 3.4, 3.5 e 3.6.

Acurácia
$$(H,D) = \frac{1}{|D|} \sum_{i=1}^{|D|} \frac{|Y_i \cap Z_i|}{|Y_i \cup Z_i|}$$
 (3.4)

$$Precis\~ao\ (H,D) = \frac{1}{|D|} \sum_{i=1}^{|D|} \frac{|Y_i \cap Z_i|}{|Z_i|}$$
 (3.5)

$$Revocação (H,D) = \frac{1}{|D|} \sum_{i=1}^{|D|} \frac{|Y_i \cap Z_i|}{|Y_i|}$$
(3.6)

No trabalho de Boutell et al. (2004), é utilizada uma versão mais generalizada da acurácia apresentada na Equação 3.4. Nessa versão, é utilizado um parâmetro $\alpha \geq 0$ chamado de "taxa de perdão" (forgiveness rate). Essa taxa reflete o quanto penalizar erros cometidos na predição dos rótulos de classes. Quanto menor o valor de α , mais os erros são tolerados, e quanto maior o valor de α , a penalização de erros é maior. O cálculo é apresentado na Equação 3.7.

$$Acurácia (H,D) = \frac{1}{|D|} \sum_{i=1}^{|D|} \left(\frac{|Y_i \cap Z_i|}{|Y_i \cup Z_i|} \right)^{\alpha}$$
(3.7)

3.4 Considerações Finais

Este capítulo apresentou os conceitos de classificação multirrótulo. Foram apresentadas as principais técnicas utilizadas para a resolução desse tipo de problema de classificação, bem como algumas medidas utilizadas para avaliação de classificadores multirrótulo. O próximo capítulo apresenta os conceitos de classificação hierárquica multirrótulo e algumas medidas de avaliação utilizadas na literatura. Também é feita uma revisão bibliográfica de trabalhos que vem sendo desenvolvidos para tratar problemas de classificação hierárquica multirrótulo.

Capítulo

Classificação Hierárquica Multirrótulo

Apresentados os conceitos de classificação hierárquica e classificação multirrótulo, será tratada agora a combinação desses dois problemas de classificação. Em um problema de classificação hierárquica multirrótulo, um exemplo pode pertencer a múltiplas classes ao mesmo tempo e essas classes são organizadas de maneira hierárquica. Esse tipo de problema é muito comum, por exemplo, na classificação de documentos quanto ao tema tratado ou na classificação de genes e proteínas quanto a suas funções.

4.1 Conceitos Fundamentais

Segundo Blockeel et al. (2006), o problema de classificação hierárquica multirrótulo pode ser definido da seguinte maneira:

Dado: um espaço de exemplos X e uma hierarquia de classes (C, \leq_h) , no qual C é um conjunto de classes e \leq_h é uma ordem parcial estruturada como uma árvore enraizada, representando o relacionamento de superclasse (para todo $c_1, c_2 \in C : c_1 \leq_h c_2$ se e somente se c_1 é uma superclasse de c_2); um conjunto T de exemplos (x_i, Y_i) com $x_i \in X$ e $Y_i \subseteq C$ tal que $c \in Y_i \Rightarrow \forall c' \leq_h c : c' \in Y_i$; e algum critério de qualidade q (que tipicamente recompensa modelos com alta acurácia preditiva e baixa complexidade).

Encontrar: uma função $f: X \to 2^C$ (na qual 2^C é o conjunto potência de C), tal que $c \in f(x) \Rightarrow \forall c' \leq_h c: c' \in f(x)$ e f maximiza q.

Um exemplo de problema de classificação hierárquica multirrótulo é ilustrado na Figura 4.1, na qual a hierarquia é representada por uma árvore. Uma reportagem de jornal, por exemplo, pode abordar assuntos relacionados à ciência da computação e esportes

coletivos e, dessa maneira, ser classificada nas classes "Ciências/Computação" e "Esportes/Coletivos/Futebol". Quando é feita uma predição nos nós internos da árvore, essa predição gera uma subárvore. No caso da classificação hierárquica simples-rótulo, essa subárvore é reduzida a um caminho.

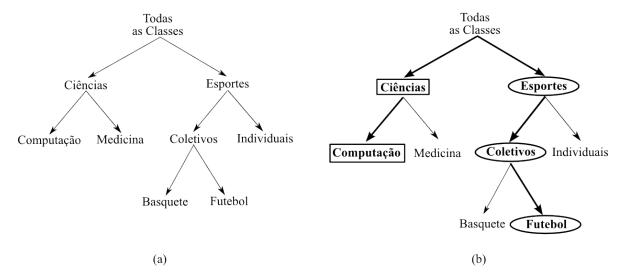


Figura 4.1: Problema hierárquico multirrótulo estruturado como uma árvore. (a) Hierarquia de classes (b) Predições gerando uma subárvore

É importante que não haja confusão entre problemas de classificação hierárquica simples-rótulo e problemas de classificação multirrótulo. Pode-se pensar em um problema hierárquico simples-rótulo como sendo naturalmente um problema multirrótulo, devido ao fato de que um ramo da hierarquia contém mais de uma classe. Quando se atribui a um exemplo às classes "Esportes/Coletivos/Futebol", está se atribuindo um ramo da hierarquia que contém três classes. Entretanto, um problema hierárquico é considerado multirrótulo quando são atribuídos mais de um ramo da hierarquia a um exemplo.

O critério de qualidade q pode, mas não necessariamente precisa, estar baseado na distância entre as classes da hierarquia. Esse critério pode ser, por exemplo, a precisão média com que todas as diferentes classes são preditas, ou pode levar em consideração que os erros de classificação em níveis mais elevados (próximos da raiz da hierarquia) são piores do que aqueles em níveis mais profundos (Struyf et al., 2005).

O problema de classificação hierárquica multirrótulo pode ser tratado por meio de duas estratégias: através do aprendizado de um classificador binário para cada classe da hierarquia, ou através do desenvolvimento de técnicas específicas para tratar tais problemas. Na primeira estratégia, o problema de classificação é dividido em um conjunto de sub-problemas binários, e então, para cada um desses sub-problemas, é utilizado um classificador binário separadamente, seguindo o método um-contra-todos. Segundo Blockeel et al. (2006), essa estratégia apresenta algumas desvantagens:

• Ela é menos eficiente, porque o algoritmo de aprendizado tem que ser executado pelo menos |C| vezes, sendo |C| o número de classes, que pode chegar a centenas ou milhares;

- O aprendizado pode ocorrer com um número muito escasso de exemplos. Dentre centenas de classes, podem haver classes com poucos exemplos, e muitos algoritmos de aprendizado têm problemas com essa situação;
- As relações hierárquicas entre as classes não são levadas em consideração, ou seja, predições não são automaticamente transferidas aos nós ancestrais;
- O modelo induzido identifica características relevantes para uma classe, não conseguindo identificar características relevantes para todo o conjunto de dados.

Devido a essas limitações, muitos trabalhos existentes na literatura seguem a segunda estratégia, apresentando técnicas específicas para tratar problemas de classificação hierárquica multirrótulo. Essas técnicas podem ser desenvolvidas segundo as abordagens *Top-Down* ou *One-Shot*.

Muitos trabalhos utilizam a abordagem *One-Shot*, considerando todas as classes de uma única vez na hora de treinar o classificador. Essa abordagem, entretanto, também apresenta algumas desvantagens. Dentre essas desvantagens, pode-se citar o aumento da complexidade da classificação, que considera todas as classes de uma única vez. Além disso, como citado no Capítulo 2, essa abordagem geralmente faz suas predições diretamente nos nós-folha da hierarquia, e uma quantidade muito grande de classes e o desbalanceamento podem prejudicar seu desempenho.

Alguns trabalhos utilizam a abordagem *Top-Down*, distinguindo as classes nível a nível na hierarquia. Essa abordagem se parece mais com a maneira com que os seres humanos fariam a tarefa de classificação. Descer na hierarquia de classes nível a nível, e a cada nó discriminar apenas algumas classes, é mais fácil do que discriminar centenas de classes de uma única vez (Kiritchenko et al., 2004). A desvantagem dessa abordagem é que erros de classificação são propagados para os níveis inferiores da hierarquia.

A próxima seção apresenta alguns trabalhos encontrados na literatura que tratam do problema de classificação hierárquica multirrótulo.

4.2 Técnicas Investigadas na Literatura

Esta seção apresenta algumas técnicas investigadas na literatura para tratar problemas hierárquicos multirrótulo.

Em Clare e King (2001), é apresentada uma técnica baseada em árvores de decisão (AD) para classificação hierárquica multirrótulo. Os autores propõem uma variação simples do algoritmo C4.5 (Quinlan, 1993), em que são feitas modificações na fórmula da entropia de classes, utilizando a soma das entropias de todas as classes. A entropia pode ser definida como uma medida da quantidade de incerteza de um conjunto de dados. Pode-se pensar em entropia como a quantidade de informação necessária para descrever um exemplo do conjunto de dados, o que é equivalente à quantidade de bits necessária para descrever todas as classes a que um exemplo pertence.

Para essa estimativa, os autores somaram o número de bits necessários para a pertinência ou não pertinência de cada classe. Quando há |L| classes e a pertinência de cada classe c_i tem probabilidade $p(c_i)$, o número total de bits necessários para descrever um exemplo é dado pela Equação 4.1.

$$-\sum_{i=1}^{|L|} ((p(c_i)\log_2 p(c_i)) + (q(c_i)\log_2 q(c_i)))$$
(4.1)

tal que:

- $p(c_i)$ é igual à probabilidade de um exemplo x_i pertencer à classe c_i ;
- $q(c_i) = 1 p(c_i)$ é igual à probabilidade de um exemplo x_i não pertencer à classe c_i .

Com essa nova informação, as folhas da árvore de decisão podem ser um conjunto de rótulos, ou seja, a saída da classificação de um novo exemplo pode ser um conjunto de classes. Dessa maneira, a árvore não prediz uma única classe, mas um vetor booleano de variáveis de classes.

Para tratar o problema hierárquico, os autores utilizaram classificadores individuais para cada nível da hierarquia. Essa estratégia tem a desvantagem de fragmentar a estrutura hierárquica, resultando em dados esparsos (muitas classes com poucos exemplos). Para contornar esse problema, foi utilizado um método de reamostragem baseado em bootstrap (Efron, 1979).

No trabalho de Blockeel et al. (2002), foi utilizado um algoritmo para classificação hierárquica multirrótulo chamado CLUS. Esse algoritmo gera uma única árvore de decisão para toda a hierarquia de classes, e é baseado na noção de *Predictive Clustering Trees* (PCTs) (Blockeel et al., 1998), em que árvores de decisão são vistas como uma hierarquia de grupos (*clusters*).

O algoritmo para indução de árvores de decisão é essencialmente um algoritmo *Top-Down Induction of Decision Trees* (TDIDT). A idéia geral é particionar recursivamente o conjunto de dados em grupos, de maneira que a variância intra-grupos seja minimizada. A heurística utilizada para selecionar os testes que são incluídos como nós da árvore é baseada nessa variância. Na árvore de decisão produzida por esse algoritmo, os nósfolha não contêm classes, e cada nó, assim como os nósfolha, corresponde a um grupo. Para calcular a distância entre dois grupos, é utilizada uma função que calcula o valor da distância entre os valores dos protótipos (*prototype*) desses dois grupos. O valor do protótipo de um grupo de exemplos pode ser considerado como uma descrição parcial dos exemplos. Na árvore de decisão final, cada nó-folha conterá o valor do protótipo dos exemplos cobertos por aquela folha.

No trabalho de (Clare e King, 2003), a técnica proposta por Clare e King (2001) foi estendida para predizer classes em vários níveis da hierarquia, atribuindo um custo de classificação errada maior para os níveis mais altos. As modificações para considerar a hierarquia de classes envolveram mudanças na maneira de ler e armazenar a hierarquia,

no teste de pertinência de uma classe, na maneira de encontrar a melhor classe ou classes para representar um nó e nos cálculos da entropia. Atribuir uma classe a um exemplo significa atribuir ao exemplo todas as classes ancestrais dessa classe. Quando um nó é representado por classes do nível mais geral da hierarquia, o valor da entropia para aquele nó é menor do que quando o nó é representado por classes mais específicas. Entretanto, é preferível a utilização de classes mais específicas nos resultados, pois estas fornecem maior precisão e informação a respeito do exemplo sendo classificado. Os autores levaram isso em consideração para computar o valor da entropia. A técnica proposta, chamada C4.5H, foi utilizada nesta pesquisa e será explicada com mais detalhes no Capítulo 5.

Em Kiritchenko et al. (2004), foi feito um trabalho de atribuição de funções biológicas a genes por meio da classificação de artigos do MedLine (MedLine database, 2008) que descrevem esses genes. Para a classificação, foram considerados como rótulos de classes os códigos da Gene Ontology (GO) (Ashburner et al., 2000) associados a cada artigo. A técnica proposta foi baseada na abordagem Top-Down. O algoritmo primeiramente treina um classificador para as classes do primeiro nível da hierarquia. Para tanto, qualquer técnica de AM que tenha como saída uma pontuação de probabilidade pode ser utilizada. Uma pontuação de probabilidade pode representar a probabilidade atual de um exemplo pertencer a uma dada classe. Em seguida, o classificador gera um conjunto de pontuações para cada novo exemplo e apenas as classes para as quais um exemplo tem uma pontuação maior que um dado limite (threshold) são atribuídas àquele exemplo. Se nenhuma das classes têm uma pontuação maior do que o dado limite, o exemplo é atribuído ao nó raiz e o processo de classificação é interrompido. Caso contrário, o processo de classificação é repetido para todas as classes atribuídas ao exemplo, classificando-o em classes dos níveis mais profundos da hierarquia. Assim, o algoritmo permite que sejam feitas predições em vários níveis da hierarquia.

No trabalho de Struyf et al. (2005) foi utilizado o conceito de PCTs. Cada exemplo é rotulado com um vetor v que representa um conjunto válido de classes. Considerando a hierarquia de classes apresentada na Figura 4.2, supondo que um dado exemplo i seja classificado com um conjunto válido de classes $C_i = [1, 2, 2.2]$, e assumindo que $v_{i,k}$ corresponde a uma classe na posição k em um percurso pré-ordem na hierarquia, o vetor representando C_i seria dado por $v_i = [1, 1, 0, 1, 0]$.

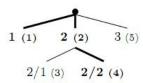


Figura 4.2: Exemplo de uma hierarquia de classes

A distância entre dois conjuntos de classes C_i e C_j pode ser calculada por meio da Equação 4.2 utilizando a distância Euclidiana.

$$d(C_i, C_j) = d_{Euclidiana}(v_i, v_j) = \sqrt{\sum_k w_k \times (v_{i,k} - v_{j,k})^2}$$
(4.2)

Os relacionamentos hierárquicos entre as classes podem ser levados em consideração por meio da utilização de valores apropriados para o peso w_k . Os autores diminuíram seu valor exponencialmente conforme aumentava a profundidade da hierarquia ($w_k = w_0^{profundidade(c_k)}$).

O protótipo p_V (prototype) de um conjunto de vetores V, que corresponde à distância Euclidiana, é dado pelo vetor média de V. Esse cálculo é apresentado na Equação 4.3.

$$p_V = \frac{\sum_{v_i \in V} v_i}{|V|} \tag{4.3}$$

Em Blockeel et al. (2006) é feito um estudo sobre métodos baseados em AD para problemas de classificação hierárquica multirrótulo na área de genômica funcional. É comparada uma técnica de indução de uma única árvore de decisão (que faz predições para todas as classes juntas) com uma técnica que induz um conjunto de árvores de decisão (uma para cada classe). O estudo mostrou que a primeira técnica se saiu melhor em todos os conjuntos de dados analisados. Ela é mais rápida para ser aprendida e aplicada, mais fácil de ser interpretada e tem um desempenho preditivo similar ou superior ao de um conjunto de árvores de decisão.

Para tornar possível, utilizando uma árvore de decisão, a predição de um conjunto de classes para um dado exemplo, algumas modificações tiveram que ser feitas no procedimento de aprendizado. O algoritmo desenvolvido foi chamado de Clus-HMC. Primeiramente, os rótulos de classes dos exemplos foram representados como vetores de classes booleanos. A i-ésima posição do vetor é 1 se o exemplo pertence à classe c_i e 0 caso contrário. Segundo os autores, a média aritmética de um conjunto de tais vetores contém, em sua i-ésima posição, a proporção de exemplos do conjunto de dados que pertencem à classe c_i . A variância de um conjunto de exemplos pode então ser definida como a distância Euclidiana quadrática média entre cada rótulo dos exemplos e o rótulo médio do conjunto. A heurística utilizada para escolher o melhor teste para ser colocado em um nó da árvore consiste na minimização da variância média dos subconjuntos criados.

Diferente de uma árvore de decisão tradicional utilizada para a tarefa de classificação simples-rótulo, um nó-folha não armazena apenas a classe majoritária para aquela folha. Nesse caso, os nós-folha armazenam a média \overline{v} dos vetores dos exemplos daquela folha. Ou seja, para cada classe c_i , a proporção de exemplos pertencentes à classe c_i é armazenada. Um exemplo que chega a um nó-folha será predito como pertencente a uma classe c_i se o i-ésimo componente de \overline{v} estiver acima de um limiar t_i . Escolhendo $t_i \leq t_j$ sempre que $c_i \leq_h c_j$, os autores garantem que sempre que uma classe é predita, suas superclasses também são preditas. A técnica Clus-HMC foi utilizada nesta pesquisa e será explicada com mais detalhes no Capítulo 5.

No trabalho de Barutcuoglu et al. (2006), é utilizada uma hierarquia de SVMs. Os

classificadores, após treinados, têm suas predições combinadas utilizando aprendizado Bayesiano, para obter o mais provável conjunto consistente de predições. O método Bayesiano utilizado permite que todos os níveis de classificadores sejam influenciados uns pelos outros. Assim, as predições feitas ficam consistentes com os relacionamentos hierárquicos.

Inicialmente, o algoritmo permite que ocorram inconsistências e então as explora. Classificadores independentes são treinados para cada classe da hierarquia. Em seguida é desenvolvido um esquema Bayesiano de combinação hierárquica para permitir uma correção de erros colaborativa entre todos os nós. Quando for obtido um conjunto inconsistente de predições pelos classificadores independentes, ele é utilizado para encontrar o mais provável conjunto de rótulos consistentes. O conjunto de treinamento utilizado foi a hierarquia de processos biológicos da GO. Os resultados dos experimentos mostraram uma melhora do desempenho da combinação hierárquica.

No trabalho de Rousu et al. (2006) é apresentado um algoritmo baseado em kernels para a classificação hierárquica multirrótulo de textos. O modelo de classificação é uma variação de um framework chamado Maximum Margin Markov Network (Taskar et al., 2003; Tsochantaridis et al., 2005), em que a hierarquia de classificação é representada como uma árvore de Markov. O treinamento de um único modelo de classificação mostrou-se tão eficiente quanto o treinamento de classificadores SVMs independentes para cada nó da hierarquia.

No trabalho de Esuli et al. (2006), a classificação hierárquica multirrótulo de textos é feita por meio de um algoritmo chamado TreeBoost.MH, que consiste de uma variação hierárquica do algoritmo AdaBoost.MH (Schapire e Singer, 1999), membro da família boosting de algoritmos de aprendizado (Freund e Schapire, 1999). O algoritmo faz uma seleção de atributos e uma seleção de exemplos de treino negativos localmente, considerando a topologia do esquema de classificação. A distribuição de pesos que o algoritmo boosting atualiza a cada iteração também é atualizada localmente. O TreeBoost.MH é definido pelos autores como um algoritmo recursivo que utiliza o AdaBoost.MH como base, e é chamado recursivamente na estrutura da árvore.

No trabalho de Cesa-Bianchi et al. (2006), o grafo que representa os relacionamentos entre as classes da hierarquia é visto como um floresta. As árvores que formam essa floresta são interpretadas como uma taxonomia de classes. Os autores assumem que cada exemplo do conjunto de dados é rotulado com um conjunto de classes, sendo que esse pode ser vazio. Sempre que um exemplo é rotulado com um nó i, ele também é rotulado com os nós do caminho da raiz da árvore até o nó i. Durante o processo de rotulação, um exemplo pode ser rotulado com nós pertencentes a mais de um caminho na floresta, ou com nós pertencentes a um caminho que não atinge um nó-folha.

O algoritmo proposto incrementalmente treina um classificador para cada nó da hierarquia de classes. Uma classificação final para um exemplo é então obtida percorrendo todos os classificadores de maneira *Top-Down*. A classificação final é consistente com a

hierarquia de classes.

Em Alves et al. (2008), é proposta a utilização de um Sistema Imunológico Artificial (AIS) para a predição de funções de proteínas descritas na (GO) (Ashburner et al., 2000). No trabalho, é utilizado um algoritmo chamado Multi-label Hierarchical Classification with an Artificial Immune System (MHC-AIS). O objetivo do algoritmo é encontrar um conjunto de regras que sejam tanto hierárquicas quanto multirrótulo, de maneira que uma única regra de classificação possa atribuir mais de uma classe a uma proteína. A fase de treinamento do algoritmo é dividida em dois procedimentos básicos, chamados Sequential Covering (SC) e Rule Evolution (RE). Esses procedimentos produzem regras de classificação candidatas, em que cada regra é composta por duas partes: uma regra antecessora (IF part), representada por um vetor de condições, e uma regra predecessora (THEN part), representada por um subconjunto de classes preditas. As classes envolvidas no problema são termos da GO representando funções de proteínas.

Em Vens et al. (2008) são apresentadas algumas abordagens que utilizam AD (baseadas no conceito de *Predictive Clustering Trees*) em problemas de classificação hierárquica multirrótulo. As abordagens são utilizadas e comparadas em bases de dados relacionadas à genômica funcional. Os autores compararam o desempenho de induzir uma única AD, que faz predições para todas as classes da hierarquia de uma única vez (chamada de HMC), com outras duas abordagens, que fazem o aprendizado de uma AD para cada classe. A primeira, chamada de SC, define uma tarefa de classificação simples-rótulo independente para cada classe, ignorando assim os relacionamentos hierárquicos entre as classes. A segunda, chamada de HSC, explora os relacionamentos para induzir uma AD para cada classe. Os experimentos mostraram que a abordagem HMC supera as abordagens HSC e SC com relação à acurácia, tamanho do modelo e tempo de indução.

O trabalho de Valentini (2009) utiliza combinação de classificadores (ensemble). Classificadores base são associados a cada classe do problema e treinados para reconhecer uma única classe da hierarquia. Esses classificadores trocam informações para que seja obtida uma classificação final baseada na combinação de suas predições. As informações são trocadas em um fluxo bidirecional, no qual predições positivas para um classe influenciam suas classes ancestrais e predições negativas influenciam suas classes descendentes. Dessa maneira, a consistência da estrutura hierárquica é mantida.

Tabela 4.1 apresenta, de forma sumarizada, os trabalhos revisados neste capítulo, e a Tabela 4.2 apresenta os trabalhos organizados de acordo com a abordagem utilizada para a resolução do problema hierárquico multirrótulo.

Tabela 4.1: Trabalhos que tratam problemas hierárquicos multirrótulo

Artigo	Descrição
(Clare e King, 2001)	Variação do algoritmo C4.5, utilizando a soma da entropia de
	todas as classes. Utiliza classificadores individuais para cada nível
	da hierarquia.
(Blockeel et al., 2002)	Utiliza um algoritmo baseado na noção de predictive clustering
	trees, onde árvores de decisão são vistas como uma hierarquia de
	clusters. Gera uma única árvore de decisão para a hierarquia.
(Clare e King, 2003)	Extensão de Clare e King (2001), predizendo classes em todos os
	níveis da hierarquia, atribuindo um custo de classificação errada
	maior para níveis mais altos. Utiliza um único classificador.
(Kiritchenko et al., 2004)	Utiliza uma técnica baseada na técnica Top-Down, utilizando téc-
	nicas de AM que têm como saída uma pontuação de probabilidade.
(Struyf et al., 2005)	Utiliza o mesmo conceito de predictive clustering trees utilizado
	por Blockeel et al. (2002), mas foi introduzida uma nova medida
	de distância específica para classificação hierárquica multirrótulo.
	É gerada uma única árvore de decisão para toda a hierarquia.
(Barutcuoglu et al., 2006)	Utiliza uma hierarquia de classificadores SVMs. Os classificadores
	são treinados e então suas predições são combinadas utilizando
	aprendizado Bayesiano.
(Rousu et al., 2006)	Propõe um algoritmo baseado em kernel para a classificação hie-
	rárquica multirrótulo. A hierarquia de classificação é representada
	como uma árvore de Markov. É treinado um único classificador
	para toda a hierarquia.
(Esuli et al., 2006)	Utiliza um algoritmo chamado <i>TreeBoost.MH</i> , que consiste de uma
	variação hierárquica do algoritmo AdaBoost.MH, membro da fa-
	mília de algoritmos boosting.
(Blockeel et al., 2006)	Compara métodos baseados em árvore de decisão. Compara um
	método que induz uma única árvore de decisão para todas as clas-
	ses com um método que induz uma árvore de decisão para cada
	classe.
(Cesa-Bianchi et al., 2006)	Treina um classificador incrementalmente para cada nó da hierar-
	quia de classes, baseado no método <i>Top-Down</i> .
(Alves et al., 2008)	Utiliza um sistema imunológico inteligente para a tarefa de clas-
	sificação. O algoritmo tem como objetivo encontrar um conjunto
	de regras que sejam tanto hierárquicas quanto multirrótulo.
(Vens et al., 2008)	Compara abordagens baseadas na noção de predictive clustering
(vens et al., 2000)	trees. Compara três abordagens de indução de árvores de deci-
	são. A primeira faz predições para todas as classes da hierarquia
	de uma única vez. A segunda define uma tarefa de classificação
	simples-rótulo independente para cada classe, ignorando relacio-
	namentos hierárquicos. A terceira induz uma árvore de decisão
	para cada classe, porém considera os relacionamentos hierárqui-
	cos.
(Valentini, 2009)	Utiliza combinação de classificadores. As saídas de classificado-
	res base, especializados em um única classe, são combinadas para
	obter uma classificação final.

Artigo	Abordagem Utilizada
(Clare e King, 2001)	Predição hierárquica com algoritmos de classificação não hierár-
	quicos
(Blockeel et al., 2002)	Abordagem One-Shot
(Clare e King, 2003)	Abordagem One-Shot
(Kiritchenko et al., 2004)	Abordagem Top-Down
(Struyf et al., 2005)	Abordagem One-Shot
(Barutcuoglu et al., 2006)	Utiliza classificadores binários para cada classe. Suas predições
	são combinadas com aprendizado Bayesiano
(Rousu et al., 2006)	Abordagem One-Shot
(Esuli et al., 2006)	Abordagem Top-Down
(Blockeel et al., 2006)	Abordagens Top-Down e One-Shot
(Cesa-Bianchi et al., 2006)	Abordagem Top-Down
(Alves et al., 2008)	Abordagem One-Shot
(Vens et al., 2008)	Predição hierárquica com algoritmos de classificação não hierár-
	quicos, Abordagens Top-Down e One-Shot
(Valentini, 2009)	Combinação de classificadores

Tabela 4.2: Abordagens para tratar problemas hierárquicos multirrótulo

Uma vez treinado um classificador hierárquico multirrótulo, este deve ser avaliado com relação a sua capacidade de predizer as classes de novos exemplos. A próxima seção apresenta algumas medidas de avaliação propostas na literatura.

4.3 Medidas de Avaliação

Apesar da maioria dos trabalhos apresentados na literatura avaliarem, de alguma maneira, o desempenho da classificação hierárquica multirrótulo, parece não haver ainda medidas bem definidas, ou adotadas por um grande número de trabalhos, para essa avaliação. Esta seção apresenta algumas medidas utilizadas na literatura especificamente para a avaliação de classificadores hierárquicos multirrótulo.

4.3.1 Medidas Baseadas nas Relações de Ancestralidade e Descendência

Medidas baseadas em ancestralidade e descendência consideram os ancestrais e os descendentes das classes preditas no momento dos cálculos da avaliação. Duas medidas baseadas nesse conceito foram utilizadas nesta pesquisa: Precisão e Revocação Hierárquicas e *Hierarchical Loss Function*.

Precisão Hierárquica e Revocação Hierárquica

No trabalho de Kiritchenko et al. (2004), foram propostas duas medidas de avaliação baseadas nas medidas convencionais de precisão e revocação, levando em consideração os relacionamentos hierárquicos entre as classes. Essas medidas foram chamadas de precisão

e revocação hierárquicas e levam em consideração classificações nos nós internos e nósfolha.

Cada exemplo pertence não apenas à sua classe, mas também a todos os ancestrais dessa classe na estrutura hierárquica. Dessa maneira, dado um exemplo qualquer (x_i, Y_i) , com x pertencente ao conjunto X de exemplos, Y_i o conjunto de classes preditas para o exemplo x_i , e Y_i' o conjunto de classes verdadeiras do exemplo x_i , os conjuntos Y_i e Y_i' podem ser estendidos para conterem suas correspondentes classes ancestrais da seguinte maneira: $\widehat{Y}_i = \bigcup_{y_k \in Y_i} Ancestrais(y_k)$ e $\widehat{Y}_i' = \bigcup_{y_l \in Y_i'} Ancestrais(y_l)$.

A precisão e revocação hierárquicas (hP e hR) são calculadas utilizando as Equações 4.4 e 4.5, respectivamente.

$$hP = \frac{\sum_{i} |\widehat{Y}_{i} \cap \widehat{Y}'_{i}|}{\sum_{i} |\widehat{Y}_{i}|}$$

$$(4.4)$$

$$hR = \frac{\sum_{i} |\widehat{Y}_{i} \cap \widehat{Y}_{i}'|}{\sum_{i} |\widehat{Y}_{i}'|} \tag{4.5}$$

Essas medidas contam o número de classes preditas corretamente, juntamente com o número de classes ancestrais dessas classes preditas corretamente, assumindo que exemplos também pertencem aos ancestrais de suas classes corretas (Kiritchenko et al., 2004). A Figura 4.3 apresenta um exemplo do cálculo dessas medidas, onde os círculos sólidos representam as classes reais de um exemplo e os círculos em negrito com uma seta representam as classes atribuídas ao exemplo pelo classificador.

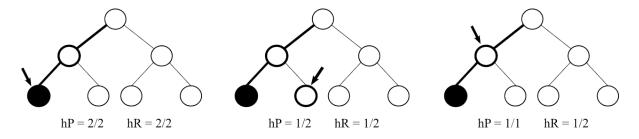


Figura 4.3: Representação gráfica das medidas precisão e revocação hierárquicas. (Kiritchenko et al., 2004)

Na figura, todos os nós do caminho da raiz até a classe predita para um exemplo também estão em negrito, pois as classes ancestrais também são atribuídas ao exemplo. O caminho da raiz até a classe verdadeira também é mostrado em negrito. Os valores de hP e hR são mostrados na figura para três diferentes predições.

A precisão e a revocação hierárquicas utilizadas sozinhas não são suficientes para a avaliação de classificadores. Sendo assim, as medidas hP e hR devem ser combinadas em uma extensão hierárquica da medida F-measure, chamada de F_{β} – Hierárquica, apresentada na Equação 4.6. Na Equação, β refere-se à importância atribuída aos valores de hP e hR. Quando aumenta-se o valor de β , aumenta-se o peso atribuído ao valor de hR, e quando diminui-se β , aumenta-se o peso atribuído ao valor de hP.

$$F_{\beta} - Hier\'{a}rquica = \frac{(\beta^2 + 1) \times hP \times hR}{\beta^2 \times hP + hR}$$
(4.6)

Hierarchical Loss Function

No trabalho de Cesa-Bianchi et al. (2006), foi proposta uma nova medida de avaliação chamada de Hierarchical Loss Function (H-Loss). Essa medida utiliza a noção intuitiva de que sempre que um erro de classificação é cometido em um nó da hierarquia de classes, não devem haver penalizações adicionais para erros cometidos na subárvore desse nó. Ou seja, se um erro é cometido em um nó i, erros adicionais cometidos na subárvore rotulada por i não são importantes. Por exemplo, se um classificador classificar um documento erroneamente como pertencendo à classe esportes, o classificador não deve ser penalizado novamente por classificar erroneamente o mesmo documento em uma subclasse futebol.

Os autores utilizaram a seguinte notação para a definição dessa medida: uma classificação multirrótulo atribuída a um dado exemplo x é qualquer subconjunto do conjunto $\{1,...,L\}$ formado por todas as classes, incluindo o conjunto vazio. Esse subconjunto pode ser representado por um vetor $v = (v_1,...,v_L) \in \{0,1\}^L$, onde uma classe i pertence ao subconjunto de classes do exemplo x se e somente se $v_i = 1$.

Antes de definir formalmente a função H-Loss, é preciso introduzir duas medidas utilizadas para medir a discrepância entre uma predição multirrótulo feita por um classificador $(\hat{y} = (\hat{y}_1, ..., \hat{y}_L))$, e o verdadeiro conjunto de classes de um exemplo $(v = (v_1, ..., v_L))$. A primeira delas é chamada de zero-one loss e pode ser vista na Equação 4.7. A segunda é chamada de symmetric difference loss, mostrada na Equação 4.8.

$$l_{0/1}(\widehat{y}, v) = \{ \exists i : \widehat{y}_i \neq v_i \}$$

$$(4.7)$$

$$l_{\Delta}(\widehat{y}, v) = \{\widehat{y}_1 \neq v_1\} + \dots + \{\widehat{y}_L \neq v_L\}$$
(4.8)

Essas duas equações são baseadas apenas no conjunto de classes $\{1, ..., L\}$, não levando em consideração a estrutura hierárquica envolvida. A função H-Loss que leva em consideração os relacionamentos hierárquicos é definida na Equação 4.9, na qual ANC(i) denota o conjunto de ancestrais do nó i.

$$l_H(\widehat{y}, v) = \sum_{i=1}^{L} \{ \widehat{y}_i \neq v_i \land \widehat{y}_j = v_j, j \in ANC(i) \}$$

$$(4.9)$$

Segundo os autores, essa função também pode ser definida da seguinte maneira: dada uma estrutura hierárquica G, essa estrutura pode ser considerada uma floresta, composta por árvores definidas sobre o conjunto de classes do problema. Uma classificação multirrótulo $v \in \{0,1\}^L$ respeita essa estrutura G se e somente se v for a união de um ou mais caminhos de G, onde cada caminho começa em uma raiz e necessariamente termina em uma folha. Assim, todos os caminhos em G, de uma raiz até uma folha, são examinados.

Sempre que um nó i é encontrado, tal que $\hat{y}_i \neq v_i$, o valor 1 é adicionado à função H-Loss, e todas as predições na subárvore enraizada no nó i são descartadas. Dada essa definição, pode-se dizer que $l_{0/1} \leq l_H \leq l_{\Delta}$. A Figura 4.4 apresenta a aplicação da função H-Loss e alguns conceitos relacionados.

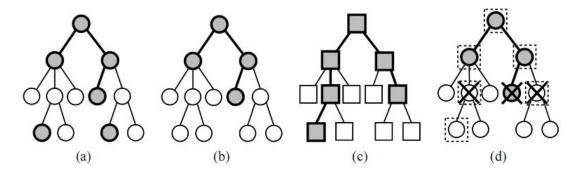


Figura 4.4: Representação gráfica da função H-Loss. (Cesa-Bianchi et al., 2006)

Na Figura 4.4, os nós de cor cinza representam a classificação multirrótulo sendo considerada. Na Figura 4.4.a, pode-se observar uma classificação que não respeita a estrutura de G, já na Figura 4.4.b essa estrutura é respeitada. A Figura 4.4.c apresenta uma classificação multirrótulo que respeita G, e é considerada aqui como a classificação verdadeira de um dado exemplo. Na Figura 4.4.d, pode-se observar a aplicação da função H-Loss, considerando as classificações multirrótulo (b) e (c). Somente os nós marcados com um X são considerados no cálculo da função. Portanto, o valor da função calculada entre as predições (b) e (c) é 3. Conforme definição, os valores das funções zero-one loss e symmetric difference loss entre as predições (b) e (c) são 1 e 4, respectivamente. Quanto menor for o valor da função H-Loss, melhor é o desempenho do classificador.

4.3.2 Medidas Baseadas em Distância

Como classes que estão mais perto umas das outras na hierarquia tendem a ser mais próximas entre si do que de outras classes, esse método considera a distância entre a classe verdadeira e a classe predita na hora de medir o desempenho do classificador.

Nesta pesquisa, foram utilizadas medidas propostas no trabalho de Sun e Lim (2001), chamadas Micro/Macro Precisão Hierárquica e Micro/Macro Revocação Hierárquica. Segundo os autores, as medidas Micro Precisão e Revocação atribuem igual importância a todos os exemplos, enquanto as medidas Macro Precisão e Revocação atribuem igual importância a todas as classes.

As medidas macro precisão/revocação hierárquicas primeiramente medem o desempenho obtido em cada classe do problema separadamente, e então obtém uma média desses desempenhos. As medidas micro precisão/revocação hierárquicas, por outro lado, medem o desempenho médio obtido em cada exemplo do conjunto de dados. Assim, as macro medidas são consideradas como a média do desempenho por classe, enquanto as micro medidas são consideradas a média do desempenho por exemplo do conjunto de dados

(Yang, 1999).

O cálculo das métricas é realizado computando-se, para cada classe, a contribuição dos exemplos erroneamente atribuídos àquela classe. Para o cálculo dessa contribuição, é necessário que se defina uma distância aceitável entre duas classes, dada por Dis_{θ} , que deve ser maior que 0. A distância Dis_{θ} é igual ao número de arestas que separam uma classe predita de uma classe verdadeira na estrutura hierárquica.

A Contribuição de um exemplo x_j a uma classe y_i é formalmente definida pelas Equações 4.10 e 4.11, nas quais, $x_j.agd$ and $x_j.lbd$ são respectivamente as classes preditas e verdadeiras do exemplo x_j .

• Se x_j é um Falso Positivo:

$$Con(x_j, y_i) = \sum_{y' \in x_i, lbd} (1.0 - \frac{Dis(y', y_i)}{Dis_{\theta}})$$
 (4.10)

• Se x_i é um Falso Negativo:

$$Con(x_j, y_i) = \sum_{y' \in x_j, aqd} (1.0 - \frac{Dis(y', y_i)}{Dis_{\theta}})$$
 (4.11)

Após seu cálculo, a contribuição de um exemplo x_j é então restringida à faixa de valores [-1.1]. Esse refinamento, denotado por $RCon(x_j, y_i)$, é definido na Equação 4.12.

$$RCon(x_j, y_i) = min(1, max(-1, Con(x_j, y_i)))$$

$$(4.12)$$

Para todos os exemplos, a contribuição total de falsos positivos $(FpCon_i)$ e falsos negativos $(FnCon_i)$ é definida nas Equações 4.13 e 4.14.

$$FpCon_i = \sum_{x_j \in FP_i} RCon(x_j, y_i)$$
(4.13)

$$FnCon_i = \sum_{x_j \in FN_i} RCon(x_j, y_i)$$
(4.14)

Após os cálculos das contribuições de cada exemplo, são calculados os valores da Precisão e da Revocação Hierárquicas de cada classe. Os cálculos são apresentados nas Equações 4.15 e 4.16

$$Pr_{i}^{CD} = \frac{max(0, |TP_{i}| + FpCon_{i} + FnCon_{i})}{|TP_{i}| + |FP_{i}| + FnCon_{i}}$$
(4.15)

$$Re_i^{CD} = \frac{max(0, |TP_i| + FpCon_i + FnCon_i)}{|TP_i| + |FN_i| + FpCon_i}$$

$$(4.16)$$

Os valores estendidos da Precisão e Revocação Hierárquicas (Micro Precisão e Revocação Hierárquicas) são apresentados nas Equações 4.17 e 4.18, em que m representa o número de classes do problema.

$$\hat{Pr}^{\mu CD} = \frac{\sum_{i=1}^{m} (max(0, |TP_i| + FpCon_i + FnCon_i))}{\sum_{i=1}^{m} (|TP_i| + |FP_i| + FnCon_i)}$$
(4.17)

$$\hat{R}e^{\mu CD} = \frac{\sum_{i=1}^{m} (max(0, |TP_i| + FpCon_i + FnCon_i))}{\sum_{i=1}^{m} (|TP_i| + |FN_i| + FpCon_i)}$$
(4.18)

De acordo com o valor escolhido para Dis_{θ} , as contribuições $FpCon_i$ e $FnCon_i$ podem ser negativas. Portanto, uma função max é aplicada ao numerador das equações 4.17 e 4.18 para fazer com que seus valores não sejam menores que 0. Como $FpCon_i \leq |FP_i|$, quando $|TP_i| + |FP_i| + FnCon_i \leq 0$, o numerador $max(0, |TP_i| + FpCon_i + FnCon_i) = 0$, e $\hat{Pr}^{\mu CD}$ pode ser considerado 0 nesse caso. A mesma regra é aplicada ao cálculo de $\hat{Re}^{\mu CD}$ (Sun e Lim, 2001).

Nesta pesquisa, considerou-se o valor de Dis_{θ} como igual a 2. Assim, quando um exemplo tem uma classe predita e uma verdadeira separadas por duas arestas na estrutura hierárquica, a contribuição desse exemplo à classe erroneamente predita é igual a 0. Dessa maneira, para o primeiro nível da hierarquia, os valores da Micro Precisão e Revocação Hierárquicas se igualam aos valores das medidas hP e hR (Precisão Hierárquica e Revocação Hierárquica), apresentadas anteriormente.

Além dos cálculos da Micro Precisão e Revocação Hierárquicas, pode-se ainda obter os valores estendidos da Macro Precisão e Revocação Hierárquicas, por meio das Equações 4.19 e 4.20. Nas equações, m refere-se ao número de classes envolvidas.

$$\hat{Pr}^{M_{CD}} = \frac{\sum_{i=1}^{m} Pr_i^{CD}}{m} \tag{4.19}$$

$$\hat{Re}^{M_{CD}} = \frac{\sum_{i=1}^{m} Re_i^{CD}}{m} \tag{4.20}$$

Assim como as medidas Precisão e Revocação Hierárquicas, utilizadas por Kiritchenko et al. (2004), as medidas Micro/Macro Precisão e Revocação Hierárquicas também podem ser combinadas na medida F_{β} – Hierárquica, calculada por meio da Equação 4.6.

Medidas baseadas em distância possuem a desvantagem de não considerarem o fato de que classificações nos níveis mais profundos da hierarquia são mais difíceis e levam a informações mais específicas do que classificações nos níveis mais altos. É possível contornar essa desvantagem, tornando mais altos os custos de classificações erradas em níveis mais elevados do que em níveis mais profundos da hierarquia.

A distância entre duas classes pode ser definida como uma função de dois fatores: o número de arestas entre a classe predita e a classe verdadeira, e a profundidade das classes verdadeira e predita na hierarquia de classes. Uma maneira de se calcular essa função é atribuir pesos a cada aresta da árvore que representa a hierarquia de classes. O erro de classificação associado à diferença entre a classe verdadeira e a classe predita é dado pela soma dos pesos de todas as arestas do caminho entre essas duas classes.

De maneira a implementar o princípio de que os custos de erros de classificação em

níveis mais elevados devem ser maiores do que em níveis mais profundos, os pesos das arestas dos níveis mais profundos tendem a ser menores do que nos níveis mais elevados. A fim de implementar esse princípio, foram utilizados, neste trabalho, valores de pesos utilizados no trabalho de Holden e Freitas (2006), que decrementa o valor dos pesos exponencialmente conforme aumenta a profundidade das arestas. Os valores utilizados, do primeiro ao quarto nível das hierarquias, foram 0.26, 0.13, 0.07 e 0.04.

Um problema com a atribuição de pesos às arestas acontece quando a árvore é muito desbalanceada, fazendo com que a profundidade da árvore varie significantemente para diferentes nós-folha. Quando isso ocorre, um erro de classificação envolvendo um par de classes predita e verdadeira perto do nó raiz recebe uma penalização menor do que um erro de classificação envolvendo um par de classes em níveis mais distantes da raiz.

Problemas com atribuições de pesos às arestas da hierarquia também já foram apontados no trabalho de Lord (2003). Segundo os autores, quando duas classes estão localizadas em diferentes subárvores da hierarquia e o único ponto de ligação entre elas é o nó raiz, o fato de uma classe estar em um nível mais profundo que outra não significa necessariamente que a classe do nível mais profundo represente informação mais significativa que a classe do nível mais elevado. Portanto, considerar apenas a profundidade, sem considerar as informações associadas às classes, pode ser um problema em alguns casos.

4.3.3 Medidas Baseadas em Similaridade

Medidas baseadas em similaridade utilizam conhecimento semântico entre as classes para definir como penalizar os erros de classificação. Intuitivamente, se um classificador A atribui erroneamente um exemplo a uma classe a, porém essa classe é similar às classes corretas do exemplo, a classificação atribuída pelo classificador A deve ser menos penalizada do que uma outra classificação atribuída por um classificador B, que atribui ao mesmo exemplo uma classe b que não é nem um pouco similar às classes verdadeiras do exemplo.

A similaridade entre duas classes y_i e y_k , denotada por $CS(y_i, y_k)$, foi definida no trabalho de Sun e Lim (2001) com a utilização da distância cosseno entre os dois vetores de características das classes. Os vetores de características de uma classe podem ser obtidos por meio do somatório dos vetores de características de todos os exemplos atribuídos àquela classe. Os vetores de características dos exemplos são aqueles utilizados pelos algoritmos de AM para a indução de classificadores. Assim, sendo t_n 's as características dos exemplos, com N igual ao número total de características, e w_n 's e v_n 's pesos atribuídos a essas características, os vetores de características de duas classes podem ser representados por:

- $y_i = \{w_1t_1, w_2t_2, \dots, w_N, t_N\};$
- $y_k = \{v_1t_1, v_2t_2, \dots, v_N, t_N\}.$

A similaridade entre duas classes pode então ser calculada por meio da Equação 4.21. Utilizando a similaridade entre as classes, pode-se definir uma similaridade média entre as classes, denotada por ACS, e definida na Equação 4.22, na qual m refere-se ao número de classes.

$$CS(y_i, y_k) = \frac{\sum_{n=1}^{N} (w_N \times v_N)}{\sqrt{\sum_{n=1}^{N} w_n^2 \times \sum_{n=1}^{N} v_n^2}}$$
(4.21)

$$ACS = \frac{2 \times \sum_{i=1}^{m} \sum_{k=i+1}^{m} CS(y_i, y_k)}{m \times (m-1)}$$
(4.22)

Baseado na similaridade entre as classes, é possível verificar quão similares são as classes verdadeiras $x_j.lbd$ e as classes atribuídas $x_j.agd$ erroneamente a um exemplo. Assim, é possível verificar quanto um exemplo x_j pode contribuir para uma classe y_i nos cálculos da precisão e revocação. Os cálculos dessas contribuições são apresentados nas Equações 4.23 e 4.24.

• Se x_j é um Falso Positivo

$$Con(x_j, y_i) = \frac{\sum_{y' \in x_j.lbd} (CS(y', y_i) - ACS)}{1 - ACS}$$

$$(4.23)$$

• Se x_i é um Falso Negativo:

$$Con(x_j, y_i) = \frac{\sum_{y' \in x_j.agd} (CS(y', y_i) - ACS)}{1 - ACS}$$
(4.24)

Assim como nas medidas baseadas em distância apresentadas anteriormente, a contribuição de um exemplo pode ser positiva ou negativa, dependendo de quão similares forem suas classes preditas e verdadeiras em comparação com a similaridade média das classes (ACS). Desse modo, também é necessário um refinamento das contribuições (Equação 4.12). A contribuição de todos os exemplos pertencentes ao conjunto dos FP e FN também é calculada (Equações 4.13 e 4.14).

Medidas estendidas de Precisão (Pr_i^{CS}) e Revocação (Re_i^{CS}) também podem ser obtidas como anteriormente (Equações 4.15 e 4.16), assim como as medidas Micro Precisão $(\hat{Pr}^{\mu CS})$ e Revocação $(\hat{Re}^{\mu CS})$ Hierárquicas (Equações 4.17 e 4.18) e Macro Precisão (\hat{Pr}^{MCS}) e Revocação (\hat{Re}^{MCS}) Hierárquicas (Equações 4.19 e 4.20).

As medidas Micro/Macro Precisão e Revocação Hierárquicas também podem ser combinadas na medida F_{β} – Hierárquica, calculada por meio da Equação 4.6, assim como calculado nas medidas anteriores.

As medidas baseadas em similaridade não foram utilizadas neste trabalho devido ao fato de não se ter conhecimento suficiente sobre os relacionamentos semânticos entre as classes, o que impossibilita que pesos adequados $(w_n$'s e v_n 's) sejam utilizados nos vetores de características. Além disso, as bases de dados utilizadas são biológicas, e já foram

construídas com relacionamentos semânticos entre as classes. Como conjuntos de dados biológicos são construídos com dados experimentais, classes mais similares tendem a ficar mais próximas na hierarquia.

4.3.4 Outras Medidas

Avaliação Baseada em Curvas Precisão-Revocação

No trabalho de Vens et al. (2008), é proposta uma medida de avaliação baseada na análise de curvas precisão-revocação (curvas PR). Uma curva PR reflete a precisão de um classificador como uma função de sua revocação.

Seja c um classificador que tenha como saída a probabilidade de um exemplo pertencer a uma dada classe y_i , ou seja, um número real entre 0 e 1, e que a pertinência do exemplo à classe y_i seja determinada de acordo com um certo limiar t. Assim, para valores de saída acima do limiar t o exemplo é atribuído à classe y_i , e para valores abaixo do limiar, o exemplo não é atribuído à classe.

Esse limiar t corresponde a um ponto no espaço PR. Variando o limiar, é obtida uma curva PR. A diminuição gradativa do valor de t de 1 até 0 faz com que o número de exemplos atribuídos a uma classe aumente gradativamente, aumentando o valor da medida revocação. A precisão pode aumentar ou diminuir, apesar de haver uma maior tendência para que seu valor diminua.

Para comparar o desempenho de diferentes classificadores, os autores utilizaram a área abaixo da curva PR (AUPRC). Quanto mais perto de 1 for o valor da AUPRC, melhor é o classificador. A vantagem da avaliação por meio das curvas PR é a possibilidade de se analisar vários valores de limiares t, e assim observar melhor a relação entre os valores de precisão e revocação. Além disso, a avaliação pode ser feita sem considerar um valor de limiar específico, o que resulta em uma melhor comparação entre vários classificadores.

De acordo com os autores, a utilização das medidas precisão e revocação torna a avaliação mais adequada para problemas de classificação hierárquica multirrótulo, porque, nesse tipo de problema, geralmente classes individuais possuem poucos exemplos positivos. Um exemplo disso pode ser extraído do problema de classificação funcional de genes. Nesse problema, apenas poucos genes possuem uma função particular, o que implica que, para a maioria das classes, o número de exemplos negativos supera o número de exemplos positivos. Nesse caso, é mais interessante saber o número de exemplos positivos para um conjunto de classes do que os negativos.

Curvas PR podem ser construídas individualmente para cada classe de um problema de classificação multirrótulo, fazendo positivos os exemplos pertencentes a uma dada classe e negativos os outros exemplos. Para combinar desempenhos individuais em cada classe de maneira a se obter o desempenho geral, dois métodos podem ser utilizados: área abaixo da curva PR média e área média abaixo das curvas PR. Eles são descritos a seguir.

Área abaixo da curva PR média

Nesse método, um problema multirrótulo é transformado em um problema binário da seguinte maneira: considere um classificador binário que receba como entrada uma dupla exemplo-classe e prediga se o exemplo pertence ou não à classe. Precisão pode então ser definida como a proporção de duplas preditas como positivas que são positivas e revocação é a proporção de duplas positivas que são preditas como positivas. Como citado anteriormente, um classificador que tem como saída um número real entre 0 e 1 pode ser utilizado como tal classificador, por meio do uso de um limiar. A variação desse limiar gera uma curva PR (Vens et al., 2008).

Dado um valor para o limiar, um ponto $(\overline{Prec}, \overline{Rev})$ no espaço PR é obtido, e pode ser calculado pelas Equações 4.25 e 4.26, em que i varia de 1 até o número de classes do problema.

$$\overline{Prec} = \frac{\sum_{i} V P_i}{\sum_{i} V P_i + \sum_{i} F P_i} \tag{4.25}$$

$$\overline{Rev} = \frac{\sum_{i} V P_i}{\sum_{i} V P_i + \sum_{i} F N_i}$$

$$(4.26)$$

Segundo os autores, \overline{Prec} corresponde à proporção de classes corretamente preditas como positivas que são realmente positivas, e \overline{Rev} corresponde à proporção de classes positivas que são preditas como positivas. Por meio da variação do limiar, é obtida uma curva PR média. A área abaixo dessa curva é denotada por $AU(\overline{PRC})$.

Área média abaixo das curvas PR

Esse método calcula a média ponderada das áreas abaixo das curvas PR individuais. O cálculo é mostrado na Equação 4.27.

$$\overline{AUPRC}_{w_1,\dots,w_{|Y|}} = \sum_{i} w_i \times AUPRC_i \tag{4.27}$$

Segundo os autores, há duas maneiras de se tratar esse método. A primeira delas inicia todos os pesos w_i com o valor 1/|Y|, onde Y é o conjunto de classes do problema. Essa medida foi denotada por \overline{AUPRC} . Uma segunda maneira é ponderar a contribuição de uma classe de acordo com sua frequência. Essa frequência é dada por $w_i = v_i / \sum_j v_j$, onde v_i é a frequência de uma classe y_i no conjunto de dados. A razão para isso é que em alguns domínios, classes mais frequêntes podem ser mais importantes. Os autores denotaram essa última medida por \overline{AUPRC}_w .

A avaliação por meio das curvas PR não foi utilizada neste trabalho devido ao fato de as técnicas implementadas não terem como saída valores reais, impossibilitando a utilização de limiares para construção das curvas. Além disso, as outras medidas utilizadas também são baseadas em precisão revocação, mostrando-se adequadas para os problemas hierárquicos multirrótulo.

4.4 Considerações Finais

Este capítulo apresentou os conceitos fundamentais de classificação hierárquica multirrótulo. Esse tipo de problema pode ser tratado como uma combinação dos problemas de classificação hierárquica com os problemas de classificação multirrótulo. Foi feita uma revisão bibliográfica dos principais trabalhos da literatura, e também apresentadas algumas medidas de avaliação específicas para problemas hierárquicos multirrótulo. O próximo capítulo apresenta as técnicas de classificação da literatura que foram utilizadas neste trabalho, e também as duas técnicas propostas nesta pesquisa.

Capítulo

5

Técnicas Propostas, Investigadas e Comparadas

Neste capítulo são apresentadas as técnicas propostas, investigadas e comparadas nesta pesquisa. Foram estudadas e utilizadas duas técnicas baseadas na abordagem *One-Shot*. A primeira delas, chamada Clus-HMC, utiliza o conceito de *Predictive Clustering Trees* (PCT), no qual são construídas Árvores de Decisão (AD) em que cada nó da árvore corresponde a um grupo de exemplos do conjunto de dados. A segunda técnica é uma extensão hierárquica multirrótulo do algoritmo de indução de AD C4.5, chamada C4.5H. Nessa extensão, o cálculo da entropia foi modificado para considerar várias classes ao mesmo tempo e também os relacionamentos hierárquicos entre elas.

Além das técnicas baseadas na abordagem *One-Shot*, foram investigadas e implementadas três técnicas baseadas na abordagem *Top-Down*. A primeira delas, chamada *Hierarchical Multilabel Classification with Binary-Relevance* (HMC-BR), utiliza classificadores binários, sendo um classificador associado a cada classe da hierarquia. A segunda técnica, chamada *Hierarchical Multilabel Classification with Label-Powerset* (HMC-LP), utiliza um método de combinação de classes, no qual todas as classes associadas a um exemplo são combinadas em uma única classe, e a terceira técnica, chamada *Hierarchical Multilabel Classification with Cross-Training* (HMC-CT), utiliza um método de decomposição de classes, criando vários problemas de classificação hierárquica simples-rótulo.

Nas técnicas da abordagem *Top-Down*, são utilizados cinco algoritmos de Aprendizado de Máquina (AM) como classificadores base. Os algoritmos utilizados foram o algoritmo de indução de AD C4.5 (Quinlan, 1993), Máquinas de Vetores de Suporte (SVMs) (Vapnik, 1999), o algoritmo de extração de regras de decisão Ripper (Cohen, 1995), Redes Bayesianas (BayesNet) (Friedman et al., 1997), e o algoritmo baseado em exemplos KNN

(Aha et al., 1991).

5.1 Utilização de PCT - Clus-HMC

Nessa técnica de classificação hierárquica multirrótulo, são construídas AD com a utilização de um *framework* chamado PCT (Blockeel et al., 1998). Neste *framework*, AD são vistas como uma hierarquia de grupos (*clusters*), na qual o nó raiz contém todos os exemplos de treinamento, e é recursivamente particionado em grupos menores, a medida que se percorre a AD em direção às folhas. As AD baseadas em PCT podem ser aplicadas tanto para a tarefa de agrupamento quanto classificação (Blockeel et al., 1998).

O procedimento utilizado para a construção de PCT é apresentado no Algoritmo 1. Seu funcionamento é similar a outros algoritmos de indução de AD como o Cart (Breiman et al., 1984) ou o C4.5. O procedimento PCT(I) recebe como entrada um conjunto de exemplos de treinamento I. O procedimento MelhorAtributo(I) (linha 1 do procedimento PCT(I)) é então chamado para buscar o melhor atributo a ser utilizado como regra em um nó da AD. O procedimento MelhorAtributo(I) é apresentado no Algoritmo 2.

```
Algoritmo 1: Procedimento para a Construção Predictive Clustering Trees (Vens et al., 2008)
```

Se um bom atributo t^* for encontrado pelo procedimento MelhorAtributo(I), então o procedimento PCT(I) cria um novo nó interno t^* e é chamado recursivamente para construir uma subárvore (grupo) utilizando a partição de exemplos P^* obtida pela atributo t^* nos exemplos de treinamento. Para a seleção do melhor atributo t^* , o procedimento MelhorAtributo(I) verifica quanto cada atributo contribui para a redução da variância dos exemplos (linha 4 do Algoritmo 2). A redução da variância de um conjunto de exemplos maximiza a homogeneidade do conjunto e melhora o desempenho preditivo da AD (Vens et al., 2008). Se nenhum atributo reduzir significantemente a variância dos exemplos, então é criado um nó-folha que é rotulado com um caso representativo, ou protótipo, dos exemplos.

Para a utilização de PCT para a tarefa de classificação hierárquica multirrótulo, o algoritmo Clus-HMC foi utilizado, no qual a variância e o protótipo de um grupo de

Algoritmo 2: Procedimento para Encontrar o Melhor Atributo em uma PCT (Vens et al., 2008)

```
procedimento MelhorAtributo(I)
Entrada: I
Saída: melhor atributo t^* e uma partição de exemplos P^*

1 (t^*, h^*, P^*) = (nenhum, 0, \emptyset)

2 para cada possível atributo t faça

3 P \leftarrow \text{partição obtida por } t \text{ em } I

4 h = Var(I) - \sum_{I_k \in P} \frac{|I_k|}{|I|} Var(I_k)

5 \text{se } (h > h^*) \wedge Aceitável(t, P) então

6 L(t^*, h^*, P^*) = (t, h, P)

7 retorna (t^*, P^*)
```

exemplos são definidos como nos trabalhos de Blockeel et al. (2002) e Blockeel et al. (2006).

Primeiramente, os rótulos dos exemplos são representados como vetores que contém componentes booleanos. Para um dado exemplo, o i-ésimo componente de seu vetor de classes recebe o valor 1 se o exemplo pertence à classe c_i , e 0 caso contrário. O vetor que contém a média aritmética, ou protótipo, de um conjunto de vetores V, denotado por \overline{v} , contém, como i-ésimo componente, a proporção de exemplos do conjunto que pertencem à classe c_i . A variância de um conjunto de exemplos Y é dada pela distância quadrática média entre cada vetor de classes v_j dos exemplos e o vetor de proporção de classes \overline{v} . Essa variância é apresentada na Equação 5.1. O cálculo do protótipo de V é apresentado na Equação 5.2

$$Var(Y) = \frac{\sum_{j} d(v_j, \overline{v})^2}{|Y|}$$
(5.1)

$$\overline{v} = \frac{\sum_{v_j \in V} v_j}{|V|} \tag{5.2}$$

Para considerar a profundidade das classes na hierarquia, ou seja, classes em níveis mais profundos representam informações mais específicas do que classes em níveis mais elevados, foi utilizada uma distância Euclidiana ponderada entre as classes. O cálculo dessa distância é apresentado na Equação 5.3, em que $v_{j,i}$ é o *i*-ésimo componente do vetor de classes v_j de um exemplo x_j , e os pesos w(c) decrescem conforme aumenta a profundidade das classes na hierarquia ($w(c) = w_0^{profundidade(c)}$, com $0 < w_0 < 1$) (Vens et al., 2008).

$$d(v_1, v_2) = \sqrt{\sum_i w(c_i) \times (v_{1,i} - v_{2,i})^2}$$
(5.3)

Para exemplificar o cálculo da distância entre dois exemplos, será considerado o exemplo de hierarquia de classes apresentado na Figura 5.1, e dois exemplos (x_1, Y_1) e (x_2, Y_2) ,

com $Y_1 = \{Política, Ciências, Ciências/Biologia\}$ e $Y_2 = \{Ciências\}$ as classes atribuídas a cada exemplo. O cálculo da distância pode ser feito considerando um vetor de classes que representa a pertinência das classes Política, Ciências, Ciências/Computação, Ciências/Biologia e Artes nessa ordem. Assim, os vetores de pertinência dos exemplos x_1 e x_2 são, respectivamente, $\{1, 1, 0, 1, 0\}$ e $\{0, 1, 0, 0, 0\}$. O cálculo da distância entre os vetores é apresentado na Equação 5.4. Apesar de "Política" ser uma ciência, apenas com o objetivo de exemplificação, "Política" e "Ciência" serão consideradas classes distintas.

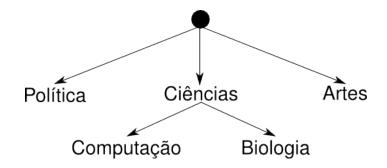


Figura 5.1: Exemplo Simples de Hierarquia de Classes

$$d(\{1, 1, 0, 1, 0\}, \{0, 1, 0, 0, 0\})$$

$$= \sqrt{w_0 \times (1 - 0)^2 + w_0 \times (1 - 1)^2 + w_0^2 \times (0 - 0)^2 + w_0^2 \times (1 - 0)^2 + w_0 \times (0 - 0)^2}$$

$$= \sqrt{w_0 + w_0^2}$$
(5.4)

A heurística utilizada para a escolha do melhor atributo para ser colocado em um nó da árvore é, como mostrado anteriormente, a maximização da redução da variância de um conjunto de exemplos. No caso de problemas simples-rótulo, essa heurística é idêntica à heurística utilizada em algoritmos de indução de árvores de regressão, como Cart, e equivalente ao índice Gini utilizado pelo Cart em problemas de classificação (Vens et al., 2008).

Diferente de uma AD comum, que armazena em um nó-folha a classe majoritária para aquela folha, em uma PCT um nó-folha armazena a média \overline{v} dos vetores de classes dos exemplos daquela folha, ou seja, a função que obtém o protótipo de um grupo de exemplos (linha 7 do Algoritmo 1) retorna o vetor \overline{v} . A Figura 5.2 apresenta uma árvore de classificação para o exemplo da Figura 5.1.

A proporção de exemplos em um nó-folha que pertencem a uma classe c_i é denotada por $\overline{v_i}$, e pode ser interpretada como a probabilidade de que, se um exemplo chegar até o nó-folha, seja atribuído à classe c_i . Quando um exemplo chega em um nó-folha percorrendo a AD, se o valor de $\overline{v_i}$ estiver acima de um dado limiar l_i , o exemplo é atribuído à classe c_i . Para garantir a integridade da estrutura hierárquica, ou seja, para garantir que quando uma classe for predita suas superclasses também sejam preditas, os valores de limiares devem ser escolhidos de maneira que $l_i \leq l_k$ sempre que $c_i \leq_h c_k$, ou seja, sempre que c_i

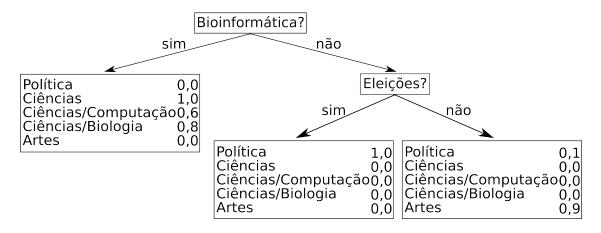


Figura 5.2: Exemplo de PCT

for uma superclasse de c_k .

Na Figura 5.2, os nós de decisão representam a presença ou ausência dos termos "Bioinformática" e "Eleições" em documentos sendo classificados. Assim, se o termo "Bioinformática" aparecer em um documento, existe uma probabilidade de 1.0 de o documento ser atribuído à classe "Ciências", 0.6 de o documento ser atribuído à classe "Ciências/Computação" e 0.8 de o documento ser atribuído à classe "Ciências/Biologia". Se o documento não possui o termo "Bioinformática", é feito um teste para verificar se o documento possui o termo "Eleições". Se sim, o documento é classificado na classe "Política" com uma probabilidade de 0.1. Se o termo "Eleições" não aparece no documento, o mesmo é classificado nas classes "Artes" com probabilidade igual a 0.9, e "Política" com probabilidade igual a 0.1.

Dependendo do contexto, os valores de limiares podem ser escolhidos de maneira a fazer com que o classificador tenha um alto valor para a medida Precisão, a um custo de um baixo valor para a medida Revocação, ou o contrário, um alto valor de Revocação e um baixo valor de Precisão. Neste trabalho, a fim de comparar os resultados do Clus-HMC com os obtidos por outras técnicas, não foram utilizados valores de limiares. Deste modo, qualquer classe que tenha um valor de $\overline{v_i}$ acima de 0 é atribuída a um exemplo.

A escolha da não utilização de valores de limiares foi baseada no fato de que quando a técnica Clus-HMC atribui um valor a uma posição $\overline{v_i}$, essa atribuição significa que o classificador atribuiu a um dado exemplo a classe que corresponde à posição $\overline{v_i}$. O valor atribuído indica apenas a probabilidade daquela classe ser atribuída ao exemplo. No trabalho de Vens et al. (2008), foram utilizadas curvas Precisão-Revocação para a avaliação dos classificadores baseados em PCT, por meio da utilização de vários valores de limiares. Esse tipo de avaliação não foi utilizada neste trabalho porque outras técnicas utilizadas não têm como saída valores reais, mas valores booleanos (0 e 1).

A função Aceitável(t, P) (linha 5 do Algoritmo 2) verifica, para um dado atributo t, se o número de exemplos em cada subconjunto da partição P é, no mínimo, igual a um dado parâmetro, e se a redução da variância do subconjunto de exemplos é significante de acordo com um teste estatístico F (teste-F) (Vens et al., 2008).

A linguagem de programação Java foi utilizada para a implementação da técnica Clus-HMC. O programa está livremente disponível para download em http://www.cs.kuleuven.be/~dtai/clus/. Para sua utilização, deve ser construído um arquivo de extensão .s, cujo nome deve ser o mesmo nome do conjunto de dados a ser utilizado. Nesse arquivo, são informados parâmetros como: o arquivo que contém o conjunto de treinamento (File), o arquivo que contém o conjunto de teste (TestSet), a porcentagem de exemplos a serem utilizados no procedimento de poda da árvore (PruneSet), o caractere utilizado para separar as classes de suas subclasses ou superclasses (HSeparator), e qual conjunto de dados deve ser utilizado para a obtenção das predições, que serão escritas em um arquivo de saída (WritePredictions). A Figura 5.3 apresenta um exemplo desse arquivo. Parâmetros que não são informados são considerados como valores padrão da técnica. Os arquivos que contém os conjuntos de treinamento e teste devem estar no formato .arff.

[Data]
File = Data.train.arff
PruneSet = 0.33
TestSet = Data.test.arff

[String]
HSeparator = .

[Output]
WritePredictions = {Test}

Figura 5.3: Exemplo de Arquivo de Configuração do Clus-HMC

5.2 Extensão do Algoritmo C4.5 - C4.5H

A técnica C4.5H é uma extensão hierárquica multirrótulo do conhecido algoritmo de indução de AD C4.5 (Quinlan, 1993), apresentado de maneira simples no Algoritmo 3. Essa extensão foi desenvolvida por Amanda Clare em sua tese de doutorado (Clare, 2003), e consistiu das seguintes modificações no algoritmo C4.5:

- Leitura e armazenamento da hierarquia de classes;
- Testes para verificar a pertinência de uma dada classe;
- Encontrar a melhor classe ou classes para representar um nó da AD;
- Cálculo da entropia.

Algoritmo 3: Algoritmo Básico C4.5

```
procedimento GeraÁrvore(X)
Entrada: Conjunto de dados X
Saída: Árvore

1 se CritérioParada(X) não for satisfeito então

2 para cada atributo a do conjunto de atributos A faça

3 Cálcula o ganho de informação para o atributo a em X

4 Obtém o atributo melhorA que tem o maior ganho de informação para cada Subconjunto X_i de X formado utilizando melhorA faça

5 para cada Subconjunto X_i de X formado utilizando melhorA faça

6 Arvore_i \leftarrow GeraÁrvore(X_i)

7 retorna (nó(melhorA, \cup_i \{Arvore_i\}))

8 senão

9 retorna nóFolha(X)
```

5.2.1 Leitura e Armazenamento da Hierarquia de Classes

Para a utilização da técnica C4.5H, a hierarquia de classes deve ser armazenada em um arquivo, juntamente com os nomes dos atributos e seus possíveis valores. A hierarquia de classes deve ser fornecida como entrada para o algoritmo por meio de um arquivo de extensão .classes, no qual as classes já estão estruturadas no formato de uma árvore. O arquivo deve utilizar espaços com a finalidade de mostrar os relacionamentos entre subclasses e superclasses por meio de indentações, sendo indentações para as subclasses maiores do que para as suas superclasses. A Figura 5.4 apresenta um exemplo desse arquivo.

```
Ciências
 Ciências.Computação
   Ciências.Computação.Inteligência Artificial
     Ciências. Computação. Inteligência Artificial. Aprendizado de Máquina
     Ciências.Computação.Inteligência Artificial.Mineração de Dados
   Ciências.Computação.Sistemas Distribuídos
 Ciências.Biologia
   Ciências.Biologia.Genética
     Ciências.Biologia.Genética.Proteômica
   Ciências.Biologia.Fisiologia
     Ciências.Biologia.Fisiologia.Fisiologia Animal
 Política.Comportamento Político
    Política.Comportamento Político.Estudos Eleitorais
    Política. Comportamento Político. Partidos Políticos
 Política. Sociologia Política
 Política. Teoria Política
    Política. Teoria Política. Teoria Política Clássica
    Política. Teoria Política. Teoria Política Moderna
```

Figura 5.4: Exemplo de Arquivo com Hierarquia de Classes para o C4.5H

As classes de cada exemplo são armazenadas da mesma maneira que na técnica Clus-HMC, ou seja, um vetor de classes para cada exemplo, no qual cada posição do vetor corresponde a uma classe da hierarquia.

Além do arquivo com a estrutura hierárquica, devem ser construídos mais três arquivos. Um arquivo, de extensão .names, deve conter todos os possíveis valores de classes e também os nomes de todos os atributos e seus valores. Os exemplos de treinamento devem ser armazenados em um arquivo de extensão .data e os exemplos de teste em um arquivo de extensão .test.

5.2.2 Testes para Verificar a Pertinência de uma Dada Classe

Como as classes atribuídas a um exemplo são representadas por um vetor, o teste para verificar se um dado exemplo pertence a uma dada classe consiste em verificar se a posição do vetor correspondente àquela classe possui um valor diferente de 0. Caso seja necessário verificar a pertinência envolvendo classes de um nível específico da hierarquia, basta que seja utilizada uma máscara que esconda as classes que pertençam a outros níveis.

5.2.3 Encontrar a Melhor Classe ou Classes para Representar um Nó da AD

De maneira similar à técnica Clus-HMC, os nós-folha da AD da técnica C4.5H são rotulados com um conjunto de classes. As classes que rotulam os nós-folha da árvore são, nessa técnica, as classes mais específicas do conjunto de dados. Essa foi uma escolha feita pelos autores, já que classes mais específicas fornecem informações mais interessantes e úteis sobre os exemplos que rotulam.

Quando a hierarquia de classes é percorrida de maneira Top-Down, uma classe nunca será mais frequente na hierarquia do que sua classe pai (Clare, 2003). Baseado nessa informação, os autores optaram por inicialmente encontrar o conjunto mais frequente de classes do primeiro nível da hierarquia. Dado esse conjunto de classes L e sua frequência F, cada classe em L que possui subclasses é especializada e substituída por uma ou mais de suas subclasses.

5.2.4 Cálculo da Entropia

Para lidar com o problema hierárquico multirrótulo, modificações foram feitas no cálculo da entropia de um conjunto de exemplos. No algoritmo C4.5, a escolha de um atributo para ser colocado em um nó da AD, ou seja, o atributo que melhor classifica um dado conjunto de exemplos, é feita considerando o ganho de informação. O ganho de informação é dado pela diferença entre a entropia de todo o conjunto de exemplos a ser classificado pelo atributo e a soma ponderada da entropia dos subconjuntos formados pelo particionamento do conjunto de exemplos de acordo com os valores do atributo. O cálculo do ganho de informação é apresentado na Equação 5.5, em que A é o atributo considerado, S é o conjunto de exemplos de treinamento considerado, e S_v é o subconjunto de S cujos exemplos têm valor v para o atributo A.

$$ganhoInformação = entropia(S) - \sum_{v \in A} \frac{|S_v|}{S} \times entropia(S_v)$$
 (5.5)

O cálculo da entropia, para o caso de problemas simples-rótulo, é apresentado na Equação 5.6, em que $p(c_i)$ é a probabilidade, ou frequência relativa, da classe c_i no conjunto de dados S.

$$entropia(S) = -\sum_{i=1}^{N} p(c_i) \log_2 p(c_i)$$
(5.6)

A entropia pode ser definida como a quantidade de incerteza de um conjunto de dados. Pode-se pensar em entropia como a quantidade de informação necessária para descrever um exemplo do conjunto de dados. Para problemas multirrótulo, pode-se definir entropia como a quantidade de bits necessária para descrever todas as classes a que um exemplo pertence (Clare, 2003). Para essa estimativa, os autores somaram o número de bits necessários para representar a pertinência ou não pertinência de cada classe. Supondo um caso geral, no qual há N classes e a pertinência de cada classe c_i tem probabilidade $p(c_i)$, o número total de bits necessários para descrever um exemplo é dado pela Equação 5.7, em que $p(c_i)$ é igual à probabilidade do exemplo pertencer à classe c_i , e $q(c_i) = 1 - p(c_i)$ é igual à probabilidade do exemplo pertencer à classe c_i .

$$entropia(S) = -\sum_{i=1}^{N} ((p(c_i)\log_2 p(c_i)) + (q(c_i)\log_2 q(c_i)))$$
 (5.7)

Para exemplificar o funcionamento da Equação 5.7, será considerado um problema que contém quatro classes $\{a,b,c,d\}$. Assim, um exemplo que pertence às classes b e d tem seu vetor de classes representado por 0101. A quantidade de bits desse vetor pode ser maior do que a realmente necessária para representar as pertinências das classes se for considerado, por exemplo, que todos os exemplos pertencem à classe b. Nesse caso, o segundo bit não é necessário, pois assume-se previamente que o exemplo pertence à classe b. Considerando agora que 75% dos exemplos pertençam à classe b, se tem o conhecimento de que um exemplo qualquer tem uma maior probabilidade de pertencer à classe b do que não pertencer. Nesse caso, pode-se utilizar a Equação 5.7 para o cálculo da quantidade mínima de bits necessária para representar a pertinência ou não pertinência da classe b:

$$-((p(b)\log_2 p(b)) + (q(b)\log_2 q(b)))$$

$$= -((0,75 \times \log_2 0,75) + (0,25 \times \log_2 0,25))$$

$$= 0,81.$$

O cálculo acima mostra que é necessário 81% de 1 bit para representar a informação necessária para se saber se um exemplo pertence ou não à classe b, ou seja, ao invés de se utilizar 1 bit por classe, o que é necessário é a quantidade de informação extra para

descrever a pertinência ou não de uma classe (Clare, 2003).

Com esse novo cálculo para a entropia, quando um conjunto de dados é particionado de acordo com um dado atributo, um novo ganho de informação pode ser calculado como a soma ponderada da entropia de cada partição, como calculado na Equação 5.5, com a diferença de que a soma ponderada agora significa que se um exemplo aparece duas ou mais vezes em um subconjunto, é porque ele pertence a duas ou mais classes, e é contado duas ou mais vezes.

Para considerar a estrutura hierárquica, mais uma modificação foi necessária no cálculo da entropia. Como a predição de subclasses de uma dada classe fornece mais informações do que a predição apenas da classe pai, essa informação pode ser incorporada no cálculo da entropia.

Supondo que um documento é classificado na classe "Ciência", e que essa classe possua duas subclasses, "Computação" e "Biologia". Pode-se dizer que uma predição que diz que o documento pertence apenas à classe "Ciência" é equivalente a uma predição que diz que o documento pertence à classe "Computação" ou à classe "Biologia", porém ser saber a qual delas. Portanto, pode-se dizer que predizer uma classe mais geral deve ter o mesmo custo de predizer todas as suas subclasses. Dessa maneira, os autores definiram uma nova fórmula para a entropia, apresentada na Equação 5.8.

$$entropia(S) = -\sum_{i=1}^{N} ((p(c_i)\log_2 p(c_i)) + (q(c_i)\log_2 q(c_i)) - \alpha(c_i)\log_2 tam \acute{A}rvore(c_i)) \quad (5.8)$$

Além dos mesmos componentes da Equação 5.7, essa nova definição de entropia tem componentes definidos como (Clare, 2003):

- $\alpha(c_i) = 0$, se $p(c_i) = 0$ (Constante definida pelo usuário. Por padrão seu valor é 1).
- $tam \acute{A}rvore(c_i) = 1 + o$ número de classes descendentes da classe c_i (1 é adicionado para representar a classe c_i).
- $(p(c_i)\log_2 p(c_i)) + (q(c_i)\log_2 q(c_i)) = \text{incerteza na escolha das classes.}$
- $\log_2 tam \acute{A}rvore(c_i)$ = incerteza na especificidade das classes, e representa também o tamanho da árvore que tem como raiz a classe em questão.

O valor de α permite que seja ponderada a parte da fórmula da entropia relativa à especificidade das classes. Seu valor padrão é 1, o que significa que a incerteza na escolha das classes e a incerteza na especificidade das classes tem o mesmo peso. O aumento do valor de α leva a uma predição de classes da hierarquia mais específicas, enquanto a diminuição de seu valor leva à predição de classes mais gerais. O valor de α pode ser definido como 0 se uma determinada classe nunca é predita, pois não há necessidade de se obter informações a respeito de sua hierarquia (Clare, 2003).

Assim como na técnica Clus-HMC, valores de limiares podem ser utilizados no processo de predição de classes, já que, na técnica C4.5H, a pertinência das classes é definida por valores reais. Porém, a fim de comparar a técnica C4.5H com as técnicas da abordagem *Top-Down*, não foram utilizados valores de limiares neste trabalho. Deste modo, qualquer classe que tenha um valor acima de 0 é atribuída a um exemplo.

A linguagem C foi utilizada para a implementação da técnica C4.5H. O programa pode ser obtido apenas por meio de uma solicitação aos autores, devido ao fato de o programa C4.5 original possuir restrições de distribuição. O programa utilizado neste trabalho foi modificado por Costa (2008) para que faça suas predições obrigatoriamente nos nós-folha da hierarquia, já que esse foi o tipo de classificação adotada nesta pesquisa.

5.3 Classificação Binária - HMC-BR

Nessa técnica de classificação hierárquica multirrótulo, chamada Hierarchical Multilabel Classification with Binary Relevance (HMC), são utilizados classificadores binários, seguindo a abordagem de classificação Top-Down. A técnica utiliza N classificadores, sendo N o número total de classes presentes na hierarquia de classes do problema. Cada classificador é associado a uma classe e treinado para resolver um problema de classificação binário. Quando um classificador é treinado, a classe associada a ele é considerada a classe positiva, e todas as outras classes do problema são consideradas negativas. Desse modo, cada classificador torna-se especializado em distinguir uma determinada classe de todas as outras. Esse método foi originalmente proposto para a tarefa de classificação multirrótulo não hierárquica, sendo chamado de um-contra-todos ou Binary-Relevance. A exemplificação do método um-contra-todos foi apresentada no Capítulo 3 e pode ser vista na Figura 3.3.

Em um problema hierárquico multirrótulo com N classes, o i-ésimo classificador da técnica HMC-BR considera os exemplos que pertencem a i-ésima classe como exemplos positivos para aquela classe, e os exemplos que não pertencem a i-ésima classe como exemplos negativos. Para a predição das classes de um novo exemplo, todas as classes para as quais os classificadores individuais apresentarem uma saída positiva são atribuídas a ele.

Supondo um problema de classificação que possui quatro níveis hierárquicos, sendo 2/3/4/5 o número de classes em cada um dos níveis da hierarquia. Como cada classificador é associado com uma classe da hierarquia, dois classificadores são treinados para as classes do primeiro nível, três classificadores são treinados no segundo nível, quatro classificadores no terceiro e cinco classificadores no quarto nível. No primeiro nível, o problema de classificação é transformado em dois problemas de classificação binários, sendo um para cada classe. O mesmo acontece nos outros níveis da hierarquia.

O processo de treinamento dos classificadores no primeiro nível ocorre da mesma forma que em um processo de classificação multirrótulo não hierárquica, utilizando o método um-contra-todos. Do segundo nível em diante, quando um classificador é treinado para uma dada classe c, o processo de treinamento é feito considerando apenas os exemplos que pertencem à classe pai da classe c. Esse processo é repetido até um nó-folha ser encontrado.

Quando o processo de treinamento é terminado, é obtida uma hierarquia de classificadores, e a classificação de novos exemplos é feita seguindo a abordagem Top-Down. Começando na primeira classe do primeiro nível, quando um exemplo é atribuído a uma classe c, o algoritmo de classificação é chamado recursivamente para submeter o exemplo a outros classificadores, a fim de que sejam preditas quais são as subclasses da classe c as quais o exemplo pertence. O processo é repetido para todas as classes da hierarquia, até um nó-folha ser encontrado. O Algoritmo 4 apresenta o processo de classificação da técnica HMC-BR.

```
Algoritmo 4: Processo de Classificação da Técnica HMC-BR
  procedimento HMC-BR(x, Y)
  Entrada: exemplo x, conjunto de Classes Y
  Saída: Classes
1 Classes \leftarrow \emptyset
2 para cada classe y_i do conjunto de classes Y faça
     se exemplo x predito como pertencente à classe y_i então
3
         se não for o último nível da hierarquia então
4
             Obtém conjunto Y_i de classes filhas da classe y_i
5
             Classes \leftarrow Classes \cup \{y_i\} \cup HMC-BR(x, Y_i)
6
7
             Classes \leftarrow Classes \cup \{y_i\}
8
9 retorna Classes
```

A técnica HMC-BR possui algumas desvantagens. Ela assume que todas as classes são independentes umas das outras, o que nem sempre é verdade. O fato de se ignorar possíveis correlações entre as classes do problema pode levar à obtenção de um classificador com pouca capacidade de generalização. Outra desvantagem é que a árvore de classificação gerada torna-se complexa, pois um classificador é gerado para cada classe da hierarquia. No caso de classificadores que geram regras, como é o caso do algoritmo C4.5 e do Ripper, é gerado um conjunto de regras para cada classe, tornando a interpretabilidade do modelo mais difícil do que nas técnicas Clus-HMC e C4.5H. O tempo de indução de todo o modelo de classificação também é alto, devido ao fato do grande número de classes e classificadores envolvidos. Por outro lado, o processo de classificação acontece de uma maneira mais fácil, pois a discriminação de classes nível a nível é um processo mais parecido com o que um humano faria. Ainda, o fato de cada classificação.

5.4 Combinação de Classes - HMC-LP

Essa técnica de classificação, proposta nesta pesquisa, e chamada *Hierarchical Multilabel Classification with Label-Powerset* (HMC-LP), utiliza um processo de combinação de classes com a finalidade de transformar o problema de classificação hierárquica multirrótulo em um problema de classificação hierárquica simples-rótulo e considerar as correlações entre as classes, a fim de contornar a desvantagem da técnica HMC-BR mencionada anteriormente.

A técnica HMC-LP é uma adaptação hierárquica de uma técnica multirrótulo não hierárquica chamada *Label-Powerset*, utilizada nos trabalhos de Tsoumakas e Vlahavas (2007) e Boutell et al. (2004). Um exemplo de funcionamento da técnica *Label-Powerset* pode ser visto na Figura 3.5, Capítulo 3. Para cada exemplo do conjunto de dados, a técnica proposta combina todas as classes atribuídas a um exemplo, em um nível específico, a fim de formar uma nova e única classe.

Dado um exemplo que pertence às classes A.D e A.E, e outro exemplo pertencente às classes B.F, B.G, C.H e C.I, sendo que A.D, A.E, B.F, B.G, C.H e C.I são estruturadas hierarquicamente em dois níveis, tal que $A \leq_h D$, $A \leq_h E$, $B \leq_h F$, $B \leq_h G$, $C \leq_h H$ e $C \leq_h I$, com A, B e C pertencentes ao primeiro nível da hierarquia e D, E, F, G, H e I pertencentes ao segundo nível, o resultado da combinação das classes dos dois exemplos seria uma nova estrutura hierárquica $C_A.C_{DE}$ e $C_{BC}.C_{FGHI}$ respectivamente. O processo de combinação das classes dos dois exemplos é ilustrado na Figura 5.5.

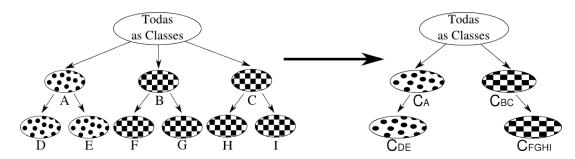


Figura 5.5: Processo de Combinação de Classes da Técnica HMC-LP

Como pode ser visto na Figura 5.5, após o processo de combinação de classes, o problema original é transformado em um problema hierárquico simples-rótulo. Durante as fases de treinamento e teste, a abordagem *Top-Down* é utilizada, usando um ou mais classificadores multiclasse por nível da hierarquia. Ao final do processo de classificação, as classes originais do problema são recuperadas. O procedimento de combinação de classes é apresentado no Algoritmo 5.

A técnica HMC-LP considera as possíveis correlações entre as classes, porém a combinação de classes pode aumentar consideravelmente a quantidade de classes envolvidas no problema. Isso acontece quando há muitas combinações multirrótulo possíveis no conjunto de dados, fazendo com que após a combinação das classes, as novas classes formadas possuam poucos exemplos positivos, fazendo com que o conjunto de dados de treinamento

Algoritmo 5: Processo de Combinação de Classes da Técnica HMC-LP

```
procedimento CombinaClasses(X, Y)
Entrada: conjunto de exemplos X, conjunto de classes Y
Saída: NovasClasses

1 para cada nivel\ j da hierarquia de classes faça

2 para cada subconjunto\ Y_i do conjunto\ Y, atribuido\ a\ um\ exemplo\ x_i\ no\ nivel\ j
faça

3 Obtém uma nova classe y_{i,j} para o exemplo x_i a partir de Y_i
4 NovasClasses_{i,j} \leftarrow y_{i,j}
5 retorna NovasClasses
```

torne-se esparso. O tempo de indução, no entanto, diminui consideravelmente se comparado com a técnica HMC-BR, pois menos classificadores precisam ser treinados, já que classificadores multiclasse são utilizados.

5.5 Decomposição de Classes - HMC-CT

A técnica de decomposição de classes, chamada *Hierarchical Multilabel Classification* with Cross Training (HMC-CT), utiliza um processo de decomposição de classes para modificar o problema hierárquico multirrótulo original, transformando-o em um conjunto de problemas hierárquicos simples-rótulo. Para essa transformação, todos os exemplos multirrótulo do conjunto de dados são decompostos em um conjunto de exemplos simples-rótulo.

No processo de decomposição da técnica, para cada exemplo, cada possível classe é considerada como a classe positiva em sequência, utilizando o conjunto de dados multirrótulo mais de uma vez durante o processo de treinamento. Exemplificando, se um conjunto de dados possui exemplos multirrótulo que pertencem às classes A, B e C, quando um classificador para a classe A é treinado, todos os exemplos multirrótulo que têm a classe A incluída em seu conjunto de classes, tornam-se exemplos simples-rótulo para a classe A, e o mesmo procedimento acontece com as classes B e C. A técnica foi originalmente proposta por Shen et al. (2004) para a tarefa de classificação multirrótulo não hierárquica, e foi chamada de Cross-Training (Capítulo 3, Figura 3.7).

Para essa nova variação hierárquica da técnica *Cross-Training*, o processo de decomposição de classes é aplicado a todos os níveis da hierarquia, e a abordagem *Top-Down* é utilizada durante as fases de treinamento e teste dos classificadores. A Figura 5.6 apresenta um exemplo do processo de decomposição de classes aplicado pela técnica HMC-CT em um conjunto de dados. Quando um exemplo pertence a mais de uma classe, essas classes são separadas por uma barra (/). O processo de classificação da técnica é apresentado no Algoritmo 6.

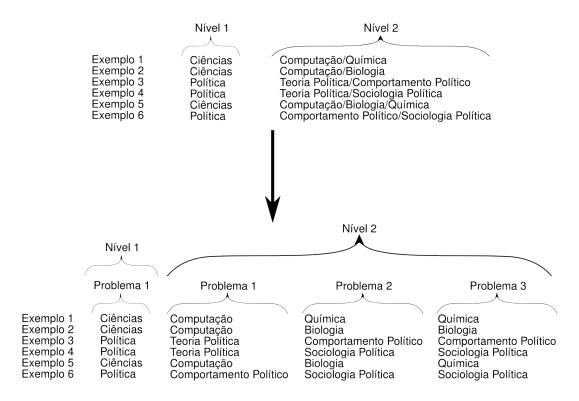


Figura 5.6: Processo de Decomposição de Classes da Técnica HMC-CT

```
Algoritmo 6: Processo de Classificação da Técnica HMC-CT
  procedimento HMC-CT(x, C)
  Entrada: exemplo x, classificadores C
  Saída: Classes
1 Classes \leftarrow \emptyset
2 para cada classificador c_i do conjunto de classificadores C faça
     Predição de uma classe y_i para o exemplo x usando o classificador c_i
     se não for o último nível da hierarquia então
4
         Obtém conjunto C_i de classificadores filhos do classificador c_i treinados com
5
         exemplos das classe y_i
         Classes \leftarrow Classes \cup \{y_i\} \cup HMC-CT(x, C_i)
6
7
      senão
         Classes \leftarrow Classes \cup y_i
8
9 retorna Classes
```

É importante que fique claro que a técnica HMC-CT é diferente da técnica HMC-BR. Na técnica HMC-CT, não é associado um classificador para cada classe, transformando o problema original em um problema binário. Ao invés disso, são utilizados classificadores multiclasse, pois todas as classes participam do processo de treinamento. A diferença é que, para um dado exemplo x, se esse exemplo pertence a duas classes A e B, o processo de treinamento ocorre duas vezes, uma vez considerando o exemplo como pertencente à classe B.

Um problema dessa técnica é o tempo de indução, que aumenta muito se comparado com as outras técnicas. Isso acontece porque o processo de treinamento ocorre várias

vezes, utilizando os mesmos exemplos várias vezes, considerando todas as classes as quais os exemplos pertencem.

5.6 Considerações Finais

Neste capítulo foram apresentadas as técnicas de classificação hierárquica multirrótulo utilizadas e também as técnicas propostas nesta pesquisa. Foram apresentadas duas técnicas de classificação baseadas na abordagem *One-Shot*, chamadas Clus-HMC, que constrói uma AD estruturada como uma hierarquia de grupos, e C4.5H, que é uma extensão hierárquica multirrótulo do algoritmo de indução de AD C4.5. Outras três técnicas, baseadas na abordagem *Top-Down*, também foram apresentadas. A primeira, chamada HMC-BR, utiliza classificadores binários. A segunda é uma técnica proposta neste trabalho, chamada HMC-LP, que utiliza um procedimento de combinação de classes. A terceira é chamada HMC-CT, e é uma técnica também proposta neste trabalho, que utiliza um processo de decomposição de classes para a tarefa de classificação. O próximo capítulo apresenta os experimentos realizados nesta pesquisa de mestrado.

Capítulo **6**

Experimentos

Neste capítulo, são apresentados os experimentos com as técnicas de classificação hierárquica multirrótulo apresentadas no Capítulo 5. Os experimentos foram realizados utilizando dez conjuntos de dados biológicos. As três técnicas de classificação que utilizam a abordagem *Top-Down* (HMC-BR, HMC-LP e HMC-CT) foram analisadas utilizando cinco algoritmos de classificação (C4.5, Ripper, BayesNet, SVM e KNN), e comparadas com as duas técnicas da abordagem *One-Shot* (C4.5H e Clus-HMC).

Este capítulo está organizado da seguinte maneira: a Seção 6.1 apresenta os conjuntos de dados utilizados nos experimentos; na Seção 6.2 é apresentada a metodologia utilizada durante os experimentos; os resultados são apresentados e discutidos na Seção 6.3, por fim, na Seção 6.5, são apresentadas algumas considerações finais com relação aos resultados obtidos.

6.1 Conjuntos de Dados

Os conjuntos de dados utilizados nos experimentos são dados biológicos do campo da genômica funcional relacionados ao organismo Saccharomyces cerevisiae ou Yeast (levedura). Eles foram primeiramente utilizados por Amanda Clare em sua tese de doutorado, e podem ser obtidos, em seu formato original, por meio do endereço http://www.aber.ac.uk/en/cs/research/cb/dss/yeastdata/. Nesta pesquisa, foram utilizadas as versões modificadas pelo grupo de pesquisa em Aprendizado de Máquina da Universidade Católica de Leuven, na Bélgica, que incluiu mais classes na estrutura hierárquica. Esses conjuntos modificados podem ser obtidos por meio do endereço http://www.cs.kuleuven.be/~dtai/clus/hmcdatasets/.

A hierarquia de classes dos conjuntos de dados está estruturada de acordo com o es-

quema do catálogo FunCat (http://mips.gsf.de/projects/funcat), desenvolvido pelo MIPS (Mewes et al., 2002). O esquema de anotação do FunCat provê descrições funcionais de proteínas de organismos eucariotos, procariotos, plantas e animais, consistindo de 28 categorias principais que cobrem campos como transporte celular, metabolismo e comunicação celular. A hierarquia é estruturada como um árvore, possuindo até seis níveis de profundidade e um total de 1632 classes funcionais. A Tabela 6.1 apresenta as categorias principais do primeiro nível da hierarquia estruturada de acordo com o esquema FunCat.

Tabela 6.1: Categorias Principais do Esquema FunCat (Ruepp et al., 2004)

Metabolismo	as do Esquema PunCat (Ruepp et al., 20
01	Metabolismo
02	Energia
04	Armazenamento
Vias Metabólicas	
10	Ciclo celular e processamento do DNA
11	Transcrição
12	Síntese de proteínas
14	Destino Proteico
16	Proteínas com funções atreladas
18	Regulação de atividades proteicas
Transporte	
20	Transporte celular
Percepção e resposta a estímulos	
30	Comunicação celular
32	Defesa celular
34	Interação com o ambiente celular
36	Interação com o ambiente (sistêmico)
38	Elementos Transponíveis, virais e plasmídeos
Processos de Desenvolvimento	
40	Destino celular
41	Desenvolvimento (sistêmico)
42	Biogêneses de componentes celulares
43	Diferenciação de tipos celulares
45	Diferenciação de tecidos
47	Diferenciação de organismos
Localização	
70	Localização subcelular
73	Localização de tipo celular
75	Localização de tecidos
77	Localização de órgãos
78	Expressão Ubiquitous
Proteínas não caracterizadas	
98	Classificação ainda não clara
99	Proteínas não classificadas

O Yeast é um fungo muito utilizado na fermentação de açúcar para a produção de etanol, e também na fermentação de trigo e cevada para a produção de bebidas alcoólicas.

Ele é um dos organismo modelo utilizados na Biologia, e tem sido utilizado em muitos estudos durante anos (Vens et al., 2008). Doze conjuntos de dados são disponibilizados pela Universidade Católica de Leuven, porém dez foram utilizados. Dois conjuntos de dados não foram utilizados porque possuíam uma quantidade muito grande de atributos, requerendo recursos de memória muito elevados, impossibilitando o uso nos experimentos, com os recursos computacionais disponíveis. A seguir são descritos os dez conjuntos de dados utilizados:

- Sequence estatísticas de sequências que dependem da sequência de aminoácidos da proteína codificada por um gene. Dentre essas características, estão o comprimento da sequência e o peso molecular. Alguns dos atributos foram obtidos com a utilização do programa ProtParam (Wilkins et al., 1999), outros foram obtidos do MIPS e outros foram calculados pelos autores. O conjunto de dados possui 478 atributos representando, em sua maioria, valores reais, com exceção de alguns poucos que possuem valores discretos (Clare, 2003);
- Phenotype contém dados relativos ao fenótipo do organismo, que representam o crescimento ou não crescimento de mutantes do organismo Yeast. O gene é removido ou inativado e o organismo resultante é cultivado utilizando diferentes características para determinar a quais delas o organismo é sensível ou resistente. Os autores coletaram os dados das fontes EUROFAN, MIPS e TRIPLES (Oliver, 1996; Mewes et al., 2002; Kumar et al., 2000). Os dados possuem 69 atributos de valores discretos (Clare, 2003);
- CellCycle dados de *microarray* de Spellman et al. (1998), consistindo de 77 atributos com valores reais obtidos a partir de quatro experimentos. Os dados originais foram preparados e utilizados por Amanda Clare (Clare, 2003);
- Church dados de *microarray* do laboratório *Church* obtidos originalmente de Roth et al. (1998). Os dados apresentam 27 atributos reais;
- **Derisi** dados de *microarray* obtidos de **DeRisi** et al. (1997), consistindo de 63 atributos reais, originalmente obtidos por Amanda Clare (Clare, 2003);
- **Eisen** dados de *microarray* consistindo de 79 atributos de valores reais. Os dados foram originalmente utilizados no trabalho de Eisen et al. (1998);
- Gasch1 dados de *microarray* consistindo de 173 atributos de valores reais, originalmente obtidos de Gasch et al. (2000);
- Gasch2 dados de *microarray* consistindo de 52 atributos de valores reais, originalmente obtidos de Gasch et al. (2001);
- SPO esses dados de *microarray* possuem 80 atributos, que em sua maioria são valores reais. Originalmente, os dados são provenientes do trabalho de Chu et al. (1998);

• Expr - esse conjunto de dados é formado pela concatenação de todos os conjuntos de dados de *microarray* descritos anteriormente (Clare, 2003). Os dados possuem 551 atributos que, em sua maioria, possuem valores reais.

A Tabela 6.2 apresenta as características dos conjuntos de dados descritos.

T 1 1 C Δ	\sim		1	a .		1	T 1	\circ
Tabela 6.7.	Car	acterísticas	dos	$(\Box \cap n)$	untag	de	Llados	()riginais
Tabela 0.2.	-care		uos	\sim	unios	uc	Dagos	Originals

Conjunto de Dados	Referência	Núm. Exemplos	Núm. Atributos
Expr	(Clare, 2003)	3788	551
CellCycle	(Spellman et al., 1998)	3766	77
Chuch	(Roth et al., 1998)	3764	27
Derisi	(DeRisi et al., 1997)	3733	63
Eisen	(Eisen et al., 1998)	2425	79
Gasch1	(Gasch et al., 2000)	3733	173
Gasch2	(Gasch et al., 2001)	3788	52
Phenotype	(Clare, 2003)	1592	69
Sequence	(Clare, 2003)	3932	478
SPO	(Chu et al., 1998)	3711	80

6.2 Metodologia

Nesta seção é descrita a metodologia adotada para a condução dos experimentos desta pesquisa. São fornecidas informações sobre a preparação dos dados para os experimentos, as ferramentas de software utilizadas, como foi conduzida a avaliação e os testes estatísticos empregados.

6.2.1 Preparação dos Dados

Durante a fase de preparação dos dados, foram feitos alguns ajustes nos conjuntos de dados. Esses ajustes consistiram na formatação adequada dos dados para serem fornecidos como entrada para os algoritmos de classificação, a imputação de valores ausentes nos conjuntos de dados, a retirada de alguns exemplos e a seleção de apenas algumas das classes da hierarquia de classes.

Formatação dos dados

Para que os dados pudessem ser utilizados pelas diferentes técnicas de classificação utilizadas, eles tiveram que ser formatados adequadamente. Nas técnicas C4.5H e Clus-HMC, os dados precisam ser colocados no formato .arff. Já nas técnicas da abordagem Top-Down, os dados foram formatados para ficaram no formato .csv.

Quando os dados são formatados no formato .arff, as classes atribuídas a um exemplo são colocadas na forma "1.2.4.6@3.7.9.10", que indica que o exemplo pertence a duas classes: "1.2.4.6" e "3.7.9.10". Ou seja, as classes são separadas por uma "@". Os níveis hierárquicos são separados por um ponto. Assim, a classe "1" é uma superclasse de "2", que por sua vez é superclasse de "4", que é superclasse de "6".

6.2 Metodologia 83

Nas técnicas da abordagem Top-Down, as classes de cada exemplo são representadas por um vetor binário, no qual cada posição corresponde a uma classe. A i-ésima posição corresponde a i-ésima classe da hierarquia e recebe o valor 1 se o exemplo pertence à classe e 0 caso contrário. A Figura 6.1 apresenta um exemplo desse vetor.

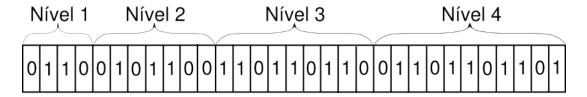


Figura 6.1: Exemplo de Vetor de Classes do Conjunto de Dados

Em todas as técnicas, as saídas são formatadas para que a representação das classes seja a mesma. Ao final do processo de classificação, cada exemplo tem suas classes preditas representadas em um vetor de classes, como o da Figura 6.1. Essa é a representação utilizada pelos algoritmos que fazem a avaliação das técnicas.

Imputação de Valores Ausentes

Muitos exemplos dos conjuntos de dados utilizados possuem alguns de seus valores ausentes. Sendo assim, métodos simples de imputação desses valores foram utilizados nesta pesquisa.

O método de imputação de valores ausentes escolhido assume que se dois exemplos são similares e um deles tem um valor ausente para algum atributo, há uma grande possibilidade de que o valor ausente em um exemplo seja similar ao valor do outro exemplo. A similaridade entre os exemplos foi calculada, neste trabalho, por meio da distância Euclidiana entre os exemplos. Assim, o método utilizado obtém os 50 exemplos mais similares ao exemplo que possui um valor faltando e os utiliza para obter o valor desse atributo.

Para obter as similaridades entre os exemplos, foi utilizada a função daisy (Kaufman e Rousseeuw, 1990) do pacote da linguagem R chamado cluster (Maechler et al., 2005). A função daisy calcula as dissimilaridades entre os exemplos com atributos de valores numéricos utilizando, por padrão, a distância Euclidiana entre os exemplos. Caso os exemplos possuam algum atributo categórico, é utilizada a distância de Gower, também conhecida como coeficiente de Gower (Gower, 1971). Com a utilização da distância de Gower, uma normalização específica é aplicada em cada atributo.

Para estimar valores ausentes de atributos numéricos, o valor de um atributo é obtido por meio do cálculo da mediana dos valores desse atributo nos 50 exemplos mais similares ao exemplo que possui o valor ausente. Para atributos categóricos, o valor ausente é substituído pelo valor mais frequente nos 50 exemplos mais similares ao exemplo que possui um valor ausente (moda).

Diminuição dos Conjuntos de Dados

Os conjuntos de dados originais possuíam uma quantidade de exemplos e classes muito grande, o que levou a um custo computacional muito elevado na realização dos experimentos. Além disso, problemas de memória foram verificados quando se utilizava os conjuntos de dados completos. Sendo assim, optou-se pela diminuição dos conjuntos de dados utilizados.

Para a diminuição do número de classes dos conjuntos de dados, foram escolhidas, aleatoriamente, quatro classes do primeiro nível da hierarquia de classes, e todos os conjuntos de dados foram limitados a possuírem no máximo quatro níveis hierárquicos. Assim, foram escolhidas, no primeiro nível, as classes "01", "02", "10" e "11" do esquema FunCat.

Além da redução do número de classes, também foram realizados procedimentos para garantir que, para cada classe, houvesse exemplos suficientes para a fase de treinamento dos classificadores. Qualquer classe que tinha menos do que 10 exemplos atribuídos a ela foi incorporada por sua classe pai. Outra modificação realizada nos conjuntos de dados foi a eliminação de exemplos que não eram classificados nos nós-folha da hierarquia. Isso foi feito devido ao tipo de predição escolhida para esta pesquisa, que é obrigatória em nós-folha. Assim, quando é feita uma predição, há a garantia de que todos os exemplos sejam classificados em nós-folha da hierarquia.

A Tabela 6.3 apresenta as características multirrótulo dos conjuntos de dados após os procedimentos descritos. A Tabela 6.4 apresenta a quantidade de classes nos quatro níveis da hierarquia.

Tabela 6.3: Características Multirrótulo dos Conjuntos de Dados

		J										
		1	N. Ex.	N. M.	N. Médio de Ex. por Classe				N. Médio de Classes por Ex.			
	N. Atributos	Total 1	Multirrótulo	Nível 1	Nível 2	Nível 3	Nível 4	Nível 1	Nível 2	Nível 3	Nível 4	
Expr	551	2444	1451	611.00	111.09	29.74	17.40	1.34	1.66	1.41	0.96	
CellCycle	77	2445	1451	611.25	111.13	29.74	17.40	1.34	1.66	1.42	0.95	
Church	27	2441	1449	610.25	110.95	29.70	17.39	1.34	1.66	1.40	0.93	
Derisi	63	2438	1449	609.50	110.81	29.65	17.30	1.35	1.66	1.42	0.94	
Eisen	79	1579	988	394.75	71.77	21.37	13.10	1.37	1.71	1.51	1.05	
Gasch1	173	2444	1450	611.00	111.09	29.74	17.40	1.34	1.65	1.40	0.93	
Gasch2	52	2454	1456	613.50	111.54	29.87	17.45	1.34	1.65	1.42	0.97	
Phenotype	69	1059	634	264.75	48.13	13.60	8.17	1.38	1.70	1.38	0.89	
Sequence	478	2480	1477	620.00	112.72	30.22	17.60	1.34	1.66	1.42	0.94	
SPO	80	2419	1439	604.75	109.95	29.45	17.21	1.35	1.66	1.40	0.93	

6.2 Metodologia 85

Tabela 6.4: Número de Classes por Nível

Conjunto de Dados	Número de Classes							
Conjunto de Dados	Nível 1	Nível 2	Nível 3	Nível 4				
Expr	4	22	70	84				
CellCycle	4	22	70	84				
Church	4	22	70	84				
Derisi	4	22	70	84				
Eisen	4	22	66	78				
Gasch1	4	22	70	84				
Gasch2	4	22	70	84				
Phenotype	4	22	66	76				
Sequence	4	22	70	84				
SPO	4	22	70	84				

6.2.2 Ferramentas de Software

Para a implementação das técnicas de classificação desta pesquisa, foi utilizada a linguagem e ambiente de desenvolvimento R (R Development Core Team, 2008). O R é uma importante ferramenta de análise estatística. Além disso, é um software livre, distribuído de acordo com os termos da Free Software Foundation e GNU General Public License, e pode ser instalado em uma variedade de plataformas UNIX, Windows e MacOS. Por possuir o código aberto, o R permite que desenvolvedores implementem suas próprias funcionalidades e as incluam na linguagem. As funcionalidades já existentes e o grande pacote de funções disponíveis fazem dessa ferramenta uma boa alternativa às ferramentas comerciais de AM.

O R possui muitos pacotes disponíveis que são voltados para a resolução de problemas de AM. Para a implementação das técnicas da abordagem *Top-Down* nesta pesquisa, foram utilizados dois pacotes, o e1071 (Dimitriadou et al., 2008) e o RWeka (Hornik et al., 2009). O pacote e1071 possui várias funcionalidades estatísticas, e também os códigos das técnicas de aprendizado estatístico SVM e *Naive Bayes*. Nesta pesquisa, o pacote e1071 foi utilizado para a geração dos classificadores SVM. O pacote RWeka é uma interface R para os métodos de AM da ferramenta Weka (Hall et al., 2009). Nesta pesquisa, o pacote RWeka foi utilizado para a geração dos classificadores Redes Bayesianas, Ripper, C4.5 e KNN.

6.2.3 Avaliação

A avaliação das técnicas de classificação foi feita nível a nível da hierarquia, ou seja, foi reportado um valor de desempenho preditivo para cada nível separadamente. Para essa avaliação, as classes reais e preditas de cada exemplo foram representadas como um vetor de classes como o da Figura 6.1. Assim, para a avaliação do desempenho das técnicas no primeiro nível hierárquico, basta obter a parte do vetor referente às classes do primeiro nível. Para a avaliação do desempenho no segundo nível, basta obter a parte do vetor correspondente às classes do primeiro e do segundo níveis. O mesmo procedimento é aplicado para a avaliação nos outros níveis da hierarquia.

Diferente de técnicas simples-rótulo, para a avaliação de classificadores multirrótulo, a representação de matrizes de confusão se torna inviável. Se um exemplo é erroneamente classificado, torna-se difícil o preenchimento da matriz de confusão, pois o erro pode ser registrado em todas as possíveis classes as quais o exemplo pode pertencer, o que distorceria os resultados, ou pode ser registrado em apenas uma das classes, o que faria com que alguns erros não fossem registrados (Clare, 2003). Assim, todas as medidas de avaliação utilizadas fazem a comparação dos vetores de classes reais e dos vetores de classes preditas no momento de computar o desempenho dos classificadores.

Para a comparação das técnicas da abordagem One-Shot com as técnicas da abordagem Top-Down, modificações foram feitas nos vetores de classes preditas obtidos com a utilização das técnicas da abordagem One-Shot. Como mencionado nas seções 5.1 e 5.2, a pertinência de um exemplo a uma determinada classe é representada por valores reais, onde cada valor representa a probabilidade de um exemplo x_j ser atribuído à classe c_i . A Figura 6.2 apresenta um exemplo de vetor de classes preditas obtido com a utilização das técnicas da abordagem One-Shot. No vetor da figura, um exemplo x_j tem uma probabilidade de 0.8 de pertencer à classe c_2 do vetor, uma probabilidade de 0.4 de pertencer à classe c_3 , e assim por diante. Para atribuir as classes do vetor a um exemplo, pode-se utilizar um valor de limiar. Exemplificando, se for escolhido um valor de limiar igual a 0.6, só são atribuídas a um exemplo classes que possuírem, em suas correspondentes posições do vetor, um valor maior ou igual a 0.5. No caso da Figura 6.2, somente as classes c_2 , c_4 , c_6 , c_{10} , c_{11} e c_{16} .



Figura 6.2: Exemplo de Vetor de Classes Obtido pelas Técnicas da Abordagem One-Shot

Neste trabalho, todos os valores reais obtidos pelas técnicas da abordagem *One-Shot* foram convertidos para o valor 1. Assim, nas predições obtidas pelas técnicas, qualquer classe que teve um valor maior que 0 atribuído à sua correspondente posição no vetor de classes preditas de um determinado exemplo foi atribuída ao exemplo. A escolha da não utilização de limiares foi devido ao fato de que os valores atribuídos às posições dos vetores de classes preditas indicam apenas as probabilidades das classes serem atribuídas aos exemplos, e não excluem nenhuma classe do conjunto de classes preditas. Além disso, não há uma decisão correta no momento de se decidir qual limiar utilizar para a classificação.

Suponha que, após o processo de divisão dos exemplos, e obtenção da árvore de decisão final das técnicas C4.5H e Clus-HMC, um determinado nó-folha possua 50 exemplos pertencentes a uma classe A e 50 exemplos pertencentes a uma classe B. Não existe decisão correta para a classificação de um novo exemplo que chegue a essa folha, pois ele pode pertencer a ambas as classes A e B, somente à classe A, ou somente à classe B.

6.2 Metodologia 87

Pode haver ainda o caso de um nó-folha possuir 97 exemplos da classe A e 3 exemplos da classe B. Nesse caso, os exemplos da classe B podem representar apenas ruído ou podem representar informação verdadeira e importante. Assim, não se sabe se a decisão correta é classificar um novo exemplo que chegue a esse nó-folha como pertencente a ambas as classes A e B, às classes A com probabildade 0.97 e B com probabilidade 0.03, ou somente à classe A. Devido a essas questões, optou-se pela não utilização de valores de limiares nesta pesquisa.

Diferentes medidas de avaliação foram utilizadas nesta pesquisa: medidas baseadas nas relações de ancestralidade e descendência (Precisão e Revocação Hierárquicas e *Hierarchical Loss Function*), e medidas baseadas em distância dependentes e não dependentes de profundidade (Micro/Macro Precisão Hierárquicas e Micro/Macro Revocação Hierárquicas).

A motivação para a escolha de diferentes medidas de avaliação está no fato de que todas as medidas fazem considerações diferentes no momento de avaliar os desempenhos dos classificadores. Assim, é possível observar como as diferentes medidas se comportam. É possível ver nos resultados que, dependendo da medida utilizada, os desempenhos dos classificadores variam.

Para considerar a profundidade da hierarquia no momento do cálculo do desempenho, foram utilizados pesos nas medidas baseadas em distância. Os pesos utilizados foram 0.26, 0.13, 0.07 e 0.04. Assim, arestas que conectam o nó raiz da hierarquia aos nós do primeiro nível recebem o valor 0.26. Arestas que conectam classes do primeiro nível com classes do segundo nível recebem o valor 0.13, e assim por diante. Esses pesos foram originalmente utilizados no trabalho de Holden e Freitas (2006).

6.2.4 Testes Estatísticos

Para a análise dos resultados obtidos nos experimentos, foram utilizados testes estatísticos. Testes estatísticos têm sido muito utilizados na comunidade de AM para validação de experimentos e comparação de resultados.

Inicialmente, pensou-se em utilizar o teste t de Student pareado, que consiste em verificar se a diferença média entre os desempenhos de dois algoritmos em vários conjuntos de dados é significantemente diferente de 0. Porém, o teste t mostra-se mais adequado quando são comparados dois classificadores, o que não é o caso dos experimentos desta pesquisa, em que são comparadas várias técnicas. Uma alternativa para a comparação de vários classificadores é a utilização da correção de Bonferroni, que divide o valor de α (nível de significância) pelo número de testes realizados, tornando seu valor mais restritivo e diminuindo a chance de erros no teste (Abdi, 2007). Assim, vários classificadores podem ser comparados dois a dois.

O teste t com a correção de Bonferroni foi utilizado nos experimentos preliminares com classificação multirrótulo realizados no início desta pesquisa. Porém, nos experimentos com classificação hierárquica multirrótulo, ele não foi utilizado devido a algumas

desvantagens. Dentre essas desvantagens, pode-se citar:

 A correção de Bonferroni faz com que o teste t torne-se muito conservador, fazendo com que possam não ser detectadas diferenças estatisticamente significantes onde elas realmente ocorrem.

- O teste t exige que os dados sejam distribuídos de acordo com uma distribuição normal, a não ser que muitos conjuntos de dados sejam utilizados (30 ou mais). Nesse caso, pode-se supor que a distribuição dos dados seja normal.
- O teste t, assim como o cálculo da média sobre os conjuntos de dados, é afetado por *outliers* que enfraquecem o teste estatístico aumentando o erro padrão estimado (Demšar, 2006).

Devido às desvantagens do teste t, optou-se pela utilização dos testes de Friedman (Friedman, 1937, 1940) e Nemenyi (Nemenyi, 1963)¹, pois esses são mais adequados para as comparações realizadas nesta pesquisa, envolvendo vários algoritmos e vários conjuntos de dados (Demšar, 2006).

O teste de Friedman é um teste não paramétrico que utiliza um *ranking* dos algoritmos para cada conjunto de dados separadamente. O algoritmo que obtiver o melhor desempenho em um conjunto de dados recebe a posição 1 do *ranking*, o segundo melhor recebe a posição 2 e assim por diante.

Seja r_i^j a posição no ranking do j-ésimo algoritmo, de um total de k algoritmos, no i-ésimo conjunto de dados, de um total de N conjuntos de dados. O teste de Friedman compara o ranking médio dos algoritmos, dado por $R_j = \frac{1}{N} \sum_i r_i^j$. Sob a hipótese nula, que considera que todos os algoritmos sendo comparados são equivalentes e têm seus valores de R_j iguais, a estatística de Friedman, dada pela Equação 6.1, é distribuída de acordo com χ_F^2 com k-1 graus de liberdade (Demšar, 2006).

$$\chi_F^2 = \frac{12N}{k(k+1)} \left[\sum_j R_j^2 - \frac{k(k+1)^2}{4} \right]$$
 (6.1)

Segundo Iman e Davenport (1980), a estatística χ_F^2 é muito conservadora. Os autores derivaram uma nova estatística distribuída de acordo com a distribuição F, com k-1 e (k-1)(N-1) graus de liberdade. Essa estatística é apresentada na Equação 6.2. Tabelas da distribuição F e seus valores críticos podem ser encontradas em vários livros de estatística (Demšar, 2006).

$$F_F = \frac{(N-1)\chi_F^2}{N(k-1) - \chi_F^2} \tag{6.2}$$

Para verificar se há diferença estatística entre um número k de algoritmos, o teste de Friedman verifica se os rankings médios dos algoritmos são significantemente diferentes

¹Os testes utilizados nesta pesquisa foram implementados por Thiago Covões, do Laboratório de Computação Bioinspirada do ICMC-USP de São Carlos.

do ranking médio R_j , sob a hipótese nula. Assim, após o cálculo do valor de F_F para k algoritmos e N conjuntos de dados, é obtido um valor crítico na tabela da distribuição F com k-1 e (k-1)(N-1) graus de liberdade, para um determinado valor de α . Esse valor é comparado com o valor de F_F , e se o valor obtido na tabela for menor, a hipótese nula é rejeitada.

Se a hipótese nula for rejeitada, ou seja, se for detectada diferença estatística entre os algoritmos, o teste de Nemenyi é aplicado para detectar entre quais algoritmos essa diferença ocorre. Assim, a diferença de desempenho entre dois algoritmos é estatisticamente significante se seus respectivos rankings médios diferirem, no mínimo, de um valor crítico CD. O cálculo desse valor é apresentado na Equação 6.3. Os valores de q_{α} são obtidos da tabela 6.5.

$$[H]CD = q_{\alpha} \sqrt{\frac{k(k+1)}{6N}} \tag{6.3}$$

Tabela 6.5: Valores Críticos para o Teste de Nemenyi Bicaudal (Demšar, 2006)

20000100 0.00	00 0110	roos pa				J	(-	5 01115011	, =000)
Num. Classificadores	2	3	4	5	6	7	8	9	10
$q_{0,05}$	1,960	2,343	2,569	2,728	2,850	2,949	3,031	3,102	3,164
$q_{0,10}$	1,645	2,052	2,291	2,459	2,589	2,693	2,780	2,855	2,920

6.3 Resultados Experimentais

Nesta seção são apresentados os resultados experimentais obtidos pelas técnicas investigadas nos dez conjuntos de dados utilizados. Os resultados são apresentados em tabelas,

em que cada tabela mostra a comparação das abordagens *Top-Down* e *One-Shot* em um nível específico das hierarquias de classes.

Em todas as técnicas da abordagem Top-Down, os parâmetros dos classificadores convencionais foram utilizados com seus valores padrões, exceto no algoritmo SVM, em que o kernel Gaussiano foi utilizado com os parâmetros c=100 e $\gamma=0.01$, e no algoritmo KNN, em que o número de vizinhos utilizado foi igual a 5. Esses valores foram escolhidos por já terem sido utilizados em outros trabalhos envolvendo esses algoritmos, apresentando bons resultados.

Como não é adequado avaliar classificadores considerando as medidas precisão e revocação separadamente, as análises foram feitas utilizando a medida F1 hierárquica apresentada no Capítulo 4, Equação 4.6, que combina as medidas de precisão e revocação. Nessa medida, o valor de β utilizado foi 1, o que faz com que as medidas de precisão e revocação tenham o mesmo peso no momento do cálculo. Os valores obtidos para essas duas medidas, individualmente, podem ser vistos no Apêndice B.

Os resultados das comparações das abordagens *Top-Down* e *One-Shot* são apresentados nas Tabelas 6.6 a 6.29. Os melhores resultados para cada conjunto de dados são mostrados em negrito, e os desvios padrão são mostrados entre parênteses. Os resultados em negrito apenas indicam que uma técnica obteve um melhor resultado para um conjunto de dados específico, e não indica que a técnica obteve o melhor desempenho geral. A comparação das técnicas foi feita considerando todos os conjuntos de dados, a fim de verificar quais técnicas obtiveram o melhor desempenho de forma geral.

6.3.1 Resultados no Primeiro Nível das Hierarquias

Pode-se observar na Tabela 6.6 que os melhores resultados, considerando a medida H-Loss, foram obtidos pelas técnicas HMC-BR e HMC-LP utilizando o algoritmo SVM. Apesar das diferenças de desempenho, praticamente não houve diferenças estatisticamente significantes entre as duas técnicas.

O menor número de erros cometidos pela técnica HMC-BR pode ser devido ao menor número de classes presentes no primeiro nível da hierarquia. Com apenas quatro classes nesse nível, o processo de classificação binária torna-se mais simples. Nas outras técnicas da abordagem *Top-Down*, as estratégias utilizadas para modificar o problema multirrótulo podem ter prejudicado a classificação. O processo de combinação de classes da técnica HMC-LP aumenta consideravelmente o número de classes, e o processo de decomposição de classes da técnica HMC-CT pode aumentar o número de classes preditas, aumentando a chance de erros.

Considerando a medida F1 hierárquica, observa-se que o melhor desempenho no primeiro nível foi obtido, na maioria dos conjuntos de dados, pela técnica HMC-CT, novamente devido a sua característica de predizer um maior número de classes. A medida F1 hierárquica combina as medidas precisão e revocação hierárquicas em uma única medida. Pode-se observar nas tabelas B.1 e B.5 do Apêdice B que, apesar dos menores valores

de precisão obtidos, se comparados com as outras técnicas da abordagem *Top-Down*, a técnica HMC-CT obteve os melhores valores para a medida revocação, o que contribuiu para seu melhor desempenho.

O melhor desempenho da técnica HMC-CT, na maioria dos conjuntos de dados, também é observado considerando as medidas micro/macro F1 hierárquicas e micro/macro F1 hierárquicas ponderadas. Novamente, se forem observados os valores de precisão e revocação das tabelas do Apêndice B, pode-se observar que, apesar de valores de precisão um pouco menores, os valores de revocação obtidos pela técnica HMC-CT são bem maiores que os obtidos pelas outras técnicas, o que faz com que a relação entre os valores de precisão e revocação resulte em maiores valores para a medida F1.

Na maioria dos conjuntos de dados, observa-se portanto que, no primeiro nível da hierarquia, considerando as medidas F1 hierárquicas, prevaleceu o desempenho da técnica HMC-CT utilizando os algoritmos SVM e Ripper. Além da predição de um maior número de classes, as características dos dados também pareceram favorecer a estratégia utilizada por essa técnica. Com apenas quatro classes no primeiro nível, um maior número médio de exemplos por classe e um menor número médio de classes por exemplo, a utilização do conjunto de dados mais de uma vez durante o treinamento pode fazer com que o conjunto de dados torne-se menos esparso. Essas características dos conjuntos de dados podem ser observadas nas Tabelas 6.3 e 6.4.

O fato da técnica HMC-CT ter cometido mais erros do que as outras técnicas, considerando a medida H-Loss, nada influi no desempenho da técnica considerando as outras medidas, já que as diferentes medidas fazem considerações diferentes. É importante observar que o fato de a técnica obter um maior valor para a medida revocação pode fazer com que mais erros sejam cometidos, pois quanto maior o número de classes preditas, maior a probabilidade de se cometer erros.

Tabela 6.6: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

H-Loss no Primeiro Nível da Hierarquia

Conjunto de Dados			n Top-Down		1	n One-Shot
Conjunto de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HM0
	KNN	1.22(0.03)	$1.24 \ (0.07)$	1.28 (0.03)		
	C4.5	$1.30 \ (0.04)$	$1.33 \ (0.06)$	$1.42 \ (0.06)$		
Expr	Ripper	$1.16 \ (0.06)$	1.25 (0.04)	$1.34 \ (0.03)$	1.40 (0.05)	1.29 (0.04
	BayesNet	$1.18 \ (0.06)$	$1.12 \ (0.07)$	$1.13 \ (0.08)$		
	SVM	1.16 (0.06)	1.18 (0.04)	1.45 (0.02)		
	KNN	1.20 (0.05)	1.22 (0.05)	1.25 (0.05)		
	C4.5	1.35 (0.06)	1.42(0.04)	1.54 (0.04)		
CellCycle	Ripper	1.24 (0.03)	1.29(0.05)	1.46 (0.06)	1.52 (0.06)	1.43 (0.02
	BayesNet	1.19(0.06)	1.19 (0.05)	1.21(0.03)		
	SVM	1.16 (0.06)	$1.13\ (0.07)$	1.16 (0.04)		
	KNN	1.38 (0.03)	1.43 (0.02)	1.45 (0.02)		
	C4.5	1.35(0.07)	1.44 (0.02)	1.52 (0.04)		
Church	Ripper	1.31 (0.04)	1.43 (0.02)	1.56 (0.06)	2.14 (0.08)	1.96 (0.1
	BayesNet	1.45 (0.02)	1.35 (0.02)	1.44 (0.04)	, ,	,
	SVM	$1.34\ (0.06)$	1.36(0.03)	1.44(0.05)		
	KNN	1.37 (0.03)	1.42 (0.04)	1.44 (0.03)		
	C4.5	1.35 (0.04)	1.52 (0.03)	1.63 (0.07)		
Derisi	Ripper	1.31 (0.02)	1.41 (0.01)	1.50 (0.05)	1.61 (0.05)	1.49 (0.04
	BayesNet	1.42 (0.04)	1.39 (0.10)	1.42 (0.05)		- (
	SVM	1.36 (0.05)	1.41 (0.04)	1.44 (0.05)		
	KNN	1.21 (0.06)	1.24 (0.04)	1.23 (0.07)		
	C4.5	1.38 (0.04)	1.37 (0.03)	1.49 (0.10)		
Eisen	Ripper	1.22 (0.04)	1.25 (0.04)	1.34 (0.09)	1.52 (0.04)	1.39 (0.04
	BayesNet	1.17 (0.07)	1.20 (0.06)	1.16 (0.04)	1.02 (0.01)	1.00 (0.0
	SVM	1.16 (0.06)	1.10 (0.04)	1.17 (0.09)		
	KNN	1.20 (0.06)	1.22 (0.03)			
	C4.5	1.28 (0.04)	1.22 (0.03) $1.33 (0.03)$	1.24 (0.04) $1.47 (0.04)$		
Gasch1	Ripper	1.28 (0.04) $1.21 (0.08)$	1.33 (0.03) $1.33 (0.03)$	1.47 (0.04) $1.36 (0.08)$	1.46 (0.05)	1.36 (0.03
Gasciii	BayesNet	1.21 (0.08) $1.25 (0.06)$	1.35 (0.05) $1.25 (0.05)$	1.28 (0.04)	1.40 (0.03)	1.50 (0.00
	SVM	1.06 (0.04)	1.03 (0.04)	1.28 (0.04) $1.09 (0.05)$		
		• • • • • • • • • • • • • • • • • • • •			<u> </u>	
	KNN	1.25 (0.04)	1.32 (0.03)	$1.30 \ (0.07)$		
Gasch2	C4.5	1.33 (0.02)	1.41 (0.04)	1.57 (0.06)	1 59 (0.05)	1 46 (0.0
Gasciiz	Ripper	1.25 (0.06) 1.31 (0.03)	1.32 (0.04)	1.50 (0.07)	1.53 (0.05)	1.46 (0.04
	BayesNet	` /	1.34 (0.04)	1.34 (0.04)		
	SVM	1.25 (0.05)	1.20 (0.02)	1.30 (0.03)	<u> </u>	
	KNN	1.38 (0.07)	1.44 (0.02)	1.54 (0.03)		
DI .	C4.5	1.39 (0.07)	1.39 (0.07)	1.59 (0.09)	0.01 (0.00)	0.00 (0.1
Phenotype	Ripper	1.35 (0.05)	1.37 (0.02)	$1.60 \ (0.03)$	2.31 (0.03)	2.08 (0.1
	BayesNet	1.63 (0.09)	1.74 (0.08)	1.76 (0.06)		
	SVM	1.32 (0.05)	1.40 (0.06)	1.57 (0.05)		
	KNN	1.36 (0.06)	1.38 (0.05)	1.45 (0.05)		
a	C4.5	1.32 (0.03)	1.37 (0.07)	1.45 (0.05)		4.05 (5.1
Sequence	Ripper	1.20 (0.04)	1.34 (0.04)	1.45 (0.03)	1.48 (0.04)	1.38 (0.06
	BayesNet	1.17 (0.03)	1.30 (0.04)	1.20 (0.04)		
	SVM	1.09 (0.06)	1.15 (0.06)	1.37 (0.04)		
	KNN	$1.40 \ (0.05)$	$1.44 \ (0.02)$	$1.44 \ (0.05)$		
	C4.5	1.39(0.04)	$1.47 \ (0.05)$	$1.63 \ (0.05)$		
SPO	Ripper	$1.27 \ (0.05)$	$1.33 \ (0.05)$	$1.41 \ (0.03)$	1.63 (0.04)	1.51 (0.06
	BayesNet	1.42 (0.06)	1.39 (0.07)	1.42 (0.04)		
	SVM	1.36(0.04)	1.38(0.05)	1.40(0.07)		

Tabela 6.7: Comparação das Abordagens Top-Downe One-Shot Utilizando a Medida F1

Hierárquica no Primeiro Nível da Hierarquia

Conjunto de Dados		0	n Top-Down		Abordagen	
- injunio de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	53.45 (0.7)	52.37(2.2)	$57.01 \ (0.5)$		
	C4.5	53.86 (1.6)	53.00(2.2)	59.94 (1.1)		
Expr	Ripper	$56.21\ (1.9)$	49.81 (1.7)	61.56 (1.0)	50.05 (1.5)	51.73 (1.6
	BayesNet	$61.92\ (1.9)$	60.44(2.3)	59.48(2.9)		
	SVM	49.66 (2.2)	51.28 (1.7)	56.50 (1.0)		
	KNN	54.99(2.0)	53.54(1.9)	58.00 (1.8)		
	C4.5	51.64(2.2)	50.00(1.7)	58.32(1.2)		
CellCycle	Ripper	$51.64\ (0.6)$	47.65(2.2)	59.08(2.1)	47.51 (1.7)	47.13 (1.6
	BayesNet	56.60(2.6)	53.64(2.0)	58.03(1.4)		
	SVM	57.45 (2.1)	58.50 (2.7)	61.85 (1.6)		
	KNN	44.76 (1.9)	43.94 (0.6)	53.80 (0.9)		
	C4.5	44.40(2.0)	45.06(0.4)	54.69(1.5)		
Church	Ripper	46.14 (1.6)	40.77(0.7)	54.74(2.0)	47.63 (1.2)	47.27 (1.2
	BayesNet	55.32(0.5)	44.48 (1.3)	49.16 (3.0)		
	SVM	44.36 (1.8)	45.56(1.2)	55.85 (1.6)		
	KNN	47.89 (1.7)	46.23 (1.2)	52.37 (1.4)		
	C4.5	40.59 (6.7)	46.11 (1.1)	55.33 (0.9)		
Derisi	Ripper	47.15 (1.5)	42.76 (0.5)	56.76 (1.6)	45.51 (1.0)	46.23 (2.3
	BayesNet	54.35 (1.5)	45.30 (3.5)	49.56 (2.0)		
	SVM	47.48 (1.4)	46.77 (1.5)	54.76 (1.9)		
	KNN	55.90 (2.1)	54.14 (1.3)	59.83 (2.2)		
Eisen	C4.5	52.53 (1.5)	52.13 (1.2)	58.35 (2.5)		
	Ripper	54.13 (2.1)	50.20 (1.9)	61.59 (1.9)	48.78 (1.7)	50.66 (1.5
	BayesNet	60.40 (1.9)	54.18 (2.5)	59.85 (1.7)	, ,	(
	SVM	58.17 (2.4)	59.85 (1.2)	$62.02 \ (2.9)$		
	KNN	55.80 (2.7)	54.63 (1.5)	59.03 (1.5)		
	C4.5	55.39 (1.7)	52.85 (1.5)	59.44 (1.6)		
Gasch1	Ripper	53.66 (2.4)	46.43 (1.0)	61.40 (1.7)	48.90 (1.9)	49.53 (1.6
	BayesNet	60.30 (1.1)	54.99 (1.2)	53.19 (1.7)		
	SVM	59.96 (2.1)	60.60 (1.7)	63.44 (2.1)		
	KNN	52.97 (1.4)	49.88 (1.2)	56.77 (2.2)		
	C4.5	49.20 (2.2)	49.80 (1.8)	57.21 (0.9)		
Gasch2	Ripper	49.92 (3.3)	45.61 (1.8)	57.99 (1.2)	46.82 (1.6)	47.42 (1.3
	BayesNet	55.65 (1.4)	48.01 (1.1)	55.31 (1.6)	(210)	(
	SVM	55.31 (1.4)	55.80 (1.0)	57.59 (0.7)		
	KNN	40.29 (3.9)	44.08 (1.3)	50.76 (2.6)		
	C4.5	40.69 (4.3)	46.16 (2.3)	53.70 (2.6)		
Phenotype	Ripper	44.08 (3.4)	44.55 (1.0)	54.11 (1.1)	48.90 (0.9)	47.43 (1.7
r nenety pe	BayesNet	46.69 (2.6)	44.21 (2.0)	47.95 (2.5)	10.00 (0.0)	11.10 (1.1
	SVM	46.06 (3.0)	45.69 (1.9)	54.23 (1.6)		
	KNN		49.58 (1.7)			
	C4.5	50.73 (2.6) 53.24 (0.9)	49.58 (1.7) $51.62 (2.4)$	53.67 (2.1) 59.22 (1.3)		
Sequence	Ripper	53.24 (0.9)	45.52 (2.4) $45.52 (1.4)$	59.73 (0.8)	48.31 (1.7)	49.11 (2.7
Sequence	BayesNet	61.02 (0.8)	54.21 (1.8)	57.04 (1.8)	10.01 (1.7)	40.11 (4.1
	SVM	45.37 (2.7)	52.69 (2.4)	59.05 (0.9)		
	KNN	46.05 (1.9)	44.94 (1.2) 47.78 (1.5)	52.43 (2.0)		
SDO	C4.5	47.14 (6.0)	` ′	55.02 (1.3)	4E 97 (1 0)	45 00 /1 0
SPO	C4.5 Ripper BayesNet	47.14 (6.0) 47.34 (2.9) 52.69 (1.7)	47.78 (1.3) 45.72 (1.8) 47.87 (1.8)	58.78 (1.2) 48.71 (1.8)	45.27 (1.0)	45.83 (1.9

Tabela 6.8: Comparação das Abordagens *Top-Down* e *One-Shot* Utilizando a Medida Micro F1 Hierárquica no Primeiro Nível da Hierarquia

Conjunto de Dados		_	n Top-Down		Abordagen	
conjunto de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	53.45 (0.7)	52.37(2.2)	$57.01 \ (0.5)$		
	C4.5	53.86 (1.6)	53.00(2.2)	59.94(1.1)		
Expr	Ripper	56.21 (1.9)	$49.81\ (1.7)$	61.56 (1.0)	50.05 (1.5)	51.73 (1.6
	BayesNet	$61.92\ (1.9)$	60.44 (2.3)	59.48 (2.9)		
	SVM	49.66 (2.2)	51.28 (1.7)	56.50 (1.0)		
	KNN	54.99(2.0)	53.54(1.9)	58.00(1.8)		
	C4.5	51.64(2.2)	50.00(1.7)	58.32(1.2)		
CellCycle	Ripper	$51.64 \ (0.6)$	47.65 (2.2)	59.08(2.1)	$47.51\ (1.7)$	47.13 (1.6
	BayesNet	56.60(2.6)	53.64(2.0)	58.03(1.4)		
	SVM	57.45 (2.1)	58.50 (2.7)	61.85 (1.6)		
	KNN	44.76(1.9)	43.94(0.6)	$53.80 \ (0.9)$		
	C4.5	44.40(2.0)	45.06 (0.4)	54.69(1.5)		
Church	Ripper	$46.14\ (1.6)$	40.77(0.7)	54.74(2.0)	$47.63\ (1.2)$	47.27 (1.2
	BayesNet	55.32 (0.5)	44.48 (1.3)	49.16(3.0)		
	SVM	44.36 (1.8)	45.56(1.2)	$55.85 \ (1.6)$		
	KNN	47.89 (1.7)	46.23 (1.2)	52.37 (1.4)		
	C4.5	40.59 (6.7)	$46.11\ (1.1)$	55.33(0.9)		
Derisi	Ripper	47.15(1.5)	42.76(0.5)	$56.76 \ (1.6)$	$45.51\ (1.0)$	46.23 (2.3
	BayesNet	54.35(1.5)	45.30(3.5)	49.56(2.0)		
	SVM	47.48 (1.4)	46.77(1.5)	54.76(1.9)		
	KNN	55.90 (2.1)	54.14 (1.3)	59.83 (2.2)		
Eisen	C4.5	52.53 (1.5)	52.13 (1.2)	58.35 (2.5)		
	Ripper	54.13(2.1)	50.20(1.9)	61.59(1.9)	48.78 (1.7)	50.66 (1.5
	BayesNet	60.40(1.9)	54.18(2.5)	59.85 (1.7)		
	SVM	58.17(2.4)	59.85(1.2)	$62.02\ (2.9)$		
	KNN	55.80 (2.7)	54.63 (1.5)	59.03 (1.5)		
	C4.5	55.39 (1.7)	52.85 (1.5)	59.44 (1.6)		
Gasch1	Ripper	53.66 (2.4)	46.43 (1.0)	61.40 (1.7)	48.90 (1.9)	49.53 (1.6
	BayesNet	60.30 (1.1)	54.99(1.2)	53.19 (1.7)		
	SVM	59.96(2.1)	60.60(1.7)	63.44 (2.1)		
	KNN	52.97 (1.4)	49.88 (1.2)	56.77 (2.2)		
	C4.5	49.20 (2.2)	49.80 (1.8)	57.21 (0.9)		
Gasch2	Ripper	49.92 (3.3)	45.61 (1.8)	57.99 (1.2)	46.82 (1.6)	47.42 (1.3
	BayesNet	55.65 (1.4)	48.01 (1.1)	55.31 (1.6)		
	SVM	55.31 (1.4)	55.80 (1.0)	57.59(0.7)		
	KNN	40.29 (3.9)	44.08 (1.3)	50.76 (2.6)		
	C4.5	40.69 (4.3)	46.16 (2.3)	53.70 (1.6)		
Phenotype	Ripper	44.08 (3.4)	44.55 (1.0)	54.11 (1.1)	48.90 (0.9)	47.43 (1.7
	BayesNet	46.69 (2.6)	44.21 (2.0)	47.95 (2.5)		
	SVM	46.06 (3.0)	45.69 (1.9)	54.23 (1.6)		
	KNN	50.73 (2.6)	49.58 (1.7)	53.67 (2.1)		
	C4.5	53.24 (0.9)	51.62 (2.4)	59.22 (1.3)		
Sequence	Ripper	53.07 (1.9)	45.52 (1.4)	59.73 (0.8)	48.31 (1.7)	49.11 (2.7
-	BayesNet	$61.02 \ (0.8)$	54.21 (1.8)	57.04 (1.8)	` ′	`
	SVM	45.37 (2.7)	52.69 (2.4)	59.05 (0.9)		
	KNN	46.05 (1.9)	44.94 (1.2)	52.43 (2.0)		
	C4.5	47.14 (6.0)	47.78 (1.5)	55.02 (1.3)		
SPO	Ripper	47.34 (2.9)	45.72 (1.8)	58.78 (1.2)	45.27 (1.0)	45.83 (1.9
	BayesNet	52.69 (1.7)	47.87 (1.8)	48.71 (1.8)	(1.0)	(1.0
	SVM	50.12 (1.6)	49.68 (1.9)	54.66 (1.5)		

Tabela 6.9: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Macro F1 Hierárquica no Primeiro Nível da Hierarquia

Conjunto de Dados		_	n Top-Down		Abordagen	
conjunio de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	51.90(1.2)	51.24(2.8)	55.48 (1.4)		
	C4.5	$51.40\ (1.9)$	$50.51\ (2.8)$	56.99 (0.9)		
Expr	Ripper	54.09(2.0)	48.35(2.2)	59.23 (0.9)	45.66 (1.6)	48.72 (0.7)
	BayesNet	61.03 (2.4)	59.05 (2.4)	59.19 (3.1)		
	SVM	38.47 (1.7)	39.99 (1.8)	45.33 (1.7)		
	KNN	52.28(2.8)	50.24(2.1)	55.22(2.7)		
	C4.5	48.79(3.0)	45.90(1.8)	$55.21\ (1.2)$		
CellCycle	Ripper	48.70(1.0)	45.08(2.9)	55.54(2.1)	43.50(1.9)	43.16 (1.2
	BayesNet	54.96(2.9)	51.27(1.9)	56.23 (1.5)		
	SVM	54.42 (1.8)	55.90 (3.1)	$58.88 \ (1.5)$		
	KNN	41.72 (2.5)	41.10 (2.1)	48.76 (2.1)		
	C4.5	40.61(2.3)	42.08(1.0)	49.58(2.3)		
Church	Ripper	41.52(3.6)	30.59(3.4)	47.23(7.1)	45.27(1.6)	44.30 (0.8
	BayesNet	$53.44\ (1.7)$	42.85 (0.5)	47.84(2.7)		
	SVM	40.85(1.4)	43.57(2.0)	51.94(2.5)		
	KNN	47.20 (1.5)	44.80 (1.5)	50.45 (0.9)		
	C4.5	39.96 (5.6)	43.76 (0.9)	52.24 (1.4)		
Derisi	Ripper	46.05 (1.7)	35.64 (1.7)	54.84 (1.1)	42.08 (1.3)	43.15 (3.3
	BayesNet	53.90 (1.1)	45.15 (3.4)	50.17 (1.8)		
	SVM	45.65 (1.8)	44.50 (1.2)	51.98 (2.0)		
	KNN	55.81 (2.2)	54.43 (1.0)	59.55 (1.6)		
Eisen	C4.5	49.66 (1.4)	48.89 (1.9)	54.11 (2.3)		
	Ripper	52.02 (3.0)	48.95 (1.5)	59.04 (1.2)	45.26 (2.5)	48.38 (1.5
	BayesNet	59.26 (2.5)	54.47 (2.9)	59.50 (1.3)	,	,
	SVM	56.19 (1.8)	58.19 (1.1)	59.15 (2.3)		
	KNN	55.20 (2.3)	54.07 (2.6)	57.71 (1.4)		
	C4.5	53.17 (1.5)	50.84 (2.1)	56.92 (1.6)		
Gasch1	Ripper	51.40 (2.0)	43.21 (1.7)	59.25 (1.3)	44.78 (1.5)	47.65 (2.4
	BayesNet	60.19 (1.3)	55.40 (0.8)	54.92 (1.4)		-1100 (-11
	SVM	57.90 (2.1)	57.48 (2.9)	61.27 (1.9)		
	KNN	49.75 (2.1)	47.40 (2.2)	53.62 (2.5)		
	C4.5	45.04 (2.1)	45.80 (2.2)	53.89 (0.6)		
Gasch2	Ripper	45.83 (3.3)	39.64 (3.7)	54.71 (0.9)	41.73 (1.5)	43.07 (1.1
Gasciiz	BayesNet	53.97 (1.8)	45.88 (1.8)	53.57 (1.7)	41.75 (1.5)	45.07 (1
	SVM	52.65 (0.9)	52.20 (2.0)	54.66 (0.4)		
				` ′		
	KNN	30.57(3.2)	35.69 (2.8)	42.60 (5.1)		
Phenotype	C4.5 Ripper	30.71 (3.4) 33.61 (2.3)	40.52 (4.1) 30.63 (8.3)	49.09 (2.3) 45.09 (1.3)	47.34 (0.8)	45.78 (2.3
1 henotype	BayesNet	45.50 (2.9)	44.53 (1.9)	47.28 (3.1)	47.34 (0.6)	40.10 (2.0
	SVM	39.59 (6.3)	38.30 (3.1)	48.86 (3.0)		
				` '		
	KNN	48.00 (2.1)	47.97 (2.2)	50.45 (1.4)		
C	C4.5	48.72 (1.4)	47.69 (2.8)	55.12 (1.6)	49.44.(9.0)	49.75 (0.5
Sequence	Ripper	47.20 (3.0)	40.85 (1.9)	54.65 (2.4)	43.44(2.0)	43.75 (2.3
	BayesNet	57.90 (1.3)	51.93 (2.0)	52.79 (2.2)		
	SVM	40.09 (2.3)	49.30 (2.4)	61.39 (1.7)		
	KNN	41.98 (1.6)	40.45 (1.1)	47.33 (1.7)		
	C4.5	43.73 (5.1)	43.73 (1.6)	51.39 (1.3)		
SPO	Ripper	43.70 (3.1)	42.23 (2.9)	55.80 (2.6)	40.28 (1.2)	40.77 (1.5)
	BayesNet	51.31 (2.3)	47.40 (2.6)	49.32 (2.9)		
	SVM	44.53(2.1)	45.23(2.3)	49.62(1.6)		

Tabela 6.10: Comparação das Abordagens *Top-Down* e *One-Shot* Utilizando a Medida Micro F1 Hierárquica Ponderada no Primeiro Nível da Hierarquia

Contractor do Dodos		Abordager	n Top-Down		Abordagen	n One-Shot
Conjunto de Dados		HMC-BR	HMC-LP	$\operatorname{HMC-CT}$	C4.5H	Clus-HM0
	KNN	53.45 (0.7)	52.37 (2.2)	57.01 (0.5)		
	C4.5	53.86(1.6)	53.00(2.2)	59.94 (1.1)		
Expr	Ripper	$56.21\ (1.9)$	49.81(1.7)	61.56 (1.0)	50.05(1.5)	51.73 (1.0
	BayesNet	$61.92\ (1.9)$	60.44(2.3)	59.48(2.9)		
	SVM	49.66 (2.2)	51.28 (1.7)	56.50 (1.0)		
	KNN	54.99 (2.0)	53.54 (1.9)	58.00 (1.8)		
	C4.5	51.64(2.2)	50.00(1.7)	58.32(1.2)		
CellCycle	Ripper	51.64 (0.6)	47.65(2.2)	59.08(2.1)	$47.51\ (1.7)$	47.13 (1.
	BayesNet	56.60(2.6)	53.64(2.0)	58.03(1.4)		
	SVM	57.45 (2.1)	58.50 (2.7)	61.85 (1.6)		
	KNN	44.76 (1.9)	43.94 (0.6)	53.80 (0.9)		
	C4.5	44.40(2.0)	45.06(0.4)	54.69(1.5)		
Church	Ripper	46.14(1.6)	40.77(0.7)	54.74(2.0)	47.63(1.2)	47.27 (1.
	BayesNet	55.32(0.5)	44.48(1.3)	49.16(3.0)		
	SVM	44.36 (1.8)	45.56(1.2)	55.85 (1.6)		
	KNN	47.89 (1.7)	46.23 (1.2)	52.37 (1.4)		
Derisi	C4.5	40.59 (6.7)	46.11 (1.1)	55.33 (0.9)		
	Ripper	47.15(1.5)	42.76(0.5)	56.76 (1.6)	45.51 (1.0)	46.23 (2.
	BayesNet	54.35(1.5)	45.30(3.5)	49.56(2.0)		
	SVM	47.48(1.4)	46.77(1.5)	54.76(1.9)		
	KNN	55.90 (2.1)	54.14 (1.3)	59.83 (2.2)		
Eisen	C4.5	52.53 (1.5)	52.13 (1.2)	58.35 (2.5)		
	Ripper	54.13 (2.1)	50.20 (1.9)	61.59 (1.9)	48.78 (1.7)	50.66 (1.
	BayesNet	60.40 (1.9)	54.18 (2.5)	59.85 (1.7)	, ,	
	SVM	58.17 (2.4)	59.85 (1.2)	$62.02\ (2.9)$		
	KNN	55.80 (2.7)	54.63 (1.5)	59.03 (1.5)		
	C4.5	55.39 (1.7)	52.85 (1.5)	59.44 (1.6)		
Gasch1	Ripper	53.66 (2.4)	46.43 (1.0)	61.40 (1.7)	48.90 (1.9)	49.53 (1.
	BayesNet	60.30 (1.1)	54.99 (1.2)	53.19 (1.7)	, ,	•
	SVM	59.96 (2.1)	60.60 (1.7)	63.44 (2.1)		
	KNN	52.97 (1.4)	49.88 (1.2)	56.77 (2.2)		
	C4.5	49.20 (2.2)	49.80 (1.8)	57.21 (0.9)		
Gasch2	Ripper	49.92 (3.3)	45.61 (1.8)	57.99 (1.2)	46.82 (1.6)	47.42 (1.
	BayesNet	55.65 (1.4)	48.01 (1.1)	55.31 (1.6)		
	SVM	55.31 (1.4)	55.80 (1.0)	57.59 (0.7)		
	KNN	40.29 (3.9)	44.08 (1.3)	50.76 (2.6)		
	C4.5	40.69 (4.3)	46.16 (2.3)	53.70 (1.6)		
Phenotype	Ripper	44.08 (3.4)	44.55 (1.0)	54.11 (1.1)	48.90 (0.9)	47.43 (1.
	BayesNet	46.69 (2.6)	44.21 (2.0)	47.95 (2.5)	, ,	•
	SVM	46.06 (3.0)	45.69 (1.9)	54.23 (1.6)		
	KNN	50.73 (2.6)	49.58 (1.7)	53.67 (2.1)		
	C4.5	53.24 (0.9)	51.62 (2.4)	59.22 (1.3)		
Sequence	Ripper	53.07 (1.9)	45.52 (1.4)	59.73 (0.8)	48.31 (1.7)	49.11 (2.
-	BayesNet	$61.02 \ (0.8)$	54.21 (1.8)	57.04 (1.8)	` ′	`
	SVM	45.37 (2.7)	52.69 (2.4)	59.05 (0.9)		
	KNN	46.05 (1.9)	44.94 (1.2)	52.43 (2.0)		
	C4.5	47.14 (6.0)	47.78 (1.5)	55.02 (1.3)		
SPO	Ripper	47.34 (2.9)	45.72 (1.8)	58.78 (1.2)	45.27 (1.0)	45.83 (1.
	BayesNet	52.69 (1.7)	47.87 (1.8)	48.71 (1.8)	(-)	(
	SVM	50.12 (1.6)	49.68 (1.9)	54.66 (1.5)		

Tabela 6.11: Comparação das Abordagens Top-Down e One-Shot Utilizando a Medida

Macro F1 Hierárquica Ponderada no Primeiro Nível da Hierarquia

Conjunto de Dados		_	n Top-Down		Abordagen	
		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	51.90(1.2)	51.24(2.8)	55.48 (1.4)		
	C4.5	$51.40\ (1.9)$	$50.51\ (2.8)$	56.99 (0.9)		
Expr	Ripper	54.09(2.0)	48.35(2.2)	59.23 (0.9)	45.66 (1.6)	48.72 (0.7)
	BayesNet	$61.03\ (2.4)$	59.05(2.4)	59.19 (3.1)		
	SVM	38.47 (1.7)	39.99 (1.8)	45.33 (1.7)		
	KNN	52.28(2.8)	50.24(2.1)	55.22(2.7)		
	C4.5	48.79(3.0)	45.90(1.8)	$55.21\ (1.2)$		
CellCycle	Ripper	48.70(1.0)	45.08(2.9)	55.54(2.1)	43.5(1.9)	43.16 (1.2
	BayesNet	54.96(2.9)	51.27(1.9)	56.23(1.5)		
	SVM	54.42 (1.8)	55.90 (3.1)	58.88 (1.5)		
	KNN	41.72 (2.5)	41.10 (2.1)	48.76 (2.1)		
	C4.5	40.61(2.3)	42.08 (1.0)	49.58(2.3)		
Church	Ripper	41.52(3.6)	30.59 (3.4)	47.23 (7.1)	45.27 (1.6)	44.30 (0.8
	BayesNet	53.44(1.7)	42.85(0.5)	47.84 (2.7)		
	SVM	40.85(1.4)	43.57(2.0)	51.94(2.5)		
	KNN	47.20 (1.5)	44.80 (1.5)	50.45 (0.9)		
Derisi	C4.5	39.96 (5.6)	43.76 (0.9)	52.24 (1.4)		
	Ripper	46.05 (1.7)	35.64 (1.7)	54.84 (1.1)	42.08 (1.3)	43.15 (3.3
	BayesNet	53.90 (1.1)	45.15 (3.4)	50.17 (1.8)	, ,	`
	SVM	45.65 (1.8)	44.50 (1.2)	51.98 (2.0)		
	KNN	55.81 (2.2)	54.43 (1.0)	59.55 (1.6)		
Eisen	C4.5	49.66 (1.4)	48.89 (1.9)	54.11 (2.3)		
	Ripper	52.02 (3.0)	48.95 (1.5)	59.04 (1.2)	45.26 (2.5)	48.38 (1.5
	BayesNet	59.26 (2.5)	54.47 (2.9)	59.50 (1.3)	20120 (210)	
	SVM	56.19 (1.8)	58.19 (1.1)	59.15 (2.3)		
	KNN	55.20 (2.3)	54.07 (2.6)	57.71 (1.4)		
	C4.5	53.17 (1.5)	50.84 (2.1)	56.92 (1.6)		
Gasch1	Ripper	51.40 (2.0)	43.21 (1.7)	59.25 (1.3)	44.78 (1.5)	47.65 (2.4
Gasciii	BayesNet	60.19 (1.3)	55.40 (0.8)	54.92 (1.4)	44.70 (1.0)	41.00 (2.5
	SVM	57.90 (2.1)	57.48 (2.9)	61.27 (1.9)		
	1			` ′		
	KNN C4.5	49.75 (2.1)	47.40 (2.2)	53.62 (2.5) 53.89 (0.6)		
Gasch2	Ripper	45.04 (2.1) 45.83 (3.3)	45.80 (2.0) 39.64 (3.7)		41.73 (1.5)	43.07 (1.1
Gasciiz	BayesNet	43.83 (3.3) 53.97 (1.8)	45.88 (1.8)	54.71 (0.9) 53.57 (1.7)	41.73 (1.5)	45.07 (1.1
	SVM	52.65 (0.9)	52.20 (2.0)	54.66 (0.4)		
				` ′		
	KNN	30.57 (3.2)	35.69 (2.8)	42.60 (5.1)		
TDI .	C4.5	30.71 (3.4)	40.52 (4.1)	49.09 (2.3)	45.84 (0.0)	4F FO (O)
Phenotype	Ripper	33.61 (2.3)	30.63 (8.3)	45.09 (1.3)	$47.34 \ (0.8)$	45.78 (2.3
	BayesNet SVM	45.50 (2.9)	44.53 (1.9)	47.28 (3.1)		
	1	39.59 (6.3)	38.30 (3.1)	48.86 (3.0)		
	KNN	48.00 (2.1)	47.97 (2.2)	50.45 (1.4)		
G.	C4.5	48.72 (1.4)	47.69 (2.8)	55.12 (1.6)	49.44.(2.0)	40 FF (0.6
Sequence	Ripper	47.20 (3.0)	40.85 (1.9)	54.65 (2.4)	43.44(2.0)	43.75 (2.3
	BayesNet	57.90 (1.3)	51.93 (2.0)	52.79 (2.2)		
	SVM	40.09 (2.3)	49.30 (2.4)	61.39 (1.7)		
	KNN	41.98 (1.6)	40.45 (1.1)	47.33 (1.7)		
	C4.5	43.73 (5.1)	43.73 (1.6)	51.39 (1.3)		
SPO	Ripper	43.70 (3.1)	42.23 (2.9)	55.80 (2.6)	40.28 (1.2)	40.77 (1.5
	BayesNet	$51.31\ (2.3)$	47.40(2.6)	49.32(2.9)		
	SVM	44.53(2.1)	45.23(2.3)	49.62(1.6)		

6.3.2 Resultados no Segundo Nível das Hierarquias

No segundo nível da hierarquia de classes, considerando a medida H-Loss (Tabela 6.12), o desempenho da técnica HMC-BR, utilizando o algoritmo KNN, mostrou-se superior ao desempenho das outras técnicas. Porém, não foram detectadas diferenças estatisticamente significantes entre as técnicas HMC-BR e HMC-LP. Diferenças estatisticamente significantes foram encontradas nos resultados da comparação da maioria dos algoritmos da técnica HMC-CT com os algoritmos das outras técnicas.

Considerando a medida F1 hierárquica (Tabela 6.13), os melhores desempenhos foram obtidos pelas técnicas HMC-BR e HMC-CT. Segundo sugerido pelo *ranking* médio dos algoritmos, os melhores resultados foram obtidos pela técnica HMC-BR utilizando o algoritmo BayesNet, seguida pela técnica HMC-CT utilizando o algoritmo SVM. Apesar disso, não houve diferença estatisticamente significante entre o primeiro e o segundo colocados.

O pior desempenho da técnica HMC-LP, na medida F1 hierárquica, pode ser devido ao grande número de classes formadas após o processo de combinação de classes aplicado pela técnica, fazendo com que muitas classes ficassem com poucos exemplos.

Considerando as medidas micro F1 hierárquicas não ponderada e ponderada (Tabelas 6.14 e 6.16), o desempenho das técnicas se inverte. Nessas medidas, os melhores desempenhos foram obtidos, na maioria dos casos, pelas técnicas HMC-LP e HMC-BR. Isso pode ter ocorrido porque essas medidas fazem considerações diferentes da medida F1 hierárquica. Elas são baseadas nas distâncias entre as classes preditas e verdadeiras, e consideram as contribuições dos exemplos Falso Positivos (FP) e Falso Negativos (FN) em seus cálculos, influenciando no resultado final da avaliação. Na medida micro F1 hierárquica, o melhor desempenho foi obtido pela técnica HMC-LP utilizando os algoritmos C4.5 e SVM, e na medida micro F1 hierárquica ponderada, o melhor desempenho foi obtido pela técnica HMC-BR com o algoritmo BayesNet, seguida pela técnica HMC-LP com o algoritmo SVM. Diferenças estatisticamente significantes foram constatadas, em sua maioria, entre os resultados das técnicas HMC-CT, C4.5H, Clus-HMC e os resultados das outras técnicas.

Na medida macro F1 hierárquica (Tabela 6.15), os melhores desempenhos foram obtidos, de acordo com o que foi indicado pelo *ranking* médio de desempenho dos algoritmos, pela técnica HMC-CT utilizando o algoritmo SVM, e pela técnica HMC-BR utilizando o algoritmo BayesNet. Já na medida macro F1 hierárquica ponderada (Tabela 6.17), os melhores desempenhos foram obtidos pela técnica HMC-LP utilizando os algoritmos Ripper e SVM, e pela técnica HMC-CT utilizando o algoritmo SVM. Entre esses algoritmos não foi detectada diferença estatisticamente significante.

Tabela 6.12: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

H-Loss no Segundo Nível da Hierarquia

Conjunto de Dados			n Top-Down		1	n One-Shot
Conjunto de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HM0
	KNN	$0.65 \ (0.06)$	$0.68 \ (0.02)$	$1.30 \ (0.10)$		
	C4.5	$1.10 \ (0.05)$	$1.00 \ (0.04)$	2.65 (0.30)		
Expr	Ripper	0.83 (0.10)	$0.85 \ (0.06)$	2.35 (0.10)	1.08 (0.04)	0.89 (0.02)
	BayesNet	$1.07 \ (0.04)$	0.99 (0.08)	1.13 (0.10)		
	SVM	0.59 (0.05)	0.82 (0.04)	1.69 (0.10)		
	KNN	$0.72\ (0.03)$	0.77(0.04)	1.37 (0.06)		
	C4.5	1.03 (0.09)	0.99(0.07)	2.96 (0.20)		
CellCycle	Ripper	$0.72\ (0.05)$	$0.86 \; (0.04)$	2.22(0.10)	1.12 (0.05)	0.89 (0.09
	BayesNet	0.85 (0.06)	$0.78 \; (0.06)$	1.21 (0.10)		
	SVM	0.86 (0.03)	$0.86 \ (0.05)$	1.40 (0.10)		
	KNN	0.63 (0.04)	0.67 (0.05)	1.53 (0.06)		
	C4.5	$0.56\ (0.04)$	0.75(0.04)	2.02(0.20)		
Church	Ripper	0.59 (0.08)	0.78 (0.04)	1.78(0.10)	3.44 (0.24)	2.52(0.3)
	BayesNet	0.82 (0.03)	0.57 (0.03)	0.85 (0.09)		
	SVM	0.59 (0.04)	0.67 (0.05)	1.66 (0.07)		
	KNN	0.71 (0.06)	0.77 (0.06)	1.46 (0.09)		
	C4.5	$0.55 \ (0.20)$	0.94 (0.08)	2.69 (0.20)		
Derisi	Ripper	0.63 (0.07)	0.82 (0.02)	1.91 (0.20)	1.16 (0.08)	0.97 (0.00
	BayesNet	0.82 (0.07)	0.68 (0.07)	0.84 (0.05)	Ì	`
	SVM	0.76 (0.08)	0.79(0.05)	1.63 (0.10)		
	KNN	0.71 (0.08)	0.76 (0.04)	1.39 (0.10)		
Eisen	C4.5	1.12 (0.05)	1.01 (0.09)	2.65 (0.20)		
	Ripper	0.79 (0.10)	0.78 (0.10)	2.39 (0.10)	1.15 (0.05)	1.01 (0.03
	BayesNet	0.94 (0.03)	0.76 (0.04)	1.15 (0.06)		- (
	SVM	0.84 (0.08)	0.83 (0.09)	1.40 (0.10)		
	KNN	0.68 (0.08)	0.77 (0.07)	1.34 (0.08)		
	C4.5	1.11 (0.05)	1.04 (0.06)	2.71 (0.20)		
Gasch1	Ripper	0.76 (0.04)	0.83 (0.03)	2.41 (0.30)	1.12 (0.07)	0.92 (0.0
Caboni	BayesNet	0.95 (0.07)	0.89 (0.07)	0.91 (0.04)	(0.01)	0.02 (0.0
	SVM	0.73 (0.06)	0.74 (0.02)	1.23 (0.06)		
	KNN	0.67 (0.04)	0.67 (0.08)	1.34 (0.03)		
	C4.5	$0.80 \ (0.04)$	0.97 (0.08)	2.69 (0.03)		
Gasch2	Ripper	0.66 (0.06)	0.83 (0.02)	2.18 (0.09)	1.10 (0.09)	0.94 (0.03
Gascii2	BayesNet	0.86 (0.09)	0.70 (0.04)	1.08 (0.10)	1.10 (0.03)	0.34 (0.0.
	SVM	0.94 (0.04)	0.86 (0.03)	1.44 (0.06)		
	KNN	0.48 (0.10)	0.69 (0.07)			
				1.55 (0.30)		
Phenotype	C4.5 Ripper	0.52 (0.20) 0.59 (0.10)	0.76 (0.10) 0.89 (0.06)	2.09 (0.50) 2.26 (0.10)	5.19 (0.2)	3.03 (1.0
пенотуре	BayesNet	1.03 (0.10)	0.89 (0.00)	1.73 (0.30)	3.19 (0.2)	5.05 (1.0
	SVM	0.62 (0.07)	0.80 (0.03)	$2.01 \ (0.30)$		
	KNN	0.64 (0.07)	0.68 (0.03)	$1.40 \ (0.10)$		
Soguenes	C4.5 Ripper	1.02 (0.07)	0.99 (0.09)	2.59 (0.20)	1.04 (0.09)	0.01 /0.1
Sequence		0.70 (0.09) 1.06 (0.07)	0.80 (0.04)	2.44 (0.10)	1.04 (0.08)	0.91 (0.1
	BayesNet SVM	, ,	1.00 (0.08)	1.21 (0.10)		
	SVM	0.37 (0.02)	0.72 (0.03)	1.52 (0.03)		
	KNN	0.67 (0.04)	0.71 (0.07)	1.44 (0.09)		
an o	C4.5	0.77 (0.20)	0.94 (0.08)	2.68 (0.10)		0.05 (5.1
SPO	Ripper	0.60 (0.07)	0.78 (0.04)	1.96 (0.06)	1.17 (0.12)	0.91 (0.03
	BayesNet	0.87 (0.05)	0.67 (0.07)	0.77 (0.09)		
	SVM	0.93(0.07)	$0.86 \ (0.09)$	1.54 (0.08)		

Tabela 6.13: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida F1

Hierárquica no Segundo Nível da Hierarquia

Conjunto de Dados		_	n Top-Down		Abordagen	
conjunto de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	40.89(1.1)	39.60(1.5)	41.26 (0.8)		
	C4.5	39.35(1.1)	$37.84\ (1.9)$	38.79 (0.9)		
Expr	Ripper	42.37(1.1)	32.27(1.5)	41.80 (1.3)	35.57(1.4)	37.36 (1.7
	BayesNet	47.46 (2.0)	44.36 (1.7)	43.20 (2.1)		
	SVM	32.92 (1.4)	32.28 (1.4)	38.94 (0.5)		
	KNN	41.24(2.1)	38.91(2.0)	41.05 (1.4)		
	C4.5	36.12(1.3)	34.02 (0.8)	36.49(1.3)		
CellCycle	Ripper	37.97(0.1)	29.33(1.5)	39.69(1.6)	33.19(1.3)	33.02 (1.1
	BayesNet	42.39(1.9)	37.85(1.3)	$42.01 \ (0.8)$		
	SVM	43.22 (1.9)	43.36 (2.7)	44.69 (1.0)		
	KNN	32.28 (1.6)	30.47 (0.3)	36.38 (1.1)		
	C4.5	32.46(1.5)	31.29(0.4)	35.78(1.6)		
Church	Ripper	33.22(1.1)	23.35(0.3)	36.87(1.6)	26.17(1.3)	28.75 (1.7)
	BayesNet	$42.02\ (1.1)$	31.40(1.7)	36.29(2.2)		
	SVM	31.83 (2.2)	31.68 (0.4)	37.63 (1.9)		
	KNN	34.47 (1.2)	32.04 (0.8)	35.62 (1.1)		
	C4.5	28.11(5.2)	31.87(1.0)	35.17(0.7)		
Derisi	Ripper	33.44(1.5)	24.94(0.7)	38.54 (0.9)	31.39(0.8)	32.36 (2.5
	BayesNet	$41.34\ (0.6)$	32.00(3.1)	36.52(1.8)		
	SVM	$34.01\ (1.1)$	32.45(1.3)	37.37(1.2)		
	KNN	43.91 (1.6)	40.70 (1.5)	43.45 (2.2)		
Eisen	C4.5	37.27 (1.6)	36.46 (1.6)	37.43 (2.1)		
	Ripper	41.01 (1.9)	33.61 (2.4)	41.48 (1.7)	34.30 (1.9)	35.79 (2.0
	BayesNet	$46.69\ (1.9)$	38.97 (2.0)	43.99 (1.5)		
	SVM	45.04(2.2)	45.68 (1.2)	45.52(2.3)		
	KNN	44.00 (1.8)	41.36 (1.2)	43.65 (0.9)		
	C4.5	40.17 (1.6)	36.74 (0.9)	38.84 (0.7)		
Gasch1	Ripper	40.65 (1.1)	29.45 (0.8)	41.51 (2.4)	35.10 (1.0)	35.08 (1.2
	BayesNet	46.90 (1.0)	39.50 (1.4)	39.77 (1.5)	, ,	·
	SVM	47.10 (1.2)	46.05 (2.2)	47.57 (2.0)		
	KNN	40.09 (1.2)	36.98 (1.1)	40.50 (1.7)		
	C4.5	36.15 (1.8)	34.60 (1.7)	36.60 (0.9)		
Gasch2	Ripper	37.24 (2.7)	27.72 (1.8)	39.66 (1.0)	32.83 (0.8)	34.00 (1.4
	BayesNet	42.39 (0.7)	33.80 (1.3)	40.69 (1.2)	, ,	`
	SVM	40.78 (0.7)	40.90 (1.2)	40.70 (1.1)		
	KNN	26.25 (2.8)	29.11 (1.2)	33.77 (1.8)		
	C4.5	27.26 (2.8)	30.24 (2.3)	35.66 (1.4)		
Phenotype	Ripper	29.12 (2.4)	24.78 (1.1)	35.72 (1.4)	25.50 (0.4)	27.87 (2.4
V -	BayesNet	31.74 (1.1)	29.59 (1.5)	28.60 (1.2)	` ,	`
	SVM	31.49 (2.7)	29.72 (1.9)	36.00 (1.8)		
	KNN	39.53 (2.0)	37.37 (1.4)	39.03 (1.3)		
	C4.5	38.34 (0.4)	36.62 (1.5)	37.92 (1.1)		
Sequence	Ripper	39.08 (1.0)	28.29 (0.7)	39.89 (1.2)	34.40 (1.2)	35.01 (1.5
1	BayesNet	45.69 (1.0)	37.91 (1.5)	40.42 (1.3)	,	- (
	SVM	33.21 (2.6)	36.35 (1.9)	42.09 (0.5)		
	KNN	32.55 (1.4)	30.78 (0.7)	35.60 (1.2)		
	C4.5	34.18 (5.2)	32.99 (1.7)	34.63 (0.9)		
SPO	Ripper	35.09 (2.0)	28.36 (1.5)	40.01 (0.7)	31.37 (1.4)	32.59 (1.6
0	BayesNet	40.23 (1.1)	34.63 (1.6)	36.08 (1.5)	52.51 (1.4)	52.55 (1.0
	SVM	35.15 (1.3)	34.71 (1.6)	36.98 (0.9)		

Tabela 6.14: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Micro F1 Hierárquica no Segundo Nível da Hierarquia

Conjunto de Dados		_	n Top-Down		Abordagen	
- injunio de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	$31.03 \ (0.5)$	32.05 (1.2)	$30.90 \ (0.5)$		
	C4.5	30.93 (0.5)	$33.20 \ (0.7)$	24.54 (0.9)		
Expr	Ripper	32.07(1.2)	$26.81\ (1.7)$	27.00(1.6)	30.32(1.1)	32.12 (0.6
	BayesNet	34.21 (1.2)	$35.30 \ (1.0)$	30.94 (1.2)		
	SVM	27.70 (1.3)	26.29 (0.7)	25.89 (0.8)		
	KNN	31.76(1.0)	$32.46 \ (0.7)$	30.88 (0.9)		
CellCycle	C4.5	$30.71\ (0.8)$	32.52(1.1)	23.74(0.3)		
	Ripper	29.16 (0.4)	24.96(1.1)	27.17(1.1)	29.16 (0.9)	$30.53 \ (0.7)$
	BayesNet	31.75 (1.4)	$30.34\ (1.2)$	28.89 (0.6)		
	SVM	33.63 (0.8)	35.17 (1.1)	32.17 (0.6)		
	KNN	25.63(1.2)	27.89(0.4)	28.83 (0.6)		
	C4.5	24.79(0.9)	29.25 (0.6)	26.83(1.0)		
Church	Ripper	25.50(1.1)	19.74 (0.4)	25.18(1.4)	16.32(1.1)	18.82 (1.7
	BayesNet	28.94 (0.5)	25.46 (0.9)	24.63 (1.2)		
	SVM	$24.91 \ (0.6)$	28.10 (0.7)	28.71 (1.2)		
	KNN	28.75 (0.9)	29.93 (0.8)	28.98 (0.5)		
	C4.5	22.97(5.1)	$30.52\ (0.3)$	23.46 (0.6)		
Derisi	Ripper	26.72(0.8)	$21.31\ (0.3)$	26.59 (0.5)	27.86(0.3)	29.57 (1.2
	BayesNet	29.39(0.9)	25.53(1.5)	$25.61\ (1.1)$		
	SVM	28.42 (0.8)	30.49 (0.9)	29.83 (0.8)		
	KNN	32.03 (1.2)	32.98 (0.8)	31.12 (0.8)		
Eisen	C4.5	31.33 (1.0)	33.11 (0.4)	24.21 (1.5)		
	Ripper	30.20 (1.0)	27.87 (1.6)	27.01 (1.3)	30.09 (1.0)	31.71 (0.5
	BayesNet	33.07(1.5)	31.56(1.1)	29.52 (0.8)		
	SVM	33.66 (0.9)	$35.64\ (0.8)$	31.88 (0.8)		
	KNN	32.71 (1.2)	33.27 (0.9)	31.13 (0.9)		
	C4.5	32.38 (0.8)	33.59 (1.1)	24.11 (0.7)		
Gasch1	Ripper	30.47 (1.4)	24.54 (0.4)	27.01 (1.9)	28.91 (0.8)	31.24 (0.8
	BayesNet	31.75(0.4)	32.35(0.6)	27.44(0.8)		
	SVM	34.47(1.0)	$35.84\ (0.9)$	32.87(0.8)		
	KNN	31.27 (0.6)	31.11 (0.7)	31.30 (1.2)		
	C4.5	28.35 (1.7)	32.26 (0.9)	23.74 (0.5)		
Gasch2	Ripper	28.00 (1.8)	23.29 (0.9)	26.64 (0.5)	28.76 (1.1)	30.10 (0.3
	BayesNet	30.24(0.4)	27.09(0.7)	28.21 (0.9)		
	SVM	33.06(0.4)	$34.57 \; (0.5)$	31.04(0.5)		
	KNN	22.81 (2.6)	26.09 (1.2)	26.11 (1.1)		
	C4.5	22.42 (2.8)	27.60 (1.5)	26.75 (0.7)		
Phenotype	Ripper	24.67 (2.6)	21.41 (0.6)	22.52 (0.6)	15.64 (0.4)	20.47 (3.2
	BayesNet	28.76 (1.8)	25.38 (1.5)	23.16 (0.7)	, ,	
	SVM	26.30 (2.2)	27.61(1.4)	28.03 (1.4)		
	KNN	29.30 (1.5)	30.08 (0.9)	27.92 (0.8)		
	C4.5	31.21 (1.2)	32.48 (1.3)	24.01 (0.9)		
Sequence	Ripper	29.81 (1.7)	24.09 (0.7)	25.93 (0.7)	30.06 (1.1)	30.95 (1.2
-	BayesNet	34.70 (0.4)	32.83 (1.5)	28.59 (0.3)	` /	`
	SVM	27.05 (2.4)	28.06 (1.2)	27.98 (0.6)		
	KNN	27.58 (0.9)	29.05 (0.8)	29.23 (0.6)		
	C4.5	27.91 (4.5)	31.34 (0.8)	22.79 (0.6)		
SPO	Ripper	26.30 (1.7)	23.19 (1.0)	27.65 (0.8)	27.80 (0.5)	29.68 (0.9
	BayesNet	29.37 (1.2)	27.90 (0.7)	25.72 (0.7)		_0.00 (0.0
	SVM	30.70 (0.8)	32.04 (0.9)	29.87 (0.6)		

Tabela 6.15: Comparação das Abordagens *Top-Down* e *One-Shot* Utilizando a Medida Macro F1 Hierárquica no Segundo Nível da Hierarquia

Conjunto de Dados		_	m Top-Down		Abordagem	
Conjunto de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	10.26 (1.6)	09.66(2.0)	$09.03 \ (0.6)$		
	C4.5	$08.21 \ (0.4)$	07.77(0.4)	08.76 (0.1)		
Expr	Ripper	10.51 (2.1)	$07.68 \; (0.5)$	09.11 (0.1)	$07.02 \ (0.2)$	07.49 (0.1
	BayesNet	10.78 (1.3)	10.87 (1.1)	09.43 (1.1)		
	SVM	06.24 (0.3)	06.15 (0.3)	06.97 (0.2)		
	KNN	08.58(1.2)	$07.73 \ (0.3)$	08.49 (0.4)		
CellCycle	C4.5	$07.66 \ (0.7)$	$07.06 \ (0.2)$	08.49(0.2)		
	Ripper	07.89 (0.7)	06.93 (0.4)	09.06 (1.0)	06.69 (0.3)	06.64 (0.2
	BayesNet	$08.53 \ (0.3)$	$07.88 \ (0.3)$	08.77(0.3)		
	SVM	10.55 (1.7)	09.44 (1.5)	09.83 (1.6)		
	KNN	06.41 (0.4)	06.32 (0.3)	07.50 (0.3)		
	C4.5	06.24 (0.3)	06.47(0.1)	07.62 (0.3)		
Church	Ripper	07.49(0.6)	$04.70 \ (0.5)$	07.26(1.1)	06.96 (0.2)	06.85 (0.5
	BayesNet	$08.58 \ (0.6)$	06.59 (0.07)	$07.36 \ (0.4)$		
	SVM	$06.51 \ (0.4)$	$06.70 \ (0.3)$	07.99(0.4)		
	KNN	07.26 (0.2)	07.20 (0.9)	07.76 (0.1)		
	C4.5	06.14(0.8)	06.73(0.1)	08.03 (0.2)		
Derisi	Ripper	07.08(0.2)	05.53(0.2)	08.43 (0.1)	06.47(0.2)	06.99 (1.3
	BayesNet	08.29(0.1)	06.94 (0.5)	07.71 (0.3)		
	SVM	07.02(0.3)	06.84 (0.2)	07.99(0.3)		
	KNN	09.71 (1.2)	09.70 (2.0)	09.47 (0.9)		
Eisen	C4.5	08.18 (1.0)	07.52 (0.3)	08.35 (0.4)		
	Ripper	09.13 (2.1)	08.07 (0.9)	09.26 (0.5)	06.96 (0.4)	07.44 (0.5
	BayesNet	09.26 (0.6)	08.38 (0.4)	09.83 (1.3)	, ,	`
	SVM	10.68 (1.5)	11.28 (2.3)	10.75 (1.7)		
	KNN	12.43 (2.1)	10.83 (1.0)	09.75 (1.4)		
	C4.5	09.56 (1.0)	08.23 (0.9)	08.83 (0.3)		
Gasch1	Ripper	11.20 (1.7)	07.40 (1.0)	09.63 (1.0)	06.88(0.2)	07.33 (0.3
	BayesNet	09.74 (0.9)	08.73 (0.5)	08.54 (0.3)	, ,	`
	SVM	13.71 (2.4)	12.14 (2.3)	09.91 (1.0)		
	KNN	08.82 (1.4)	07.88 (1.2)	08.25 (0.4)		
	C4.5	07.06 (0.4)	07.04 (0.3)	08.32 (0.1)		
Gasch2	Ripper	07.54 (0.5)	06.09 (0.5)	08.41 (0.1)	06.42 (0.2)	06.62 (0.5
	BayesNet	08.30 (0.2)	07.05 (0.3)	08.24 (0.2)	, ,	`
	SVM	09.07 (0.9)	09.11 (1.4)	08.52 (0.2)		
	KNN	04.99 (0.7)	05.49 (0.4)	06.55 (0.8)		
	C4.5	04.72 (0.5)	06.23 (0.6)	07.55 (0.3)		
Phenotype	Ripper	05.17 (0.3)	04.71 (1.3)	06.93 (0.2)	07.78 (0.5)	07.04 (0.
V 1	BayesNet	06.99 (0.4)	06.85 (0.3)	07.27 (0.5)	` ,	`
	SVM	06.09 (0.9)	05.89 (0.4)	07.51 (0.4)		
	KNN	07.38 (0.3)	07.38 (0.3)	07.95 (0.5)		
	C4.5	07.49 (0.2)	07.33 (0.4)	08.51 (0.3)		
Sequence	Ripper	07.79 (1.6)	06.28 (0.3)	08.40 (0.3)	06.68 (0.3)	06.73 (0.
**** · · ·	BayesNet	09.16 (0.2)	07.98 (0.3)	08.12 (0.3)	(()	(0
	SVM	10.52 (1.8)	08.96 (1.1)	09.44 (0.2)		
	KNN	06.45 (0.2)	06.22 (0.2)	07.28 (0.2)		
	C4.5	$06.45 \ (0.2)$ $06.73 \ (0.8)$	06.22 (0.2) 06.72 (0.2)	07.28 (0.2) 07.90 (0.2)		
SPO	Ripper	07.24 (1.1)	06.72 (0.2) $06.49 (0.4)$	07.90 (0.2)	06.19 (0.2)	06.27 (0.3
D1 O	BayesNet	07.24 (1.1) $07.89 (0.3)$	00.49 (0.4) $07.29 (0.4)$	07.59 (0.4)	00.13 (0.2)	00.27 (0.2
	SVM	07.89 (0.3) 06.85 (0.3)	07.29 (0.4) $06.96 (0.3)$	07.63 (0.2)		

Tabela 6.16: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Micro F1 Hierárquica Ponderada no Segundo Nível da Hierarquia

Conjunto de Dados		_	n Top-Down		Abordagen	
J		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	29.52 (0.6)	28.35 (1.4)	28.79(1.3)		
_	C4.5	29.31 (1.8)	28.53 (1.4)	25.48 (2.1)		
Expr	Ripper	32.85 (1.7)	31.40 (1.3)	29.34 (1.7)	26.58 (0.7)	$27.91 \ (0.8)$
	BayesNet	35.87 (2.3)	35.45 (3.1)	29.19 (2.4)		
	SVM	30.88 (1.2)	32.35 (2.0)	30.29 (1.3)		
	KNN	31.18(2.0)	29.48(1.9)	29.07 (0.9)		
	C4.5	27.28(1.3)	26.19(1.4)	23.74(1.5)		
CellCycle	Ripper	29.62(1.3)	$30.31\ (1.4)$	28.28 (1.9)	25.50 (0.7)	26.12 (0.6
	BayesNet	34.19 (2.4)	30.68 (1.6)	29.52 (0.5)		
	SVM	33.14 (2.0)	34.64 (3.4)	32.59 (2.0)		
	KNN	23.55(1.1)	23.16(0.3)	26.77(0.7)		
	C4.5	23.40(1.2)	23.84(0.3)	25.27(1.3)		
Church	Ripper	25.26(2.2)	28.18 (0.8)	27.94(2.2)	17.13 (1.1)	18.94 (1.4
	BayesNet	31.19 (0.6)	23.14(0.9)	23.00(1.2)		
	SVM	23.54 (1.2)	24.09 (0.9)	26.78 (1.8)		
	KNN	25.85 (1.1)	24.38 (0.6)	25.69 (1.2)		
	C4.5	22.49(4.7)	$24.51 \ (0.8)$	22.43(1.3)		
Derisi	Ripper	25.64(0.4)	28.57(0.5)	27.88(1.7)	24.53 (0.3)	25.44 (0.9
	BayesNet	28.98 (1.8)	24.47(2.3)	24.14 (0.8)		
	SVM	24.90 (0.5)	24.94(1.2)	25.74(1.3)		
	KNN	32.37 (2.1)	29.68 (1.1)	30.28 (2.8)		
Eisen	C4.5	27.80 (0.8)	27.88 (1.0)	24.54(2.7)		
	Ripper	31.63(2.4)	29.89 (1.5)	30.26 (2.0)	26.56 (1.0)	27.70 (0.4
	BayesNet	35.31(3.0)	29.87(2.0)	30.29(2.6)		
	SVM	34.36(2.3)	36.78(1.4)	33.16(3.5)		
	KNN	32.32 (2.4)	30.54 (0.6)	29.91 (1.7)		
	C4.5	30.53 (1.5)	28.51 (1.1)	25.48 (1.9)		
Gasch1	Ripper	31.97 (2.6)	30.34 (1.6)	28.95 (2.2)	25.74 (0.7)	27.10 (0.9
	BayesNet	33.38(1.7)	29.90(1.1)	25.61 (0.8)		
	SVM	38.00(2.4)	38.86(2.1)	34.88(2.6)		
	KNN	29.52 (1.8)	27.35 (1.4)	28.87 (2.3)		
	C4.5	26.59 (0.9)	26.54 (1.2)	23.53 (1.0)		
Gasch2	Ripper	28.91 (2.5)	30.12 (1.8)	28.21 (1.6)	25.11 (0.8)	26.02 (0.3
	BayesNet	29.46 (1.3)	26.63 (0.9)	26.39 (1.0)	, ,	
	SVM	30.48 (1.2)	32.16 (0.9)	28.37(0.5)		
	KNN	23.05 (3.5)	25.26 (1.2)	26.04 (1.7)		
	C4.5	22.06 (2.5)	25.79 (1.4)	26.39 (1.0)		
Phenotype	Ripper	24.52 (1.4)	30.99 (0.7)	26.98 (0.9)	17.70 (0.9)	19.81 (2.0
· -	BayesNet	27.42 (1.7)	24.52 (1.9)	21.86 (1.1)	, , ,	`
	SVM	26.37 (2.3)	27.33 (1.4)	27.29 (1.7)		
	KNN	27.51 (1.9)	26.31 (0.9)	25.90 (1.0)		
	C4.5	29.11 (0.6)	28.62 (2.2)	25.47 (1.8)		
Sequence	Ripper	30.91 (1.8)	28.52 (0.6)	28.07 (1.1)	26.31 (0.8)	27.22 (1.0
4	BayesNet	36.53 (1.9)	31.30 (1.2)	29.43 (2.3)		(2.0
	SVM	29.70 (3.2)	33.77 (1.9)	32.72 (0.7)		
	KNN	24.66 (0.9)	23.73 (0.4)	25.25 (1.0)		
	C4.5	24.66 (0.9) 25.43 (3.8)	25.73 (0.4) 25.06 (0.8)	` ,		
SPO	Ripper	` ′	` ,	21.17 (0.7)	24 30 (0.4)	25 67 (0.5
DI O		26.49 (2.3)	29.92 (1.2) 25.72 (0.0)	28.95 (1.1)	24.39 (0.4)	25.67 (0.7
	BayesNet	27.44(0.6)	25.72(0.9)	23.87(0.9)	I	

Tabela 6.17: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Macro F1 Hierárquica Ponderada no Segundo Nível da Hierarquia

Conjunto de Dados		_	em Top-Down		Abordagen	
James de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	14.25 (1.3)	13.36 (2.5)	13.24 (2.8)		
	C4.5	12.68 (1.8)	10.54(2.3)	11.70 (1.9)		
Expr	Ripper	16.08 (3.5)	19.28 (3.7)	15.24 (3.1)	08.08 (1.4)	08.45 (1.3)
	BayesNet	16.97 (3.0)	19.05 (4.2)	14.56 (3.8)		
	SVM	10.04 (0.4)	20.79 (3.5)	22.31 (1.3)		
	KNN	12.22(1.9)	$11.51\ (1.2)$	13.32(2.0)		
	C4.5	09.38 (1.8)	08.13(1.7)	10.65 (1.2)		
CellCycle	Ripper	12.40(2.1)	18.83(2.2)	13.03(1.3)	06.87 (0.6)	07.06 (1.0
	BayesNet	$16.57\ (1.6)$	16.02 (4.0)	16.35 (2.6)		
	SVM	17.05 (2.6)	19.89 (4.6)	17.30 (1.8)		
	KNN	07.88(1.2)	06.95 (0.9)	10.53 (1.3)		
	C4.5	$07.04\ (1.3)$	$06.90 \ (0.2)$	09.50(1.3)		
Church	Ripper	$10.53\ (1.8)$	20.02 (1.0)	19.87(3.1)	09.73 (2.3)	07.93 (0.5)
	BayesNet	12.46(2.1)	$08.51\ (1.3)$	09.56 (1.6)		
	SVM	07.56 (0.8)	07.97 (1.1)	10.44 (1.0)		
	KNN	08.79 (1.9)	08.70 (1.7)	10.08 (0.9)		
	C4.5	10.24(2.4)	08.05(1.6)	09.14(1.1)		
Derisi	Ripper	12.43(2.0)	20.58(2.6)	17.52(2.3)	06.47 (0.2)	08.40 (1.4
	BayesNet	13.08(3.4)	13.83(3.5)	12.56 (4.8)		
	SVM	08.02(0.3)	09.10(1.2)	09.40(1.3)		
	KNN	14.95 (3.6)	15.30 (2.9)	16.16 (3.6)		
Eisen	C4.5	11.90 (1.4)	09.38 (1.4)	11.23 (2.0)		
	Ripper	15.36 (2.6)	16.49 (3.4)	17.24 (1.9)	08.31 (1.8)	08.37 (1.1
	BayesNet	19.42 (2.7)	16.52 (2.7)	17.96 (3.7)	, ,	`
	SVM	18.47 (4.1)	24.36 (4.6)	19.34 (3.4)		
	KNN	19.87 (3.6)	17.35 (1.4)	15.86 (3.8)		
	C4.5	13.75 (3.3)	11.60 (2.4)	12.39 (2.0)		
Gasch1	Ripper	18.11 (3.1)	19.56 (3.8)	14.45 (0.4)	08.32 (1.2)	07.68 (0.6
	BayesNet	15.86 (1.7)	16.49 (2.4)	13.23 (3.6)	,	(
	SVM	24.47 (3.6)	24.19 (4.9)	23.74 (3.6)		
	KNN	12.11 (1.6)	11.28 (2.5)	13.38 (1.9)		
	C4.5	08.98 (2.0)	08.45 (1.1)	10.29 (0.6)		
Gasch2	Ripper	12.66 (1.9)	18.90 (2.3)	13.85 (2.1)	06.70 (0.4)	07.09 (0.6
Guscii	BayesNet	11.33 (1.3)	12.84 (0.7)	11.60 (0.8)	00170 (011)	01.00 (01.
	SVM	14.31 (2.0)	14.72 (2.0)	11.63 (1.3)		
	KNN	07.43 (2.3)	10.40 (2.6)	13.50 (3.3)		
	C4.5	06.24 (1.4)	10.56 (3.0)	11.85 (2.2)		
Phenotype	Ripper	07.23 (0.5)	21.31 (1.3)	23.27 (1.5)	10.50 (0.9)	08.29 (1.0
r nonety pe	BayesNet	11.40 (3.4)	11.50 (1.9)	12.58 (6.5)	10.00 (0.0)	00.20 (1.0
	SVM	08.95 (1.8)	12.39 (1.2)	11.26 (1.6)		
	KNN	09.22 (1.9)	09.19 (0.8)	09.98 (1.3)		
	C4.5	10.94 (1.9)	10.44 (3.4)	11.94 (1.5)		
Sequence	Ripper	13.24 (1.4)	15.48 (1.1)	12.60 (1.3)	07.11 (1.1)	07.93 (1.1
Sequence	BayesNet	16.28 (3.1)	13.43 (1.1)	14.19 (3.3)	J (1.1)	01.00 (1.1
	SVM	12.96 (2.2)	21.81 (1.5)	25.20 (1.7)		
	1					
	KNN C4.5	06.46 (0.2)	06.90 (0.6)	08.63 (0.8)		
		$08.03\ (1.8)$	07.12 (0.6)	08.38 (0.5)		00.00 (0.0
CDO	Dinna	10 49 (9.0)	10 99 / 9 9\		0707/161	
SPO	Ripper BayesNet	10.48 (2.0) 10.44 (1.5)	19.33 (2.2) 10.74 (1.4)	14.76 (3.9) 13.56 (2.6)	07.07 (1.6)	06.36 (0.2

6.3.3 Resultados no Terceiro Nível das Hierarquias

No terceiro nível das hierarquias, os melhores resultados, considerando a medida H-Loss (Tabela 6.18), foram obtidos pela técnica HMC-LP utilizando o algoritmo Ripper. Porém, não foram detectadas diferenças estatisticamente significantes entre essa técnica e a técnica HMC-BR, com exceção da diferença entre os algoritmos Ripper e BayesNet. Como já citado anteriormente, o maior número de erros das técnicas HMC-CT, C4.5H e Clus-HMC deve-se ao maior número de predições realizadas.

A avaliação utilizando a medida F1 hierárquica (Tabela 6.19) apresentou os melhores resultados para a técnica HMC-BR com o algoritmo BayesNet. As diferenças entre o desempenho desse algoritmo e os desempenhos dos algoritmos das técnicas HMC-LP mostraram-se estatisticamente significantes apenas na comparação com os algoritmos C4.5 e Ripper. Na comparação com os algoritmos utilizados na técnica HMC-CT, os resultados mostraram-se estatisticamente significantes na maioria dos casos. Também foi detectada diferença estatisticamente significante na comparação com as técnicas C4.5H e Clus-HMC.

Assim como no nível anterior, quando são consideradas as distâncias entre as classes, utilizando as medidas micro F1 hierárquicas não ponderada e ponderada (Tabelas 6.20 e 6.22), a técnica HMC-LP apresentou resultados melhores que a técnica HMC-BR. Na medida micro F1 hierárquica, os melhores resultados foram obtidos pela técnica HMC-LP utilizando os algoritmos C4.5 e SVM. Entre esses dois algoritmos, não foi detectada diferença estatisticamente significante. Foram detectadas diferenças estatisticamente significantes entre os resultados desses algoritmos e os resultados da maioria dos algoritmos da técnica HMC-CT, e também nos resultados da técnica C4.5H.

Na medida micro F1 hierárquica ponderada, a técnica HMC-LP, utilizando o algoritmo SVM, obteve o melhor desempenho. Foram detectadas diferenças estatisticamente significantes entre os resultados desse algoritmo e os resultados de todos os algoritmos da técnica HMC-CT. A diferença entre os resultados da técnica HMC-LP também mostrouse estatisticamente significante na comparação com os resultados das técnicas C4.5H e Clus-HMC.

Considerando as medidas macro F1 hierárquicas não ponderada e ponderada (Tabelas 6.21 e 6.23), alguns dos melhores resultados também foram obtidos pela técnica HMC-CT. Na medida macro F1 hierárquica não ponderada, os testes estatísticos sugerem um melhor desempenho da técnica HMC-CT utilizando o algoritmo SVM. Na medida macro F1 hierárquica ponderada, os melhores desempenhos foram obtidos pela técnica HMC-LP utilizando os algoritmos Ripper e SVM. As diferenças de desempenho entre esses dois algoritmos e alguns algoritmos das técnicas HMC-BR e HMC-LP mostraram-se estatisticamente significantes. A comparação com os resultados das técnicas C4.5H e Clus-HMC também apresentou diferença estatisticamente significante.

Tabela 6.18: Comparação das Abordagens *Top-Down* e *One-Shot* Utilizando a Medida H-Loss no Terceiro Nível da Hierarquia

Conjunto de Dados		Abordage	m Top-Down		Abordagen	n One-Shot
Conjunto de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HM
	KNN	$0.31 \ (0.04)$	$0.28 \; (0.02)$	1.44(0.09)		
	C4.5	0.55 (0.07)	$0.40 \ (0.02)$	4.22(0.40)		
Expr	Ripper	0.35 (0.07)	$0.18 \; (0.05)$	3.66 (0.40)	$0.74 \ (0.04)$	$0.50 \ (0.07)$
	BayesNet	$0.78 \; (0.07)$	0.33(0.04)	1.35 (0.20)		
	SVM	$0.12\ (0.01)$	0.13 (0.02)	1.26 (0.20)		
	KNN	0.28 (0.04)	0.25 (0.04)	1.36 (0.10)		
	C4.5	$0.46 \ (0.05)$	0.33(0.13)	4.60 (0.40)		
CellCycle	Ripper	$0.30 \ (0.02)$	$0.13\ (0.03)$	2.99(0.20)	0.79(0.1)	0.46 (0.03)
	BayesNet	$0.38 \ (0.03)$	0.19(0.04)	1.56 (0.20)		
	SVM	$0.37 \ (0.05)$	0.29 (0.02)	1.57(0.10)		
	KNN	0.30 (0.03)	0.22 (0.05)	1.41 (0.20)		
	C4.5	0.29 (0.05)	0.21 (0.06)	2.06 (0.20)		
Church	Ripper	0.23 (0.10)	0.10 (0.006)	1.45 (0.20)	7.95 (0.7)	5.38 (1.0
	BayesNet	0.76 (0.07)	0.18 (0.04)	1.36 (0.40)	,	`
	SVM	0.29 (0.10)	0.19 (0.03)	1.63 (0.20)		
	KNN	0.24 (0.03)	0.21 (0.03)	1.27 (0.07)	<u> </u>	
	C4.5	0.12 (0.05)	0.30 (0.07)	3.88 (0.20)		
Derisi	Ripper	0.21 (0.06)	0.11 (0.02)	2.08 (0.10)	0.74 (0.05)	0.52 (0.0
Derisi	BayesNet	0.79 (0.07)	0.15 (0.04)	1.43 (0.30)	(0.00)	0.02 (0.0
	SVM	0.32 (0.04)	0.25 (0.04)	1.60 (0.09)		
	KNN	0.33 (0.05)	0.28 (0.04)	1.45 (0.03)		
Eisen	C4.5	0.46 (0.09)	0.35 (0.04) $0.35 (0.07)$	4.15 (0.30)		
	Ripper	0.37 (0.03)	0.14 (0.04)	3.50 (0.20)	0.74 (0.1)	0.47 (0.0
	BayesNet	0.56 (0.06)	0.22 (0.06)	1.61 (0.30)	0.14 (0.1)	0.47 (0.0
	SVM	0.36 (0.05)	0.30 (0.03)	1.62 (0.10)		
	KNN					
	C4.5	0.35 (0.02)	0.27 (0.04)	1.58 (0.04)		
Gasch1		0.52 (0.06)	$0.36 \ (0.05)$	4.48 (0.50)	0.90 (0.07)	0.44 (0.0
Jasciii	Ripper BayesNet	0.39 (0.06)	$0.14 \ (0.02)$	3.61 (0.40)	0.89 (0.07)	0.44 (0.0
	SVM	1.06 (0.10) 0.34 (0.04)	0.26 (0.04) 0.25 (0.03)	1.42 (0.20) 1.39 (0.09)		
	KNN	0.29 (0.02)	0.23 (0.02)	1.33 (0.09)		
~	C4.5	0.41 (0.10)	0.33 (0.09)	4.32 (0.60)	()	
Gasch2	Ripper	0.31 (0.06)	0.13 (0.01)	2.86 (0.30)	0.77(0.1)	0.52 (0.0
	BayesNet	0.67 (0.06)	0.21 (0.04)	1.67 (0.10)		
	SVM	0.40 (0.04)	0.29 (0.03)	1.49 (0.07)		
	KNN	$0.08 \; (0.03)$	0.19 (0.04)	$1.31\ (0.30)$		
	C4.5	$0.10\ (0.02)$	$0.23 \ (0.10)$	1.92 (0.50)		
Phenotype	Ripper	$0.10\ (0.03)$	$0.10 \ (0.03)$	2.34 (0.20)	8.35 (0.5)	2.94(2.4)
	BayesNet	$0.31\ (0.03)$	$0.21 \ (0.04)$	$1.11 \ (0.30)$		
	SVM	0.16 (0.06)	0.19 (0.07)	1.81 (0.50)		
	KNN	0.37 (0.09)	0.30 (0.03)	1.79 (0.20)		
	C4.5	$0.46 \ (0.07)$	$0.37 \ (0.03)$	4.69(0.30)		
Sequence	Ripper	$0.34\ (0.06)$	$0.14\ (0.02)$	3.89(0.50)	$0.64 \ (0.07)$	0.47 (0.0
	BayesNet	0.62 (0.10)	$0.30 \ (0.05)$	1.44(0.30)		
	SVM	0.16 (0.04)	0.17 (0.02)	1.28 (0.10)		
	KNN	0.21 (0.02)	0.19 (0.03)	1.27 (0.08)		
	C4.5	0.34 (0.10)	0.32(0.05)	4.01 (0.30)		
SPO	Ripper	0.26 (0.04)	$0.16\ (0.04)$	2.36 (0.20)	0.80 (0.09)	0.51 (0.03
	BayesNet	0.61 (0.09)	0.22(0.05)	1.09 (0.10)		•
	SVM	0.34 (0.05)	0.29 (0.06)	1.42 (0.08)		

Tabela 6.19: Comparação das Abordagens $\mathit{Top-Down}$ e $\mathit{One-Shot}$ Utilizando a Medida F1

Hierárquica no Terceiro Nível da Hierarquia

Conjunto de Dados		9	n Top-Down		Abordagen	
Conjunto de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HM0
	KNN	34.50 (1.2)	33.15(1.2)	29.92 (0.8)		
	C4.5	32.07(1.0)	$30.21\ (1.7)$	25.02(1.0)		
Expr	Ripper	$35.31\ (1.3)$	24.69(1.3)	27.86 (1.1)	26.90(1.1)	28.99 (1.5
	BayesNet	39.92 (1.6)	36.76 (1.4)	31.38 (1.5)		
	SVM	26.26 (1.1)	24.77 (1.3)	28.89 (0.7)		
	KNN	35.19(2.0)	32.55(1.9)	30.18 (1.2)		
	C4.5	28.97(1.3)	$27.01 \ (0.5)$	22.94 (0.8)		
CellCycle	Ripper	31.67 (0.2)	22.22(1.2)	26.90 (1.2)	25.32(0.9)	25.30 (0.5)
	BayesNet	$35.96 \ (1.3)$	31.07(1.1)	30.93 (0.7)		
	SVM	35.94 (1.6)	35.94 (2.7)	33.06 (0.6)		
	KNN	26.28(1.8)	24.26 (0.7)	25.49(0.9)		
	C4.5	26.68 (1.6)	24.98 (0.6)	23.99 (1.3)		
Church	Ripper	27.35 (0.6)	16.22 (0.4)	26.19 (1.2)	$16.38 \ (0.8)$	18.81 (1.
	BayesNet	$35.85 \ (1.1)$	$25.31\ (1.4)$	26.21 (1.7)		
	SVM	25.94 (1.8)	25.23(0.2)	26.28 (1.7)		
	KNN	28.16 (1.2)	25.84 (0.8)	24.59 (0.4)		
	C4.5	22.71(4.4)	25.17(0.9)	22.58 (0.6)		
Derisi	Ripper	27.46(1.6)	17.75 (0.5)	26.29 (0.6)	$23.43 \ (0.5)$	24.37(2.3)
	BayesNet	$34.33\ (0.5)$	26.16(3.1)	25.55(1.5)		
	SVM	27.34 (0.9)	$25.61\ (1.2)$	25.99(1.0)		
	KNN	37.65 (1.3)	33.97 (1.5)	31.68 (1.7)		
Eisen	C4.5	30.10 (1.3)	29.25 (1.6)	23.97 (1.6)		
	Ripper	34.54(1.9)	26.60(2.8)	27.72(1.3)	26.39(1.6)	27.51 (2.
	BayesNet	40.09(1.8)	32.00(2.0)	32.62 (1.6)		
	SVM	38.39(2.1)	38.53(1.3)	33.78 (2.1)		
	KNN	37.48 (1.7)	34.53 (0.9)	32.00 (0.8)		
	C4.5	32.28 (1.5)	28.98 (0.7)	25.07 (0.3)		
Gasch1	Ripper	33.64 (0.9)	22.13 (0.8)	27.72 (2.0)	26.63 (0.7)	26.99 (0.
	BayesNet	38.51 (1.1)	32.28(1.4)	28.61 (1.1)		
	SVM	40.44 (1.4)	38.67(2.1)	36.12 (1.6)		
	KNN	33.55 (1.0)	30.49 (1.1)	29.08 (1.5)		
	C4.5	29.47 (1.4)	27.50 (1.1)	23.43 (0.8)		
Gasch2	Ripper	31.05 (2.6)	20.46 (1.5)	27.10 (0.4)	24.55(0.7)	26.17 (1.3
	BayesNet	35.92 (0.7)	27.90 (1.3)	29.54 (0.8)		
	SVM	33.63(0.6)	33.70(1.4)	29.44 (0.9)		
	KNN	21.01 (2.1)	22.28 (1.2)	23.54 (1.4)		
	C4.5	22.24 (2.3)	23.49 (2.2)	24.55 (1.1)		
Phenotype	Ripper	23.48 (2.5)	17.30 (0.9)	23.79 (1.1)	15.95 (0.3)	19.93 (2.
	BayesNet	24.45 (0.6)	23.63 (1.4)	20.37 (0.9)		
	SVM	25.74(2.1)	22.89 (1.6)	24.70 (1.1)		
	KNN	33.51 (1.7)	30.36 (1.1)	27.03 (0.8)		
	C4.5	30.85 (0.4)	29.08 (1.3)	23.83 (0.6)		
Sequence	Ripper	32.35 (0.8)	21.15 (0.3)	26.10 (1.2)	26.34 (1.1)	26.88 (1.5
-	BayesNet	38.51 (1.3)	30.55 (1.1)	29.14 (0.8)	` '	,
	SVM	28.14 (2.5)	29.08 (1.6)	31.81 (0.3)		
	KNN	26.68 (1.3)	24.84 (0.4)	24.78 (0.9)		
	C4.5	28.11 (4.6)	26.21 (1.7)	22.07 (0.7)		
SPO	Ripper	28.78 (1.3)	20.93 (1.2)	27.52 (0.4)	23.46 (1.4)	24.9 (1.4
*	BayesNet	33.68 (0.8)	28.74 (1.5)	26.62 (1.1)	(1.1)	(
	SVM	28.01 (1.1)	27.93 (1.2)	25.90 (0.5)		

Tabela 6.20: Comparação das Abordagens *Top-Down* e *One-Shot* Utilizando a Medida Micro F1 Hierárquica no Terceiro Nível da Hierarquia

Conjunto de Dados			n Top-Down		Abordagen	n One-Shot
Conjunto de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HM0
	KNN	22.84 (0.4)	24.71 (1.0)	17.78 (0.5)		
	C4.5	23.29(0.5)	26.78 (0.4)	11.27 (0.6)		
Expr	Ripper	23.51 (0.9)	19.06 (1.6)	12.98 (1.0)	$21.30 \ (0.6)$	24.13 (0.5)
	BayesNet	24.79 (1.2)	26.76 (0.8)	18.03 (0.5)		
	SVM	19.32 (0.9)	17.96 (0.4)	14.97 (0.4)		
	KNN	23.33(0.5)	24.86 (0.6)	$18.16 \ (0.6)$		
	C4.5	23.29(0.5)	26.45 (1.0)	$10.51 \ (0.3)$		
CellCycle	Ripper	$21.21 \ (0.3)$	$17.41 \ (0.8)$	13.57 (0.7)	19.81 (0.8)	23.06 (0.5)
	BayesNet	23.19 (1.1)	22.07 (0.9)	16.42 (0.5)		
	SVM	25.31 (0.7)	27.51 (0.5)	19.12 (0.5)		
	KNN	18.42 (0.9)	20.97(0.6)	16.22 (0.5)		
	C4.5	17.56 (0.7)	22.55 (0.9)	$13.70 \ (0.4)$		
Church	Ripper	17.80 (1.0)	13.17 (0.3)	13.99(1.1)	$07.24 \ (0.6)$	09.13 (1.
	BayesNet	19.92 (0.6)	17.69 (0.6)	$13.41 \ (0.9)$		
	SVM	17.61 (0.4)	21.17 (0.8)	15.97 (0.7)		
	KNN	21.20 (0.7)	23.54 (0.7)	16.60 (0.4)		
	C4.5	16.40 (4.0)	$25.01\ (0.6)$	$10.71 \ (0.4)$		
Derisi	Ripper	$19.11 \ (0.6)$	14.46 (0.3)	13.82(0.1)	19.21 (0.5)	22.30 (0.
	BayesNet	20.41 (0.4)	17.83(1.1)	$13.63 \ (0.6)$		
	SVM	21.54 (0.7)	24.74 (0.3)	$16.60 \ (0.5)$		
	KNN	23.44 (0.93)	25.06 (0.6)	17.70 (0.7)		
Eisen	C4.5	23.63(0.75)	26.33(0.5)	10.95 (0.8)		
	Ripper	21.88 (0.82)	19.87(1.3)	12.95 (0.5)	20.53 (0.9)	23.64 (0.
	BayesNet	$23.81 \ (0.72)$	23.14(0.9)	$16.81 \ (0.7)$		
	SVM	25.00(0.66)	27.39 (0.6)	18.64 (0.4)		
	KNN	24.27 (1.1)	25.55 (0.8)	18.01 (0.8)		
	C4.5	24.57(0.6)	$27.37 \ (0.8)$	11.11 (0.3)		
Gasch1	Ripper	22.14(0.9)	17.24 (0.4)	13.14(1.4)	19.39(0.7)	23.54 (0.
	BayesNet	21.67(0.6)	23.73(0.4)	15.09(0.5)		
	SVM	25.50 (0.8)	27.02 (0.8)	19.98 (0.8)		
	KNN	23.31 (0.8)	24.24 (0.5)	17.93 (0.9)		
	C4.5	20.63 (1.5)	26.24 (0.8)	10.72(0.4)		
Gasch2	Ripper	20.14(1.3)	16.07(0.7)	13.38 (0.2)	19.66 (0.8)	22.48 (0.
	BayesNet	21.49(0.3)	19.63(0.6)	$15.70 \ (0.4)$		
	SVM	25.43(0.6)	27.63 (0.4)	17.94(0.2)		
	KNN	16.21 (1.9)	18.96 (1.1)	15.00 (0.9)		
	C4.5	15.85 (1.9)	19.91 (0.9)	14.24 (0.7)		
Phenotype	Ripper	17.46 (1.7)	14.25 (0.4)	11.53 (0.5)	$07.01 \ (0.2)$	12.31 (3.
	BayesNet	$20.64\ (1.2)$	18.61 (1.0)	13.91 (1.0)		
	SVM	18.98 (1.8)	20.26(1.1)	15.37(1.1)		
	KNN	21.15 (1.1)	22.52 (0.7)	14.59 (0.5)		
	C4.5	23.56 (0.9)	26.16 (1.0)	10.56 (0.4)		
Sequence	Ripper	21.60 (1.3)	16.96 (0.5)	12.15 (0.5)	21.70 (0.6)	23.35 (0.
-	BayesNet	24.77 (0.3)	25.66 (1.4)	16.25 (0.4)	` /	`
	SVM	19.57 (1.8)	19.22 (0.9)	16.56 (0.4)		
	KNN	20.30 (0.6)	22.63 (0.8)	16.43 (0.4)		
	C4.5	20.42 (3.8)	25.62 (0.7)	10.26 (0.4) $10.26 (0.4)$		
SPO	Ripper	18.83 (1.4)	15.86 (0.7)	14.33 (0.6)	18.90 (0.3)	22.21 (0.
	BayesNet	20.54 (0.7)	20.54 (0.5)	14.88 (0.6)	_===== (0.0)	(0.
	SVM	23.85 (0.8)	25.77 (0.7)	17.09 (0.5)		

Tabela 6.21: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Macro F1 Hierárquica no Terceiro Nível da Hierarquia

Conjunto de Dados		Abordage	em Top-Down		Abordagen	One-Shot
Jonjunto de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	$03.34 \ (0.7)$	03.18 (1.3)	$02.37 \ (0.10)$		
	C4.5	$02.42 \ (0.4)$	$02.31 \ (0.5)$	$02.37 \ (0.04)$		
Expr	Ripper	$03.34 \ (0.7)$	02.08(0.1)	$02.47 \ (0.04)$	$01.90 \ (0.07)$	02.03 (0.03
	BayesNet	$03.45 \ (0.7)$	$03.02 \ (0.4)$	$02.56 \ (0.30)$		
	SVM	01.67 (0.1)	01.66 (0.07)	01.89 (0.07)		
	KNN	02.64 (0.5)	02.30 (0.5)	02.30 (0.10)		
	C4.5	02.29(0.6)	$02.12 \ (0.4)$	$02.30\ (0.05)$		
CellCycle	Ripper	$02.40 \ (0.4)$	01.88(0.1)	$02.59 \ (0.30)$	01.91 (0.2)	01.94 (0.3
	BayesNet	$02.38 \ (0.2)$	$02.13 \ (0.08)$	$02.36 \ (0.06)$		
	SVM	$03.26 \ (0.8)$	03.04 (1.1)	02.90 (0.60)		
	KNN	01.74 (0.1)	01.71 (0.08)	02.03 (0.08)		
	C4.5	01.69 (0.09)	01.96 (0.4)	02.06 (0.09)		
Church	Ripper	02.07(0.2)	01.27(0.1)	01.97(0.30)	$01.88 \; (0.06)$	02.03 (0.4
	BayesNet	$02.29\ (0.2)$	$01.78 \ (0.02)$	01.99(0.10)		
	SVM	$01.70 \ (0.06)$	$01.81 \ (0.08)$	$02.16 \ (0.10)$		
	KNN	01.96 (0.06)	01.95 (0.2)	02.10 (0.04)		
	C4.5	01.73 (0.3)	01.85 (0.1)	02.17 (0.06)		
Derisi	Ripper	01.92 (0.07)	01.50 (0.05)	$02.28 \ (0.04)$	01.75 (0.05)	01.80 (0.1
	BayesNet	02.24 (0.04)	01.88 (0.1)	02.09 (0.07)		
	SVM	01.95(0.2)	01.85 (0.05)	02.16 (0.08)		
	KNN	02.81 (0.3)	02.91 (0.5)	02.77 (0.30)		
	C4.5	02.52 (0.6)	02.25 (0.3)	02.35 (0.10)		
Eisen	Ripper	02.53 (0.5)	02.79 (0.2)	02.68 (0.20)	01.96 (0.1)	02.10 (0.0
Disen	BayesNet	02.73 (0.4)	02.49 (0.3)	02.99 (0.40)	(-)	- (
	SVM	03.63 (0.8)	$03.96 \ (1.1)$	02.99 (0.40)		
	KNN	04.05 (1.3)	03.28 (0.3)	03.07 (0.90)		
	C4.5	02.81 (0.4)	02.46 (0.4)	02.37 (0.06)		
Gasch1	Ripper	03.36 (0.8)	02.03 (0.3)	02.58 (0.30)	01.96 (0.2)	01.98 (0.01
	BayesNet	02.59 (0.1)	02.51 (0.1)	02.29 (0.05)	,	(
	SVM	05.10 (1.0)	03.76 (1.2)	03.30 (0.60)		
	KNN	03.05 (1.2)	02.39 (0.4)	02.23 (0.10)		
	C4.5	01.89 (0.1)	02.06 (0.2)	02.24 (0.02)		
Gasch2	Ripper	02.12 (0.2)	01.76 (0.2)	02.28 (0.03)	01.74 (0.06	01.79 (0.0
	BayesNet	02.28 (0.09)	01.93 (0.1)	02.23 (0.07)	()	(
	SVM	02.67 (0.3)	03.19 (0.8)	02.63 (0.50)		
		. ,	. ,	, ,		
	KNN	01.61.(0.4)	01.55 (0.1)	02.07 (0.50)		
	KNN C4.5	01.61 (0.4) 01.33 (0.1)	01.55 (0.1) 01.76 (0.2)	02.07 (0.50) 02.13 (0.01)		
Phenotype	C4.5	01.33 (0.1)	01.76 (0.2)	02.13 (0.01)	02.06 (0.03)	01.99 (0.0
Phenotype	C4.5 Ripper	01.33 (0.1) 01.56 (0.2)	01.76 (0.2) 01.33 (0.3)	02.13 (0.01) 01.96 (0.06)	02.06 (0.03)	01.99 (0.0
Phenotype	C4.5 Ripper BayesNet	01.33 (0.1) 01.56 (0.2) 01.98 (0.1)	01.76 (0.2) 01.33 (0.3) 01.93 (0.08)	02.13 (0.01) 01.96 (0.06) 02.05 (0.10)	02.06 (0.03)	01.99 (0.0
Phenotype	C4.5 Ripper BayesNet SVM	01.33 (0.1) 01.56 (0.2) 01.98 (0.1) 01.86 (0.5)	01.76 (0.2) 01.33 (0.3) 01.93 (0.08) 01.66 (0.1)	02.13 (0.01) 01.96 (0.06) 02.05 (0.10) 02.12 (0.10)	02.06 (0.03)	01.99 (0.0
Phenotype	C4.5 Ripper BayesNet SVM KNN	01.33 (0.1) 01.56 (0.2) 01.98 (0.1) 01.86 (0.5) 02.23 (0.3)	01.76 (0.2) 01.33 (0.3) 01.93 (0.08) 01.66 (0.1) 01.99 (0.09)	02.13 (0.01) 01.96 (0.06) 02.05 (0.10) 02.12 (0.10) 02.31 (0.40)	02.06 (0.03)	01.99 (0.0
	C4.5 Ripper BayesNet SVM KNN C4.5	01.33 (0.1) 01.56 (0.2) 01.98 (0.1) 01.86 (0.5) 02.23 (0.3) 02.03 (0.06)	01.76 (0.2) 01.33 (0.3) 01.93 (0.08) 01.66 (0.1) 01.99 (0.09) 01.98 (0.1)	02.13 (0.01) 01.96 (0.06) 02.05 (0.10) 02.12 (0.10) 02.31 (0.40) 02.29 (0.06)		
Phenotype Sequence	C4.5 Ripper BayesNet SVM KNN C4.5 Ripper	01.33 (0.1) 01.56 (0.2) 01.98 (0.1) 01.86 (0.5) 02.23 (0.3) 02.03 (0.06) 02.11 (0.4)	01.76 (0.2) 01.33 (0.3) 01.93 (0.08) 01.66 (0.1) 01.99 (0.09) 01.98 (0.1) 01.70 (0.08)	02.13 (0.01) 01.96 (0.06) 02.05 (0.10) 02.12 (0.10) 02.31 (0.40) 02.29 (0.06) 02.27 (0.10)	02.06 (0.03) 01.81 (0.08)	
	C4.5 Ripper BayesNet SVM KNN C4.5 Ripper BayesNet	01.33 (0.1) 01.56 (0.2) 01.98 (0.1) 01.86 (0.5) 02.23 (0.3) 02.03 (0.06) 02.11 (0.4) 02.41 (0.04)	01.76 (0.2) 01.33 (0.3) 01.93 (0.08) 01.66 (0.1) 01.99 (0.09) 01.98 (0.1) 01.70 (0.08) 02.16 (0.08)	02.13 (0.01) 01.96 (0.06) 02.05 (0.10) 02.12 (0.10) 02.31 (0.40) 02.29 (0.06) 02.27 (0.10) 02.20 (0.09)		
	C4.5 Ripper BayesNet SVM KNN C4.5 Ripper BayesNet SVM	01.33 (0.1) 01.56 (0.2) 01.98 (0.1) 01.86 (0.5) 02.23 (0.3) 02.03 (0.06) 02.11 (0.4) 02.41 (0.04) 04.40 (1.0)	01.76 (0.2) 01.33 (0.3) 01.93 (0.08) 01.66 (0.1) 01.99 (0.09) 01.98 (0.1) 01.70 (0.08) 02.16 (0.08) 02.42 (0.3)	02.13 (0.01) 01.96 (0.06) 02.05 (0.10) 02.12 (0.10) 02.31 (0.40) 02.29 (0.06) 02.27 (0.10) 02.20 (0.09) 03.18 (0.60)		
	C4.5 Ripper BayesNet SVM KNN C4.5 Ripper BayesNet SVM KNN	01.33 (0.1) 01.56 (0.2) 01.98 (0.1) 01.86 (0.5) 02.23 (0.3) 02.03 (0.06) 02.11 (0.4) 02.41 (0.04) 04.40 (1.0) 01.75 (0.06)	01.76 (0.2) 01.33 (0.3) 01.93 (0.08) 01.66 (0.1) 01.99 (0.09) 01.98 (0.1) 01.70 (0.08) 02.16 (0.08) 02.42 (0.3) 01.68 (0.05)	02.13 (0.01) 01.96 (0.06) 02.05 (0.10) 02.12 (0.10) 02.31 (0.40) 02.29 (0.06) 02.27 (0.10) 02.20 (0.09) 03.18 (0.60) 01.97 (0.07)		
Sequence	C4.5 Ripper BayesNet SVM KNN C4.5 Ripper BayesNet SVM KNN C4.5	01.33 (0.1) 01.56 (0.2) 01.98 (0.1) 01.86 (0.5) 02.23 (0.3) 02.03 (0.06) 02.11 (0.4) 02.41 (0.04) 04.40 (1.0) 01.75 (0.06) 01.98 (0.5)	01.76 (0.2) 01.33 (0.3) 01.93 (0.08) 01.66 (0.1) 01.99 (0.09) 01.98 (0.1) 01.70 (0.08) 02.16 (0.08) 02.42 (0.3) 01.68 (0.05) 01.82 (0.06)	02.13 (0.01) 01.96 (0.06) 02.05 (0.10) 02.12 (0.10) 02.31 (0.40) 02.29 (0.06) 02.27 (0.10) 02.20 (0.09) 03.18 (0.60) 01.97 (0.07) 02.14 (0.05)	01.81 (0.08)	02.03 (0.4
	C4.5 Ripper BayesNet SVM KNN C4.5 Ripper BayesNet SVM KNN	01.33 (0.1) 01.56 (0.2) 01.98 (0.1) 01.86 (0.5) 02.23 (0.3) 02.03 (0.06) 02.11 (0.4) 02.41 (0.04) 04.40 (1.0) 01.75 (0.06)	01.76 (0.2) 01.33 (0.3) 01.93 (0.08) 01.66 (0.1) 01.99 (0.09) 01.98 (0.1) 01.70 (0.08) 02.16 (0.08) 02.42 (0.3) 01.68 (0.05)	02.13 (0.01) 01.96 (0.06) 02.05 (0.10) 02.12 (0.10) 02.31 (0.40) 02.29 (0.06) 02.27 (0.10) 02.20 (0.09) 03.18 (0.60) 01.97 (0.07)		01.99 (0.05 02.03 (0.4 01.70 (0.06

Tabela 6.22: Comparação das Abordagens *Top-Down* e *One-Shot* Utilizando a Medida Micro F1 Hierárquica Ponderada no Terceiro Nível da Hierarquia

Conjunto de Dados			n Top-Down		1	n One-Shot
Conjunto de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	21.76 (1.6)	20.72(1.2)	16.37 (0.5)		
	C4.5	21.05(1.1)	20.39(0.7)	11.73 (0.9)		
Expr	Ripper	24.61 (1.1)	22.73(0.9)	13.98 (0.9)	18.18 (0.6)	19.59 (0.3
	BayesNet	25.54(2.0)	$26.22\ (2.2)$	$17.86 \ (1.8)$		
	SVM	23.68 (1.4)	23.97 (1.4)	20.37 (1.2)		
	KNN	23.45 (2.5)	21.78(1.5)	16.99(0.6)		
	C4.5	$19.44 \ (0.8)$	19.00(0.6)	10.79(0.7)		
CellCycle	Ripper	21.48(1.0)	22.27(1.0)	14.00(1.2)	17.20 (0.6)	18.45 (0.2)
	BayesNet	24.93(2.1)	$22.41\ (1.4)$	16.85 (0.4)		
	SVM	24.46 (1.8)	25.22 (2.8)	18.80 (1.3)		
	KNN	16.49 (0.9)	16.66 (0.4)	15.55 (0.7)		
	C4.5	$16.44 \ (0.7)$	17.25 (0.2)	13.52 (0.8)		
Church	Ripper	18.10 (1.4)	$21.86 \ (0.9)$	17.73(1.6)	07.29 (0.5)	09.19 (1.3
	BayesNet	21.46(1.0)	16.92(1.2)	$14.11 \ (0.6)$		
	SVM	16.79 (1.0)	17.32 (0.5)	15.10 (0.8)		
	KNN	18.42 (0.8)	17.76 (0.6)	14.61 (0.7)		
	C4.5	16.03(3.5)	18.05 (0.5)	10.55 (0.7)		
Derisi	Ripper	$18.10 \ (0.4)$	21.45 (0.6)	15.20(0.5)	16.82 (0.3)	18.20 (0.5
	BayesNet	19.60(1.2)	17.42(1.8)	13.62 (0.5)		
	SVM	17.99 (0.5)	18.22 (0.7)	$14.56 \ (0.6)$		
	KNN	23.77 (2.1)	22.03 (1.9)	17.20 (1.2)		
Eisen	C4.5	20.04(0.6)	19.91(0.7)	11.30 (1.1)		
	Ripper	24.48(2.1)	21.91(1.8)	14.84 (1.0)	18.20 (1.0)	19.64 (0.6
	BayesNet	26.13(3.0)	21.73(1.7)	17.75(1.6)		
	SVM	26.12(1.8)	27.47 (1.7)	18.94 (1.8)		
	KNN	23.58 (1.5)	22.33 (0.3)	17.16 (1.1)		
	C4.5	21.33(0.7)	20.16(0.9)	11.54 (0.8)		
Gasch1	Ripper	23.34(2.2)	22.54(1.6)	13.83(1.3)	17.03 (0.5)	19.21 (1.0
	BayesNet	22.29(1.6)	21.35 (0.4)	$15.46 \ (0.7)$		
	SVM	29.89 (2.1)	29.48 (1.9)	21.03 (1.7)		
	KNN	21.80 (2.0)	19.99 (0.2)	16.47 (1.2)		
	C4.5	18.90 (0.3)	19.09 (0.7)	10.68 (0.6)		
Gasch2	Ripper	21.25(2.3)	22.93(1.5)	$14.34 \ (0.8)$	17.03 (0.6)	18.40 (0.3
	BayesNet	20.60(1.1)	19.46 (0.7)	14.95 (0.5)		
	SVM	22.13(0.4)	$23.10 \ (0.8)$	16.19 (0.4)		
	KNN	16.99 (2.6)	18.90 (1.1)	16.00 (2.0)		
	C4.5	16.19 (2.1)	18.92 (0.8)	14.39 (0.6)		
Phenotype	Ripper	17.61 (0.9)	24.49 (0.6)	$16.31\ (0.8)$	07.94 (0.5)	12.15 (2.7
	BayesNet	19.08 (1.4)	17.58 (0.9)	13.28(1.0)		
	SVM	19.49 (1.8)	19.92 (0.5)	15.16 (1.3)		
	KNN	19.19 (1.3)	18.46 (0.5)	13.53 (0.3)		
	C4.5	20.71 (0.6)	20.79 (1.7)	11.61 (1.0)		
Sequence	Ripper	22.54 (1.8)	20.87 (0.7)	12.95 (0.4)	18.07 (0.5)	20.27 (0.8
	BayesNet	26.61 (2.5)	23.12 (1.1)	16.69 (0.9)		
	SVM	24.06 (3.0)	25.99 (1.3)	22.38(0.7)		
	KNN	17.63 (0.4)	17.34 (0.4)	14.21 (0.4)		
	C4.5	18.54 (2.7)	18.30 (0.6)	09.80 (0.3)		
SPO	Ripper	19.61 (1.9)	22.46 (1.2)	15.10 (0.7)	16.62 (0.3)	18.16 (0.3
	BayesNet	19.29 (0.2)	18.31 (0.8)	14.38 (0.6)	<u> </u>	`
	SVM	19.21 (0.6)	19.46 (0.7)	14.87 (0.7)		

Tabela 6.23: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Macro F1 Hierárquica Ponderada no Terceiro Nível da Hierarquia

Conjunto de Dados			em Top-Down		-	n One-Shot
Conjunto de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	07.48(1.2)	$07.30\ (1.6)$	07.00(1.7)		
	C4.5	06.95(1.4)	$06.94\ (1.9)$	04.08 (1.4)		
Expr	Ripper	09.48(1.4)	12.30(2.2)	$05.41 \ (0.8)$	05.06 (1.8)	05.30 (1.3)
	BayesNet	09.57(2.0)	10.02 (2.7)	08.16 (3.0)		
	SVM	04.98 (0.9)	15.34 (2.4)	20.82 (3.6)		
	KNN	06.37(1.6)	$07.14\ (1.6)$	06.75(1.5)		
	C4.5	05.85(1.9)	04.85 (0.8)	03.49 (0.2)		
CellCycle	Ripper	$06.51\ (1.0)$	12.90 (2.1)	05.93(1.1)	03.67 (0.8)	04.46 (0.9)
	BayesNet	06.83(1.6)	07.55 (1.5)	09.14(2.2)		
	SVM	09.40 (1.2)	11.47 (1.5)	08.72 (1.7)		
	KNN	03.58 (0.8)	03.71 (0.5)	05.51 (1.6)		
	C4.5	03.65 (0.9)	03.79(0.7)	$03.73 \ (0.8)$		
Church	Ripper	04.38(0.9)	$17.68 \ (1.5)$	12.55 (1.4)	03.75 (1.3)	03.54 (0.9
	BayesNet	04.52(1.2)	04.86(1.1)	09.09(2.1)		
	SVM	04.94(1.4)	$05.34\ (0.9)$	04.26 (1.3)		
	KNN	03.64 (1.1)	03.82 (1.2)	05.08 (0.8)		
	C4.5	04.19 (1.6)	04.89 (1.6)	02.60 (0.3)		
Derisi	Ripper	04.50 (0.9)	15.22 (2.3)	08.13 (1.1)	03.16 (1.3)	05.24 (0.7
	BayesNet	05.79 (0.9)	06.03 (2.0)	09.85 (2.4)	, ,	•
	SVM	05.72 (1.7)	04.90 (1.3)	03.36 (1.4)		
	KNN	06.86 (1.8)	08.35 (1.7)	08.14 (2.3)		
	C4.5	05.46 (0.6)	06.05 (1.7)	04.37 (1.3)		
Eisen	Ripper	09.44 (2.4)	11.07 (4.1)	07.19 (1.8)	04.87 (2.1)	05.59 (1.5
	BayesNet	09.29 (1.6)	08.35 (1.6)	11.49 (2.9)	,	
	SVM	10.61 (3.0)	13.95 (3.4)	08.05 (0.5)		
	KNN	09.89 (1.6)	08.81 (0.8)	09.08 (2.1)		
	C4.5	07.11 (1.5)	05.87 (1.0)	03.55 (0.9)		
Gasch1	Ripper	09.31 (3.7)	13.41 (2.5)	05.59 (1.1)	03.89 (0.7)	04.71 (1.5
Cascili	BayesNet	08.14 (1.0)	06.84 (0.7)	11.35 (2.0)	00.00 (0.1)	01.71 (1.0
	SVM	13.41 (2.3)	16.55 (3.4)	15.96 (2.9)		
	KNN	06.87 (2.6)	07.14 (1.1)	07.25 (3.4)		
	C4.5	06.37 (2.0) $06.32 (2.0)$	04.16 (0.3)	02.94 (0.6)		
Gasch2	Ripper	07.40 (2.6)	15.94 (1.1)	08.62 (1.6)	03.65 (1.1)	04.94 (2.2
Gasciiz	BayesNet	06.82 (1.6)	06.37 (0.7)	07.54 (1.7)	05.05 (1.1)	04.34 (2.2
	SVM	07.69 (1.2)	09.73 (1.6)	06.53 (1.3)		
				` ′		
	KNN	03.06 (0.9)	07.27 (1.7)	08.43 (5.0)		
Dhamatana	C4.5	02.84 (0.4)	07.53 (1.1)	06.22 (2.5)	09.44 (0.4)	09.05 /1.5
Phenotype	Ripper	03.13 (0.8)	18.47 (1.7)	21.51 (3.2)	03.44 (0.4)	03.85 (1.3
	BayesNet SVM	04.26 (1.7) 05.00 (1.8)	03.86 (0.9) 08.49 (1.3)	03.66 (1.1) 04.60 (0.6)		
	KNN	04.04 (0.8)	04.33 (1.6)	04.12 (0.9)		
G	C4.5	05.25 (1.2)	06.95 (2.0)	04.07 (0.7)	09.00 (1.9)	05 71 (0.5
Sequence	Ripper	06.94 (1.5)	11.68 (1.4)	03.85 (0.7)	03.80 (1.3)	05.71 (0.5
	BayesNet	09.07 (1.5)	08.32 (0.9)	08.97 (2.0)		
	SVM	06.94 (1.5)	17.19 (1.9)	20.94 (2.0)		
	KNN	03.12 (1.1)	04.06 (1.4)	03.99 (1.1)		
	C4.5	04.95 (1.1)	03.70 (1.0)	$02.66 \ (0.5)$		
SPO	Ripper	05.62 (0.9)	$15.34 \ (2.5)$	06.67 (1.5)	03.97 (2.2)	03.42 (1.2)
	BayesNet	05.45 (0.7)	04.19 (1.1)	06.48 (1.3)		
	SVM	05.62(0.9)	05.48(1.2)	04.30 (0.8)		

6.3.4 Resultados no Quarto Nível das Hierarquias

A menor quantidade de erros de predição no quarto nível das hierarquias (Tabela 6.24) foi cometida, assim como no terceiro nível, pela técnica HMC-LP. Isso ocorreu porque a medida H-Loss não considera erros cometidos nas classes filhas das classes preditas erroneamente. Sendo assim, erros não são propagados para os níveis inferiores, o que faz com que, no geral, o valor dessa medida diminua conforme o nível da hierarquia se torna mais profundo. O melhor desempenho na medida H-Loss foi obtido pela técnica HMC-LP utilizando o algoritmo Ripper. Foram detectadas diferenças estatisticamente significantes entre a maioria dos algoritmos da técnica HMC-CT e os algoritmos das outras técnicas da abordagem *Top-Down*, bem como entre as técnicas C4.5H, Clus-HMC e os algoritmos utilizados nas técnicas da abordagem *Top-Down*.

Considerando a medida F1 hierárquica (Tabela 6.25), os melhores resultados foram obtidos pela técnica HMC-BR utilizando o algoritmo BayesNet. As diferenças entre a técnica HMC-BR e as técnicas HMC-LP e HMC-CT mostraram-se estatisticamente significantes na maioria dos algoritmos, bem como a diferença entre a técnica HMC-BR e as técnicas C4.5H e Clus-HMC.

Conforme pode ser observado nas tabelas relativas às medidas micro F1 hierárquicas não ponderada e ponderada (Tabelas 6.26 e 6.28), novamente os melhores desempenhos foram obtidos pela técnica HMC-LP, o que sugere que a técnica apresenta um bom desempenho quando avaliada por medidas baseadas em distância, pois a técnica apresentou os melhores desempenhos no segundo, terceiro e quarto níveis da hierarquia. Na medida micro F1 hierarquia, o melhor desempenho foi obtido pela técnica HMC-LP utilizando o algoritmo C4.5, apresentando diferenças estatisticamente significantes na comparação com a maioria dos outros algoritmos, com exceção dos algoritmos da técnica HMC-BR. Na medida micro F1 hierarquia ponderada, a técnica HMC-LP, utilizando o algoritmo SVM, obteve o melhor desempenho, apresentando resultados estatisticamente superiores aos apresentados pela maioria do outros algoritmos, com exceção dos algoritmos da técnica HMC-BR.

Nas medidas macro F1 hierárquicas não ponderada e ponderada (Tabelas 6.27 e 6.29), novamente os algoritmos das técnicas HMC-LP e HMC-CT também apresentaram alguns dos melhores resultados, devido às características de avaliação dessas medidas, conforme já mencionado. Na medida macro F1 hierárquica, o melhor desempenho foi obtido pela técnica HMC-BR utilizando o algoritmo SVM. Porém, a diferença dos resultados não mostrou-se estatisticamente significante na maioria dos casos. Na medida macro F1 hierárquica ponderada, os melhores resultados foram obtidos pela técnica HMC-LP utilizando o algoritmo Ripper. A diferença entre os resultados dos algoritmos, no entanto, mostrou-se estatisticamente significante em apenas alguns casos.

Tabela 6.24: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

H-Loss no Quarto Nível da Hierarquia

Conjunto de Dados			em Top-Down		-	n One-Shot
Conjunto de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HM0
	KNN	0.14 (0.02)	$0.12 \ (0.02)$	$0.90 \ (0.09)$		
	C4.5	$0.21 \ (0.04)$	$0.13 \ (0.02)$	2.85 (0.20)		
Expr	Ripper	0.14 (0.02)	$0.04\ (0.01)$	2.29 (0.20)	0.38 (0.05)	0.23 (0.03)
	BayesNet	$0.31\ (0.03)$	$0.12 \ (0.02)$	$0.82 \ (0.04)$		
	SVM	0.07 (0.02)	0.05 (0.009)	0.79 (0.20)		
	KNN	0.15 (0.03)	0.13 (0.03)	0.89 (0.10)		
	C4.5	0.15 (0.03)	$0.11 \ (0.01)$	3.12(0.20)		
CellCycle	Ripper	$0.12 \ (0.02)$	$0.03\ (0.01)$	$2.01\ (0.20)$	0.41 (0.01)	0.20 (0.01
	BayesNet	0.14 (0.03)	$0.08 \; (0.01)$	$0.68 \; (0.05)$		
	SVM	0.14 (0.03)	0.12 (0.02)	1.03 (0.08)		
	KNN	0.10 (0.01)	0.08 (0.03)	0.90 (0.10)		
	C4.5	0.11 (0.03)	0.07 (0.02)	1.21(0.10)		
Church	Ripper	0.09(0.03)	$0.007\ (0.009)$	1.09(0.10)	5.82 (0.5)	3.85 (1.0)
	BayesNet	0.32 (0.05)	$0.08 \; (0.04)$	0.70(0.20)		
	SVM	0.12 (0.03)	$0.06 \ (0.01)$	0.99(0.20)		
	KNN	0.09 (0.01)	0.09 (0.02)	0.80 (0.09)		
	C4.5	0.05 (0.03)	0.11 (0.01)	2.61 (0.30)		
Derisi	Ripper	0.08 (0.02)	$0.01\ (0.005)$	1.36 (0.20)	0.32 (0.03)	0.20 (0.04
	BayesNet	0.28 (0.06)	0.08 (0.02)	0.72(0.10)		
	SVM	$0.11\ (0.01)$	$0.10 \ (0.02)$	1.03 (0.07)		
	KNN	0.18 (0.01)	0.11 (0.02)	1.10 (0.10)		
	C4.5	0.19 (0.05)	0.12 (0.03)	2.89 (0.40)		
Eisen	Ripper	0.16 (0.04)	$0.05\ (0.04)$	2.18 (0.10)	0.40 (0.08)	0.20 (0.03
	BayesNet	0.28 (0.04)	0.08 (0.02)	1.16 (0.10)	, ,	`
	SVM	0.17 (0.04)	0.13 (0.03)	1.14 (0.20)		
	KNN	0.16 (0.04)	0.10 (0.02)	1.10 (0.10)		
	C4.5	0.22 (0.04)	0.12 (0.03)	2.90 (0.30)		
Gasch1	Ripper	0.13 (0.02)	0.03 (0.008)	2.37 (0.40)	0.50 (0.07)	0.20 (0.02
	BayesNet	0.36 (0.07)	0.09 (0.02)	0.74 (0.10)		,
	SVM	0.15 (0.03)	0.11 (0.02)	0.93 (0.05)		
	KNN	0.14 (0.03)	0.09 (0.02)	0.97 (0.10)		
	C4.5	0.14 (0.04)	0.11 (0.02)	2.93 (0.50)		
Gasch2	Ripper	0.13 (0.03)	0.03 (0.01)	1.59 (0.20)	0.37 (0.09)	0.23 (0.05
	BayesNet	0.20 (0.03)	0.08 (0.01)	0.81 (0.10)		- ((
	SVM	0.15 (0.01)	0.12 (0.03)	0.96 (0.10)		
	KNN	0.02 (0.007)	0.04 (0.03)	0.74 (0.30)		
	C4.5	0.03 (0.02)	0.06 (0.05)	$1.31 \ (0.50)$		
Phenotype	Ripper	0.04 (0.02)	0.001 (0.003)	1.35 (0.20)	4.82 (0.39)	1.56 (1.24
	BayesNet	0.03 (0.03)	0.05 (0.05)	0.50 (0.10)	(0.00)	
	SVM	0.06 (0.04)	0.08 (0.02)	1.09 (0.30)		
	KNN	0.17 (0.02)	0.09 (0.02)	1.34 (0.20)		
	C4.5	0.17 (0.02) $0.20 (0.02)$	$0.09 \ (0.02)$ $0.12 \ (0.02)$	3.09 (0.04)		
Sequence	Ripper	$0.12 \ (0.02)$	0.03 (0.02)	2.35 (0.20)	0.34 (0.06)	0.23 (0.01
Sequence	BayesNet	$0.12 \ (0.02)$ $0.22 \ (0.03)$	0.08 (0.02)	0.86 (0.10)	0.01 (0.00)	0.23 (0.01
	SVM	0.06 (0.02)	0.05 (0.02)	0.78 (0.07)		
	KNN	0.09 (0.02)	0.06 (0.004)	0.86 (0.10)		
	C4.5	0.09 (0.02) $0.13 (0.07)$	$0.06 \ (0.004)$ $0.10 \ (0.03)$	0.86 (0.10) $2.68 (0.30)$		
SPO		$0.13 \ (0.07)$ $0.09 \ (0.01)$	0.10 (0.03) 0.03 (0.01)		0.36 (0.1)	0.21 (0.05
DI O	Ripper	, ,	0.03 (0.01)	$1.46 (0.20) \\ 0.53 (0.10)$	0.36 (0.1)	0.21 (0.05
	BayesNet	0.21(0.03)	0.08 / 0.01	0 23 (0 10)		

Tabela 6.25: Comparação das Abordagens Top-Down e One-Shot Utilizando a Medida F1 Hierárquica no Quarto Nível da Hierarquia

2 1 5 1		Abordagem	Top-Down		Abordagen	n One-Shot
Conjunto de Dados		$\mathrm{HMC}\text{-}\mathrm{BR}$	HMC-LP	$\operatorname{HMC-CT}$	C4.5H	Clus-HMC
	KNN	31.48 (1.2)	30.04 (1.2)	24.88 (0.7)		
	C4.5	28.91 (0.7)	26.95(1.7)	19.92 (0.8)		
Expr	Ripper	31.95(1.5)	21.20(1.2)	21.67(1.0)	23.34(1.1)	25.40 (1.6)
	BayesNet	$36.40\ (1.3)$	33.26(1.4)	26.80(1.5)		
	SVM	23.55 (1.0)	21.63 (1.2)	22.69 (1.3)		
	KNN	32.30 (2.2)	29.52 (1.5)	25.02 (1.3)		
	C4.5	25.78(1.2)	23.98(0.4)	17.94 (0.8)		
CellCycle	Ripper	28.71 (0.4)	19.05(1.1)	20.43(1.1)	22.21 (0.8)	21.97 (0.6)
	BayesNet	33.04(1.3)	27.99(1.0)	25.19(0.6)		
	SVM	32.54 (1.6)	32.28 (2.5)	27.76 (0.7)		
	KNN	23.72 (1.8)	21.28 (0.9)	20.99 (1.1)		
	C4.5	24.35(1.9)	22.30(1.0)	19.25 (1.2)		
Church	Ripper	24.72(0.8)	13.17(0.4)	19.70(1.0)	13.46 (0.7)	15.65 (1.4)
	BayesNet	$32.67 \ (1.0)$	22.11(1.1)	21.63(1.2)		
	SVM	23.57(1.8)	22.61 (0.4)	21.38 (1.8)		
	KNN	25.45 (1.2)	23.25 (1.1)	19.70 (0.3)		
	C4.5	20.47(3.9)	22.37(1.1)	17.47(0.4)		
Derisi	Ripper	24.81(1.7)	14.72(0.4)	19.69 (0.6)	20.37(0.5)	21.21 (2.0
	BayesNet	$31.41\ (0.3)$	23.42(3.5)	20.93 (1.7)		
	SVM	24.52 (0.8)	22.55(1.2)	21.35 (0.7)		
	KNN	34.93 (1.1)	31.04 (1.8)	26.22 (1.4)		
	C4.5	27.10 (1.2)	26.13 (1.4)	19.06 (1.5)		
Eisen	Ripper	31.54 (1.9)	23.58 (3.2)	21.56 (1.2)	23.49 (1.5)	24.35 (2.3
	BayesNet	36.93 (1.7)	29.08 (2.0)	26.24 (1.7)	,	`
	SVM	35.63 (1.9)	35.26 (1.4)	28.50 (2.1)		
	KNN	34.32 (1.8)	31.33 (1.0)	26.79 (0.8)		
	C4.5	28.93 (1.5)	25.70 (0.5)	19.95 (0.2)		
Gasch1	Ripper	30.60 (0.9)	18.64 (0.7)	21.30 (1.7)	23.21 (0.7)	23.46 (0.7
	BayesNet	35.01 (1.3)	28.91 (1.4)	24.25 (1.4)	,	
	SVM	37.26 (1.4)	35.03 (1.9)	30.10 (1.3)		
	KNN	30.56 (0.9)	27.55 (1.1)	23.79 (1.4)		
	C4.5	26.46 (1.3)	24.47 (1.1)	18.41 (0.8)		
Gasch2	Ripper	28.17 (2.7)	17.10 (1.6)	20.88 (0.6)	21.12 (0.7)	22.99 (1.5
	BayesNet	32.69 (0.6)	25.13 (1.4)	24.48 (1.5)	,	
	SVM	30.36 (0.7)	30.17 (1.7)	24.39 (0.9)		
	KNN	18.60 (1.7)	19.21 (1.0)	18.86 (1.5)		
	C4.5	20.14 (2.2)	20.22 (2.1)	19.36 (0.9)		
Phenotype	Ripper	21.18 (2.7)	14.19 (0.8)	17.86 (1.2)	12.90 (0.3)	16.73 (2.5
	BayesNet	21.62 (0.7)	20.33 (1.3)	16.91 (1.5)		
	SVM	23.26 (2.0)	19.77 (1.5)	19.79 (0.7)		
	KNN	30.76 (1.5)	27.58 (1.0)	21.82 (0.5)		
	C4.5	27.74 (0.6)	26.15 (1.3)	18.65 (0.3)		
Sequence	Ripper	29.61 (0.8)	18.06 (0.3)	20.49 (1.4)	23.12 (1.2)	23.57 (1.3
	BayesNet	35.54 (1.1)	27.40 (0.9)	24.21 (0.9)	20:12 (1:2)	_5.51 (1.6
	SVM	26.06 (2.3)	26.01 (1.5)	25.66 (0.5)		
	KNN	24.04 (1.4)	22.33 (0.4)	19.88 (0.7)		
	IXININ	44.U4 (1.4)	` '	` ′		
		25 25 (4.5)	93 37 (1 5)	17/18 (0.6)		
SDO	C4.5	25.25 (4.5)	23.37 (1.5)	17.48 (0.6)	20 27 (1 2)	21 04 (1 5
SPO		25.25 (4.5) 25.67 (1.2) 30.58 (0.9)	23.37 (1.5) 17.49 (1.0) 25.87 (1.5)	17.48 (0.6) 21.08 (0.5) 22.95 (1.1)	20.27 (1.3)	21.84 (1.3

Tabela 6.26: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Micro F1 Hierárquica no Quarto Nível da Hierarquia

	_	-		-	
	HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
KNN	19.28 (0.3)	21.05 (0.7)	$12.28 \ (0.4)$		
	` ′	` ,	` '		
				16.78 (0.4)	19.98 (0.6
	` ′	` ′	` '		
SVM	16.04 (0.7)	14.41 (0.3)	09.93 (0.6)		
KNN	19.46 (0.4)	21.16 (0.7)	12.37 (0.5)		
C4.5	19.69 (0.5)	23.12(1.0)	$06.64 \ (0.3)$		
Ripper	17.70 (0.2)	14.07 (0.7)	$08.31 \ (0.5)$	15.39 (0.6)	19.07 (0.4
BayesNet	19.29 (0.9)	$18.24 \ (0.8)$	11.15 (0.4)		
SVM	21.49 (0.7)	23.63 (0.4)	13.49 (0.5)		
KNN	15.21 (0.7)	17.75 (0.7)	11.11 (0.5)		
C4.5	$14.46 \ (0.7)$	$18.80 \ (0.9)$	$09.03 \ (0.3)$		
Ripper	14.48 (0.7)	$10.36 \ (0.2)$	08.77(0.9)	04.76 (0.4)	06.18 (1.1
BayesNet	16.22 (0.5)	$14.10 \ (0.4)$	$09.23 \ (0.7)$		
SVM	14.42 (0.3)	17.40 (0.7)	10.81 (0.6)		
KNN	17.89 (0.6)	20.06 (0.7)	11.00 (0.3)		
C4.5	13.54(3.5)	21.92 (0.6)	$06.80 \ (0.3)$		
Ripper	15.79(0.6)	11.49(0.3)	08.54 (0.3)	15.05 (0.4)	18.50 (0.8
BayesNet	16.66 (0.3)	14.35 (0.9)	09.29 (0.5)		
SVM	18.36 (0.6)	21.56 (0.2)	11.28 (0.4)		
KNN	18.95 (0.7)	21.12 (0.5)	11.80 (0.5)		
C4.5	19.87 (0.6)	22.74 (0.3)	06.94 (0.5)		
Ripper	18.14 (0.9)	16.35 (1.2)	08.09 (0.3)	15.93 (0.7)	19.39 (0.5
BayesNet	19.48(0.5)	19.19 (0.9)	11.07 (0.5)		
SVM	20.98(0.5)	23.54 (0.3)	12.95(0.3)		
KNN	20.58 (1.1)	21.92 (0.6)	12.44 (0.6)		
C4.5	, ,	` ,	` ′		
Ripper	18.51 (0.8)	13.89 (0.4)	08.16 (0.9)	14.77 (0.8)	19.55 (0.6
BayesNet	17.73 (0.7)	19.69 (0.4)	10.81 (0.5)	, ,	·
SVM	21.61 (0.7)	22.59(0.8)	14.11 (0.6)		
KNN	19.82 (0.8)	20.67 (0.4)	12.12 (0.9)		
	, ,	` ′	` '		
		` ′		15.19 (0.9)	18.64 (0.4
BayesNet		16.19 (0.5)	10.79 (0.6)	, ,	`
SVM	21.76 (0.7)	24.02 (0.3)	12.42 (0.2)		
KNN	13.48 (1.8)	15.62 (1.0)	10.01 (0.6)		
	` /	` ′	` ′		
	` ′	, ,	` ′	04.71 (0.1)	09.05 (2.6
**	, ,	, ,	, ,	, ,	`
SVM	15.93 (1.5)	16.58 (0.8)	10.13 (0.5)		
KNN		18 89 (0.6)	. ,		
	, ,	, ,	, ,		
	, ,	, ,	` '	17.25 (0.7)	19.16 (0.8
				125 (07)	10.10 (0.0
SVM	17.04 (1.0)	15.37 (0.7)	11.38 (0.2)		
~	101 (1.0)				
IZNINI	17 04 (0 5)	10.90 (0.0)			
KNN	17.04 (0.5)	19.29 (0.8)	10.71 (0.3)		
C4.5	16.83 (3.1)	22.34 (0.7)	06.56 (0.2)	14 51 (0.5)	10 40 (0
	, ,	, ,	, ,	14.51 (0.5)	18.40 (0.4
	KNN C4.5 Ripper BayesNet SVM	Abordager HMC-BR KNN 19.28 (0.3) C4.5 19.84 (0.4) Ripper 19.71 (0.6) BayesNet 20.65 (0.8) SVM 16.04 (0.7) KNN 19.46 (0.4) C4.5 19.69 (0.5) Ripper 17.70 (0.2) BayesNet 19.29 (0.9) SVM 21.49 (0.7) KNN 15.21 (0.7) C4.5 14.46 (0.7) Ripper 14.48 (0.7) BayesNet 16.22 (0.5) SVM 14.42 (0.3) KNN 17.89 (0.6) C4.5 13.54 (3.5) Ripper 15.79 (0.6) BayesNet 16.66 (0.3) SVM 18.36 (0.6) KNN 18.95 (0.7) C4.5 19.87 (0.6) Ripper 18.14 (0.9) BayesNet 19.48 (0.5) SVM 20.98 (0.5) KNN 20.58 (1.1) C4.5 20.78 (0.5) Ripper 18.51 (0.8) BayesNet 17.73 (0.7) SVM 21.61 (0.7) KNN 19.82 (0.8) C4.5 17.20 (1.2) Ripper 16.82 (1.0) BayesNet 17.79 (0.2) SVM 21.76 (0.7) KNN 19.82 (0.8) C4.5 17.20 (1.2) Ripper 16.82 (1.0) BayesNet 17.79 (0.2) SVM 21.76 (0.7) KNN 13.48 (1.8) C4.5 13.17 (1.6) Ripper 14.50 (1.4) BayesNet 17.47 (1.1) SVM 15.93 (1.5) KNN 17.61 (0.9) C4.5 19.98 (0.7) Ripper 18.11 (1.2) BayesNet 21.27 (0.4)	Abordagem Top-Down HMC-BR	Abordagem Top-Down	HMC-BR

Tabela 6.27: Comparação das Abordagens $\mathit{Top-Down}$ e $\mathit{One-Shot}$ Utilizando a Medida

Comiumto do Dodos		Abordage	em Top-Down		Abordagen	One-Shot
Conjunto de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	$02.24\ (0.6)$	02.14(0.9)	$01.26 \ (0.09)$		
	C4.5	$01.51 \ (0.6)$	01.34 (0.3)	$01.26 \ (0.02)$		
Expr	Ripper	02.05 (0.7)	$01.11 \ (0.08)$	$01.31 \ (0.02)$	01.01 (0.03)	01.08 (0.01
	BayesNet	$02.29\ (1.0)$	01.74 (0.4)	$01.36 \ (0.10)$		
	SVM	$00.90 \ (0.05)$	01.00 (0.3)	$01.04\ (0.09)$		
	KNN	02.02 (0.6)	01.34 (0.5)	01.35 (0.30)		
	C4.5	01.22(0.3)	01.35 (0.7)	$01.22\ (0.02)$		
CellCycle	Ripper	01.25 (0.2)	01.00 (0.06)	01.31 (0.10)	00.96 (0.04)	01.03 (0.1)
	BayesNet	01.28(0.1)	$01.14 \ (0.04)$	$01.26 \ (0.03)$		
	SVM	01.93(0.7)	01.84 (0.7)	$01.72\ (0.60)$		
	KNN	00.92 (0.05)	00.91 (0.04)	01.08 (0.04)		
	C4.5	00.90 (0.05)	01.04 (0.24)	01.10 (0.05)		
Church	Ripper	01.12 (0.1)	00.67 (0.07)	01.05 (0.16)	01.00 (0.03)	01.08 (0.2
	BayesNet	01.22 (0.1)	00.95 (0.01)	01.06 (0.06)	, , ,	` .
	SVM	01.01 (0.2)	00.96 (0.04)	01.15 (0.05)		
	KNN	01.11 (0.1)	01.04 (0.1)	01.12 (0.02)		
	C4.5	00.94 (0.2)	00.98 (0.05)	01.16 (0.03)		
Derisi	Ripper	01.04 (0.04)	00.91 (0.2)	01.21 (0.02)	00.93 (0.03)	00.95 (0.07
	BayesNet	01.19 (0.02)	01.00 (0.07)	01.11 (0.04)	(()
	SVM	01.24 (0.5)	00.98 (0.02)	01.15 (0.04)		
	KNN	01.47 (0.2)	01.58 (0.3)	01.44 (0.1)		
	C4.5	01.51 (0.3)	01.21 (0.1)	01.27 (0.05)		
Eisen	Ripper	01.31 (0.3) $01.35 (0.3)$	01.53 (0.1)	01.44 (0.1)	01.06 (0.06)	01.14 (0.03
Lisen	BayesNet	01.47 (0.1)	01.35 (0.1)	01.77 (0.5)	01.00 (0.00)	01.14 (0.00
	SVM	02.45 (0.8)	03.02 (1.0)	01.85 (0.3)		
	KNN	03.26 (1.8)	02.26 (0.5)	01.73 (0.7)		
	C4.5	03.20 (1.8) $01.76 (0.4)$	$02.20 \ (0.3)$ $01.31 \ (0.2)$	01.26 (0.03)		
Gasch1	Ripper	01.70 (0.4) $02.01 (0.6)$	01.31 (0.2) $01.09 (0.1)$	01.44 (0.3)	01.10 (0.2)	01.06 (0.05
Gasciii	BayesNet	02.01 (0.0) $01.52 (0.2)$	01.09 (0.1)	01.22 (0.03)	01.10 (0.2)	01.00 (0.00
	SVM	$03.82 \ (0.2)$	02.66 (0.9)	02.47 (0.9)		
		, ,		, ,		
	KNN	02.29 (1.0)	01.51 (0.62)	01.19 (0.05)		
Gasch2	C4.5	01.03 (0.04)	01.11 (0.1)	01.19 (0.01)	00.09 (0.03)	01.00 (0.1
Gascii2	Ripper BayesNet	01.19 (0.1)	00.93 (0.1)	01.21 (0.02)	00.92 (0.03)	01.09 (0.1
	SVM	01.28 (0.2) 01.67 (0.3)	01.10 (0.1) 02.15 (0.6)	01.19 (0.04) 01.53 (0.5)		
	KNN	00.90 (0.3)	00.84 (0.06)	01.30 (0.3)		
TO .	C4.5	00.73 (0.08)	00.96 (0.09)	01.17 (0.05)	04 00 (0 0)	04.05 (0.0
Phenotype	Ripper	00.85 (0.1)	00.72 (0.2)	01.07 (0.03)	01.26 (0.3)	01.35 (0.3
	BayesNet	01.20 (0.2)	01.18 (0.3)	01.12 (0.07)		
	SVM	01.40 (0.5)	01.15 (0.2)	01.16 (0.07)		
	KNN	01.93 (0.3)	01.77(0.4)	01.30 (0.2)		
~	C4.5	01.66 (0.8)	01.39 (0.5)	01.22 (0.03)		/
Sequence	Ripper	01.55 (0.4)	00.90 (0.04)	01.21 (0.05)	00.96 (0.04)	01.17 (0.2)
	BayesNet	01.74 (0.6)	01.15 (0.04)	01.17 (0.05)		
	SVM	03.76 (1.0)	01.67 (0.5)	02.14 (0.8)		
	KNN	00.93 (0.03)	00.97 (0.1)	01.05 (0.04)		
	C4.5	$01.16 \ (0.5)$	$01.08 \ (0.2)$	$01.14 \ (0.03)$		
SPO	Ripper	01.09 (0.2)	01.06 (0.2)	$01.24\ (0.06)$	00.89 (0.03)	01.02 (0.2
	BayesNet	$01.14\ (0.05)$	$01.10 \ (0.1)$	$01.09 \ (0.06)$		
	SVM	01.05(0.1)	$01.00 \ (0.05)$	$01.10 \ (0.03)$		

Tabela 6.28: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Micro F1 Hierárquica Ponderada no Quarto Nível da Hierarquia

Conjunto de Dados			n Top-Down	J	Abordagen	
		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HM0
	KNN	18.90(1.2)	18.15 (1.4)	12.08 (0.6)		
_	C4.5	18.47 (0.9)	17.61 (0.7)	07.83 (0.5)		
Expr	Ripper	21.83 (0.6)	19.26 (0.6)	09.52 (0.7)	14.95 (0.9)	16.76 (1.1
	BayesNet	22.11 (2.1)	22.52 (2.1)	14.01 (2.1)		
	SVM	20.35 (0.9)	19.80 (1.3)	16.18 (0.8)		
	KNN	20.36(2.1)	$19.54\ (1.6)$	$12.45 \ (0.5)$		
	C4.5	16.74 (0.7)	$16.24 \ (0.6)$	07.07 (0.5)		
CellCycle	Ripper	$18.70 \ (0.9)$	18.65 (1.1)	09.40 (0.8)	13.49 (0.5)	15.14 (0.4)
	BayesNet	22.08(1.7)	19.17 (1.2)	12.82 (0.3)		
	SVM	21.52 (1.4)	22.35 (2.5)	14.31 (1.1)		
	KNN	$14.01 \ (0.7)$	14.05 (0.4)	$11.21 \ (0.5)$		
	C4.5	13.91 (0.8)	14.75 (0.3)	09.42 (0.5)		
Church	Ripper	15.40(1.2)	$17.91 \ (0.8)$	12.78 (1.2)	04.99(0.3)	06.44 (1.0
	BayesNet	17.32 (0.8)	$14.21 \ (0.8)$	10.58 (0.5)		
	SVM	14.19 (0.6)	14.65 (0.4)	10.94 (0.7)		
	KNN	15.84 (0.6)	15.10 (0.5)	10.26 (0.5)		
	C4.5	13.78(3.1)	15.54 (0.7)	06.80 (0.4)		
Derisi	Ripper	15.49(0.6)	17.71 (0.4)	10.49 (0.5)	13.22 (0.3)	15.08 (0.7
	BayesNet	16.22(1.0)	14.63 (1.5)	09.78 (0.3)		
	SVM	$15.20 \ (0.6)$	15.58 (0.8)	10.24 (0.4)		
	KNN	20.36 (1.7)	19.21 (2.0)	12.40 (0.8)		
	C4.5	17.32 (0.7)	17.22 (0.6)	07.36 (0.8)		
Eisen	Ripper	21.14(1.5)	18.62 (1.4)	09.82 (0.6)	14.34 (0.9)	16.75 (0.9
	BayesNet	22.67(2.3)	18.71(1.5)	12.58 (1.1)		
	SVM	23.02(1.7)	24.37 (1.5)	14.07 (1.4)		
	KNN	20.74 (1.4)	19.66 (0.4)	12.62 (0.9)		
	C4.5	18.61 (0.5)	17.55 (1.0)	07.68 (0.4)		
Gasch1	Ripper	20.33 (2.1)	19.24 (1.2)	09.26 (1.1)	13.47(0.7)	15.96 (1.5
	BayesNet	18.76 (1.9)	18.11 (0.6)	12.26 (0.7)		
	SVM	26.09 (1.8)	26.68 (1.8)	16.14 (1.6)		
	KNN	19.31 (1.4)	17.49 (0.3)	11.88 (1.2)		
	C4.5	16.45 (0.5)	16.55 (0.7)	07.03 (0.4)		
Gasch2	Ripper	18.88 (2.3)	19.34 (1.5)	09.90 (0.5)	13.50 (0.6)	15.31 (0.8
	BayesNet	18.59 (1.1)	16.71 (0.5)	11.00 (0.6)	, ,	`
	SVM	19.31 (0.6)	20.50 (1.0)	11.78 (0.4)		
	KNN	14.62 (2.6)	16.21 (1.3)	11.59 (2.0)		
	C4.5	13.56 (1.6)	16.36 (0.8)	09.90 (0.5)		
Phenotype	Ripper	14.98 (0.9)	20.35 (0.6)	11.33 (1.1)	05.38 (0.6)	09.05 (2.0
V -	BayesNet	16.47 (1.3)	14.96 (0.7)	10.24 (0.9)	,	`
	SVM	16.81 (1.9)	17.48 (0.7)	10.56 (0.8)		
	KNN	16.09 (1.0)	15.49 (0.4)	09.18 (0.2)		
	C4.5	18.05 (0.8)	18.20 (1.5)	07.72 (0.5)		
Sequence	Ripper	19.33 (1.4)	17.56 (0.8)	08.73 (0.4)	14.52 (0.4)	17.93 (1.3
	BayesNet	23.13 (2.1)	20.32 (0.8)	12.80 (0.6)	- (~)	(
	SVM	21.82 (2.4)	21.81 (1.1)	17.84 (0.7)		
	KNN	15.24 (0.3)	14.73 (0.2)	09.76 (0.3)		
	C4.5	15.24 (0.3) $15.77 (2.6)$	$15.70 \ (0.2)$	06.41 (0.2)		
SPO	Ripper	16.77 (2.6)	18.68 (1.0)	10.43 (0.7)	12.96 (0.5)	14.76 (0.5
51 0	BayesNet	16.21 (0.7)	15.47 (0.5)	10.45 (0.7)	12.00 (0.0)	14.10 (0.2
	SVM	16.21 (0.7) $16.77 (0.8)$	16.83 (0.8)	10.93 (0.8)		

Tabela 6.29: Comparação das Abordagens *Top-Down* e *One-Shot* Utilizando a Medida Macro F1 Hierárquica Ponderada no Quarto Nível da Hierarquia

Conjunto de Dados			em Top-Down			n One-Shot
Conjunio de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HM0
	KNN	06.58 (0.8)	07.95(1.7)	07.49(1.5)		
	C4.5	08.58 (2.8)	08.17 (1.5)	$03.31 \ (0.8)$		
Expr	Ripper	09.95 (0.9)	11.59 (1.8)	$04.46 \ (0.5)$	07.09(2.0)	09.13 (2.2)
	BayesNet	09.66(2.3)	09.32(2.1)	09.70(4.4)		
	SVM	03.46 (0.4)	12.30 (1.9)	26.79 (2.9)		
	KNN	06.75(1.7)	08.11(2.6)	06.99 (0.8)		
	C4.5	06.85(1.4)	05.27(0.8)	02.45 (0.4)		
CellCycle	Ripper	05.59(1.4)	11.94(2.6)	05.82 (0.4)	04.39 (1.0)	06.60 (0.6
	BayesNet	05.57(0.9)	07.10(1.7)	$11.14\ (2.2)$		
	SVM	09.92 (1.5)	13.45 (2.1)	09.03 (2.2)		
	KNN	03.56 (0.6)	04.69 (0.8)	05.54 (1.7)		
	C4.5	03.96(1.0)	05.05(0.8)	03.65(1.1)		
Church	Ripper	03.66 (0.5)	14.19 (1.0)	14.66 (1.8)	03.40(1.3)	03.77 (0.6
	BayesNet	03.76(1.6)	06.03(1.3)	11.10 (1.5)		
	SVM	04.99(2.2)	06.62(1.7)	04.17(1.1)		
	KNN	03.17 (0.9)	04.76 (0.9)	05.19 (0.9)		
	C4.5	03.90 (2.0)	05.15 (1.7)	01.86 (0.4)		
Derisi	Ripper	03.30 (0.7)	12.53 (1.4)	08.30 (0.9)	03.22(0.9)	06.27 (1.0
	BayesNet	04.68 (1.2)	05.16 (1.7)	10.11 (2.8)		
	SVM	06.12 (2.2)	05.04 (1.5)	02.79 (1.7)		
	KNN	05.55 (1.8)	07.47 (1.1)	08.13 (2.3)		
	C4.5	06.96 (0.8)	07.20 (2.5)	02.83 (1.0)		
Eisen	Ripper	09.86 (1.8)	09.92 (3.6)	06.11 (1.7)	04.50 (2.4)	08.42 (3.3
	BayesNet	07.92 (1.6)	07.46 (1.3)	12.13 (1.4)	,	(-
	SVM	11.22 (2.5)	14.65 (3.1)	09.08 (1.2)		
	KNN	09.70 (2.5)	09.62 (2.1)	08.89 (1.1)		
	C4.5	07.83 (1.8)	07.15 (1.6)	03.46 (0.2)		
Gasch1	Ripper	08.51 (2.9)	13.10 (2.1)	05.80 (1.2)	05.06 (0.9)	06.55 (0.9
	BayesNet	07.40 (3.0)	06.04 (0.7)	13.01 (2.0)	(010)	00100 (011
	SVM	11.81 (1.2)	18.31 (3.1)	17.87 (3.5)		
	KNN	06.45 (1.7)	09.09 (1.0)	06.60 (2.9)		
	C4.5	06.43 (1.7)	04.98 (1.1)	$02.12 \ (0.5)$		
Gasch2	Ripper	06.40 (1.5)	13.69 (1.1)	08.01 (2.2)	04.86 (1.4)	06.36 (2.5
Gasciiz	BayesNet	06.04 (2.2)	05.24 (0.7)	09.41 (1.7)	01.00 (1.1)	00.00 (2.0
	SVM	07.22 (1.7)	10.91 (1.5)	06.29 (1.9)		
	KNN	02.24 (0.8)	06.88 (1.7)	• • • • • • • • • • • • • • • • • • • •		
	C4.5	02.24 (0.8) $01.97 (0.5)$	08.52 (0.9)	09.41 (4.8) 06.28 (2.4)		
Phenotype	Ripper	01.97 (0.5) $02.19 (0.5)$	$14.10 \ (1.3)$	20.15 (4.3)	03.11 (1.3)	03.39 (1.6
Пенотуре	BayesNet	03.95 (1.6)	04.54 (0.9)	03.16 (1.1)	03.11 (1.3)	05.55 (1.0
	SVM	05.95 (1.6) $05.11 (1.6)$	04.34 (0.9) 09.29 (1.5)	03.89 (1.0)		
				, ,		
	KNN	04.66 (1.3)	05.65 (1.9)	03.98 (1.0)		
Coguenae	C4.5	06.54 (1.8)	08.56 (1.9)	03.04 (0.4)	05 04 (1.2)	00 69 (1
Sequence	Ripper	07.49 (2.3)	10.51 (0.6)	02.95 (0.6)	$05.04\ (1.3)$	09.68 (1.4
	BayesNet	09.17 (2.3)	08.60 (1.5)	09.45 (2.3)		
	SVM	05.85 (1.4)	15.49 (1.5)	25.38 (3.4)		
	KNN	03.03 (1.5)	03.80 (0.8)	03.30 (1.0)		
an 0	C4.5	04.60 (0.8)	04.10 (1.4)	01.98 (0.4)		
SPO	Ripper	05.00 (0.9)	13.65 (2.0)	06.66 (1.1)	04.28 (1.3)	04.51 (1.1
	BayesNet	04.74 (0.8)	04.37 (0.7)	05.56 (1.4)		
	SVM	07.39(1.6)	06.28(1.1)	04.11 (1.4)		

6.4 Considerações Gerais sobre os Desempenhos das Técnicas e sobre as Medidas de Avaliação

Pode-se observar que, na maioria dos resultados, o desempenho das técnicas diminui conforme o nível das hierarquias torna-se mais profundo. Esse tipo de desempenho é esperado, pois quanto mais profundo é um nível hierárquico, maior é a quantidade de classes envolvidas, tornando o processo de classificação mais difícil. Ainda, nas técnicas da abordagem *Top-Down*, existe o problema de propagação de erros, ou seja, erros de predição cometidos no nível mais alto da hierarquia são propagados para os níveis mais profundos, contribuindo para a diminuição do desempenho nesses níveis.

As Tabelas 6.30 a 6.35 apresentam os *rankings* médios dos algoritmos considerando todos os dez conjuntos de dados utilizados. Os números em negrito representam os algoritmos que obtiveram as três melhores posições do *ranking*. As tabelas apresentam, em cada linha, o *ranking* obtido em um nível específico da hierarquia de classes.

Tabela 6.30: Ranking Médio dos Algoritmos Utilizando a Medida H-Loss

		$_{\rm HN}$	ІС-В	R			HI	MC-L	Р			$_{\rm HN}$	AC-C	Т		CAEU	Clus-HMC
	KNN	C4.5	Rip	BN	SVM	KNN	C4.5	Rip	BN	SVM	KNN	C4.5	Rip	BN	SVM	04.511	Clus-IllviC
Nível 1	4	8	3	6	1	11	13	7	5	2	12	16	14	10	9	17	15
Nível 2	1	8	2	9	5	3	10	7	4	6	14	17	16	12	15	13	11
Nível 3	5	9	6	10	7	3	8	1	2	4	14	17	16	13	15	12	11
Nível 4	7	9	6	10	8	3	5	1	2	4	15	17	16	12	14	13	11

Tabela 6.31: Ranking Médio dos Algoritmos Utilizando a Medida F1 Hierárquica

		HN	IC-B	R			HN	MC-L	Р		HMC-CT					C4.5H	Clus-HMC
	KNN	C4.5	Rip	BN	SVM	KNN	C4.5	Rip	BN	SVM	KNN	C4.5	Rip	BN	SVM	04.511	Clus-IIIVIC
Nível 1	10	13	11	4	8	14	12	17	9	7	5	3	1	6	2	16	15
Nível 2	9	13	8	1	6	12	14	17	11	7	5	10	3	4	2	16	15
Nível 3	3	9	2	1	4	8	13	17	6	7	12	15	11	10	5	16	14
Nível 4	4	7	2	1	3	8	9	17	5	6	12	16	15	11	10	14	13

Tabela 6.32: Ranking Médio dos Algoritmos Utilizando a Medida Micro F1 Hierárquica

				J			0										
		HMC-BR HMC-LP									HN	ЛС-С		C4.5H	Clus-HMC		
	KNN	C4.5	Rip	BN	SVM	KNN	C4.5	Rip	BN	SVM	KNN	C4.5	Rip	BN	SVM	04.511	Clus-IIIVIC
Nível 1	10	13	11	4	8	14	12	17	9	7	5	3	1	6	2	16	15
Nível 2	7	11	12	3	5	4	1	17	10	2	8	16	15	14	6	13	9
Nível 3	7	8	10	5	4	3	1	14	6	2	13	17	16	15	12	11	9
Nível 4	6	7	10	5	4	3	1	11	8	2	14	17	16	15	13	12	9

Tabela 6.33: Ranking Médio dos Algoritmos Utilizando a Medida Micro F1 Hierárquica Ponderada

1 Ollder	<u>aua</u>																
		HN	ІС-В	R			H	MC-L	Р			HN	AC-C	Т		C4.5H	Clus-HMC
	KNN	C4.5	Rip	BN	SVM	KNN	C4.5	Rip	BN	SVM	KNN	C4.5	Rip	BN	SVM	04.011	Clus-IIWIC
Nível 1	10	13	11	4	8	14	12	17	9	7	5	3	1	6	2	16	15
Nível 2	8	11	6	1	5	13	12	4	9	2	10	16	7	14	3	17	15
Nível 3	7	10	5	2	4	8	9	3	6	1	15	17	14	16	12	13	11
Nível 4	6	10	5	2	3	9	8	4	7	1	14	17	16	15	12	13	11

Tabela 6.34: Ranking Médio dos Algoritmos Utilizando a Medida Macro F1 Hierárquica

		HN	ис-в	R			$_{\rm HI}$	ис-L	Р			HN	ис-с	Т		C4.5H	Clus-HMC
	KNN	C4.5	Rip	BN	SVM	KNN	C4.5	Rip	BN	SVM	KNN	C4.5	Rip	BN	SVM	04.511	Clus-IIIviC
Nível 1	9	14	11	1	10	12	13	17	7	8	6	4	2	5	3	16	15
Nível 2	9	13	10	2	6	12	14	17	11	8	7	4	3	5	1	16	15
Nível 3	7	13	10	2	4	11	14	17	12	8	5	6	3	9	1	16	15
Nível 4	4	12	8	3	1	9	13	16	14	5	7	10	6	11	2	17	15

Tabela 6.35: Ranking Médio dos Algoritmos Utilizando a Medida Macro F1 Hierárquica

Ponderada

	e ondereda																
		HN	ІС- В	R		HMC-LP HMC-CT			C4.5H Clus-HMC								
	KNN	C4.5	Rip	BN	SVM	KNN	C4.5	Rip	BN	SVM	KNN	C4.5	Rip	BN	SVM	04.511	Clus-IIWIC
Nível 1	9	14	11	1	10	12	13	17	7	8	6	4	2	5	3	16	15
Nível 2	12	14	8	5	10	13	15	1	6	2	9	11	4	7	3	16	17
Nível 3	13	14	9	6	4	11	12	1	8	2	10	17	7	3	5	16	15
Nível 4	15	14	11	12	4	8	13	1	10	2	7	17	6	3	5	16	9

Analisando o desempenho das técnicas com a medida H-Loss, percebe-se que, nos dois primeiros níveis, o melhor desempenho foi obtido pela técnica HMC-BR, seguida pela técnica HMC-LP. Isso pode ter ocorrido porque nesses níveis a quantidade de classes é menor, facilitando o processo de classificação binária. No terceiro e quarto níveis da hierarquia, prevaleceu o desempenho da técnica HMC-LP, o que pode ser devido à maneira com que os erros foram cometidos. A técnica HMC-BR cometeu menos erros nas classes do primeiro nível, enquanto na técnica HMC-LP, a maioria dos erros foi cometida nas classes do primeiro nível. Dessa maneira, como a medida H-Loss não considera os erros cometidos nas classes filhas das classes preditas erroneamente, os valores da medida foram menores para a técnica HMC-LP, pois menos erros foram computados nos níveis mais profundos.

As técnicas HMC-CT, Clus-HMC e C4.5H cometeram mais erros em todos os níveis hierárquicos, devido ao grande número de predições realizadas para cada exemplo. Como pode ser observado nas tabelas do Apêndice B, essas técnicas possuem os maiores valores para as medidas de revocação hierárquica, e valores baixos para as medidas de precisão hierárquica. Altos valores para as medidas de revocação e baixos valores de precisão podem ser um indicativo de que as técnicas cometem mais erros, pois o número de FP aumenta e o número de Verdadeiros Positivos (VP) diminui.

Considerando a medida F1 hierárquica, pode-se observar que o melhor desempenho no primeiro nível foi obtido pela técnica HMC-CT, novamente como reflexo do maior número de classes preditas por essa técnica. Apesar de ser menos precisa no primeiro nível, cometendo mais erros, como constatado pela medida H-Loss, a técnica HMC-CT obteve uma cobertura maior sobre os exemplos do conjunto de dados, obtendo um maior valor de revocação hierárquica, como pode ser visto na Tabela B.5. Assim, essa técnica conseguiu um melhor equilíbrio entre as medidas de precisão e revocação, obtendo um melhor desempenho para a medida F1.

Como já observado para o primeiro nível da hierarquia, o menor número de classes nesse nível também parece ter favorecido a técnica, pois ao utilizar os dados mais de uma vez durante o treinamento, em um número reduzido de classes (quatro classes no primeiro nível), os dados podem ter se tornado menos esparsos, o que melhorou o desempenho da classificação.

No segundo, terceiro e quarto níveis da hierarquia, os melhores resultados foram obtidos pela técnica HMC-BR, pois essa obteve um equilíbrio melhor entre as medidas de precisão e revocação hierárquicas, como pode ser observado nas tabelas do Apêndice B. É importante lembrar que as medidas precisão e revocação hierárquicas apenas consideram a intersecção entre o conjunto de classes preditas e de classes verdadeiras dos exemplos, juntamente com as classes ancestrais dessas classes. Elas não consideram distâncias entre as classes e também não consideram que predições em níveis mais profundos são mais importantes. Essas considerações são feitas pelas medidas micro/macro precisão e revocação hierárquicas e micro/macro precisão e revocação hierárquicas ponderada. Devido a essas diferenças, as medidas apresentaram resultados diferentes umas das outras.

Na medida micro F1 hierárquica, os melhores resultados foram obtidos, no primeiro nível, pela técnica HMC-CT. Esse desempenho foi devido aos mesmos motivos anteriormente citados para a medida F1 hierárquica. Como já mencionado anteriormente, as medidas F1 hierárquica e micro F1 hierárquica ponderada e não ponderada apresentam os mesmos valores no primeiro nível das hierarquias. Isso ocorre devido à escolha do valor $Dis_{\theta} = 2$ no cálculo das micro medidas. Com essa escolha, a distância de duas arestas entre uma classe predita e uma verdadeira é considerada uma distância aceitável para não haver contribuições de FP e FN. Não havendo essas contribuições, o cálculo das micro medidas se iguala ao cálculo das medidas precisão e revocação hierárquicas.

Ainda considerando a medida micro F1 hierárquica, no segundo, terceiro e quarto níveis, os melhores resultados foram obtidos pela técnica HMC-LP. O fato dessa medida considerar as distâncias entre as classes da hierarquia pode ter contribuído para a inversão no desempenho das técnicas, já que considerando a medida F1 hierárquica, a técnica HMC-BR obteve o melhor desempenho nesses níveis.

Por meio de análises dos erros das técnicas, constatou-se que a técnica HMC-LP cometeu mais erros nas subárvores rotuladas pelas classes "01" e "10" do que a técnica HMC-BR. Essas subárvores possuem a maioria das classes da estrutura hierárquica e,

proporcionalmente, são as que possuem um maior número de classes que possuem filhos que são nós-folha no terceiro nível (a hierarquia de classes possui quatro níveis). As subárvores rotuladas pelas classes "02" e "11' possuem menos classes e, proporcionalmente, existem mais classes que possuem filhos que são nós-folha no último nível. Além disso, a subárvore rotulada pela classe "02" não possui o quarto nível da hierarquia. Devido a essas características, a técnica HMC-LP pode ter cometido mais erros em classes que possuem filhos que são nós-folha no terceiro nível.

Para calcular as contribuições de FP e FN, as micro/macro medidas fazem um somatório das distâncias entre todas as classes verdadeiras de um exemplo e a classe atribuída a ele, isso quando o exemplo é um FP para essa classe. Quando o exemplo é classificado como um FN para uma dada classe, são calculadas as distâncias de todas as classes atribuídas ao exemplo e essa classe. Assim, na avaliação da técnica HMC-LP, pode ter acontecido, em muitos casos, o cálculo de distâncias entre classes localizadas em nós-folha do terceiro nível das subárvores "01" e "10", e classes que são nós-folha no terceiro nível das subárvores "02" e "11", ou o contrário. De qualquer forma, pode ter ocorrido que as distâncias entre as classes preditas e verdadeiras, na classificação produzida pela técnica HMC-LP, sejam menores. Como a medida considera que classes mais próximas são mais similares, as classificações são menos penalizadas, resultando em um melhor desempenho da técnica.

Analisando a medida micro F1 hierárquica ponderada no segundo, terceiro e quarto níveis da hierarquia, observam-se variações nos desempenhos dos algoritmos, porém o desempenho da técnica HMC-LP ainda prevalece. Essas variações são devido aos pesos atribuídos às arestas que conectam as classes, de maneira a implementar o princípio de que classes localizadas em níveis mais profundos da hierarquia são mais importantes do que classes localizadas em níveis mais elevados.

Outro fato que pode ter influenciado a medida micro F1 hierárquica ponderada é o desbalanceamento dos conjuntos de dados. Como citado anteriormente, existem subárvores que não possuem quatro níveis hierárquicos. Assim, pode ocorrer de erros de predição envolvendo classes que são nós-folha de níveis mais próximos à raiz serem menos penalizados. Essa situação é ilustrada pela Figura 6.3.

Considere dois exemplos que pertencem à classe "11.04.03.01" e duas predições para esses exemplos, sendo um exemplo predito como sendo da classe "11.02.03.04" e outro exemplo predito como pertencente à classe "11.06.01". No último caso, a predição recebe uma penalização menor, pois o caminho na árvore entre as classes predita e verdadeira é menor. Nesse caso, essa penalização não foi justa, pois a predição foi feita em um nó mais próximo à raiz devido ao fato daquela subárvore não possuir mais folhas.

Todas as técnicas da abordagem *Top-Down* apresentaram, em pelo menos um conjunto de dados, bons resultados para a medida macro F1 hierárquica, principalmente no segundo, terceiro e quarto níveis da hierarquia. Isso pode ter ocorrido porque essa medida considera que todas as classes da hierarquia têm a mesma importância, independente da

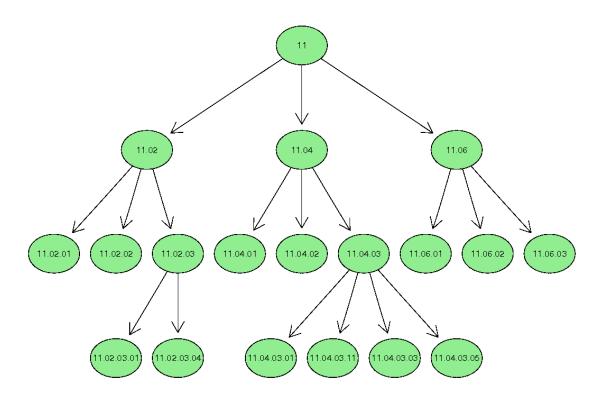


Figura 6.3: Hierarquia de Classes

frequência com que elas aparecem, já que avalia os classificadores calculando a média dos desempenhos de cada classe. Em alguns conjuntos de dados, erros foram cometidos majoritariamente em algumas subárvores. Assim, ao considerar a média dos desempenhos sobre todas as classes, pode-se ter favorecido a avaliação de algumas técnicas em alguns conjuntos de dados. No primeiro nível da hierarquia, ainda prevaleceu o desempenho da técnica HMC-CT.

Assim como na medida micro F1 hierárquica ponderada, a utilização de pesos nas arestas também influenciou a avaliação da medida macro F1 hierárquica ponderada, fazendo com que os desempenhos dos algoritmos variassem.

As técnicas baseadas na abordagem *One-Shot* tiveram piores desempenhos devido ao fato de predizerem uma quantidade maior de classes do que as técnicas da abordagem *Top-Down*. Esse maior número de classes preditas é devido ao fato de não terem sido considerados limiares no processo de predição, conforme explicado no Capítulo 5. As tabelas relativas às medidas de revocação hierárquicas, apresentadas no Apêndice B, refletem essa característica das técnicas da abordagem *One-Shot*, pois, em muitos casos, essas técnicas obtiveram valores para essas medidas mais altos do que os valores obtidos pelas técnicas da abordagem *Top-Down*. O maior número de classes preditas também é uma característica da técnica HMC-CT, pois o processo de decomposição de classes faz com que os exemplos de treinamento sejam utilizados várias vezes, aumentando o número de classificadores, e aumentando assim o número de predições.

A não utilização de limiares pode ter prejudicado essas técnicas, bem como a téc-

nica HMC-CT. Com a utilização de vários limiares, melhores desempenhos poderiam ser obtidos, pois as medidas de precisão e revocação variariam, e uma melhor relação entre elas poderia ser ajustada, aumentando os valores da medida F1 hierárquica. Algoritmos que têm como saída valores reais poderiam ser utilizados na técnica HMC-CT com esse objetivo.

Apesar do pior desempenho preditivo, é possível observar que, em alguns casos, as técnicas da abordagem *One-Shot* obtiveram desempenhos melhores que as técnicas da abordagem *Top-Down* nos últimos níveis hierárquicos, principalmente na comparação com a técnica HMC-CT. Isso pode ter ocorrido porque, na abordagem *Top-Down*, erros cometidos nos primeiros níveis da hierarquia são propagados para os últimos níveis. Essa característica não está presente nas técnicas da abordagem *One-Shot*.

É importante considerar que, embora o desempenho preditivo das técnicas da abordagem *One-Shot* tenha sido menor, essas técnicas produzem um modelo final menos complexo e mais interpretável. As técnicas C4.5H e Clus-HMC produzem árvores de decisão que podem ser utilizadas para gerar um conjunto de regras, tornando o modelo mais interpretável para especialistas da área investigada. Nesta pesquisa, tais pontos não foram levados em consideração, porém uma avaliação multi-objetivo poderia ser utilizada para levar em consideração a interpretabilidade do modelo gerado.

6.5 Considerações Finais

Neste capítulo foram apresentados os experimentos realizados nesta pesquisa. Foram mostrados os resultados obtidos com três técnicas baseadas na abordagem *Top-Down*, chamadas *HMC-Binary-Relevance* (HMC-BR), *HMC-Label-Powerset* (HMC-LP) e *HMC-Cross-Training* (HMC-CT), e também os resultados obtidos com a utilização de duas técnicas baseadas na abordagem *One-Shot*, chamadas C4.5H e Clus-HMC. Os resultados das técnicas foram comparados com a utilização de dez conjuntos de dados biológicos, e a avaliação foi feita com a utilização de várias medidas de avaliação, a fim de analisar o comportamento de cada uma delas.

De acordo com os resultados obtidos, foi possível observar que as variações propostas para a abordagem *Top-Down* (HMC-LP e HMC-CT) levaram a melhores resultados em vários conjuntos de dados. As técnicas baseadas na abordagem *One-Shot*, no entanto, também poderiam ter obtido melhores resultados se fossem utilizados valores de limiares no processo de classificação. Esses valores também poderiam ser utilizados na técnica HMC-CT, por meio da utilização de classificadores que têm valores reais como saída.

Testes estatísticos foram realizados de maneira a validar os experimentos e comparar seus resultados. Os detalhes das análises estatísticas feitas para os resultados reportados neste capítulo podem ser verificados no Apêndice C.

Capítulo

Conclusão

Esta dissertação apresentou um estudo comparativo entre técnicas de classificação para problemas hierárquicos multirrótulo. Nesses tipos de problemas, as classes envolvidas são estruturas em uma hierarquia, ou seja, classes podem ser subclasses ou superclasses de outras classes. Adicionalmente, nesses problemas, duas ou mais classes podem ser atribuídas a um exemplo simultaneamente, dificultando a tarefa de classificação. Problemas com essas características são comuns no campo da Bioinformática. Por isso, nesta pesquisa, foram utilizados conjuntos de dados biológicos para a aplicação das técnicas de classificação investigadas. Mais especificamente, os dados utilizados são relativos à classificação funcional de proteínas.

As técnicas de classificação investigadas dividem-se em duas abordagens: *Top-Down* e *One-Shot*. A abordagem *Top-Down* realiza a classificação discriminando as classes da hierarquia nível a nível, utilizando uma estratégia "Dividir para Conquistar". Essa estratégia permite que sejam desenvolvidas técnicas *Top-Down* utilizando algoritmos de classificação convencionais. Nesta pesquisa, foram implementadas três técnicas baseadas nessa abordagem. A primeira é chamada HMC-BR, e as outras duas, propostas neste trabalho, são chamadas HMC-LP e HMC-CT. Nessas técnicas, foram utilizados cinco algoritmos de Aprendizado de Máquina: KNN, C4.5, Ripper, BayesNet e SVM.

Na abordagem *One-Shot*, são desenvolvidos algoritmos que, ao contrário da abordagem *Top-Down*, consideram todas as classes simultaneamente para induzir um classificador. Dessa forma, um único classificador é utilizado para o processo de classificação, não sendo possível a utilização de classificadores convencionais. Nesta pesquisa, foram utilizadas duas técnicas baseadas nessa abordagem, chamadas C4.5H e Clus-HMC.

Os experimentos foram realizados com o objetivo de comparar as técnicas das abordagens *Top-Down* e *One-Shot*. Para isso, foram utilizados dez conjuntos de dados de proteínas do organismo *Saccharomyces cerevisiae*, também conhecido como *Yeast*. Es-

126 7 Conclusão

ses conjuntos de dados são estruturados de acordo com o esquema do catálogo FunCat (http://mips.gsf.de/projects/funcat), desenvolvido pelo MIPS.

Os resultados dos experimentos foram avaliados com a utilização de medidas específicas para problemas hierárquicos multirrótulo, baseadas nas relações de ancestralidade e descendência entre as classes e nas distâncias que separam as classes na hierarquia. A profundidade dos níveis em que se localizam as classes também é considerada por algumas dessas medidas. As medidas utilizadas são chamadas de *Hierarchical Loss Function*, Precisão e Revocação Hierárquicas, Micro/Macro Precisão e Revocação Hierárquicas e Micro/Macro Precisão e Revocação Hierárquicas Ponderadas, além também das medidas F1 hierárquicas, que combinam as medidas de precisão e revocação.

Todos os resultados dos experimentos foram reportados nível a nível das hierarquias, e testes estatísticos foram utilizados para analisar a significância estatística nos casos em que os resultados das técnicas envolvidas nas comparações mostraram-se diferentes. Os testes estatísticos aplicados foram o teste Friedman e o teste Nemenyi, que consideram comparações envolvendo vários algoritmos e vários conjuntos de dados.

Este Capítulo está organizado da seguinte maneira: a Seção 7.1 apresenta um resumo dos principais resultados obtidos nos experimentos; a Seção 7.2 destaca as contribuições desta pesquisa de mestrado, e, por fim, a Seção 7.3 apresenta os trabalhos futuros que podem ser realizados.

7.1 Principais Resultados

No primeiro nível das hierarquias, as técnicas HMC-BR e HMC-LP, utilizando o algoritmo SVM, cometeram menos erros preditivos, e por isso obtiveram um melhor desempenho quando analisadas com a medida H-Loss. As analises estatísticas, no entanto, mostraram não haver diferenças estatisticamente significantes na comparação da maioria dos algoritmos dessas duas técnicas. Por meio da avaliação com as outras medidas, ainda no primeiro nível da hierarquia, os melhores resultados foram obtidos com a utilização da técnica HMC-CT com os algoritmos SVM e Ripper, mostrando diferenças estatisticamente significantes na comparação com a maioria dos algoritmos.

A análise dos resultados no segundo nível hierárquico, utilizando a medida H-Loss, mostrou que a técnica HMC-BR, utilizando o algoritmo KNN, obteve o melhor desempenho, porém sem diferenças estatisticamente significantes entre as técnicas HMC-BR e HMC-LP. Já na medida F1 hierárquica, os melhores resultados foram obtidos pelas técnicas HMC-BR usando o algoritmo BayesNet, e pela técnica HMC-CT com a utilização do algoritmo SVM. Entre esses dois algoritmos, os testes estatísticos não detectaram diferença estatisticamente significante.

Ainda no segundo nível, a medida micro F1 hierárquica apontou o melhor desempenho da técnica HMC-LP utilizando os algoritmos C4.5 e SVM. Na medida micro F1 hierárquica ponderada, o melhor desempenho foi obtido pela técnica HMC-BR com o

algoritmo BayesNet, seguida pela técnica HMC-LP com o algoritmo SVM. Os testes estatísticos apontaram diferenças estatisticamente significantes, em sua maioria, entre os resultados das técnicas HMC-CT, C4.5H, Clus-HMC e os resultados das outras técnicas. Considerando as medidas macro F1 hierárquicas ponderada e não ponderada, os melhores resultados foram obtidos pela técnica HMC-CT utilizando o algoritmo SVM, e pela técnica HMC-LP utilizando os algoritmos Ripper e SVM. Entre esses algoritmos não foi detectada diferença estatisticamente significante.

No terceiro nível, os melhores resultados, considerando a medida H-Loss, foram obtidos pela técnica HMC-LP utilizando o algoritmo Ripper. Na comparação com a técnica HMC-BR, porém, não foram detectadas diferenças estatisticamente significantes, exceto na comparação com o algoritmo Bayes-Net, obteve o melhor desempenho na avaliação da medida F1 hierárquica. A comparação com as técnicas HMC-LP e HMC-CT mostraram diferenças estatisticamente significantes. Também foi detectada diferença estatisticamente significante na comparação com as técnicas C4.5H e Clus-HMC. Na medida micro F1 hierárquica, os melhores resultados foram obtidos pela técnica HMC-LP utilizando os algoritmos C4.5 e SVM. Entre esses dois algoritmos, não foi detectada diferença estatisticamente significante. A técnica HMC-LP, utilizando o algoritmo SVM, obteve o melhor desempenho na medida micro F1 hierárquica ponderada. Nas medidas macro F1 hierárquicas ponderada e não ponderada, alguns dos melhores resultados também foram obtidos pela técnica HMC-CT, principalmente utilizando os algoritmos Ripper e SVM.

Assim como no terceiro nível, a menor quantidade de erros de predição no quarto nível das hierarquias foi cometida pela técnica HMC-LP, principalmente com a utilização do algoritmo Ripper. Os resultados mostraram-se estatisticamente significantes na comparação com a maioria dos outros algoritmos. Na medida F1 hierárquica, os melhores resultados foram obtidos pela técnica HMC-BR com o algoritmo BayesNet, mostrando diferenças estatisticamente significantes na comparação com a maioria dos algoritmos das técnicas HMC-LP e HMC-CT, bem como na comparação com as técnicas C4.5H e Clus-HMC.

Nas medidas micro F1 hierárquicas não ponderada e ponderada, novamente os melhores desempenhos foram obtidos pela técnica HMC-LP, principalmente com a utilização dos algoritmos C4.5 e SVM, apresentando resultados estatisticamente superiores aos apresentados pela maioria do outros algoritmos, com exceção dos algoritmos da técnica HMC-BR. Nas medidas macro F1 hierárquicas não ponderada e ponderada, novamente os algoritmos das técnicas HMC-LP e HMC-CT também apresentaram alguns dos melhores resultados, principalmente com a utilização dos algoritmo SVM e Ripper. Porém, a diferença dos resultados não mostrou-se estatisticamente significante na maioria dos casos.

Com base nos resultados apresentados, pode-se observar que as variações hierárquicas das técnicas de classificação multirrótulo não hierárquicas podem levar a melhores resultados. Esses melhores resultados são constatatos principlamente com a utilização da técnica HMC-Label-Powerset, proposta nesta pesquisa.

128 7 Conclusão

7.2 Contribuições do Trabalho

Uma das importantes contribuições deste trabalho foi a proposta de duas novas técnicas de classificação hierárquica multirrótulo, chamadas HMC-LP e HMC-CT. Não é do conhecimento dos autores a existência de trabalhos prévios que propuseram tais variações hierárquicas de técnicas de classificação multirrótulo não hierárquicas. Os experimentos mostraram que, em muitos casos, essas variações apresentaram um desempenho superior ao das outras técnicas investigadas.

O estudo da avaliação de classificadores hierárquicos multirrótulo também pode ser considerado uma importante contribuição. Ainda não parece haver na literatura um consenso sobre quais são as melhores medidas de avaliação para problemas hierárquicos multirrótulo. A pesquisa bibliográfica feita nesta pesquisa, reunindo várias medidas diferentes, pode ajudar na escolha da medida mais adequada a ser utilizada, bem como na elaboração de variações dessas medidas ou mesmo na criação de medidas novas.

Contribuições também foram feitas no campo da Bioinformática, as quais foram relatadas por meio de duas publicações. Em Cerri et al. (2009), foram comparadas diferentes técnicas de classificação multirrótulo não hierárquica utilizando diferentes medidas de avaliação, com o objetivo de mostrar a importância das técnicas de classificação multirrótulo para a classificação de dados biológicos. A importância de técnicas de classificação hierárquica multirrótulo para o problema de classificação funcional de proteínas também foi mostrada em Cerri e Carvalho (2009), no qual foram comparadas diferentes técnicas de classificação utilizando árvores de decisão.

Em uma menor escala, pode-se considerar que as revisões bibliográficas desta pesquisa são uma contribuição para o estudo de técnicas de classificação hierarquia multirrótulo. Além da revisão de medidas de avaliação específicas, também foram levantados os principais trabalhos da área e as técnicas desenvolvidas. O mesmo foi feito com trabalhos que envolvem técnicas de classificação multirrótulo não hierárquicas.

7.3 Trabalhos Futuros

Muitos estudos ainda podem ser realizados como trabalhos futuros envolvendo problemas de classificação hierárquica multirrótulo. A investigação de técnicas de classificação para problemas de classificação opcional em nós-folha é um deles. Nesses problemas, o nível de classificação mais profundo associado a cada exemplo de entrada é automaticamente definido pelo classificador, sem a obrigatoriedade dos exemplos terem que ser sempre classificados em classes representadas por nós-folha.

Nas técnicas da abordagem *Top-Down*, uma importante melhora que pode ser incorporada é um mecanismo de correção de propagação de erros que são cometidos devido à estratégia empregada pela abordagem. Esse mecanismo poderia detectar, em cada passo da estratégia *Top-Down*, os exemplos que foram previamente classificados de forma incor-

7.3 Trabalhos Futuros 129

reta, de maneira a fazer uma reclassificação dos mesmos posteriormente. A incorporação desse tipo de mecanismo melhoraria o desempenho das predições realizadas por técnicas da abordagem *Top-Down*.

Técnicas de combinação de classificadores também podem ser exploradas para melhorar o desempenho das técnicas da abordagem *Top-Down*. Estratégias de *ensemble* já foram empregadas no desenvolvimento de técnicas *Top-Down* simples-rótulo, como no trabalho de Costa et al. (2008), melhorando o desempenho preditivo das mesmas.

O desenvolvimento de técnicas baseadas na abordagem *One-Shot* também apresenta um grande potencial para pesquisa. Embora tenham um processo de desenvolvimento mais complexo, técnicas dessa abordagem geralmente produzem um modelo final mais interpretável e simples do que as técnicas da abordagem *Top-Down*, principalmente se forem desenvolvidas técnicas que produzem árvores de decisão ou regras de classificação, como é o caso das técnicas C4.5H e Clus-HMC, utilizadas nesta pesquisa.

A consideração de outros tipos de estruturas hierárquicas, como as hierarquias estruturadas como DAGs (Grafos Acíclicos Direcionados), também é um campo de futuros estudos. Nessa estrutura, um nó pode ter mais de um pai na hierarquia, o que dificulta o processo de classificação. Para considerar hierarquias estruturadas como DAGs, modificações devem ser feitas nas técnicas de classificação.

A avaliação de classificadores hierárquicos multirrótulo também apresenta boas oportunidades para estudos futuros. Apesar de, nesta pesquisa, várias medidas tenham sido utilizadas, muitas considerações ainda podem ser feitas com relação ao desempenho das mesmas. Além disso, novas medidas de avaliação podem ser desenvolvidas, e modificações podem ser incorporadas às medidas existentes.

Por fim, além da análise de diferentes medidas de avaliação, também pode ser feita uma análise de como diferentes técnicas de classificação são influenciadas por diferentes características hierárquicas e multirrótulo dos conjuntos de dados. Um estudo desse tipo pode ajudar no aperfeiçoamento das técnicas existentes e no desenvolvimento de novas técnicas.

130 7 Conclusão

Referências Bibliográficas

- Abdi, H. (2007). Bonferroni and Sidak corrections for multiple comparisons, páginas 175–208. Sage. Encyclopedia of Measurement and Statistics. Citado na página 87.
- Aha, D. W., Kibler, D., e Albert, M. K. (1991). Instance-based learning algorithms. *Machine Learning*, 6(1):37–66. Citado nas páginas 5 e 64.
- Alves, R., Delgado, M., e Freitas, A. (2008). Multi-label hierarchical classification of protein functions with artificial immune systems. In *III Brazilian Symposium on Bioinformatics*, volume 5167 of *LNBI*, páginas 1–12, Berlin, Heidelberg. Springer-Verlag. Citado nas páginas 50, 51, e 52.
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., e Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, 25(1):25–29. Citado nas páginas 4, 47, e 50.
- Barutcuoglu, Z., Schapire, R. E., e Troyanskaya, O. G. (2006). Hierarchical multi-label prediction of gene function. *Bioinformatics*, 22(7):830–836. Citado nas páginas 48, 51, e 52.
- Bayes, T. (1763). An essay towards solving a problem in doctrine of chances. *Philos. Trans. R. Soc. London*, 53:293–315. Citado na página 14.
- Blockeel, H., Bruynooghe, M., Dzeroski, S., Ramon, J., e Struyf, J. (2002). Hierarchical multi-classification. In *KDD-2002 Workshop Notes: MRDM 2002, Workshop on Multi-Relational Data Mining*, páginas 21–35. URL: http://www.cs.kuleuven.ac.be/cgi-bin-dtai/publ_info.pl?id=39450. Citado nas páginas 4, 46, 51, 52, e 65.
- Blockeel, H., De Raedt, L., e Ramon, J. (1998). Top-down induction of clustering trees. In *Proceedings of the 15th International Conference on Machine Learning*, páginas 55–63. Morgan Kaufmann. URL: http://www.cs.kuleuven.ac.be/cgi-bin-dtai/publ_info.pl?id=20419. Citado nas páginas 46 e 64.

- Blockeel, H., Schietgat, L., Struyf, J., Dzeroski, S., e Clare, A. (2006). Decision trees for hierarchical multilabel classification: A case study in functional genomics. In Fürnkranz, J., Scheffer, T., e Spiliopoulou, M., editors, *PKDD*, volume 4213 of *LNCS*, páginas 18–29. Springer. Citado nas páginas 43, 44, 48, 51, 52, e 65.
- Boutell, M. R., Luo, J., Shen, X., e Brown, C. M. (2004). Learning multi-label scene classification. *Pattern Recognition*, 37(9):1757–1771. Citado nas páginas 5, 29, 33, 41, e 75.
- Breiman, L., Friedman, J., Olshen, R., e Stone, C. (1984). Classification and Regression Trees. Wadsworth and Brooks, Monterey, CA. Citado na página 64.
- Carvalho, A. e Freitas, A. (2009). A tutorial on multi-label classification techniques, volume Foundations of Computational Intelligence Vol. 5 of Studies in Computational Intelligence 205, páginas 177–195. Springer. Citado nas páginas 30, 31, 32, 33, 35, 36, 38, e 41.
- Cerri, R. e Carvalho, A. C. P. L. F. (2009). Comparing local and global hierarchical multilabel classification methods using decision trees. In *V Workshop em Algoritmos e Aplicações de Mineração de Dados*, páginas 75–82, Fortaleza. Citado na página 128.
- Cerri, R., Carvalho, A. C. P. L. F., e Costa, E. P. and Freitas, A. A. (2008a). Classificação hierárquica de proteínas utilizando abordagens top-down e big-bang. In *IV Workshop em Algoritmos e Aplicações de Mineração de Dados*, páginas 46–54, Campinas. Citado na página 141.
- Cerri, R., Carvalho, A. C. P. L. F., e Costa, E. P. (2008b). Classificação hierárquica de proteínas utilizando técnicas de aprendizado de máquina. In *II Workshop on Computational Intelligence*, páginas 1–6, Salvador. Citado na página 141.
- Cerri, R., Silva, R., e Carvalho, A. (2009). Comparing methods for multilabel classification of proteins using machine learning techniques. In *IV Brazilian Symposium on Bioinformatics*, volume 5676 of *LNBI*, páginas 109–120, Berlin, Heidelberg. Springer-Verlag. Citado na página 128.
- Cesa-Bianchi, N., Gentile, C., e Zaniboni, L. (2006). Incremental algorithms for hierarchical classification. *J. Mach. Learn. Res.*, 7:31–54. Citado nas páginas 49, 51, 52, 54, e 55.
- Chu, S., Derisi, J., Eisen, M., Mulholl, J., Botstein, D., Brown, P. O., e Herskowitz, I. (1998). The transcriptional program of sporulation in budding yeast. *Science*, 282:699–705. Citado nas páginas 81 e 82.
- Clare, A. (2003). Machine Learning and Data Mining for Yeast Functional Genomics. Tese de Doutorado, University of Wales, Aberystwyth. Citado nas páginas 4, 68, 70, 71, 72, 81, 82, e 86.

- Clare, A. e King, R. D. (2001). Knowledge discovery in multi-label phenotype data. In 5th European Conference on Principles of Data Mining and Knowledge Discovery (PKDD2001), volume 2168 of LNAI, páginas 42–53. Springer. Citado nas páginas 29, 38, 45, 46, 51, e 52.
- Clare, A. e King, R. D. (2003). Predicting gene function in saccharomyces cerevisiae. *Bioinformatics*, 19:42–49. Citado nas páginas 24, 25, 46, 51, e 52.
- Cohen, W. W. (1995). Fast effective rule induction. In *In Proceedings of the Twelfth International Conference on Machine Learning*, páginas 115–123. Citado nas páginas 5, 12, e 63.
- Comite, F., Gilleron, R., e Tommasi, M. (2003). Learning multi-label alternating decision trees from texts and data. In *International Conference on Machine Learning and Data Mining*, volume 2734 of *LNAI*, páginas 35–49. SV. Citado na página 38.
- Costa, E. P. (2008). Investigação de técnicas de classificação hierárquica para problemas de bioinformática. Dissertação de Mestrado, Universidade de São Paulo Instituto de Ciências Matemáticas e de Computação, São Carlos SP. Citado nas páginas 73, 141, e 142.
- Costa, E. P., Lorena, A. C., Carvalho, A. C., e Freitas, A. A. (2008). Top-down hierarchical ensembles of classifiers for predicting g-protein-coupled-receptor functions. In *III Brazilian Symposium on Bioinformatics*, volume 5167 of *LNBI*, páginas 35–46, Berlin, Heidelberg. Springer-Verlag. Citado nas páginas 24 e 129.
- Costa, E. P., Lorena, A. C., Carvalho, A. C., Freitas, A. A., e Holden, N. (2007). Comparing several approaches for hierarchical classification of proteins with decision trees. In *II Brazilian Symposium on Bioinformatics*, volume 4643 of *LNBI*, páginas 126–137, Berlin, Heidelberg. Springer-Verlag. Citado nas páginas 24 e 142.
- Cristianini, N. e Shawe-Taylor, J. (2000). An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press. Citado nas páginas 13 e 38.
- Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30. Citado nas páginas 88, 89, e 195.
- DeRisi, J. L., Iyer, V. R., e Brown, P. O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, 278(5338):680–686. Citado nas páginas 81 e 82.
- Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D., e Weingessel, A. (2008). r-cran-e1071. http://mloss.org/software/view/94/. Citado na página 85.

- ECC (1992). Enzyme Nomenclature. Academic Press, San Diego, California. Citado na página 3.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26. Citado na página 46.
- Egan, J. P. (1975). Signal detection theory and ROC-analysis. Academic Press, New York. Citado na página 20.
- Eisen, M. B., Spellman, P. T., Brown, P. O., e Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America*, 95(25):14863–14868. Citado nas páginas 81 e 82.
- Elisseeff, A. e Weston, J. (2001). Kernel methods for multi-labelled classification and categorical regression problems. In *In Advances in Neural Information Processing Systems* 14, páginas 681–687. MIT Press. Citado nas páginas 29 e 30.
- Esuli, A., Fagni, T., e Sebastiani, F. (2006). TreeBoost.mh: A boosting algorithm for multi-label hierarchical text categorization. In *In Proceedings of the 13th International Symposium on String Processing and Information Retrieval (SPIRE06)*, páginas 13–24. Citado nas páginas 49, 51, e 52.
- Fawcett, T. (2004). Roc graphs: Notes and practical considerations for researchers. Technical Report HPL-2003-4, HP Laboratories. Citado na página 20.
- Fix, E. e Hodges, J. (1951). Discriminatory analysis, non-parametric discrimination: consistency properties. Technical report, USAF Scholl of aviation and medicine, Randolph Field. 4. Citado na página 16.
- Freitas, A. A. e Carvalho, A. C. (2007). A Tutorial on Hierarchical Classification with Applications in Bioinformatics., volume 1, chapter VII, páginas 175–208. Idea Group. Research and Trends in Data Mining Technologies and Applications. Citado nas páginas 22, 23, e 24.
- Freund, Y. e Mason, L. (1999). The alternating decision tree learning algorithm. In *In Machine Learning: Proceedings of the Sixteenth International Conference*, páginas 124–133. Morgan Kaufmann. Citado na página 38.
- Freund, Y. e Schapire, R. (1999). A short introduction to boosting. *Japonese Society for Artificial Intelligence*, 14(5):771–780. Citado nas páginas 38 e 49.
- Freund, Y. e Schapire, R. E. (1995). A decision-theoretic generalization of on-line learning and an application to boosting. In *European Conference on Computational Learning Theory*, páginas 23–37. Citado nas páginas 38 e 39.

- Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32(200):675–701. Citado nas páginas 88 e 195.
- Friedman, M. (1940). A comparison of alternative tests of significance for the problem of m rankings. Annals of Mathematical Statistics, 11(1):86–92. Citado nas páginas 88 e 195.
- Friedman, N., Geiger, D., e Goldszmidt, M. (1997). Bayesian network classifiers. *Mach. Learn.*, 29(2-3):131–163. Citado nas páginas 5 e 63.
- Gasch, A. P., Huang, M., Metzner, S., Botstein, D., Elledge, S. J., e Brown, P. O. (2001). Genomic expression responses to DNA-damaging agents and the regulatory role of the yeast ATR homolog Mec1p. *Mol. Biol. Cell*, 12:2987–3003. Citado nas páginas 81 e 82.
- Gasch, A. P., Spellman, P. T., Kao, C. M., Carmel-Harel, O., Eisen, M. B., Storz, G., Botstein, D., e Brown, P. O. (2000). Genomic expression programs in the response of yeast cells to environmental changes. *Molecular biology of the cell*, 11(12):4241–4257. Citado nas páginas 81 e 82.
- Godbole, S. e Sarawagi, S. (2004). Discriminative methods for multi-labeled classification. In 8th Pacific-Asia Conference on Knowledge Discovery and Data Mining, páginas 22–30. Springer. Citado na página 41.
- Gonçalves, T. e Quaresma, P. (2003). A preliminary approach to the multilabel classification problem of portuguese juridical documents. In *EPIA*, páginas 435–444. Citado na página 29.
- Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, 27(4):857–871. Citado na página 83.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., e Witten, I. H. (2009). The weka data mining software: an update. *SIGKDD Explor. Newsl.*, 11(1):10–18. Citado na página 85.
- Holden, N. e Freitas, A. (2006). Hierarchical classification of g-protein-coupled receptors with a pso/aco algorithm. In *Proc. IEEE Swarm Intelligence Symposium (SIS-06)*, páginas 77–84. IEEE Press. Citado nas páginas 58 e 87.
- Hornik, K., Buchta, C., e Zeileis, A. (2009). Open-source machine learning: R meets Weka. *Computational Statistics*, 24(2):225–232. Citado na página 85.
- Iman, R. e Davenport, J. (1980). Approximations of the critical region of the friedman statistic. *Communications in Statistics*, páginas 571–595. Citado na página 88.
- Karalic, A. e Pirnat, V. (1991). Significance level based multiple tree classification. In *Informatica*, volume 15, página 12. Citado na página 29.

- Kaufman, L. e Rousseeuw, P. J. (1990). Finding Groups in Data: An Introduction to Cluster Analysis. Wiley-Interscience. Citado na página 83.
- Kiritchenko, S., Matwin, S., e Famili, A. F. (2004). Hierarchical text categorization as a tool of associating genes with gene ontology codes. In *Proceedings of the Second European Workshop on Data Mining and Text Mining in Bioinformatics*, páginas 30–34, Pisa, Italy. Citado nas páginas 45, 47, 51, 52, 53, e 57.
- Kumar, A., Cheung, K.-H., Ross-Macdonald, P., Coelho, P. S. R., Miller, P., e Snyder,
 M. (2000). TRIPLES: a database of gene function in Saccharomyces cerevisiae. *Nucl. Acids Res.*, 28(1):81–84. Citado na página 81.
- Larranaga, P., Calvo, B., Santana, R., Bielza, C., Galdiano, J., Inza, I., Lozano, J. A., Armananzas, R., Santafe, G., Perez, A., e Robles, V. (2006). Machine learning in bioinformatics. volume 7, páginas 86–112. Citado na página 10.
- Lauser, B. e Hotho, A. (2003). Automatic multi-label subject indexing in a multilingual environment. In *Proc. of the 7th European Conference in Research and Advanced Technology for Digital Libraries, ECDL 2003*, volume 2769, páginas 140–151. Springer. Citado na página 29.
- Lord, P.W., S. R. B. A. G. C. (2003). Investigating semantic similarity measures across the gene ontology: The relationship between sequence and annotation. *Bioinformatics*, 19(10):1275–1283. Citado na página 58.
- Luo, X. e Zincir-Heywood, N. A. (2005). Evaluation of two systems on multi-class multilabel document classification. In *International Syposium on Methodologies for Intelli*gent Systems, páginas 161–169. Citado na página 29.
- Maechler, M., Rousseeuw, P., Struyf, A., e Hubert, M. (2005). Cluster analysis basics and extensions. Rousseeuw et al provided the S original which has been ported to R by Kurt Hornik and has since been enhanced by Martin Maechler: speed improvements, silhouette() functionality, bug fixes, etc. See the 'Changelog' file (in the package source). Citado na página 83.
- MedLine database (2008). Medline database. [Online; acessado 8-Julho-2008] URL: http://www.ncbi.nlm.nih.gov/pubmed/. Citado nas páginas 4 e 47.
- Mewes, H. W., Frishman, D., Güldener, U., Mannhaupt, G., Mayer, K., Mokrejs, M., Morgenstern, B., Münsterkötter, M., Rudd, S., e Weil, B. (2002). Mips: a database for genomes and protein sequences. *Nucleic Acids Res*, 30(1):31–34. Citado nas páginas 4, 80, e 81.
- Mitchell, T. (1997). Machine Learning. McGraw Hill. Citado nas páginas 10 e 17.

- Monard, M. C. e Baranauskas, J. A. (2003). Conceitos sobre Aprendizado de Máquina, volume Sistema Inteligêntes: Fundamentos e Aplicações, chapter 4, páginas 39–56. Solange Oliveira Rezende. Citado nas páginas 8, 10, 13, e 16.
- Nemenyi, P. B. (1963). *Distribution-free multiple comparisons*. Tese de Doutorado, Princeton University. Citado nas páginas 88 e 195.
- Oliver, S. (1996). A network approach to the systematic analysis of yeast gene function. Trends in Genetics, 12(7):241 – 242. Citado na página 81.
- Pavlidis, P. e Grundy, W. N. (1999). Combining microarray expression data and phylogenetic profiles to learn functional categories using support vector machines. In *In*, páginas 44–59. Routledge. Citado na página 38.
- Quinlan, J. R. (1993). C4.5: programs for machine learning. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. Citado nas páginas 5, 11, 25, 38, 45, 63, e 68.
- R Development Core Team (2008). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Citado na página 85.
- Roth, F. P., Hughes, J. D., Estep, P. W., e Church, G. M. (1998). Finding dna regulatory motifs within unaligned noncoding sequences clustered by whole-genome mrna quantitation. *Nature biotechnology*, 16(10):939–945. Citado nas páginas 81 e 82.
- Rousu, J., Saunders, C., Szedmak, S., e Shawe-Taylor, J. (2006). Kernel-based learning of hierarchical multilabel classification models. *J. Mach. Learn. Res.*, 7:1601–1626. Citado nas páginas 49, 51, e 52.
- Ruepp, A., Zollner, A., Maier, D., Albermann, K., Hani, J., Mokrejs, M., Tetko, I., Güldener, U., Mannhaupt, G., Münsterkötter, M., e Mewes, H. W. (2004). The funcat, a functional annotation scheme for systematic classification of proteins from whole genomes. *Nucl. Acids Res.*, 32(18):5539–5545. Citado na página 80.
- Saridis, G. (1983). Parameter estimation: Principles and problems. *Automatic Control*, *IEEE Transactions on*, 28(5):634–635. Citado na página 39.
- Schapire, R. E. e Singer, Y. (1999). Improved boosting algorithms using confidence-rated predictions. In *Machine Learning*, volume 37, páginas 297–336, Hingham, MA, USA. Kluwer Academic Publishers. Citado nas páginas 39 e 49.
- Schapire, R. E. e Singer, Y. (2000). Boostexter: a boosting-based system for text categorization. In *Machine Learning*, volume 39, páginas 135–168, Hingham, MA, USA. Kluwer Academic Publishers. Citado nas páginas 39 e 40.

- Shen, X., Boutell, M., Luo, J., e Brown, C. (2004). Multi-label machine learning and its application to semantic scene classification. In *International Symposium on Electronic Imaging*, San Jose, CA. Citado nas páginas 5, 29, 35, 41, e 76.
- Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D., e Futcher, B. (1998). Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Mol Biol Cell*, 9(12):3273–97. Citado nas páginas 81 e 82.
- Struyf, J., Blockeel, H., e Clare, A. (2005). Hierarchical multi-classification with predictive clustering trees in functional genomics. In Workshop on Computational Methods in Bioinformatics at the 12th Portuguese Conference on Artificial Intelligence, volume 3808 of LNAI, páginas 272–283. Springer Berlin / Heidelberg. Citado nas páginas 44, 47, 51, e 52.
- Su, C.-Y., Lo, A., Lin, C.-C., Chang, F., e Hsu, W.-L. (2005). A novel approach for prediction of multi-labeled protein subcellular localization for prokaryotic bacteria. In CSBW '05: Proceedings of the 2005 IEEE Computational Systems Bioinformatics Conference Workshops, páginas 79–82, Washington, DC, USA. IEEE Computer Society. Citado na página 38.
- Sun, A. e Lim, E.-P. (2001). Hierarchical text classification and evaluation. In *Fourth IEEE International Conference on Data Mining*, páginas 521–528. Citado nas páginas 55, 57, e 58.
- Tan, P.-N., Steinbach, M., e Kumar, V. (2005). Introduction to Data Mining, (First Edition). Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1 edition. Citado nas páginas 8, 10, 11, 12, 14, 15, 16, e 17.
- Taskar, B., Guestrin, C., e Koller, D. (2003). Max-margin markov networks. In Thrun, S., Saul, L. K., e Schölkopf, B., editors, *Advances in Neural Information Processing Systems 16*. MIT Press. Citado na página 49.
- Thabtah, F. A., Cowling, P., Peng, Y., Rastogi, R., Morik, K., Bramer, M., e Wu, X. (2004). Mmac: A new multi-class, multi-label associative classification approach. In Fourth IEEE International Conference on Data Mining, páginas 217–224. Citado na página 39.
- Tsochantaridis, I., Joachims, T., Hofmann, T., e Altun, Y. (2005). Large margin methods for structured and interdependent output variables. *Journal of Machine Learning Research*, 6:1453–1484. Citado na página 49.
- Tsoumakas, G. e Katakis, I. (2007). Multi label classification: An overview. *International Journal of Data Warehousing and Mining*, 3(3):1–13. Citado nas páginas 29, 39, e 40.

- Tsoumakas, G. e Vlahavas, I. (2007). Random k-labelsets: An ensemble method for multilabel classification. In *Proceedings of the 18th European Conference on Machine Learning (ECML 2007)*, páginas 406–417, Warsaw, Poland. Citado nas páginas 5, 32, 33, 34, e 75.
- Valentini, G. (2009). True path rule hierarchical ensembles. volume 5519 of *LNBI*, páginas 232–241. Citado nas páginas 50, 51, e 52.
- Vapnik, V. N. (1999). The Nature of Statistical Learning Theory (Information Science and Statistics). Springer. Citado nas páginas 5 e 63.
- Vens, C., Struyf, J., Schietgat, L., Džeroski, S., e Blockeel, H. (2008). Decision trees for hierarchical multi-label classification. *Machine Learning*, 73(2):185–214. Citado nas páginas 6, 25, 50, 51, 52, 60, 61, 64, 65, 66, 67, e 81.
- Wilkins, M., Gasteiger, E., Bairoch, A., Sanchez, J., Williams, K., Appel, R., e Hochstrasser, D. (1999). Protein identification and analysis tools in the expasy server. *Methods in molecular biology (Clifton, N.J.)*, 112:531–552. cited By (since 1996) 95. Citado na página 81.
- Witten, I. H. e Frank, E. (2005). Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, 2 edition. Citado na página 18.
- Yang, Y. (1999). An evaluation of statistical approaches to text categorization. *Information Retrieval*, 1(1/2):69–90. Citado na página 56.
- Zhang, M.-L. e Zhou, Z.-H. (2005). A k-nearest neighbor based algorithm for multi-label classification. volume 2, páginas 718–721 Vol. 2. The IEEE Computational Intelligence Society. Citado nas páginas 29 e 38.
- Zhu, S., Ji, X., Xu, W., e Gong, Y. (2005). Multi-labelled classification using maximum entropy method. In SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, páginas 274–281, New York, NY, USA. ACM. Citado na página 39.



Experimentos Preliminares

Este apêndice apresenta alguns experimentos que foram realizados com classificação hierárquica simples-rótulo e classificação multirrótulo não hierárquica. Para os experimentos com classificação hierárquica foram utilizados os algoritmos implementados por Costa (2008), e para os experimentos com classificação multirrótulo foram utilizados algumas técnicas que foram vistas no Capitulo 3 (Classificação Multirrótulo).

A.1 Experimentos com Classificação Hierárquica

Os experimentos com classificação hierárquica simples-rótulo foram realizados de duas maneiras. Em um primeiro experimento, foi investigada a aplicação de algoritmos tradicionais de Aprendizado de Máquina (AM), por meio da abordagem *Top-Down*, em um problema de classificação hierárquica de proteínas. Os algoritmos foram avaliados por meio das medidas específicas de avaliação que foram apresentadas na Seção 2.2.2 (Avaliando Modelos de Classificação Hierárquica). Um segundo experimento comparou as abordagens de classificação hierárquica *Top-Down* e *One-Shot*, utilizando árvores de decisão, também para a tarefa de classificação hierárquica de proteínas. A avaliação também foi feita utilizando as medidas apresentadas na Seção 2.2.2. Os dois experimentos realizados deram origem a duas publicações:

- Classificação Hierárquica de Proteínas Utilizando Técnicas de Aprendizado de Máquina (Cerri et al., 2008b)
- Classificação Hierárquica de Proteínas Utilizando Abordagens Top-Down e Big-Bang (Cerri et al., 2008a)

A.1.1 Bases de Dados Utilizadas

Duas bases de dados foram utilizadas para os experimentos com classificação hierárquica. A primeira é formada por proteínas *G-Protein-Coupled Receptor* (GPCR) e a segunda é formada por famílias de enzimas codificadas segundo códigos da *Enzyme Commission* (EC).

Proteínas GPCR são proteínas envolvidas em sinalização. Elas expandem as paredes celulares influenciando os processos químicos dentro de uma célula pela sensibilidade aos processos químicos fora da célula. Esse tipo de proteína é particularmente importante para aplicações médicas porque acredita-se que 40% a 50% dos medicamentos atacam a atividade das GPCRs (Costa et al., 2007).

As enzimas são outro subconjunto de proteínas que funcionam como catalisadores que aceleram e tornam possíveis muitas das reações químicas dentro de uma célula, sem que elas próprias sejam alteradas durante esse processo. Elas são utilizadas para controlar tanto a velocidade de uma reação, ou o curso total das reações em uma via metabólica dentro de uma célula. A base de dados de GPCRs utilizada possui 450 atributos, 7461 exemplos e 12/54/82/50 classes por nível. A base de dados de enzimas possui 1216 atributos, 6925 exemplos e 2/21/48/87 classes por nível.

A.1.2 Teste Estatístico Aplicado

Para verificar se a diferença nos resultados é estatisticamente significante, foi utilizado o teste t de Student corrigido. O nível de significância do teste, que originalmente foi determinado como 5% ($\alpha=0.05$), foi ajustado por meio da correção de Bonferroni, que consiste em determinar um novo nível de significância baseado no nível de significância original e no número de comparações empregadas. Essa correção foi necessária, pois vários algoritmos foram considerados na comparação. O valor obtido com a correção de Bonferroni depende do número de comparações que são feitas. Após a correção, o α utilizado no experimento 1 teve valor 0.005 e no experimento 2 valor 0.025. Como foi utilizado o 5-fold cross-validation, os testes foram realizados utilizando 4 graus de liberdade.

A.1.3 Experimento 1

Para a classificação dos exemplos das bases de dados de GPCRs e enzimas, foram utilizados cinco algoritmos de classificação convencionais utilizando a abordagem *Top-Down*. Os algoritmos utilizados foram: Redes Bayesianas (BayesNet), *k-Nearest Neighbor* (KNN), Máquinas de Vetores de Suporte (SVMs), Árvores de Decisão (C4.5) e Extração de Conjunto de Regras (Ripper).

Todos os algoritmos de classificação foram utilizados na técnica Top-Down implementada por Costa (2008) utilizando a ferramenta R. As bases de dados foram divididas de acordo com o método k-fold cross-validation com valor de k igual a 5.

Os experimentos tiveram o objetivo de avaliar, por meio das medidas apresentadas na Seção 2.2.2, os algoritmos de classificação convencionais utilizando a abordagem *Top-Down*. As médias das taxas de acerto obtidas para cada nível da hierarquia são apresentadas nas tabelas A.1, A.2, A.3 e A.4 para a base de dados de GPCRs, e nas tabelas A.5, A.6, A.7 e A.8 para a base de dados de enzimas.

Conforme observado nas tabelas com os resultados da base de GPCRs, o melhor desempenho preditivo foi obtido pelo algoritmo KNN, seguido pelos algoritmos C4.5, SVM, Ripper e BayesNet. Pode-se observar também, que os algoritmos C4.5, KNN e SVM obtiveram um desempenho preditivo bem próximo. O mesmo acontece na comparação dos resultados dos algoritmos Ripper e BayesNet.

Ainda na base de dados GPCRs, todos os algoritmos, com exceções, apresentaram um melhor desempenho no quarto nível do que no terceiro. Isso ocorre porque a hierarquia GPCR tem mais classes no terceiro nível do que no quarto nível, fazendo com que alguns erros sejam propagados apenas até o terceiro nível da hierarquia.

Na base de dados de enzimas, todos os algoritmos tiveram um desempenho de quase 100% devido às características dessa base de dados. Para os dois primeiros níveis da hierarquia, o algoritmo SVM obteve um melhor desempenho, seguido pelos algoritmos KNN, C4.5, Ripper e BayesNet. No terceiro nível, essa ordem de desempenho se manteve, com exceção da medida custo uniforme, onde o algoritmo KNN obteve um melhor desempenho. No quarto nível, o algoritmo KNN obteve um melhor desempenho para todas as medidas, com exceção da medida dependente de profundidade. Os algoritmos SVM e KNN obtiveram um desempenho próximo, e o mesmo ocorreu com os algoritmos C4.5 e Ripper. Uma diferença um pouco maior pode ser observada no algoritmo BayesNet.

Tabela A.1: Resultados na base de dados de GPCRs utilizando abordagem *Top-Down* e medida baseada em custo uniforme

C4.5	KNN	SVM	BayesNet	Ripper
93.10	93.26	92.34	85.55	89.90
87.69	88.15	86.08	76.68	81.82
52.43	53.36	49.05	39.49	45.53
78.55	78.38	73.84	55.38	61.88

Tabela A.2: Resultados na base de dados de GPCRs utilizando abordagem *Top-Down* e medida baseada em distância independente de profundidade

C4.5	KNN	SVM	BayesNet	Ripper
93.10	93.26	92.34	85.55	89.90
90.21	90.53	89.00	80.75	85.58
78.27	79.11	76.60	67.94	73.54
83.91	84.92	82.93	66.39	74.02

Tabela A.3: Resultados na base de dados de GPCRs utilizando abordagem *Top-Down* e medida baseada em distância dependente de profundidade

C4.5	KNN	SVM	BayesNet	Ripper
93.10	93.26	92.34	85.55	89.90
91.05	91.32	89.97	82.09	86.83
85.28	85.87	83.64	75.46	80.52
85.48	86.30	85.11	68.46	75.36

Tabela A.4: Resultados na base de dados de GPCRs utilizando abordagem *Top-Down* e medida baseada nas relações de ancestralidade e descendência

C4.5	KNN	SVM	BayesNet	Ripper
93.10	93.26	92.34	85.55	89.90
90.21	90.52	88.99	80.72	85.58
77.70	78.36	75.47	66.66	71.81
83.04	83.24	81.51	63.02	69.24

Tabela A.5: Resultados na base de dados de enzimas utilizando abordagem *Top-Down* e medida baseada em custo uniforme

C4.5	KNN	SVM	BayesNet	Ripper
98.40	99.82	99.94	94.49	98.26
97.70	99.11	99.30	90.95	97.24
96.99	98.63	98.38	89.27	96.51
95.90	97.53	96.71	87.60	95.18

Tabela A.6: Resultados na base de dados de enzimas utilizando abordagem *Top-Down* e medida baseada em distância independente de profundidade

		<u> </u>		
C4.5	KNN	SVM	BayesNet	Ripper
98.40	99.82	99.94	94.49	98.26
98.05	99.47	99.66	92.72	97.75
97.70	99.19	99.29	91.57	97.34
97.32	98.87	98.72	90.80	96.86

Tabela A.7: Resultados na base de dados de enzimas utilizando abordagem *Top-Down* e medida baseada em distância dependente de profundidade

and buseded our disturbed depondence de protundade								
C4.5	KNN	SVM	BayesNet	Ripper				
98.40	99.82	99.94	94.49	98.26				
98.17	99.58	99.73	93.31	97.92				
97.99	99.44	99.52	92.69	97.71				
97.78	99.31	99.31	92.22	97.46				

Tabela A.8: Resultados na base de dados de enzimas utilizando abordagem *Top-Down* e medida baseada nas relações de ancestralidade e descendência

C4.5	KNN	SVM	BayesNet	Ripper
98.40	99.82	99.94	94.49	98.26
98.05	99.47	99.62	92.72	97.75
97.70	99.19	99.21	91.57	97.34
97.20	98.80	98.59	90.49	96.76

A.1.4 Experimento 2

O experimento foi conduzido de maneira a avaliar as abordagens de classificação hierárquica *Top-Down* e *One-Shot* utilizando o algoritmo de indução de árvores de decisão C4.5. Para essa avaliação, as mesmas medidas do experimento 1 foram utilizadas. As taxas de acerto médias obtidas em cada nível da hierarquia para as bases de GPCRs e enzimas, utilizando as diferentes medidas, são mostradas nas tabelas A.9 a A.16.

Quando é comparado o desempenho das abordagens *Top-Down* e *One-Shot*, observa-se que, para as duas bases de dados, a abordagem *One-Shot* obteve um pior desempenho. O teste estatístico aplicado às diferentes medidas mostrou que a diferença entre os resultados obtidos pelas abordagens é estatisticamente significante para ambas as bases nos quatro níveis da hierarquia.

O melhor desempenho da abordagem *Top-Down* pode ser devido à menor complexidade desse algoritmo, que não precisa considerar simultaneamente todas as classes da hierarquia em uma única iteração. Ainda, como a predição utilizada nos experimentos foi obrigatória nos nós-folha, a distribuição das classes na base de dados de GPCRs acabou favorecendo a abordagem *Top-Down*, pois erros cometidos no terceiro nível não se propagaram para o último nível da hierarquia.

Tabela A.9: Resultados na base de dados de GPCRs utilizando abordagens Top-Down e

One-Shot com C4.5 e medida baseada em custo uniforme

	Nível 1	Nível 2	Nível 3	Nível 4
Top-Down	93.10	87.69	52.43	78.55
One-Shot	91.13	75.23	39.72	50.19

Tabela A.10: Resultados na base de dados de GPCRs utilizando abordagens *Top-Down* e *One-Shot* com C4.5 e medida baseada em distância independente de profundidade

	Nível 1	Nível 2	Nível 3	Nível 4
$Top ext{-}Down$	93.10	90.21	78.27	83.91
One-Shot	91.13	83.41	70.79	74.33

Tabela A.11: Resultados na base de dados de GPCRs utilizando abordagens *Top-Down* e *One-Shot* com C4.5 e medida baseada em distância dependente de profundidade

	Nível 1	Nível 2	Nível 3	Nível 4
Top-Down	93.10	91.05	85.28	85.48
$One ext{-}Shot$	91.13	85.83	79.65	79.67

Tabela A.12: Resultados na base de dados de GPCRs utilizando abordagens *Top-Down* e *One-Shot* com C4.5 e medida baseada nas relações de ancestralidade e descendência

	Nível 1	Nível 2	Nível 3	Nível 4
Top-Down	93.10	90.21	77.70	83.04
$One ext{-}Shot$	91.13	83.34	69.83	71.13

Tabela A.13: Resultados na base de dados de enzimas utilizando abordagens Top-Down

e <u>One-Shot com C4.5 e medida baseada em custo uniforme</u>

	Nível 1	Nível 2	Nível 3	Nível 4
$Top ext{-}Down$	98.40	97.70	96.99	95.90
$One ext{-}Shot$	88.96	82.70	76.85	72.14

Tabela A.14: Resultados na base de dados de enzimas utilizando abordagens *Top-Down* e *One-Shot* com C4.5 e medida baseada em distância independente de profundidade

	Nível 1	Nível 2	Nível 3	Nível 4
Top-Down	98.40	98.05	97.70	97.32
$One ext{-}Shot$	88.96	86.64	85.71	86.85

Tabela A.15: Resultados na base de dados de enzimas utilizando abordagens *Top-Down* e *One-Shot* com C4.5 e medida baseada em distância dependente de profundidade

	Nível 1	Nível 2	Nível 3	Nível 4
$Top ext{-}Down$	98.40	98.17	97.99	97.78
$One ext{-}Shot$	88.96	87.42	86.87	87.22

Tabela A.16: Resultados na base de dados de enzimas utilizando abordagens *Top-Down* e *One-Shot* com C4.5 e medida baseada nas relações de ancestralidade e descendência

	Nível 1	Nível 2	Nível 3	Nível 4
$Top ext{-}Down$	98.40	98.05	97.70	97.20
$One ext{-}Shot$	88.96	86.36	85.14	84.71

A.2 Experimentos com Classificação Multirrótulo

Os experimentos com classificação multirrótulo não hierárquica foram realizados com a finalidade de comparar diferentes técnicas que utilizam algoritmos de classificação convencionais. Essas técnicas fazem parte de uma abordagem chamada abordagem independente de algoritmo. Técnicas dessa abordagem foram apresentados na Seção 3.1.1. Também foi utilizada uma técnica que usa o algoritmo KNN, chamada ML-kNN, baseada na abordagem dependente de algoritmo (Seção 3.1.2).

Para os experimentos, foram implementadas três técnicas. A primeira, chamada Um-Contra-Todos, faz uma transformação baseada em rótulos, treinando um classificador binário para cada classe do problema. A segunda técnica, chamada Label-Powerset, cria novas classes a partir de todas as combinações envolvendo mais de uma classe. A terceira técnica é chamada de Cross-Training, e consiste em considerar cada uma das possíveis classes como sendo a classe positiva em sequência. O número de classificadores k é igual ao número de classes que rotulam pelo menos um exemplo multirrótulo. A técnica RAKEL, que combina vários classificadores da técnica Label-Powerset também foi utilizada. Em cada uma das técnicas foram utilizados os algoritmos de classificação KNN, SVMs, C4.5, Ripper e BayesNet. Os experimentos foram realizados utilizando o método k-Fold Cross-Validation com um valor de k igual a 5. As técnicas foram implementadas utilizando a ferramenta R.

A.2.1 Bases de Dados Utilizadas

Para os experimentos, foi utilizada uma base de dados composta por proteínas relacionadas ao organismo *Yeast* e uma base de dados de sequências de proteínas classificadas em famílias estruturais.

O Yeast é um fungo unicelular, e sua espécie mais conhecida é chamada de Saccharomyces cerevisiae. Esse organismo é muito utilizado na fermentação de açúcar para a
produção de etanol, e também para a fermentação de trigo e cevada para a produção
de bebidas alcoólicas. Esse fungo é muitas vezes utilizado como suplemento vitamínico,
devido a sua composição rica em proteínas e por ser fonte de vitamina B e ácido fólico.
O conjunto de dados utilizado possui 2417 exemplos, 103 atributos e 14 classes.

A base de dados de famílias estruturais de proteínas possui 662 exemplos, 1186 atributos e 14 classes. As classes consideradas fazem parte de famílias de proteínas importantes na biologia, como as *oxidoredutases* (Enzimas que catalisam as reações de oxidação-redução), *isomerases* (Enzimas que catalisam a interconversão de um isômero em outro) e *hidrolases* (Enzimas que promovem a cisão de um material orgânico através da utilização de água).

A.2.2 Teste Estatístico Aplicado

Para verificar se a diferença entre os resultados dos experimentos foi estatisticamente significante, foi realizado o teste t de Student corrigido. O nível de significância do teste foi originalmente determinado como 5% ($\alpha=0.05$). Após aplicada a correção de Bonferroni, o α utilizado na comparação dos algoritmos, quando utilizados com um método específico, teve valor 0.005. Quando comparados os desempenhos dos três métodos, o valor de α foi de 0.01. Como foi utilizado o 5-fold cross-validation, os testes foram realizados utilizando 4 graus de liberdade.

A.2.3 Experimento

O experimento foi executado de maneira a comparar o desempenho das técnicas Um-Contra-Todos, Label-Powerset, Cross-Training, RAKEL e ML-kNN. Para a comparação foram utilizados os algoritmos KNN, SVMs, C4.5, Ripper e BayesNet. Esses algoritmos também foram comparados quanto ao seu desempenho quando utilizados com a mesma técnica de classificação multirrótulo. As tabelas A.17, A.18, A.19, A.20 e A.21 apresentam os resultados obtidos com a utilização das cinco técnicas na base de dados de Yeasts, juntamente com o desvio padrão calculado. As tabelas A.22, A.23, A.24, A.25 e A.26 mostram os resultados para a base de dados de famílias de proteínas. Para a avaliação dos resultados, foram utilizadas cinco medidas: Hamming-Loss, Acurácia, Precisão, Revocação e Medida-F.

Tabela A.17: Resultados Utilizando a Técnica Um-Contra-Todos na base de dados de Yeasts

~~						
		KNN	SVN	C4.5	Ripper	BayesNet
	HammingLoss	0.20 (0.006)	0.20 (0.003)	0.24 (0.005)	0.21 (0.006)	0.25 (0.012)
	Precisão	67.14(1.0)	67.85(1.0)	60.28(1.7)	69.11(2.6)	59.06(2.2)
	Revocação	60.38(1.5)	60.49(1.3)	56.07(2.4)	54.71(1.0)	60.65(1.7)
	Medida-F	60.86 (1.1)	61.15(1.0)	55.06(1.7)	57.97(1.2)	56.74(2.0)
	Acurácia	51.02 (1.1)	51.05 (0.9)	42.79(1.4)	46.32 (1.0)	45.01 (2.1)

Tabela A.18: Resultados Utilizando a Técnica Label-Powerset na base de dados de Yeasts

	KNN	SVN	C4.5	Ripper	BayesNet
HammingLoss	0.24 (0.005)	0.20 (0.003)	0.27 (0.006)	0.25 (0.007)	0.27 (0.009)
Precisão	60.51(1.1)	68.03(1.0)	53.89(1.6)	58.55(1.0)	54.27(1.4)
Revocação	54.83(1.5)	63.50(1.3)	51.33(1.4)	54.83(1.4)	50.49(2.0)
Medida-F	55.39 (1.0)	63.48(0.7)	50.01 (1.1)	55.27(1.2)	51.41 (1.6)
Acurácia	46.58(1.1)	54.55 (0.6)	40.06 (0.8)	44.71(1.5)	39.78(1.5)

Tabela A.19: Resultados Utilizando a Técnica Cross-Training na base de dados de Yeasts

	KNN	SVN	C4.5	Ripper	BayesNet
HammingLoss	0.22 (000.7)	0.22 (0.008)	0.35 (0.008)	0.30 (0.005)	0.25 (0.003)
Precisão	63.14(0.8)	63.29(1.3)	47.24(1.0)	51.32(0.4)	58.50 (0.7)
Revocação	70.96(1.4)	78.22(1.7)	77.94(1.4)	81.74 (0.4)	69.25(1.2)
Medida-F	64.28(0.9)	67.12(1.3)	56.55 (1.0)	60.78(0.2)	60.90(0.7)
Acurácia	53.61 (0.9)	$56.26\ (1.2)$	42.56 (0.9)	47.19(0.3)	48.88(0.7)

Tabela A.20: Resultados Utilizando a Técnica RAKEL na base de dados de Yeasts

	KNN	SVN	C4.5	Ripper	BayesNet
HammingLoss	0.24 (000.8)	0.19 (0.006)	0.18 (0.0083	0.21 (0.005)	0.22 (0.004)
Precisão	59.56(1.5)	70.38(1.2)	70.33(0.8)	68.35 (0.5)	64.62(1.2)
Revocação	59.48(1.2)	64.44(1.9)	$57.61\ (0.6)$	54.71(2.0)	59.94(0.8)
Medida-F	59.52(1.3)	67.27(1.5)	64.52 (0.6)	60.76(1.3)	62.21 (0.9)
Acurácia	47.49(1.6)	55.14 (1.4)	51.24 (0.4)	48.35(1.3)	48.84 (0.8)

Tabela A.21: Resultados Utilizando a Técnica ML-kNN na base de dados de Yeasts

HammingLoss	Precisão	Revocação	Medida-F	Acurácia
0.19 (0.006)	72.04(1.3)	58.43 (1.5)	64.51(1.1)	51.06 (1.0)

Tabela A.22: Resultados Utilizando a Técnica Um-Contra-Todos na base de dados de famílias de proteínas

	KNN	SVN	C4.5	Ripper	BayesNet
HammingLoss	0.002 (0.001)	0.0005 (0.0003)	0.001 (0.001)	0.001 (0.0003)	0.004 (0.0009)
Precisão	98.13 (1.2)	99.67(0.3)	99.11 (0.8)	99.11 (0.3)	96.39 (1.0)
Revocação	97.78(1.5)	99.77(0.3)	99.04 (1.1)	99.47(0.5)	98.33(0.3)
Medida-F	97.90 (1.3)	99.68 (0.2)	98.93(0.9)	99.21 (0.3)	97.00 (0.7)
Acurácia	97.73(1.4)	99.59(0.2)	98.61 (1.3)	99.04 (0.3)	96.24 (0.8)

Tabela A.23: Resultados Utilizando a Técnica Label-Powerset na base de dados de famílias de proteínas

	KNN	SVN	C4.5	Ripper	BayesNet
HammingLoss	0.003 (0.002)	0.0006 (0.0008)	0.001 (0.002)	0.003 (0.003)	0.0007 (0.0008)
Precisão	98.43(1.2)	99.79(0.3)	99.44(0.7)	98.43 (1.3)	99.95(0.1)
Revocação	98.08 (1.6)	99.67(0.5)	99.41 (0.7)	98.41 (1.2)	99.67(0.4)
Medida-F	98.20 (1.4)	99.70(0.4)	99.41(0.7)	98.40 (1.2)	99.76(0.2)
Acurácia	98.03(1.5)	99.62 (0.5)	99.32(0.7)	98.36 (1.2)	99.62 (0.4)

Tabela A.24: Resultados Utilizando a Técnica *Cross-Training* na base de dados de famílias de proteínas

	KNN	SVN	C4.5	Ripper	BayesNet
HammingLoss	0.061 (0.0075)	0.061 (0.0073)	0.061 (0.0073)	0.062 (0.0075)	0.061 (0.0072)
Precisão	77.55(2.7)	77.86(2.9)	77.83(2.9)	77.25(2.9)	78.16(2.8)
Revocação	$82.51\ (2.6)$	83.57 (1.7)	83.52(1.8)	82.82 (2.0)	83.72(1.9)
Medida-F	78.57(2.6)	79.25(2.3)	79.23(2.4)	78.60(2.5)	79.43(2.4)
Acurácia	74.64 (3.0)	75.11 (2.8)	75.11 (2.8)	74.51 (2.9)	75.29(2.8)

proteínas

	KNN	SVN	C4.5	Ripper	BayesNet
HammingLoss	0.002 (0.001)	0.0006 (0.0004)	0.001 (0.0003)	0.001 (0.0003)	0.003 (0.001)
Precisão	98.59(0.9)	99.57(0.4)	99.11 (0.3)	99.11 (0.3)	97.38 (1.1)
Revocação	98.28(0.6)	99.77(0.3)	99.47(0.5)	99.47(0.5)	98.54(0.9)
Medida-F	98.43(0.7)	99.67(0.3)	99.29(0.4)	99.29(0.4)	97.95(0.8)
Acurácia	98.23(0.6)	99.49(0.3)	99.04 (0.4)	99.04 (0.4)	97.27 (1.0)

Tabela A.25: Resultados Utilizando a Técnica RAKEL na base de dados de famílias de

Tabela A.26: Resultados Utilizando a Técnica ML-kNN na base de dados de famílias de proteínas

HammingLoss	Precisão	Revocação	Medida-F	Acurácia
0.002 (0.001)	97.81 (1.1)	97.86 (1.3)	97.99 (1.2)	97.81 (1.3)

Na base de dados de Yeasts, considerando as tabelas A.17, A.18, A.19 e A.20, os melhores desempenhos foram obtidos, na maioria dos algoritmos, pelas técnicas Cross-Training e RAKEL. Isto ocorreu porque, na técnica Cross-Training, os dados são usados mais de uma vez durante a fase de treinamento, tornando os dados menos esparsos. A técnica Label-Powerset obteve os piores resultados na maioria dos algoritmos devido ao grande número de classes criadas para o treinamento dos classificadores (198), e algumas dessas classes ficaram com poucos exemplos. A técnica RAKEL teve um desempenho melhor que a técnica Label-Powerset devido à combinação de muitos classificadores Label-Powerset, evitando o problema de dados esparsos.

Alguns dos resultados obtidos pela técnica ML-kNN (Tabela A.21), foram melhores que os obtidos pelas outras quatro técnicas. Comparada com a técnica Um-Contra-Todos, a técnica ML-kNN obteve um melhor desempenho na maioria dos algoritmos, mas não foi detectada diferença estatisticamente significante entre a técnica ML-kNN e os algoritmos SVM e KNN, com exceção da medida precisão. Na comparação com a técnica Label-Powerset, a técnica ML-kNN teve um desempenho melhor, mas não houve diferença estatisticamente significante na comparação com o algoritmo SVM, exceto na medida acurácia. Os experimentos também mostraram que o algoritmo SVM obteve os melhores resultados nessa medida. A técnica Cross-Training, usando os algoritmos SVM e KNN, obteve um desempenho melhor que a técnica ML-kNN, exceto nas medidas precisão e hamming loss. O algoritmo SVM, utilizado na técnica RAKEL, obteve um melhor desempenho do que a técnica ML-kNN na maioria das medidas, com exceção das medidas precisão e hammingloss. O algoritmo C4.5 também obteve melhores resultados que o ML-kNN nessas duas medidas, mas não foi detectada diferença estatisticamente significante.

Comparando os algoritmos considerando as técnicas separadamente, o algoritmo SVM obteve os melhores resultados nas técnicas Um-Contra-Todos e Label-Powerset. Não houve diferença estatisticamente significante na comparação com o algoritmo KNN na técnica Um-Contra-Todos. Na técnica Label-Powerset, entretanto, o desempenho do algoritmo SVM mostrou-se estatisticamente superior. Na técnica Cross-Training, o algoritmo SVM obteve os melhores resultados, exceto na medida revocação, na qual o algoritmo Ripper obteve um melhor desempenho. Em algumas outras medidas, o melhor desempenho do algoritmo SVM também não mostrou-se estatisticamente significante. Na técnica RA-KEL, o algoritmo SVM também apresentou os melhores resultados, exceto nas medidas hamming-loss e precisão, em que o algoritmo C4.5 foi melhor, porém sem diferença estatisticamente significante na medida hamming-loss. O melhor desempenho do algoritmo SVM foi estatisticamente significante.

Na base de dados de famílias de proteínas, considerando as tabelas A.22, A.23, A.24 e A.25, os melhores resultados foram obtidos, na maioria dos algoritmos, pelas técnicas Label-Powerset e RAKEL. Uma possível razão para isso é que, nesse conjunto de dados, diferente do conjunto de dados de Yeasts, há poucas combinações multirrótulo, e o número de rótulos formados para treinamento foi muito menor, igual a 15. O alto desempenho obtido pelas técnicas Label-Powerset, Um-Contra-Todos e RAKEL pode ser devido à distribuição das classes no conjunto de treinamento. Esse conjunto de dados tem menos classes por exemplo e menos exemplos multirrótulo que o conjunto de dados de Yeasts. Apesar das diferenças nos resultados das três técnicas, houve diferença estatisticamente significante apenas nos resultados do algoritmo BayesNet. O pior desempenho da técnica Cross-Training pode ser devido ao grande número de exemplos que não são multirrótulo, o que afeta o desempenho da técnica, que utiliza os dados mais desbalanceados durante a fase de treinamento.

Quando comparando a técnica ML-kNN (Tabela A.26) com as outras quatro técnicas, as diferenças de desempenho das técnicas ML-kNN, Um-Contra-Todos, *Label-Powerset* e RAKEL não foi estatisticamente significante, com exceção do algoritmo BayesNet na técnica Um-Contra-Todos para a medida *hamming-loss*. Houve diferença estatisticamente significante em todos os resultados quando foi feita a comparação das técnicas ML-kNN e *Cross-Training*.

Considerando as técnicas separadamente, o algoritmo SVM obteve resultados um pouco melhores que os outros algoritmos, com exceção dos algoritmos da técnica Cross-Training. Entretanto, praticamente não houve diferença estatisticamente significante nos resultados. Dentre os algoritmos da técnica Um-Contra-Todos, houve diferença estatisticamente significante apenas na comparação dos resultados do algoritmo BayesNet com os outros algoritmos. Na técnica Label-Powerset, não houve diferença estatisticamente significante nos resultados. Na técnica Cross-Training, houve significância estatística apenas na diferença dos resultados dos algoritmos Ripper e BayesNet para as medidas acurácia, revocação e medida-F. Na técnica RAKEL, a diferença de resultados foi estatisticamente significante entre o algoritmo KNN e os algoritmos SVM, C4.5 e Ripper para a medida revocação. Nas outras medidas, houve diferença estatisticamente significante apenas na comparação dos resultados dos algoritmos SVM e BayesNet.

A.3 Considerações Finais

Este apêndice apresentou alguns experimentos preliminares realizados utilizando técnicas de classificação hierárquica simples-rótulo e técnicas de classificação multirrótulo não hierárquica. Por meio dos experimentos realizados, foi adquirido conhecimento sobre as técnicas de classificação hierárquica e classificação multirrótulo, que foi fundamental para o desenvolvimento das técnicas de classificação hierárquica multirrótulo desta pesquisa.

Apêndice B

Valores de Precisão e Revocação Hierárquicas

Neste apêndice são apresentados os resultados obtidos utilizando as medidas precisão e revocação hierárquicas, micro/macro precisão e revocação hierárquicas e micro/macro precisão e revocação hierárquicas ponderada. Os melhores resultados para cada conjunto de dados são destacados em negrito, e os desvios padrão são mostrados entre parênteses.

As Tabelas B.1 a B.40 apresentam os resultados das comparações entre as técnicas baseadas nas abordagens *Top-Down* e *One-Shot*, considerando as medidas precisão e revocação hierárquicas e suas variações. Por meio desses resultados, foram calculados os valores das medidas F1 hierárquica, micro/macro F1 hierárquicas e micro/macro F1 hierárquicas ponderada, apresentados no Capítulo 6.

É importante lembrar que não é adequado avaliar os desempenhos dos classificadores considerando as medidas de precisão e revocação separadamente, pois um classificador precisa ser, ao mesmo tempo, preciso e cobrir um grande número de exemplos. Por isso, a medida F1 hierárquica foi utilizada para a avaliação dos classificadores no Capítulo 6, combinando as medidas de precisão e revocação. Os valores obtidos com as medidas de precisão e revocação são apresentados neste apêndice apenas para informar ao leitor sobre os comportamentos das técnicas considerando essas duas medidas.

De acordo com os resultados apresentados nas Tabelas B.1 a B.8, a técnica HMC-BR obteve os melhores resultados de precisão hierárquica em todos os níveis hierárquicos, seguida pela técnica HMC-LP, que obteve alguns dos melhores resultados no primeiro nível das hierarquias. Por outro lado, essas duas técnicas obtiveram os piores resultados para a medida revocação hierárquica, na maioria de seus algoritmos, se comparadas com as técnicas HMC-CT, C4.5H e Clus-HMC. Isso ocorreu porque essas últimas três técnicas têm como característica a predição de um maior número de classes, fazendo com que mais erros sejam cometidos, diminuindo os valores de precisão hierárquica e aumentando os valores de revocação hierárquica.

Nas medidas micro precisão hierárquicas não ponderada e ponderada, percebe-se bons resultados também da técnica HMC-LP, em muitas vezes melhores que os resultados obtidos pela técnica HMC-BR, em todos os níveis hierárquicos. Entretanto, nas medidas micro revocação hierárquicas não ponderada e ponderada, os melhores desempenhos ainda são obtidos pelas técnicas HMC-CT, C4.5H e Clus-HMC.

Como explicado no Capítulo 6, as medidas precisão e revocação hierárquicas apenas consideram a intersecção entre o conjunto de classes preditas e de classes verdadeiras dos exemplos, juntamente com as classes ancestrais dessas classes. Elas não consideram distâncias entre as classes e também não consideram que predições em níveis mais profundos são mais importantes. Essas considerações são feitas pelas medidas micro/macro precisão e revocação hierárquicas e micro/macro precisão e revocação hierárquicas ponderada. Devido a essas diferenças, as medidas apresentaram resultados diferentes umas das outras.

Nas medidas macro precisão hierárquica não ponderada e ponderada, os melhores desempenhos foram obtidos pelas técnicas HMC-BR e HMC-LP. Já na medida macro revocação hierárquica não ponderada, as técnicas HMC-CT, C4.5H e Clus-HMC obtiveram melhores resultados, com exceção do terceiro e quarto níveis, em que os desempenhos das técnicas tiveram uma variação maior. Quando pesos são atribuídos às arestas das hierarquias (macro revocação hierárquica ponderada), os desempenhos das técnicas HMC-CT, C4.5H e Clus-HMC se mantêm melhores em todos os níveis.

As variações dos resultados considerando as avaliações das micro e macro medidas ocorreu devido às diferentes considerações feitas por cada medida. Como apresentado no Capítulo 6, ao contrário das micro medidas, as macro medidas consideram que todas as classes da hierarquia têm a mesma importância, independente da frequência com que elas aparecem, já que avaliam os classificadores calculando a média dos desempenhos de cada classe. Em alguns conjuntos de dados, erros foram cometidos majoritariamente em algumas subárvores. Assim, ao considerar a média dos desempenhos sobre todas as classes, pode-se ter favorecido a avaliação de algumas técnicas em alguns conjuntos de dados.

Tabela B.1: Comparação das Abordagens Top-Down e One-Shot Utilizando a Medida

Precisão Hierárquica no Primeiro Nível da Hierarquia

Conjunto de Dados			n Top-Down		Abordagen	
conjunio de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HM0
	KNN	$57.81\ (1.3)$	57.38(3.0)	54.18 (1.0)		
	C4.5	53.78(2.4)	52.99(2.0)	49.82(1.7)		
Expr	Ripper	60.10(3.2)	$57.50\ (1.9)$	51.67 (1.7)	44.88 (1.7)	48.13 (1.9
	BayesNet	56.78(1.9)	60.06(2.8)	60.26 (2.4)		
	SVM	64.11 (4.1)	61.72 (1.9)	48.90 (1.5)		
	KNN	58.43 (2.2)	57.69 (2.2)	55.16 (2.1)		
	C4.5	52.42(3.2)	49.66(1.5)	47.13 (1.1)		
CellCycle	Ripper	57.38 (1.7)	55.50(2.7)	48.93 (1.9)	41.43 (1.7)	43.35 (1.0
	BayesNet	58.28(3.4)	59.23(1.7)	56.97(1.7)		
	SVM	59.39(2.7)	$60.72\ (2.8)$	57.76 (1.8)		
	KNN	51.22 (1.2)	49.16 (0.7)	49.04 (1.7)		
	C4.5	52.98 (2.5)	48.89 (0.8)	47.24 (1.8)		
Church	Ripper	55.27 (1.1)	48.94 (0.6)	46.40 (2.0)	34.16 (1.1)	35.46 (1.4
	BayesNet	48.98 (1.1)	52.81 (1.0)	49.23 (2.7)	, ,	·
	SVM	53.53 (1.5)	52.48 (1.3)	49.15 (1.6)		
	KNN	51.81 (1.7)	49.88 (1.7)	49.20 (1.3)		
	C4.5	54.94 (4.7)	46.36 (1.1)	45.32 (1.3)		
Derisi	Ripper	55.21 (1.6)	50.40 (0.5)	48.00 (1.5)	39.18 (1.1)	41.77 (1.8
20101	BayesNet	49.95 (1.5)	51.12 (4.8)	49.86 (1.7)	30.13 (1.1)	11111 (111
	SVM	52.53 (2.3)	50.12 (1.5)	49.31 (1.5)		
	KNN	58.19 (3.7)	57.41 (1.5)	56.12 (3.0)		
	C4.5	51.57 (1.8)	52.09 (1.1)	48.58 (3.1)		
Eisen	Ripper	58.55 (3.0)	` '	ì í	42.37 (1.8)	45.86 (1.4
Eisen	BayesNet	58.28 (3.1)	57.90 (1.9)	52.12 (3.2)	42.37 (1.6)	45.60 (1.4
	SVM	59.80 (2.8)	59.42 (2.5) 62.42 (1.9)	58.99 (1.5) 57.61 (3.2)		
	KNN	57.90 (2.6)	57.50 (1.4)	55.17 (1.7)		
~	C4.5	54.66 (2.4)	52.87 (1.2)	48.71 (1.0)		/.
Gasch1	Ripper	58.94 (4.8)	53.67 (1.2)	51.18 (2.2)	43.03 (1.7)	45.72 (1.9
	BayesNet	54.74 (1.0)	55.76 (2.5)	55.05 (1.0)		
	SVM	64.32 (2.9)	65.99 (1.9)	60.22 (2.8)		
	KNN	56.29(1.7)	53.59(1.3)	53.56(2.3)		
	C4.5	53.39(1.7)	49.79(1.7)	46.48 (1.2)		
Gasch2	Ripper	57.10(2.7)	54.10(2.0)	48.03(1.5)	41.00 (1.7)	42.66 (0.9)
	BayesNet	53.42(1.2)	$53.11\ (1.7)$	52.29(1.7)		
	SVM	55.92 (1.3)	57.74 (0.7)	53.28 (0.9)		
	KNN	54.31(2.6)	50.25 (0.8)	47.43 (0.6)		
	C4.5	53.81(2.6)	52.82(3.2)	46.89(2.6)		
Phenotype	Ripper	55.32(1.8)	53.95(1.3)	46.30(1.6)	33.88 (0.7)	34.93 (1.8
	BayesNet	44.60(2.4)	41.37(2.2)	41.94(2.0)		
	SVM	$56.48\ (1.8)$	52.38(2.6)	47.22(1.9)		
	KNN	51.80 (2.3)	51.08 (1.8)	48.78 (1.6)		
	C4.5	53.25 (0.6)	51.20 (2.3)	49.18 (1.5)		
Sequence	Ripper	59.10 (2.3)	53.10 (1.7)	49.04 (0.8)	42.39 (1.6)	45.09 (2.3
-	BayesNet	57.51 (1.5)	54.02 (2.1)	57.57 (1.0)		,
	SVM	76.98 (3.0)	62.63 (2.9)	51.22 (1.1)		
	KNN	50.46 (2.5)	49.00 (0.9)	49.32 (1.9)		
	C4.5	51.62 (2.0)	47.93 (1.9)	45.14 (1.1)		
SPO	Ripper	57.24 (2.5)	54.15 (2.1)	50.16 (1.5)	38.82 (0.8)	41.32 (2.0
D1 O	BayesNet	49.87 (2.6)	54.13 (2.1) 50.81 (2.5)	50.16 (1.5) 50.01 (2.4)	90.02 (0.0)	41.02 (2.0

Tabela B.2: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Precisão Hierárquica no Segundo Nível da Hierarquia

Conjunto de Dados		_	n Top-Down		Abordagen	
Conjunto de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HM0
	KNN	46.83(1.5)	43.89(2.0)	35.85(1.1)		
	C4.5	$39.21\ (1.8)$	37.65 (1.8)	27.46(1.1)		
Expr	Ripper	48.98(3.5)	38.72(1.7)	30.74(1.7)	29.96 (1.5)	33.30 (1.9
	BayesNet	42.50 (2.1)	45.58(2.3)	38.98 (1.5)		
	SVM	58.41 (3.8)	41.68 (1.7)	31.54 (1.1)		
	KNN	47.38 (2.4)	42.83(2.5)	35.52(1.8)		
	C4.5	37.19(2.0)	$33.86 \ (0.8)$	25.24 (0.9)		
CellCycle	Ripper	46.17(2.1)	35.76(2.1)	29.41 (1.4)	27.14 (1.2)	29.13 (0.6
	BayesNet	46.72(3.2)	43.86 (0.9)	37.19 (0.8)		
	SVM	46.82 (2.4)	46.54 (3.0)	38.06 (1.3)		
	KNN	39.43(2.0)	34.79(0.9)	30.16(1.1)		
	C4.5	41.55(2.0)	34.74(1.0)	27.22(1.6)		
Church	Ripper	44.79 (0.8)	29.78(0.4)	29.04 (1.4)	16.23 (1.0)	18.91 (1.0
	BayesNet	$38.01\ (1.0)$	$38.94\ (1.6)$	32.49(1.4)		
	SVM	40.80 (1.3)	37.29 (0.5)	29.93 (1.8)		
	KNN	40.82(1.6)	$35.23 \ (0.7)$	30.31 (0.9)		
	C4.5	44.03(2.7)	32.05(1.0)	24.73(0.9)		
Derisi	Ripper	$44.45 \ (1.1)$	31.22 (0.6)	29.82 (0.8)	25.25 (0.6)	27.88 (2.0
	BayesNet	37.67(1.0)	37.77(4.4)	$33.31\ (1.1)$		
	SVM	38.97 (2.0)	35.13 (1.5)	30.49 (0.93		
	KNN	48.67 (3.2)	43.93 (1.8)	37.03 (2.5)		
	C4.5	36.79(1.2)	36.63(1.8)	26.55(2.2)		
Eisen	Ripper	47.57(4.2)	40.53(3.2)	30.67(1.8)	28.17 (1.9)	30.99 (1.
	BayesNet	45.55(3.0)	44.71(2.7)	39.18(1.3)		
	SVM	48.32(3.0)	$49.34\ (1.9)$	38.38(2.3)		
	KNN	48.09 (1.2)	44.47 (1.1)	37.18 (1.1)		
	C4.5	39.76(1.9)	36.87 (0.9)	27.22(0.6)		
Gasch1	Ripper	47.73(3.3)	35.54(1.2)	30.38(2.5)	28.67 (0.6)	31.06 (1.5
	BayesNet	40.99(1.1)	41.68(1.8)	36.67(1.4)		
	SVM	$54.01\ (1.5)$	52.68(2.5)	41.53(2.6)		
	KNN	45.81 (1.5)	40.97 (1.4)	35.06 (1.7)		
	C4.5	40.44 (1.4)	34.70 (1.5)	25.54(0.8)		
Gasch2	Ripper	$46.54\ (2.0)$	34.62(2.0)	29.53(1.0)	26.93 (1.0)	29.29 (0.9
	BayesNet	40.04(1.9)	39.06(1.8)	34.80(1.0)		
	SVM	41.83 (0.9)	43.20 (0.8)	34.14 (1.1)		
	KNN	47.36 (4.0)	34.59 (1.4)	28.54 (1.1)		
	C4.5	48.07(2.8)	35.76(2.5)	27.90 (1.8)		
Phenotype	Ripper	47.26(3.5)	32.25(1.5)	27.20(1.3)	15.28 (0.3)	18.50 (2.3
	BayesNet	33.03(1.2)	29.14(1.9)	$22.51 \ (0.5)$		
	SVM	48.26 (2.3)	35.14(2.1)	28.42 (0.9)		
	KNN	41.99 (1.7)	39.77 (1.5)	32.28 (1.1)		
	C4.5	38.76 (0.4)	36.41 (1.8)	26.71 (1.2)		
Sequence	Ripper	47.52 (1.8)	34.18 (0.7)	28.71 (0.9)	28.80 (1.2)	30.83 (1.5
	BayesNet	43.19 (1.1)	38.59(2.0)	35.24 (0.7)		
	SVM	69.86 (2.1)	46.30 (2.5)	34.55 (0.5)		
	KNN	38.93 (1.7)	34.19 (0.7)	30.38 (1.1)		
	C4.5	39.41 (1.5)	33.09 (2.2)	24.19 (0.7)		
SPO	Ripper	47.80 (2.6)	35.52 (1.8)	30.85 (1.0)	25.21 (1.3)	28.32 (1.0
	BayesNet	38.06 (2.1)	38.58 (2.1)	33.91 (2.1)	<u> </u>	`
	SVM	37.50 (1.6)	36.53 (0.9)	30.69 (1.1)		

Tabela B.3: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Precisão Hierárquica no Terceiro Nível da Hierarquia

Conjunto de Dados			Top-Down		Abordagen	n One-Shot
Conjunto de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	40.76 (1.7)	36.95(1.6)	22.91 (0.9)		
	C4.5	32.03(1.7)	30.04(1.8)	15.77 (0.8)		
Expr	Ripper	42.47 (3.5)	$30.11\ (1.4)$	18.18 (1.1)	21.48 (0.9)	24.91 (1.6
	BayesNet	35.53(1.8)	38.05(2.1)	24.65 (1.0)		
	SVM	53.53 (3.8)	32.37 (1.6)	21.65 (0.7)		
	KNN	41.78 (2.3)	36.18(2.3)	23.17 (1.2)		
	C4.5	29.99(1.6)	27.06(0.5)	$14.18 \; (0.5)$		
CellCycle	Ripper	40.07(1.9)	27.58(1.8)	17.82 (0.9)	19.61 (0.8)	21.57 (0.3)
	BayesNet	40.71(2.6)	36.76(1.2)	24.28 (0.6)		
	SVM	40.25 (2.2)	39.12 (3.2)	25.19 (0.7)		
	KNN	33.68 (2.2)	28.08 (1.1)	18.84 (0.9)		
	C4.5	35.54(2.3)	28.17(1.2)	16.32 (1.1)		
Church	Ripper	$39.40 \ (1.5)$	21.01 (0.5)	18.86 (1.1)	09.45 (0.5)	11.43 (1.5
	BayesNet	33.25(1.3)	31.68(0.9)	20.62 (1.2)		
	SVM	34.06 (0.7)	30.00(0.6)	18.80 (1.3)		
	KNN	34.94 (1.5)	28.73 (0.8)	18.45 (0.4)		
	C4.5	38.19 (1.3)	25.38 (1.0)	14.19 (0.5)		
Derisi	Ripper	38.94 (1.1)	22.66 (0.5)	18.17 (0.4)	17.98 (0.4)	20.26 (1.8
	BayesNet	31.27 (0.6)	31.66 (4.3)	19.98 (1.2)	, ,	`
	SVM	31.63 (1.5)	27.60 (1.3)	18.83 (0.8)		
	KNN	42.74 (2.8)	37.08 (1.9)	23.87 (1.6)		
	C4.5	29.85 (1.3)	29.39 (1.8)	15.12 (1.3)		
Eisen	Ripper	41.68 (4.5)	32.44 (3.2)	18.16 (1.1)	20.60 (1.5)	23.13 (1.0
	BayesNet	39.51 (2.9)	37.31 (2.7)	25.52 (1.4)		
	SVM	42.38 (3.2)	42.28 (2.0)	25.52 (1.8)		
	KNN	42.08 (1.3)	37.73 (0.7)	24.18 (0.7)		
	C4.5	31.96 (1.6)	29.11 (0.8)	15.71 (0.3)		
Gasch1	Ripper	40.85 (2.9)	27.33 (1.2)	18.09 (1.7)	20.49 (0.5)	23.08 (1.3
Gasciii	BayesNet	32.73 (1.2)	34.44 (1.8)	22.56 (0.8)	20.10 (0.0)	20.00 (1
	SVM	48.77 (1.4)	45.02 (2.4)	28.47 (1.8)		
	KNN	39.67 (1.4)	34.10 (1.4)	22.15 (1.3)		
	C4.5	33.44 (1.5)	27.59 (1.0)	14.62 (0.6)		
Gasch2	Ripper	40.84 (2.0)	25.89 (1.7)	18.09 (0.4)	19.06 (0.8)	21.78 (0.9
	BayesNet	34.35 (1.6)	32.78 (1.8)	22.45 (0.6)	10.00 (0.0)	21.10 (0
	SVM	35.03 (0.9)	35.92 (1.2)	21.98 (0.7)		
	KNN	44.80 (4.0)	26.89 (1.1)	17.70 (0.6)		
	C4.5	44.45 (3.7)	27.80 (2.4)	17.18 (1.1)		
Phenotype	Ripper	42.27 (4.1)	22.63 (1.2)	16.14 (0.9)	09.03 (0.1)	12.75 (2.5
r nenety pe	BayesNet	26.61 (1.0)	24.42 (1.6)	14.90 (0.7)	00.00 (0.1)	12.10 (2.1
	SVM	42.82 (2.0)	26.83 (1.7)	17.47 (0.6)		
	KNN	36.22 (1.5)	32.01 (1.1)	19.43 (0.6)		
	C4.5	31.13 (0.6)	28.92 (1.4)	14.85 (0.5)		
Sequence	Ripper	41.26 (1.3)	26.00 (0.4)	16.72 (0.8)	21.16 (1.0)	22.82 (0.8
Sequence	BayesNet	37.02 (1.5)	32.01 (1.5)	22.26 (0.7)	21.10 (1.0)	22.02 (0.0
	SVM	64.69 (2.3)	37.56 (2.2)	24.25 (0.4)		
	KNN	33.32 (1.8)	27.83 (0.4)	18.56 (0.7)		
CDO	C4.5	32.94 (1.5)	26.39 (2.1)	13.77 (0.5)	17.00 (1.1)	20.00 /1 /
SPO	Ripper	41.26 (1.6)	26.59 (1.5)	18.92 (0.5)	17.92 (1.1)	20.89 (1.2
	BayesNet	32.02 (1.1)	32.99(2.0)	21.91 (1.3)		
	SVM	30.03(1.3)	29.38 (0.8)	19.13 (0.5)		

Tabela B.4: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Precisão Hierárquica no Quarto Nível da Hierarquia

Conjunto de Dados		_	n Top-Down		Abordagen	
conjunto de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HM0
	KNN	36.79(1.6)	33.02(1.4)	$17.82 \ (0.7)$		
	C4.5	28.96 (1.4)	$26.81\ (1.9)$	$12.00 \ (0.6)$		
Expr	Ripper	39.02(3.6)	25.34(1.4)	$13.34 \ (0.8)$	18.15 (0.8)	21.42 (1.5)
	BayesNet	32.19(1.7)	34.17(2.0)	19.81 (1.0)		
	SVM	49.44 (3.9)	27.13 (1.5)	15.68 (1.2)		
	KNN	37.49(2.5)	32.08(1.8)	17.87(1.1)		
	C4.5	26.69(1.5)	24.04 (0.4)	$10.62 \ (0.5)$		
CellCycle	Ripper	36.50 (1.3)	23.17(1.5)	12.67 (0.8)	16.78 (0.6)	18.39 (0.4
	BayesNet	37.76 (2.9)	$32.41\ (1.5)$	$18.31 \ (0.5)$		
	SVM	36.57 (2.3)	34.51 (3.0)	19.91 (0.7)		
	KNN	30.55 (2.1)	24.28 (1.1)	14.58 (0.9)		
	C4.5	32.35(2.5)	24.88(1.6)	12.39(0.8)		
Church	Ripper	$35.31\ (2.1)$	16.51 (0.5)	13.00 (0.8)	07.59 (0.5)	09.27 (1.1
	BayesNet	30.56(1.1)	26.49(0.6)	15.80 (0.8)		
	SVM	30.84 (1.1)	26.05 (0.6)	14.42(1.3)		
	KNN	31.37 (1.5)	25.50 (1.0)	13.74 (0.3)		
Derisi	C4.5	34.99 (1.0)	22.56 (1.3)	10.48 (0.3)		
	Ripper	35.80 (1.5)	18.20 (0.5)	12.59 (0.4)	15.32 (0.4)	17.37 (1.5
	BayesNet	29.16 (0.8)	27.92 (4.6)	15.21 (1.2)	,	`
	SVM	28.02 (1.2)	23.94 (1.1)	14.59 (0.5)		
	KNN	38.35 (2.2)	33.15 (2.1)	18.39 (1.3)		
	C4.5	26.93 (1.4)	26.30 (1.4)	11.48 (1.1)		
Eisen	Ripper	38.10 (4.6)	28.37 (3.3)	13.32 (0.9)	17.92 (1.3)	20.20 (1.7
	BayesNet	36.39 (3.2)	33.46 (2.6)	18.78 (1.4)	11.02 (1.0)	20.20 (1.1
	SVM	39.38 (3.2)	38.23 (1.8)	20.27 (1.6)		
	KNN	37.97 (1.4)	33.69 (0.6)	18.97 (0.6)		
	C4.5	28.53 (1.6)	25.84 (0.5)	11.96 (0.1)		
Gasch1	Ripper	37.53 (2.6)	23.54 (0.3) $22.59 (1.1)$	13.09 (1.3)	17.33 (0.5)	19.74 (0.9
Gasciii	BayesNet	29.81 (1.5)	30.74 (1.6)	18.01 (1.0)	17.33 (0.3)	19.74 (0.8
	SVM	45.33 (1.1)	39.78 (2.1)	22.13 (1.3)		
	KNN	35.74 (1.4)	30.39 (1.3)	16.87 (1.2)		
Gasch2	C4.5	30.22 (1.9)	24.69 (1.0)	11.00 (0.5)	15 00 (0.0)	10.00 /1.1
Gasciiz	Ripper	37.22 (2.1)	21.29 (1.7)	13.07 (0.4) 17.45 (1.4)	15.96 (0.8)	18.88 (1.1
	BayesNet SVM	31.96 (1.5) 31.76 (0.9)	29.42 (1.9) 31.92 (1.4)	17.45 (1.4) $17.15 (0.7)$		
				. ,		
	KNN	42.15 (4.1)	22.46 (0.7)	13.14 (0.8)		
Til.	C4.5	42.19 (4.6)	23.12 (2.5)	12.64 (0.6)	0= 4= (0.0)	40 70 (4.4
Phenotype	Ripper	39.13 (4.3)	17.75 (1.0)	11.20 (0.8)	$07.17 \ (0.2)$	10.50 (1.9
	BayesNet	24.78 (1.3)	20.69 (1.4)	11.92 (1.0)		
	SVM	39.23 (2.3)	22.46 (1.5)	13.13 (0.3)		
	KNN	32.85 (1.6)	28.71 (0.9)	14.65 (0.5)		
_	C4.5	27.89 (0.6)	26.12 (1.2)	11.09 (0.2)		
Sequence	Ripper	37.73 (1.5)	21.91 (0.5)	12.46 (0.9)	18.15 (1.1)	19.62 (0.9
	BayesNet	34.78 (1.7)	28.76 (1.2)	17.35 (0.9)		
	SVM	62.49 (2.6)	32.31 (1.9)	18.20 (0.4)		
	KNN	29.59 (1.9)	24.58 (0.3)	13.80 (0.5)		-
	C4.5	29.57(1.4)	23.52(1.6)	$10.44 \ (0.4)$		
SPO	Ripper	$37.66 \ (1.5)$	21.45(1.1)	$13.46 \ (0.4)$	15.10 (1.0)	18.08 (1.1
	BayesNet	29.27(1.4)	29.21 (1.9)	17.77(1.3)		
	SVM	26.89 (1.3)	25.77 (1.1)	14.85 (0.3)		

Tabela B.5: Comparação das Abordagens Top-Down e One-Shot Utilizando a Medida

Revocação Hierárquica no Primeiro Nível da Hierarquia

Conjunto de Dados			n Top-Down		Abordagem	
Conjunio de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	$49.71 \ (0.8)$	48.17(1.6)	60.20(1.6)		
	C4.5	53.97(1.2)	53.02(2.5)	75.40(3.3)		
Expr	Ripper	52.99(3.4)	43.93(1.5)	$76.22 \ (1.1)$	56.59(1.5)	55.94 (1.6)
	BayesNet	68.11(2.3)	60.85 (2.3)	58.79(3.9)		
	SVM	40.54 (1.6)	43.86 (1.6)	66.94 (0.9)		
	KNN	51.94 (2.0)	49.94 (1.7)	61.17 (1.6)		
	C4.5	50.95(1.9)	50.35(2.0)	$76.50 \ (1.9)$		
CellCycle	Ripper	47.01 (1.8)	41.75 (1.8)	74.56(2.2)	55.69 (1.6)	51.67 (2.8
	BayesNet	55.03 (1.9)	49.02(2.2)	59.16 (1.7)		
	SVM	55.64(1.6)	56.43 (2.8)	66.58(1.8)		
	KNN	39.79 (2.6)	39.72 (0.8)	59.64 (1.3)		
	C4.5	38.24 (2.0)	41.81 (0.9)	64.99 (2.0)		
Church	Ripper	39.70 (2.9)	34.94 (0.7)	66.77 (2.1)	78.63 (1.7)	70.95 (1.4
	BayesNet	63.56 (0.9)	38.43 (1.5)	49.35 (5.1)	, ,	`
	SVM	37.90 (2.2)	40.25 (1.2)	64.70 (2.1)		
	KNN	44.52 (1.7)	43.09 (1.1)	55.98 (1.6)		
	C4.5	33.55 (10.0)	45.87 (1.1)	71.12 (2.4)		
Derisi	Ripper	41.22 (2.3)	37.13 (0.6)	69.44 (2.2)	54.29 (1.1)	51.77 (3.1
Derisi	BayesNet	59.60 (1.5)	40.69 (2.7)	49.43 (4.0)	04.20 (1.1)	01.77 (0
	SVM	43.34 (1.1)	43.84 (1.5)	61.57 (2.6)		
	KNN	53.90 (2.4)	51.23 (1.3)	64.16 (2.6)		
Eisen	C4.5	53.59 (2.3)	52.17 (1.3)	73.28 (3.9)	F7 F0 (1.7)	FC C4 (0.
	Ripper	50.53 (3.8)	44.31 (2.0)	75.49 (2.1)	57.52 (1.7)	56.64 (2.4
	BayesNet	62.76 (1.7)	49.81 (2.8)	60.85 (3.4)		
	SVM	56.71 (3.0)	57.53 (1.8)	67.24 (3.4)		
	KNN	53.90 (3.4)	52.06 (2.2)	63.49 (1.5)		
	C4.5	56.15 (1.1)	52.84 (2.0)	76.35 (4.4)		
Gasch1	Ripper	49.39 (2.0)	40.90 (0.9)	76.79 (0.4)	56.67(2.8)	54.10 (2.1
	BayesNet	67.15 (1.8)	54.36 (2.4)	51.50 (2.8)		
	SVM	56.20 (2.1)	56.02 (1.7)	67.08 (2.2)		
	KNN	50.02(1.4)	46.66 (1.3)	60.41(2.4)		
	C4.5	45.96(5.1)	49.87(2.7)	$74.51 \ (2.9)$		
Gasch2	Ripper	44.42 (4.0)	39.43 (1.7)	73.20(1.0)	54.57(1.6)	53.41 (2.2
	BayesNet	58.15 (2.8)	$43.81 \ (0.8)$	58.72(1.7)		
	SVM	54.72 (1.6)	54.01 (1.5)	62.68 (0.9)		
	KNN	32.15(4.3)	39.27(1.6)	55.01 (6.5)		
	C4.5	32.97(5.5)	41.01 (1.8)	$63.61\ (7.3)$		
Phenotype	Ripper	36.86 (4.8)	37.94 (0.8)	65.14(1.5)	87.88 (1.7)	74.50 (7.5
	BayesNet	49.09(3.7)	47.47(1.7)	56.05(4.0)		
	SVM	39.02(4.0)	40.52 (1.6)	64.20 (6.2)		
	KNN	49.80 (3.6)	48.22 (2.3)	59.75 (3.8)		
	C4.5	53.27 (2.1)	52.07 (2.9)	74.47 (2.3)		
Sequence	Ripper	48.33 (3.7)	39.83 (1.1)	76.43 (2.0)	56.18 (2.0)	53.92 (3.3
	BayesNet	65.06 (2.1)	54.50 (3.1)	56.67 (3.9)	, ,	,
	SVM	32.20 (2.4)	45.47 (2.1)	69.75 (1.8)		
	KNN	42.41 (2.3)	41.51 (1.4)	56.01 (2.8)		
	C4.5	44.81 (12.0)	47.66 (1.6)	70.45 (2.2)		
SPO	Ripper	40.49 (3.8)	39.57 (1.6)	71.01 (1.9)	54.34 (2.0)	51.46 (1.8
J. 0	BayesNet	55.87 (0.6)	45.28 (1.5)	47.53 (2.0)	01.04 (2.0)	01.40 (1.0
	Dayconce	48.39 (2.3)	48.08 (2.7)	59.63 (1.5)		

Tabela B.6: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Revocação Hierárquica no Segundo Nível da Hierarquia

Conjunto de Dados			m Top-Down	J	Abordagem	
Conjunio de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	$36.31\ (1.3)$	36.08(1.1)	48.65 (1.6)		
	C4.5	39.52 (0.6)	38.05(2.0)	66.29 (3.4)		
Expr	Ripper	37.49(1.9)	27.65 (1.4)	65.54 (1.8)	43.78(1.4)	42.59(1.8)
	BayesNet	53.77(2.1)	$43.24\ (1.6)$	48.50 (3.4)		
	SVM	22.93 (1.0)	26.34 (1.2)	50.98 (2.0)		
	KNN	36.50 (1.9)	35.66(1.7)	48.65 (0.9)		
	C4.5	35.17(1.5)	34.18 (0.9)	$65.85\ (2.1)$		
CellCycle	Ripper	$32.31\ (1.1)$	24.86(1.1)	61.05(2.0)	42.71(1.6)	38.16(2.1)
	BayesNet	38.82 (0.9)	33.30(1.5)	48.32(1.9)		
	SVM	40.14 (1.4)	40.59 (2.6)	54.18 (1.6)		
	KNN	27.36 (1.7)	27.12 (0.2)	45.90 (1.7)		
	C4.5	26.70(1.9)	28.48(0.6)	52.28 (1.6)		
Church	Ripper	26.41 (1.2)	19.21 (0.3)	50.49 (1.7)	$67.74\ (0.81)$	60.36 (2.1)
	BayesNet	47.01 (1.7)	26.31(1.7)	41.33 (4.7)		
	SVM	26.15(2.6)	27.54(0.4)	50.68 (2.2)		
	KNN	29.83 (1.0)	29.40 (1.0)	43.18 (1.4)		
Derisi	C4.5	21.27 (6.2)	31.70 (1.2)	61.03 (1.7)		
	Ripper	26.89 (2.2)	20.77 (0.7)	54.47 (1.2)	41.49 (1.3)	38.56 (3.2)
	BayesNet	45.83 (0.7)	27.78 (2.3)	40.49 (3.2)	,	
	SVM	30.20 (1.0)	30.16 (1.2)	48.29 (1.7)		
	KNN	40.07 (1.3)	37.93 (1.4)	52.70 (2.8)		
	C4.5	37.85 (2.7)	36.30 (1.4)	63.82 (3.3)		
Eisen	Ripper	36.21 (2.2)	28.71 (2.0)	64.18 (1.1)	43.86 (1.7)	42.36 (2.8
	BayesNet	47.96 (1.6)	34.56 (1.9)	50.21 (2.5)	10.00 (1.1)	12.00 (2.0)
	SVM	42.28 (2.7)	42.55 (1.3)	56.00 (3.0)		
	KNN	40.59 (2.4)	38.68 (1.6)	52.89 (1.3)		
	C4.5	40.60 (1.5)	36.63 (1.2)	67.88 (3.6)		
Gasch1	Ripper	35.48 (0.7)	25.14 (0.6)	65.79 (1.6)	45.30 (2.0)	40.32 (1.3)
Gasciii	BayesNet	54.81 (1.5)	37.61 (2.0)	43.47 (2.1)	45.50 (2.0)	40.52 (1.5
	SVM	41.78 (1.5)	40.90 (2.0)	55.73 (1.6)		
				, ,		
	KNN	35.65 (1.0)	33.70 (0.9)	47.95 (1.8)		
C l- 0	C4.5	32.93 (3.8)	34.53 (2.2)	64.74 (3.7)	40.00 (0.0)	40 59 (0.9)
Gasch2	Ripper	31.11 (3.1)	23.11 (1.6)	60.40 (1.8)	42.08 (0.8)	40.53 (2.3)
	BayesNet	45.16 (2.0)	29.79 (0.9)	48.99 (1.6)		
	SVM	39.79 (0.6)	38.84 (1.5)	50.38 (1.0)		
	KNN	18.22 (2.4)	25.13 (1.1)	41.78 (5.5)		
TOI.	C4.5	19.19 (3.1)	26.19 (2.2)	50.12 (6.2)	10 (0 -)	FO FO (10 -
Phenotype	Ripper	21.22 (2.9)	20.12 (0.9)	52.04 (1.2)	77.18 (2.7)	58.70 (10.0
	BayesNet	30.56 (1.2)	30.07 (1.2)	39.29 (3.4)		
	SVM	23.43 (2.7)	25.75 (1.7)	49.38 (5.5)		
	KNN	37.38(2.7)	35.27(1.7)	49.50 (3.1)		
	C4.5	37.96 (0.9)	36.86 (1.5)	65.46 (2.0)		
Sequence	Ripper	33.28(2.0)	24.14 (0.7)	65.44 (3.3)	42.73(1.2)	40.51 (2.2)
	BayesNet	48.60(2.5)	37.34(2.2)	47.53 (3.5)		
	SVM	21.82 (2.1)	29.92 (1.5)	53.88 (1.8)		
	KNN	28.00 (1.7)	27.99 (0.8)	43.05 (2.0)		
	C4.5	31.21 (8.8)	32.91 (1.4)	$61.03\ (2.6)$		
SPO	Ripper	27.77(2.0)	23.60 (1.3)	57.02 (1.7)	41.57(2.0)	38.38 (1.7)
	BayesNet	42.74 (0.9)	31.44 (1.4)	38.61 (1.6)		
	SVM	33.14 (1.9)	33.09 (2.1)	46.54 (1.1)		

Tabela B.7: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Revocação Hierárquica no Terceiro Nível da Hierarquia

Conjunto de Dados		Abordage	m Top-Down		Abordagem	One-Shot
Conjunto de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	29.93 (1.3)	30.07 (1.1)	43.18 (1.5)		
	C4.5	32.13(0.4)	30.38(1.7)	60.77 (3.6)		
Expr	Ripper	30.38(1.9)	20.93(1.2)	59.92(1.9)	36.02 (1.5)	34.69(1.7)
	BayesNet	45.59(1.7)	35.58(1.2)	43.24(3.1)		
	SVM	17.41 (0.7)	20.06 (1.0)	43.46 (1.4)		
	KNN	30.39 (1.8)	29.59 (1.6)	43.29 (1.0)		
	C4.5	28.05(1.5)	26.96 (0.5)	$60.03\ (1.8)$		
CellCycle	Ripper	26.24 (0.9)	$18.61 \ (0.9)$	55.02(2.0)	35.76(1.4)	30.63 (1.3)
	BayesNet	32.23 (0.5)	$26.91\ (1.2)$	42.68(2.3)		
	SVM	32.48 (1.1)	33.25 (2.3)	48.14 (1.7)		
	KNN	21.57 (1.7)	21.36 (0.6)	39.45 (1.3)		
	C4.5	21.42(1.8)	22.45 (0.4)	45.37(1.7)		
Church	Ripper	20.97(0.7)	$13.21 \ (0.3)$	42.96(1.8)	61.49 (0.6	53.70 (2.7)
	BayesNet	38.91(1.3)	21.09(1.5)	36.29(4.7)		
	SVM	21.02(2.3)	21.78(0.2)	43.66(2.4)		
	KNN	23.58 (1.0)	23.49 (0.9)	36.85 (0.8)		
	C4.5	16.54(4.7)	24.97(1.1)	$55.35 \ (1.5)$		
Derisi	Ripper	21.29(2.1)	14.59 (0.4)	47.55(1.1)	33.62(1.1)	30.59 (3.0)
	BayesNet	38.07(0.5)	22.30(2.4)	35.57(3.2)		
	SVM	24.09(0.9)	23.89(1.2)	41.93(1.7)		
	KNN	33.69 (0.9)	31.35 (1.3)	47.22 (2.9)		
	C4.5	30.45 (2.3)	29.12 (1.5)	58.24 (2.9)		
Eisen	Ripper	29.64 (1.9)	22.54 (2.4)	58.63 (0.9)	36.73 (1.7)	33.94 (2.9)
	BayesNet	40.73 (1.1)	28.04 (1.8)	45.24(2.6)		
	SVM	35.20(2.4)	35.42(1.2)	50.02 (3.2)		
	KNN	33.82 (2.2)	31.84 (1.3)	47.34 (1.6)		
	C4.5	32.62 (1.5)	28.88 (0.9)	62.08 (2.8)		
Gasch1	Ripper	28.67 (0.8)	18.59 (0.5)	59.68 (1.8)	38.04 (1.7)	32.53 (1.1)
	BayesNet	46.81 (1.5)	30.40 (1.7)	39.14 (2.3)		
	SVM	34.56(1.6)	33.90(2.0)	49.47(0.9)		
	KNN	29.07 (0.8)	27.58 (0.9)	42.33 (1.8)		
	C4.5	26.60 (3.2)	27.44 (1.5)	59.18 (3.6)		
Gasch2	Ripper	25.15 (3.0)	16.91 (1.4)	54.13 (1.8)	34.50 (0.6)	32.80 (2.1
	BayesNet	37.74 (1.6)	24.28 (1.0)	43.17 (1.5)	, ,	
	SVM	32.34 (0.4)	31.75 (1.6)	44.53 (1.3)		
	KNN	13.75 (1.6)	19.03 (1.4)	35.50 (5.1)		
	C4.5	14.94 (2.2)	20.35 (2.1)	43.65 (5.6)		
Phenotype	Ripper	16.39 (2.5)	14.01 (0.8)	45.23 (1.5)	68.41 (2.8)	49.05 (11.0
V 1	BayesNet	22.63 (0.6)	22.88 (1.2)	32.33 (3.1)	,	
	SVM	18.47 (2.1)	19.96 (1.5)	42.39 (4.8)		
	KNN	31.23 (2.2)	28.87 (1.1)	44.46 (2.6)		
	C4.5	30.58 (0.7)	29.27 (1.4)	60.38 (1.9)		
Sequence	Ripper	26.66 (1.4)	17.83 (0.3)	59.60 (3.2)	34.91 (1.2)	32.70 (2.0)
1	BayesNet	40.20 (2.3)	29.26 (1.5)	42.38 (3.3)	()	- , - (=.0)
	SVM	18.01 (1.9)	23.73 (1.3)	46.28 (1.5)		
	KNN	22.27 (1.4)	22.43 (0.5)	37.30 (1.8)		
	C4.5	25.35 (7.3)	26.04 (1.5)	55.57 (2.5)		
SPO	Ripper	23.33 (7.3) $22.12 (1.3)$	17.25 (1.0)	50.59 (1.9)	34.00 (2.4)	30.84 (1.8)
51 0	BayesNet	35.52 (0.6)	25.48 (1.4)	33.99 (1.5)	04.00 (2.4)	00.04 (1.0)
	SVM	26.27 (1.4)	26.63 (1.5)	40.09 (1.0)		

Tabela B.8: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Revocação Hierárquica no Quarto Nível da Hierarquia

Conjunto de Dados		Abordage	m Top-Down		Abordagem	One-Shot
conjunio de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	27.53(1.3)	27.56(1.1)	41.27 (1.4)		
	C4.5	28.88 (0.3)	27.10(1.5)	$58.88 \; (3.5)$		
Expr	Ripper	27.20(1.9)	18.22(1.1)	58.06 (1.8)	32.68 (1.8)	$31.22 \ (1.7)$
	BayesNet	41.94(1.3)	32.43(1.1)	41.50 (3.0)		
	SVM	15.47 (0.6)	17.98 (1.1)	41.19 (1.2)		
	KNN	28.38 (1.9)	27.35 (1.3)	41.71 (1.4)		
	C4.5	24.96(1.2)	23.92(0.4)	$57.92\ (1.9)$		
CellCycle	Ripper	23.70(1.0)	16.18 (0.9)	52.95(2.1)	32.85 (1.4)	27.28 (1.2
	BayesNet	29.41 (0.4)	24.64(0.9)	40.50(2.4)		
	SVM	29.32(1.2)	30.33(2.3)	45.95(1.8)		
	KNN	19.42 (1.9)	18.95 (0.8)	37.49 (1.2)		
	C4.5	19.60 (1.9)	20.21 (0.7)	43.11 (2.0)		
Church	Ripper	19.03 (0.6)	10.95 (0.3)	40.79 (1.8)	59.13 (0.5)	51.13 (3.1
	BayesNet	35.11 (0.9)	19.00 (1.4)	34.65 (4.6)	, ,	`
	SVM	19.18 (2.3)	19.98 (0.3)	41.37 (3.0)		
	KNN	21.41 (0.9)	21.37 (1.2)	34.77 (0.8)		
	C4.5	14.79 (4.2)	22.19 (1.1)	52.50 (1.5)		
Derisi	Ripper	19.09 (2.2)	12.37 (0.4)	45.26 (1.0)	30.40 (0.9)	27.25 (3.0
	BayesNet	34.07 (0.7)	20.18 (2.9)	33.63 (3.4)	00.40 (0.0)	21.20 (0.0
	SVM	21.80 (0.7)	21.32 (1.2)	39.81 (1.4)		
					<u> </u>	
	KNN	32.13 (0.9)	29.19 (1.5)	45.85 (2.8)		
Eisen	C4.5	27.42 (2.3)	25.97 (1.4)	56.62 (2.6)	0.4.11 (1.0)	20 00 (2.0
	Ripper	27.09 (1.8)	20.19 (3.0)	56.75 (0.8)	34.11 (1.8)	30.66 (2.8
	BayesNet	37.60 (0.9)	25.73 (1.8)	43.62 (2.5)		
	SVM	32.66 (2.1)	32.75 (1.4)	48.12 (3.2)		
	KNN	31.34(2.2)	29.30(1.4)	45.62(1.8)		
	C4.5	29.36 (1.6)	25.58 (0.8)	$60.11\ (2.9)$		
Gasch1	Ripper	25.90(1.0)	15.87 (0.5)	57.46 (1.9)	35.13 (1.5)	28.95 (0.8)
	BayesNet	42.43(1.0)	27.32(1.6)	37.17(2.7)		
	SVM	31.64 (1.6)	31.30 (1.9)	47.12 (0.6)		
	KNN	26.69 (0.7)	25.21 (0.9)	40.38 (1.6)		
	C4.5	23.79(2.9)	24.28(1.4)	$56.72 \ (3.8)$		
Gasch2	Ripper	22.78(3.1)	14.30(1.4)	51.90(1.7)	31.23 (0.5)	29.40 (2.3
	BayesNet	33.54(1.5)	21.95(1.0)	41.13(0.9)		
	SVM	29.07(0.6)	$28.61\ (1.9)$	$42.21\ (1.5)$		
	KNN	11.96 (1.2)	16.79 (1.2)	33.72 (5.3)		
	C4.5	13.33 (2.0)	17.97 (1.8)	41.94 (5.8)		
Phenotype	Ripper	14.62 (2.4)	11.83 (0.7)	44.21 (1.8)	64.08 (2.8)	44.77 (11.0
	BayesNet	19.18 (0.5)	19.99 (1.2)	29.26 (3.3)	, ,	,
	SVM	16.59 (2.0)	17.66 (1.5)	40.53 (4.8)		
	KNN	28.97 (1.9)	26.53 (1.1)	42.86 (2.4)		
	C4.5	27.60 (0.9)	26.19 (1.4)	58.53 (1.5)		
Sequence	Ripper	24.42 (1.3)	15.37 (0.2)	57.65 (3.2)	31.84 (1.3)	29.51 (2.0
	BayesNet	36.41 (1.8)	26.19 (1.3)	40.32 (3.4)	31.07 (1.0)	20.01 (2.0
	SVM	16.50 (1.8)	20.19 (1.3) $21.77 (1.2)$	43.55 (1.6)		
					<u> </u>	
	KNN	20.27 (1.3)	20.46 (0.4)	35.61 (1.8)		
GDO.	C4.5	22.93 (7.1)	23.24 (1.4)	53.63 (2.3)	20.00 (0.7)	OF 00 /= -
SPO	Ripper	19.51 (1.3)	14.76 (0.9)	48.63 (2.2)	30.90 (2.5)	27.60 (1.9)
	BayesNet	32.03 (0.4)	23.23 (1.4)	32.51 (1.4)		
	SVM	23.49(1.2)	23.82(1.5)	37.88(1.1)		

Tabela B.9: Comparação das Abordagens Top-Down e One-Shot Utilizando a Medida

Micro Precisão Hierárquica no Primeiro Nível da Hierarquia

Conjunto de Dados			n Top-Down		-	n One-Shot
conjunio de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	$57.81\ (1.3)$	57.38(3.0)	54.18 (1.0)		
	C4.5	53.78(2.4)	52.99(2.0)	49.82(1.7)		
Expr	Ripper	60.10(3.2)	57.50 (1.9)	51.67 (1.7)	44.88 (1.7)	48.13 (1.9
	BayesNet	56.78(1.9)	60.06 (2.8)	60.26 (2.4)		
	SVM	64.11 (4.1)	61.72 (1.9)	48.90 (1.5)		
	KNN	58.43(2.2)	57.69(2.2)	55.16(2.1)		
	C4.5	52.42(3.2)	49.66 (1.5)	47.13(1.1)		
CellCycle	Ripper	57.38(1.7)	55.50(2.7)	48.93(1.9)	41.43 (1.7)	43.35 (1.0
	BayesNet	58.28(3.4)	59.23(1.7)	56.97(1.7)		
	SVM	59.39 (2.7)	60.72 (2.8)	57.76 (1.8)		
	KNN	51.22 (1.2)	49.16 (0.7)	49.04 (1.7)		
	C4.5	52.98(2.5)	48.89(0.8)	47.24(1.8)		
Church	Ripper	55.27(1.1)	48.94 (0.6)	46.40 (2.0)	34.16 (1.1)	35.46 (1.4
	BayesNet	48.98 (1.1)	52.81 (1.0)	49.23(2.7)		
	SVM	53.53(1.5)	52.48(1.3)	49.15(1.6)		
	KNN	51.81 (1.7)	49.88 (1.7)	49.20 (1.3)		
Derisi	C4.5	54.94 (4.7)	46.36 (1.1)	45.32 (1.3)		
	Ripper	55.21 (1.6)	50.40 (0.5)	48.00 (1.5)	39.18 (1.1)	41.77 (1.8
	BayesNet	49.95 (1.5)	51.12 (4.8)	49.86 (1.7)	, ,	`
	SVM	52.53 (2.3)	50.12 (1.5)	49.31 (1.5)		
	KNN	58.19 (3.7)	57.41 (1.5)	56.12 (3.0)		
	C4.5	51.57 (1.8)	52.09 (1.1)	48.58 (3.1)		
Eisen	Ripper	58.55 (3.0)	57.90 (1.9)	52.12 (3.2)	42.37 (1.8)	45.86 (1.4
	BayesNet	58.28 (3.1)	59.42 (2.5)	58.99 (1.5)		20100 (211
	SVM	59.80 (2.8)	62.42 (1.9)	57.61 (3.2)		
	KNN	57.90 (2.6)	57.50 (1.4)	55.17 (1.7)		
	C4.5	54.66 (2.4)	52.87 (1.2)	48.71 (1.0)		
Gasch1	Ripper	58.94 (4.8)	53.67 (1.2)	51.18 (2.2)	43.03 (1.7)	45.72 (1.9
Gasciii	BayesNet	54.74 (1.0)	55.76 (2.5)	55.05 (1.0)	45.05 (1.7)	40.72 (1.3
	SVM	64.32 (2.9)	65.99 (1.9)	60.22 (2.8)		
				` '		
	KNN	56.29 (1.7)	53.59 (1.3)	53.56 (2.3)		
Gasch2	C4.5	53.39 (1.7) 57.10 (2.7)	49.79 (1.7)	46.48 (1.2) 48.03 (1.5)	41.00 (1.7)	49.CC (0.C
Gasciiz	Ripper	` ′	54.10 (2.0)	48.03 (1.3) 52.29 (1.7)	41.00 (1.7)	42.66 (0.9
	BayesNet SVM	53.42 (1.2) 55.92 (1.3)	53.11 (1.7) 57.74 (0.7)	53.28 (0.9)		
				(/	<u> </u>	
	KNN	54.31 (2.6)	50.25 (0.8)	47.43 (0.6)		
TOI.	C4.5	53.81 (2.6)	52.82 (3.2)	46.89 (2.6)	99.00 (0.7)	04.00 (1.6
Phenotype	Ripper	55.32 (1.8)	53.95 (1.3)	46.30 (1.6)	33.88 (0.7)	34.93 (1.8
	BayesNet	44.60 (2.4)	41.37 (2.2)	41.94 (2.0)		
	SVM	56.48 (1.8)	52.38 (2.6)	47.22 (1.9)		
	KNN	51.80 (2.3)	51.08 (1.8)	48.78 (1.6)		
~	C4.5	53.25 (0.6)	51.20 (2.3)	49.18 (1.5)		
Sequence	Ripper	59.10 (2.3)	53.10 (1.7)	49.04 (0.8)	42.39 (1.6)	45.09 (2.3
	BayesNet	57.51 (1.5)	54.02 (2.1)	57.57 (1.0)		
	SVM	76.98 (3.0)	62.63 (2.9)	51.22 (1.1)		
	KNN	50.46(2.5)	49.00 (0.9)	49.32 (1.9)		
	C4.5	51.62(2.0)	47.93(1.9)	45.14 (1.1)		
SPO	Ripper	$57.24\ (2.5)$	54.15(2.1)	50.16 (1.5)	38.82 (0.8)	41.32 (2.0
	BayesNet	49.87(2.6)	$50.81\ (2.5)$	50.01(2.4)		
	SVM	52.06(2.2)	51.42 (1.6)	50.49 (2.0)		

Tabela B.10: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Micro Precisão Hierárquica no Segundo Nível da Hierarquia

Conjunto de Dados		Abordager	n Top-Down		Abordagen	Abordagem One-Shot	
Conjunto de Dados		HMC-BR	HMC-LP	$\operatorname{HMC-CT}$	C4.5H	Clus-HMC	
	KNN	36.00(1.2)	37.63(1.9)	24.13(1.0)			
	C4.5	29.79(1.3)	32.84 (0.6)	15.25 (0.8)			
Expr	Ripper	37.30(2.6)	33.95(2.8)	17.74(1.5)	22.93(1.1)	26.35 (1.4)	
	BayesNet	28.89(1.5)	36.42(1.3)	26.26 (1.0)			
	SVM	49.06 (3.9)	$35.50 \ (0.8)$	19.57 (1.2)			
	KNN	37.11 (1.2)	37.46 (0.8)	24.01 (1.3)			
	C4.5	30.96(1.8)	32.08(1.2)	$14.30 \ (0.3)$			
CellCycle	Ripper	35.42(2.1)	31.43(1.7)	18.06 (0.9)	21.01 (0.9)	24.57 (0.5)	
	BayesNet	33.97(2.5)	36.32(1.2)	23.35 (0.9)			
	SVM	36.92 (1.4)	39.15 (1.4)	24.83 (1.0)			
	KNN	31.47 (1.1)	34.17 (1.3)	21.19 (0.7)			
	C4.5	31.58(2.1)	34.93(1.4)	17.73(0.9)			
Church	Ripper	33.86(2.5)	24.87(0.4)	18.33(1.2)	$09.41 \ (0.7)$	11.41 (1.3	
	BayesNet	23.81 (0.5)	32.37 (0.8)	20.03(1.3)			
	SVM	32.40 (2.5)	35.47 (0.9)	20.24 (1.0)			
	KNN	34.88 (1.3)	34.88 (2.0)	21.78 (0.5)			
	C4.5	38.68 (3.5)	30.88 (0.5)	14.25 (0.6)			
Derisi	Ripper	35.89(2.0)	$26.50 \ (0.5)$	18.81 (0.4)	19.58 (0.4)	22.98 (0.7	
	BayesNet	24.56(1.1)	30.18(2.2)	21.30(0.7)			
	SVM	33.48(1.4)	35.05(1.3)	21.25 (0.6)			
	KNN	35.39 (1.8)	37.00 (1.0)	23.86 (1.5)			
Eisen	C4.5	29.83 (0.9)	33.32 (0.7)	14.97 (1.5)			
	Ripper	35.07 (2.4)	34.98 (1.7)	17.94 (1.1)	21.88 (1.2)	25.04 (0.3	
	BayesNet	30.60(2.5)	37.34 (1.8)	23.95(0.7)	, ,	·	
	SVM	36.54(2.1)	40.17(1.5)	24.28 (1.2)			
	KNN	36.07 (1.4)	37.12 (0.9)	23.88 (1.3)			
	C4.5	30.95 (0.9)	33.78 (1.3)	14.84 (0.5)			
Gasch1	Ripper	36.29 (2.6)	30.29 (1.0)	17.57 (1.7)	20.98 (0.8)	25.44 (1.0	
	BayesNet	25.79 (0.3)	34.25 (1.0)	23.81 (0.8)	, ,	`	
	SVM	40.65 (1.4)	43.43 (0.9)	26.21 (1.5)			
	KNN	36.53 (0.8)	36.60 (1.4)	24.13 (1.3)			
	C4.5	31.79 (2.5)	32.46 (0.9)	14.41 (0.6)			
Gasch2	Ripper	35.72 (1.8)	29.42 (1.1)	17.89 (0.6)	20.76 (1.2)	23.47 (0.5	
	BayesNet	26.76 (1.5)	31.07 (1.3)	21.66 (0.6)	,	`	
	SVM	33.86 (0.5)	37.88 (0.6)	23.07 (0.2)			
	KNN	43.26 (2.9)	32.71 (1.4)	20.19 (1.5)			
	C4.5	39.79 (2.6)	34.73 (1.8)	18.56 (1.5)			
Phenotype	Ripper	40.05 (1.4)	27.51 (0.9)	15.80 (0.4)	08.72 (0.3)	12.46 (2.5	
~ -	BayesNet	29.44 (2.1)	24.81 (1.9)	17.09 (0.6)	, ,		
	SVM	42.47 (1.4)	34.27 (1.4)	19.68 (1.3)			
	KNN	30.36 (1.0)	32.18 (1.2)	20.49 (0.9)			
	C4.5	30.68 (1.2)	32.06 (1.6)	14.82 (0.8)			
Sequence	Ripper	36.84 (2.1)	28.99 (1.4)	16.52 (0.7)	22.53 (1.0)	25.00 (0.9	
-	BayesNet	31.36 (1.6)	34.08 (2.7)	22.85 (1.0)	, ,	`	
	SVM	58.06 (2.6)	38.40 (1.6)	21.69 (0.6)			
	KNN	33.45 (1.6)	33.95 (0.8)	21.84 (0.8)			
	*****	33.13 (1.0)	` '	` ,			
	C4.5	32.59(2.8)	31.38 (1.2)	13.79 (0.5)			
SPO	C4.5 Ripper	32.59 (2.8) 36.65 (2.3)	31.38 (1.2) 30.00 (1.7)	13.79 (0.5) 19.38 (0.9)	19.56 (0.5)	23.51 (0.9	
SPO	C4.5 Ripper BayesNet	32.59 (2.8) 36.65 (2.3) 25.69 (1.9)	31.38 (1.2) 30.00 (1.7) 31.18 (1.3)	13.79 (0.5) 19.38 (0.9) 22.34 (1.2)	19.56 (0.5)	23.51 (0.9	

Tabela B.11: Comparação das Abordagens Top-Down e One-Shot Utilizando a Medida

Micro Precisão Hierárquica no Terceiro Nível da Hierarquia

Conjunto de Dados			n Top-Down		_	n One-Shot
conjunio de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	$26.91\ (1.1)$	30.30(2.1)	11.57 (0.6)		
	C4.5	21.54(1.1)	$26.34 \ (0.4)$	$06.18 \ (0.4)$		
Expr	Ripper	28.28(2.6)	25.43(2.7)	$07.42 \ (0.7)$	13.96 (0.5)	17.67 (1.1)
	BayesNet	19.57 (1.4)	27.50(1.1)	12.49 (0.5)		
	SVM	39.22 (3.8)	25.03 (0.6)	10.36 (0.4)		
	KNN	27.83(0.6)	29.69 (0.5)	11.82 (0.6)		
	C4.5	22.54(1.4)	26.08(1.2)	05.67 (0.2)		
CellCycle	Ripper	26.51 (2.2)	22.94(1.4)	07.87 (0.5)	$12.48 \ (0.6)$	16.55 (0.3)
	BayesNet	24.78(2.2)	27.36(1.2)	11.32 (0.6)		
	SVM	28.39 (1.4)	32.05 (1.0)	$12.45 \ (0.6)$		
	KNN	23.24 (1.0)	27.05 (1.4)	10.16 (0.4)		
	C4.5	22.76(1.9)	28.20 (1.3)	$07.91\ (0.3)$		
Church	Ripper	24.45(2.3)	16.98(0.4)	09.18 (0.8)	03.83 (0.3)	04.97 (0.8
	BayesNet	$16.31\ (0.6)$	23.18(0.6)	09.35 (0.9)		
	SVM	23.06(2.6)	27.59(1.1)	09.72 (0.4)		
	KNN	26.54 (0.8)	28.98 (2.1)	10.39 (0.3)		
Derisi	C4.5	29.87 (2.4)	25.52 (0.7)	05.81 (0.2)		
	Ripper	26.96 (2.1)	18.58 (0.6)	08.56 (0.1)	11.91 (0.4)	15.36 (0.5
	BayesNet	16.28 (0.3)	21.85 (2.1)	09.44 (0.5)	, ,	`
	SVM	25.20 (1.1)	29.38 (1.0)	09.97 (0.3)		
	KNN	26.09 (1.4)	28.97 (1.0)	11.36 (0.7)		
	C4.5	21.54 (1.0)	26.35 (1.1)	05.98 (0.5)		
Eisen	Ripper	25.83 (2.4)	26.12 (1.5)	07.45 (0.3)	13.13 (0.8)	16.90 (0.3
	BayesNet	21.35 (1.7)	28.64 (1.1)	11.45 (0.5)	,	,
	SVM	27.55 (2.4)	32.43 (1.5)	12.00 (0.5)		
	KNN	27.08 (1.2)	29.76 (1.0)	11.54 (0.7)		
	C4.5	22.36 (0.9)	27.63 (1.2)	06.06 (0.2)		
Gasch1	Ripper	26.79 (2.5)	22.43 (0.9)	07.47 (0.8)	12.22 (0.5)	17.18 (1.0
	BayesNet	16.20 (0.6)	25.17 (0.8)	10.59 (0.5)	(010)	(
	SVM	31.48 (0.6)	34.91 (0.8)	13.68 (0.8)		
	KNN	27.99 (1.3)	30.05 (1.4)	11.51 (0.8)		
	C4.5	22.71 (2.0)	26.30 (1.1)	05.81 (0.3)		
Gasch2	Ripper	26.64 (2.0)	21.03 (0.8)	07.90 (0.2)	12.40 (0.7)	15.59 (0.4
Gusciiz	BayesNet	18.47 (1.0)	22.92 (1.2)	10.32 (0.2)	12.10 (0.1)	10.00 (0.
	SVM	25.81 (0.7)	31.60 (0.6)	11.20 (0.1)		
	KNN	35.42 (4.2)	25.13 (1.2)	09.95 (0.9)		
	C4.5	31.74 (2.9)	26.09 (1.3)	09.95 (0.9) $08.66 (0.8)$		
Phenotype	Ripper	30.73 (2.2)	18.52 (0.6)	07.20 (0.3)	03.66 (0.1)	06.95 (2.0
1 nonety pe	BayesNet	21.65 (1.1)	19.41 (1.3)	09.28 (1.0)	00.00 (0.1)	00.00 (2.0
	SVM	33.92 (1.5)	26.19 (1.4)	09.37 (0.8)		
	KNN	21.57 (0.7)				
	C4.5	21.97 (0.7) $21.98 (0.9)$	23.92 (0.8) 25.79 (1.3)	08.94 (0.4) 05.74 (0.3)		
Sequence	Ripper	21.98 (0.9) $27.67 (1.9)$	25.79 (1.3) 21.62 (1.2)	06.81 (0.4)	14.28 (0.6)	16.78 (0.5
bequence	BayesNet	21.76 (1.9) $21.76 (1.4)$	28.11 (2.4)	10.78 (0.4)	14.20 (0.0)	10.10 (0.6
	SVM	46.05 (2.8)	27.55 (1.2)	10.78 (0.7) $11.72 (0.4)$		
	KNN	25.33 (1.1)	27.81 (1.0)	10.24 (0.4)		
CDO	C4.5	23.77(2.5)	25.75 (0.6)	05.55 (0.2)	11.60 (0.4)	15.00 (0.3
SPO	Ripper	27.36 (2.1)	21.15 (1.3)	08.73 (0.5)	11.69 (0.4)	15.62 (0.6
	BayesNet	17.34 (0.7)	23.93 (0.9)	10.87 (0.8)		
	SVM	25.14(0.9)	28.68 (1.4)	10.45 (0.4)		

Tabela B.12: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Micro Precisão Hierárquica no Quarto Nível da Hierarquia

Conjunto de Dados		Abordage	m Top-Down		Abordagen	n One-Shot
Conjunto de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	21.69(1.1)	25.32(1.6)	$07.38 \ (0.3)$		
	C4.5	17.94(1.0)	22.65 (0.6)	$03.84 \ (0.2)$		
Expr	Ripper	23.72(2.7)	20.25 (2.4)	$04.42 \ (0.4)$	10.35 (0.2)	13.77(0.9)
	BayesNet	15.66 (1.1)	23.06(1.1)	$08.21 \ (0.3)$		
	SVM	33.32 (3.8)	19.24 (0.5)	06.31 (0.5)		
	KNN	21.90 (0.5)	24.26 (0.6)	07.40 (0.4)		
	C4.5	18.54(1.3)	22.80(1.5)	03.48(0.1)		
CellCycle	Ripper	21.68(1.7)	18.22(1.2)	04.55 (0.3)	09.17 (0.4)	12.92 (0.07
	BayesNet	20.32(1.9)	21.95(1.2)	07.09(0.4)		
	SVM	23.56(1.4)	26.96 (1.1)	08.14 (0.4)		
	KNN	18.77 (0.9)	22.81 (1.4)	06.49 (0.3)		
	C4.5	18.22(1.5)	23.25 (1.3)	04.96 (0.2)		
Church	Ripper	19.34(1.4)	12.95(0.3)	05.29(0.6)	02.46 (0.2)	03.26 (0.6)
	BayesNet	13.34 (0.4)	$17.71\ (0.9)$	05.96 (0.7)		
	SVM	18.29(2.1)	21.65(1.0)	06.17(0.3)		
	KNN	21.55 (0.7)	24.19 (2.0)	06.37 (0.2)		
	C4.5	24.43 (1.8)	22.25 (0.6)	03.58 (0.1)		
Derisi	Ripper	21.94 (2.0)	14.32 (0.5)	04.90 (0.2)	08.89 (0.3)	12.07 (0.6
	BayesNet	13.49 (0.3)	17.31 (1.9)	05.97 (0.4)	, ,	•
	SVM	20.50 (0.9)	25.24 (1.2)	06.37 (0.2)		
	KNN	19.64 (0.8)	23.35 (1.0)	07.00 (0.4)		
	C4.5	17.71 (1.2)	22.85 (0.7)	03.67 (0.3)		
Eisen	Ripper	20.81 (1.8)	21.33 (1.6)	04.41 (0.2)	09.64 (0.6)	13.19 (0.2
	BayesNet	16.95 (1.5)	23.57 (1.5)	06.85 (0.3)	00.01 (0.0)	10.10 (0.2)
	SVM	22.62 (2.2)	27.63 (1.7)	07.76 (0.3)		
	KNN	21.92 (0.9)	24.96 (1.0)	07.38 (0.5)		
	C4.5	18.27 (0.8)	23.98 (1.2)	03.78 (0.1)		
Gasch1	Ripper	22.07 (2.2)	17.77 (0.9)	04.42 (0.5)	08.78 (0.6)	13.52 (0.8
Gubbili	BayesNet	13.04 (0.6)	20.85 (0.7)	07.05 (0.4)	00.70 (0.0)	13.32 (0.8)
	SVM	26.13 (0.8)	28.34 (0.8)	08.85 (0.5)		
	KNN	22.81 (1.4)	25.15 (1.0)	07.17 (0.6)		
	C4.5	18.51 (1.8)	23.36 (0.9)	03.61 (0.2)		
Gasch2	Ripper	21.72 (2.1)	16.75 (0.7)	04.68 (0.1)	09.05 (0.6)	12.31 (0.5
Gasciiz	BayesNet	15.31 (0.7)	18.87 (0.9)	06.64 (0.5)	03.03 (0.0)	12.01 (0.0
	SVM	21.71 (0.7)	27.28 (0.7)	07.25 (0.08)		
	KNN	31.41 (5.1)	20.12 (0.9)	06.11 (0.5)		
	C4.5	27.32 (2.7)	20.12 (0.9) $20.23 (1.2)$	$05.11 \ (0.3)$ $05.19 \ (0.3)$		
Phenotype	Ripper	25.87 (1.9)	13.95 (0.5)	04.09 (0.3)	02.42 (0.07)	04.97 (1.6
1 henotype	BayesNet	$19.51 \ (1.3)$	15.95 (0.5) $15.26 (1.0)$	04.09 (0.3) 06.57 (0.8)	02.42 (0.07)	04.97 (1.0
	SVM	28.80 (1.3)	20.83 (0.8)	05.79 (0.8)		
	KNN C4.5	17.22 (0.8) 17.98 (0.5)	19.74 (0.7) 22.83 (1.3)	05.48 (0.2) 03.50 (0.1)		
Sequence	Ripper	22.47 (1.7)	17.54 (0.8)	03.50 (0.1) $04.14 (0.2)$	10.70 (0.6)	12.95 (0.3
bequence	BayesNet	18.65 (1.3)	24.60 (2.0)	04.14 (0.2) $06.96 (0.5)$	10.70 (0.0)	14.30 (0.3
	SVM	18.00 (1.3) $41.57 (3.0)$	24.60 (2.0) $20.97 (0.9)$	06.96 (0.5) $07.43 (0.2)$		
					<u> </u>	
	KNN	20.26 (1.1)	23.13 (0.8)	06.17 (0.2)		
gp.o	C4.5	19.16 (2.2)	22.45 (0.3)	03.45 (0.1)	00 70 (0.1)	10.60 (0.1)
SPO	Ripper	22.61 (2.1)	16.27 (0.9)	05.06 (0.4)	08.52 (0.4)	12.30 (0.4)
	BayesNet	14.07 (0.7)	19.28 (0.9)	07.33 (0.6)		
	SVM	20.77(1.1)	24.36 (1.2)	06.72 (0.2)		

Tabela B.13: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Micro Revocação Hierárquica no Primeiro Nível da Hierarquia

Conjunto de Dados		Abordage	m Top-Down		Abordagem	One-Shot
Conjunto de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	49.71 (0.8)	48.17 (1.6)	60.20 (1.6)		
	C4.5	53.97(1.2)	53.02(2.5)	75.40(3.3)		
Expr	Ripper	52.99(3.4)	43.93(1.5)	$76.22\ (1.1)$	56.59(1.5)	55.94 (1.6
	BayesNet	68.11(2.3)	60.85 (2.3)	58.79(3.9)		
	SVM	40.54 (1.6)	43.86 (1.6)	66.94 (0.9)		
	KNN	51.94 (2.0)	49.94 (1.7)	61.17 (1.6)		
	C4.5	50.95 (1.9)	50.35(2.0)	$76.50 \ (1.9)$		
CellCycle	Ripper	47.01 (1.8)	41.75(1.8)	74.56(2.2)	55.69 (1.6)	51.67 (2.8
	BayesNet	55.03(1.9)	49.02(2.2)	59.16(1.7)		
	SVM	55.64(1.6)	$56.43\ (2.8)$	66.58(1.8)		
	KNN	39.79 (2.6)	39.72 (0.8)	59.64 (1.3)		
	C4.5	38.24 (2.0)	41.81 (0.9)	64.99 (2.0)		
Church	Ripper	39.70 (2.9)	34.94 (0.7)	66.77 (2.1)	78.63 (1.7)	70.95 (1.4
	BayesNet	63.56 (0.9)	38.43 (1.5)	49.35 (5.1)	, ,	`
	SVM	37.90 (2.2)	40.25 (1.2)	64.70 (2.1)		
	KNN	44.52 (1.7)	43.09 (1.1)	55.98 (1.6)		
	C4.5	33.55 (10.0)	45.87 (1.1)	71.12 (2.4)		
Derisi	Ripper	41.22 (2.3)	37.13 (0.6)	69.44 (2.2)	54.29 (1.1)	51.77 (3.1
	BayesNet	59.60 (1.5)	40.69 (2.7)	49.43 (4.0)	()	(-
	SVM	43.34 (1.1)	43.84 (1.5)	61.57 (2.6)		
	KNN	53.90 (2.4)	51.23 (1.3)	64.16 (2.6)		
	C4.5	53.59 (2.3)	52.17 (1.3)	73.28 (3.9)		
Eisen	Ripper	50.53 (3.8)	44.31 (2.0)	75.49 (2.1)	57.52 (1.7)	56.64 (2.4
	BayesNet	62.76 (1.7)	49.81 (2.8)	60.85 (3.4)	01.02 (1.1)	00.01 (2.
	SVM	56.71 (3.0)	57.53 (1.8)	67.24 (3.4)		
	KNN	53.90 (3.4)	52.06 (2.2)	63.49 (1.5)		
	C4.5	56.15 (1.1)	52.84 (2.0)	76.35 (4.4)		
Gasch1	Ripper	49.39 (2.0)	40.90 (0.9)	76.79 (0.4)	56.67 (2.8)	54.10 (2.1
Gasciii	BayesNet	67.15 (1.8)	54.36 (2.4)	51.50 (2.8)	00.01 (2.0)	04.10 (2.1
	SVM	56.20 (2.1)	56.02 (1.7)	67.08 (2.2)		
	KNN					
	C4.5	50.02 (1.4) 45.96 (5.1)	46.66 (1.3) 49.87 (2.7)	60.41 (2.4) 74.51 (2.9)		
Gasch2	Ripper	44.42 (4.0)	39.43 (1.7)	73.20 (1.0)	54.57 (1.6)	53.41 (2.2
Gasciiz	BayesNet	58.15 (2.8)	43.81 (0.8)	58.72 (1.7)	04.07 (1.0)	00.41 (2.2
	SVM	54.72 (1.6)	54.01 (1.5)	62.68 (0.9)		
	+	32.15 (4.3)				
	KNN	` ′	39.27 (1.6)	55.01 (6.5)		
Phenotype	C4.5 Ripper	32.97 (5.5) 36.86 (4.8)	41.01 (1.8)	63.61 (7.3) 65.14 (1.5)	87.88 (1.7)	74 50 (7 9
r nenotype	BayesNet	49.09 (3.7)	37.94 (0.8) 47.47 (1.7)	56.05 (4.0)	67.66 (1.7)	74.50 (7.2
	SVM	39.02 (4.0)	40.52 (1.6)	64.20 (6.2)		
	KNN	49.80 (3.6)	48.22 (2.3)	59.75 (3.8)		
Cognongo	C4.5	53.27 (2.1)	52.07 (2.9)	74.47 (2.3)	E6 19 (2.0)	E2 00 /2 5
Sequence	Ripper	48.33 (3.7) 65.06 (2.1)	39.83 (1.1) 54.50 (3.1)	76.43 (2.0)	56.18 (2.0)	53.92 (3.3
	BayesNet	65.06 (2.1)	54.50 (3.1) 45.47 (2.1)	56.67 (3.9) 60.75 (1.8)		
	SVM	32.20 (2.4)	45.47 (2.1)	69.75 (1.8)		
	KNN	42.41 (2.3)	41.51 (1.4)	56.01 (2.8)		
an 0	C4.5	44.81 (12.0)	47.66 (1.6)	70.45 (2.2)		
SPO	Ripper	40.49 (3.8)	39.57 (1.6)	71.01 (1.9)	54.34 (2.0)	51.46 (1.8
	BayesNet	55.87 (0.6)	45.28 (1.5)	47.53 (2.0)		
	SVM	48.39(2.3)	48.08(2.7)	59.63(1.5)		

Tabela B.14: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Micro Revocação Hierárquica no Segundo Nível da Hierarquia

Conjunto de Dados		_	m Top-Down		Abordagem	
		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	27.31 (0.9)	27.91 (0.8)	43.05 (1.5)		
_	C4.5	32.21 (0.5)	33.59 (1.1)	63.26 (3.5)		
Expr	Ripper	28.31 (2.4)	22.16 (1.1)	57.00 (3.1)	44.83 (1.8)	41.29 (1.6
	BayesNet	41.98 (1.3)	34.28 (1.3)	37.76 (2.8)		
	SVM	19.32 (0.8)	20.88 (0.6)	38.43 (1.6)		
	KNN	27.75(0.9)	28.65 (0.8)	43.43 (1.8)		
	C4.5	30.58(1.6)	32.98(1.0)	70.07 (3.6)		
CellCycle	Ripper	24.91 (1.6)	20.71 (0.8)	54.98 (1.8)	47.70(1.3)	40.38 (2.4
	BayesNet	29.84 (0.6)	26.06(1.3)	37.98(1.8)		
	SVM	30.89 (0.6)	31.95 (1.1)	45.79 (1.9)		
	KNN	21.66 (1.7)	23.58 (0.5)	45.21 (2.5)		
	C4.5	20.51 (1.5)	25.20(1.1)	55.53(4.5)		
Church	Ripper	20.50(1.1)	16.36 (0.3)	40.28(2.0)	61.59(4.1)	54.14 (2.2
	BayesNet	36.92(1.2)	20.99(1.0)	32.43(4.4)		
	SVM	20.38(1.6)	23.27(0.7)	49.44(2.5)		
	KNN	24.45 (0.8)	26.23 (0.6)	43.34 (1.2)		
Derisi	C4.5	16.97 (5.7)	30.19 (1.0)	66.57 (3.7)		
	Ripper	21.38 (1.5)	17.83 (0.4)	45.35 (1.2)	48.28 (1.0)	41.53 (2.7
	BayesNet	36.63 (1.4)	22.14 (1.1)	32.30 (3.2)		
	SVM	24.71 (1.0)	27.00 (0.9)	50.08 (2.0)		
	KNN	29.34 (2.0)	29.76 (0.8)	45.07 (3.0)		
	C4.5	33.09 (2.4)	32.91 (0.6)	64.64 (6.1)		
Eisen	Ripper	26.72 (2.3)	23.18 (1.6)	54.81 (1.8)	48.32 (1.9)	43.24 (1.6
	BayesNet	36.08 (0.8)	27.36 (1.2)	38.52 (2.2)	(-)	- (
	SVM	31.31 (1.6)	32.08 (1.5)	46.56 (2.0)		
	KNN	29.98 (1.8)	30.17 (1.5)	44.80 (0.7)		
	C4.5	33.96 (1.2)	33.42 (1.4)	64.45 (4.8)		
Gasch1	Ripper	26.36 (1.8)	20.63 (0.1)	58.86 (1.7)	46.56 (2.2)	40.58 (2.4
	BayesNet	41.33 (1.0)	30.73 (1.7)	32.40 (1.4)	,	(
	SVM	29.95 (1.3)	30.51 (0.9)	44.22 (1.9)		
	KNN	27.34 (0.6)	27.06 (0.5)	44.63 (1.4)		
	C4.5	26.04 (4.0)	32.11 (1.8)	68.10 (4.8)		
Gasch2	Ripper	23.12 (2.3)	19.27 (0.9)	52.30 (2.5)	46.91 (1.1)	42.03 (1.5
Gubeliz	BayesNet	34.98 (2.2)	24.02 (0.4)	40.48 (2.1)	10.01 (1.1)	12.00 (1.0
	SVM	32.30 (0.6)	31.81 (0.7)	47.48 (2.1)		
	KNN	15.54 (2.2)	21.71 (1.2)	37.85 (5.9)		
	C4.5	15.76 (3.0)	22.91 (1.4)	49.86 (9.7)		
Phenotype	Ripper	17.99 (3.0)	17.53 (0.5)	39.27 (2.2)	76.03 (2.1)	61.96 (9.9
1 henotype	BayesNet	28.16 (2.2)	26.00 (1.2)	36.23 (3.8)	70.05 (2.1)	01.30 (3.3
	SVM	19.10 (2.2)	23.12 (1.4)	49.67 (7.2)		
	KNN	28.39 (2.4)	28.28 (1.4)	44.21 (4.2)		
Sequence	C4.5 Ripper	31.81 (1.7)	32.97 (1.6)	63.47 (3.3) 60.54 (2.0)	45.24 (2.0)	40.63 (2.1
bequence		25.13 (2.3) 39.01 (2.1)	20.64 (0.9)	60.54 (2.0)	40.24 (2.0)	40.05 (2.1
	BayesNet SVM	39.01 (2.1) 17.66 (1.9)	31.84 (2.3) 22.11 (1.0)	38.53 (3.6) 39.46 (1.3)		
	KNN	23.51 (1.3)	25.39 (0.8)	44.36 (2.4)		
GDO.	C4.5	25.80 (8.5)	31.36 (1.4)	65.65 (1.8)	40.01 (0.1)	10.00 /3
SPO	Ripper	20.57 (1.8)	18.91 (0.7)	48.39 (2.1)	48.21 (3.1)	40.28 (1.4
	BayesNet	34.43 (1.1)	25.30 (1.3)	30.44 (1.9)		
	SVM	28.78 (1.6)	29.61 (1.5)	48.00 (1.4)		

Tabela B.15: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Micro Revocação Hierárquica no Terceiro Nível da Hierarquia

Comiumto J. D. J		Abordage	m Top-Down		Abordagem	One-Shot
Conjunto de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	19.88 (0.8)	20.89 (0.7)	38.62 (1.8)		
	C4.5	25.42 (0.8)	27.25 (0.6)	65.09(4.7)		
Expr	Ripper	20.33(2.1)	15.25(1.1)	52.68(3.5)	44.96 (1.6)	38.44(2.9)
	BayesNet	33.90(1.2)	26.09(1.1)	32.59(2.7)		
	SVM	12.84 (0.5)	14.00 (0.4)	27.07 (1.1)		
	KNN	20.08 (0.5)	21.39 (0.7)	39.38 (1.5)		
	C4.5	24.23(1.5)	26.84(0.9)	71.77 (4.1)		
CellCycle	Ripper	17.78 (1.1)	14.04 (0.6)	49.21 (2.9)	48.19 (1.0)	38.20 (2.7)
	BayesNet	21.85 (0.6)	$18.50 \ (0.9)$	30.02(1.1)		
	SVM	22.86 (0.7)	24.12 (0.8)	41.45 (2.2)		
	KNN	15.29 (1.2)	17.14 (0.6)	40.49 (3.2)		
	C4.5	14.37(1.0)	18.82 (1.1)	51.30 (4.1)		
Church	Ripper	14.04 (1.0)	10.75 (0.3)	29.42 (2.4)	$65.91\ (5.8)$	58.39 (4.9)
	BayesNet	$25.61\ (0.9)$	14.32 (0.8)	24.29(4.0)		
	SVM	14.40 (1.3)	17.17(0.6)	44.77(3.1)		
	KNN	17.65 (0.7)	19.85 (0.4)	41.23 (0.8)		
	C4.5	11.68 (4.1)	24.57(1.3)	67.76 (4.2)		
Derisi	Ripper	14.91 (1.3)	11.84 (0.3)	35.73 (1.7)	49.72(0.9)	40.76 (2.8)
	BayesNet	27.38(1.1)	15.09(0.8)	24.79 (3.0)		
	SVM	18.85 (0.9)	21.38(0.4)	49.53 (1.9)		
	KNN	21.39 (1.8)	22.09 (0.6)	40.42 (2.6)		
	C4.5	26.38 (2.6)	26.36 (0.5)	66.05 (8.0)		
Eisen	Ripper	19.26 (2.3)	16.06 (1.4)	49.59 (3.0)	47.50 (2.6)	39.33 (2.0
	BayesNet	27.05 (1.0)	19.44 (1.2)	31.72 (1.9)	,	
	SVM	23.11 (1.9)	23.78 (1.3)	42.05 (3.1)		
	KNN	22.04 (1.7)	22.42 (1.3)	41.22 (1.4)		
	C4.5	27.31 (0.9)	27.16 (1.4)	67.08 (7.6)		
Gasch1	Ripper	18.98 (1.2)	14.00 (0.2)	55.07 (2.3)	47.10 (2.9)	37.76 (3.4)
	BayesNet	32.79 (1.3)	22.49 (1.1)	26.36 (1.6)	,	,
	SVM	21.46 (1.1)	22.03 (0.8)	37.27 (1.7)		
	KNN	19.97 (0.6)	20.32 (0.4)	40.63 (1.2)		
	C4.5	19.39 (3.5)	26.27 (1.8)	70.05 (5.4)		
Gasch2	Ripper	16.32 (1.8)	13.00 (0.7)	43.87 (3.2)	47.76 (2.0)	40.48 (2.4)
	BayesNet	25.84 (1.6)	17.18 (0.3)	32.91 (2.5)	(-)	
	SVM	25.07 (0.7)	24.55 (0.5)	45.09 (1.7)		
	KNN	10.54 (1.4)	15.25 (1.2)	31.24 (4.7)		
	C4.5	10.70 (2.0)	16.13 (1.0)	42.12 (8.3)		
Phenotype	Ripper	12.35 (2.0)	11.59 (0.3)	29.07 (2.2)	83.98 (4.0)	61.64 (13.0
1 henotype	BayesNet	19.73 (1.4)	17.87 (0.8)	28.35 (3.3)	(2.0)	01.01 (10.0
	SVM	13.21 (1.6)	16.54 (1.0)	44.15 (8.0)		
	KNN	20.82 (2.0)	21.29 (0.9)	40.09 (4.5)		
	C4.5	25.44 (1.6)	26.56 (1.1)	65.45 (3.1)		
Sequence	Ripper	17.79 (1.6)	13.98 (0.5)	56.57 (4.4)	45.25 (2.1)	38.48 (3.1)
Sequence	BayesNet	28.95 (1.8)	23.74 (2.0)	33.56 (3.7)	10.20 (2.1)	30.10 (0.1)
	SVM	12.44 (1.3)	14.76 (0.7)	28.27 (1.2)		
	KNN	16.99 (0.9)	19.08 (0.9)	41.70 (2.4)		
SDO	C4.5	19.22 (7.2)	25.58 (1.9)	67.35 (2.8)	40 FO (9.4)	90 FO /1 0
SPO	Ripper	14.38 (1.2)	12.68 (0.5)	40.27 (2.8)	49.58 (3.4)	38.50 (1.6)
	BayesNet	25.21 (1.2)	18.05 (1.1)	23.76 (1.3)		
	SVM	22.75 (1.6)	23.45 (1.2)	46.96 (1.7)		

Tabela B.16: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Micro Revocação Hierárquica no Quarto Nível da Hierarquia

Conjunto de Dados		Abordage	m Top-Down		Abordagen	n One-Shot
Conjunto de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	$17.42\ (1.0)$	18.04 (0.6)	36.78 (2.2)		
	C4.5	22.27(1.1)	24.06 (0.4)	$65.17 \ (3.5)$		
Expr	Ripper	$17.12\ (1.9)$	12.54 (0.9)	51.23 (3.2)	44.31 (2.1)	36.85 (3.0)
	BayesNet	30.42 (1.2)	22.72 (1.2)	29.96 (2.8)		
	SVM	10.58 (0.3)	11.51 (0.3)	23.33 (0.9)		
	KNN	17.52 (0.4)	18.77(0.8)	37.73 (1.6)		
	C4.5	21.15(1.5)	23.47(0.7)	70.77 (3.5)		
CellCycle	Ripper	15.04 (0.9)	11.47 (0.5)	48.21 (3.1)	48.08(1.4)	36.51 (2.6)
	BayesNet	$18.43 \ (0.5)$	15.62 (0.7)	26.27 (1.2)		
	SVM	19.78 (0.7)	21.07 (0.9)	39.65 (2.1)		
	KNN	12.84 (1.1)	14.54 (0.6)	38.89 (3.2)		
	C4.5	12.05 (0.9)	15.82(1.1)	50.09 (4.4)		
Church	Ripper	$11.60 \ (0.8)$	$08.62 \ (0.2)$	25.79(2.7)	69.94 (6.2)	60.39(5.0)
	BayesNet	20.68 (0.6)	11.74 (0.7)	21.16 (3.7)		
	SVM	12.05 (1.1)	14.54 (0.5)	43.25 (3.5)		
	KNN	15.30 (0.6)	17.18 (0.5)	40.44 (1.0)		
Derisi	C4.5	09.67(3.5)	21.66(1.3)	$66.52 \ (3.3)$		
	Ripper	12.46(1.3)	09.60 (0.2)	33.24 (2.4)	49.09(1.4)	39.81 (3.2)
	BayesNet	21.78 (0.9)	12.29(0.6)	21.16 (2.7)		
	SVM	$16.64 \ (0.7)$	18.83 (0.3)	48.80 (1.8)		
Eisen	KNN	18.39 (1.6)	19.30 (0.4)	37.99 (2.9)		
	C4.5	23.02 (3.0)	22.66 (0.3)	65.26 (7.4)		
	Ripper	16.35 (2.3)	13.24 (1.3)	48.71 (3.9)	46.29(2.8)	36.61 (2.6)
	BayesNet	23.12(1.4)	16.23(1.0)	28.86 (1.8)		
	SVM	19.81 (1.8)	20.59(1.1)	39.61 (3.9)		
	KNN	19.44 (1.6)	19.59 (1.1)	39.73 (1.6)		
	C4.5	24.16 (1.1)	23.60 (1.3)	67.06 (7.6)		
Gasch1	Ripper	16.06 (1.2)	11.41 (0.2)	53.68 (2.3)	46.73(2.9)	36.00 (4.5)
	BayesNet	27.72 (1.0)	18.69 (0.9)	23.25 (1.6)		
	SVM	18.44 (0.9)	18.78(0.7)	35.08 (1.8)		
	KNN	17.53 (0.5)	17.55 (0.3)	39.43 (1.4)		
	C4.5	16.58 (3.3)	22.84 (1.5)	70.06 (5.5)		
Gasch2	Ripper	13.92 (1.7)	10.54 (0.6)	41.46 (3.1)	47.56 (1.9)	38.68 (2.8)
	BayesNet	21.33 (1.1)	14.18 (0.3)	29.06 (2.6)	, ,	, ,
	SVM	21.82 (0.9)	21.47 (0.3)	43.23 (2.3)		
	KNN	08.61 (1.2)	12.79 (1.1)	28.50 (4.3)		
	C4.5	08.78 (1.6)	13.45 (0.8)	38.22 (7.6)		
Phenotype	Ripper	10.20 (1.6)	09.39 (0.3)	25.46 (2.2)	84.17(2.3)	58.45 (13.00
	BayesNet	15.84 (1.1)	14.55 (0.7)	24.42 (3.3)		·
	SVM	11.04 (1.4)	13.78 (0.8)	41.71 (7.7)		
	KNN	18.13 (1.8)	18.13 (0.8)	38.02 (4.3)		
	C4.5	22.57 (1.7)	22.98 (0.8)	65.19 (2.7)		
Sequence	Ripper	15.23 (1.5)	11.34 (0.4)	55.33 (4.8)	44.70 (2.1)	37.01 (4.1)
-	BayesNet	24.93 (1.6)	20.27 (1.9)	30.96 (3.4)	` '	, ,
	SVM	10.74 (0.8)	12.13 (0.6)	24.42 (1.4)		
	KNN	14.76 (0.8)	16.56 (0.9)	40.56 (2.0)		
	C4.5	16.45 (6.7)	22.28 (1.7)	66.67 (3.0)		
SPO	Ripper	11.96 (1.0)	10.35 (0.4)	37.93 (3.3)	49.15 (3.9)	36.63 (1.9)
- ·	BayesNet	20.57 (1.1)	15.22 (0.9)	21.03 (1.5)	(0.0)	22.30 (2.0)
	SVM	20.00 (1.4)	20.68 (1.1)	46.17 (1.7)		

Tabela B.17: Comparação das Abordagens Top-Down e One-Shot Utilizando a Medida

Macro Precisão Hierárquica no Primeiro Nível da Hierarquia

Conjunto de Dados			n Top-Down)	n One-Shot
Conjunto de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	61.51 (1.8)	59.63(4.4)	57.16 (1.7)		
	C4.5	51.06(2.3)	50.55 (2.5)	$47.01\ (1.5)$		
Expr	Ripper	58.48(3.3)	$62.15 \ (4.0)$	50.74(1.6)	40.86 (1.8)	45.29 (1.4)
	BayesNet	55.37(2.2)	57.72(2.9)	59.08(2.7)		
	SVM	49.75 (3.0)	48.97 (3.2)	39.74(2.5)		
	KNN	60.91 (4.7)	56.59(3.7)	57.17(4.3)		
	C4.5	49.82(3.9)	45.62(1.4)	44.39(1.3)		
CellCycle	Ripper	56.37(3.7)	61.26 (5.7)	46.92(1.7)	37.71 (1.9)	39.53 (0.7)
	BayesNet	55.49(3.3)	56.64 (0.9)	54.30(1.3)		
	SVM	57.68 (2.4)	59.37 (2.7)	55.73 (1.4)		
	KNN	51.22 (2.8)	48.05 (3.5)	47.11 (3.1)		
	C4.5	50.41(3.2)	46.55(1.9)	43.52(3.2)		
Church	Ripper	53.31 (9.3)	35.80 (8.4)	42.09(10.0)	32.45(1.2)	33.23 (0.7)
	BayesNet	47.50(2.2)	51.50 (0.9)	47.85(2.6)		
	SVM	51.38 (1.8)	53.52 (3.7)	48.18 (3.8)		
	KNN	55.66 (2.8)	49.96 (2.2)	50.85 (1.2)		
Derisi	C4.5	53.56(3.1)	44.14 (1.3)	42.67(2.1)		
	Ripper	$56.57 \ (1.9)$	43.37(5.9)	49.17(1.4)	36.13 (1.1)	38.76 (2.8
	BayesNet	49.70(0.7)	49.05 (4.5)	50.24(1.2)		
	SVM	52.61 (3.1)	48.71 (1.3)	48.34 (1.8)		
	KNN	63.22 (3.6)	61.58 (2.6)	60.57 (2.1)		
Eisen	C4.5	48.83 (1.8)	48.94 (1.4)	45.06(3.1)		
	Ripper	58.04 (4.3)	62.31(2.6)	51.26 (2.6)	39.09 (2.4)	43.35 (1.8
	BayesNet	56.43(3.2)	58.76(3.0)	57.98(2.2)		
	SVM	58.79(2.4)	62.06(3.1)	$55.22\ (2.5)$		
	KNN	60.15 (3.3)	58.52 (3.2)	56.37 (2.3)		
	C4.5	52.29(2.0)	51.12(1.9)	46.39(1.1)		
Gasch1	Ripper	57.54(4.1)	53.51(5.7)	50.36(1.7)	39.32 (1.2)	43.38 (2.6
	BayesNet	54.19(1.2)	55.06(2.6)	55.52 (0.8)		
	SVM	66.19 (3.0)	64.67(3.8)	59.51 (3.1)		
	KNN	57.06 (4.3)	52.75 (2.9)	53.85 (3.7)		
	C4.5	49.92(2.4)	45.87(1.8)	43.48 (1.0)		
Gasch2	Ripper	54.37(3.0)	52.77 (9.5)	47.29(1.2)	36.49 (1.4)	38.64 (1.0
	BayesNet	51.75(1.3)	49.39(2.7)	$50.51\ (1.9)$		
	SVM	53.76 (1.1)	54.77 (2.1)	50.61 (0.8)		
	KNN	40.14 (3.0)	41.14 (3.9)	40.32 (4.7)		
	C4.5	39.03(3.3)	48.26(5.7)	44.95(3.1)		
Phenotype	Ripper	40.98(2.6)	$42.12\ (20.0)$	39.45(1.3)	$32.71\ (0.6)$	33.68 (2.1
	BayesNet	43.40(2.5)	41.87(2.0)	41.30(2.7)		
	SVM	54.67 (13.0)	45.51 (5.0)	44.01 (1.4)		
	KNN	53.13 (1.4)	51.32 (2.2)	48.79 (1.7)		
	C4.5	48.85 (1.3)	$47.21\ (2.5)$	45.60 (1.5)		
Sequence	Ripper	53.67(3.2)	54.48 (3.8)	45.71 (1.8)	38.05 (1.9)	40.20 (1.9
	BayesNet	54.05 (1.8)	51.23(1.7)	$53.13 \ (0.5)$		
	SVM	79.99 (5.1)	73.08 (3.7)	66.55 (2.7)		
	KNN	49.49 (1.7)	45.22 (1.2)	46.97 (2.1)		
	C4.5	48.65 (2.5)	43.87 (1.7)	41.93 (1.0)		
SPO	Ripper	54.91 (3.2)	62.10 (7.5)	51.47 (3.2)	34.48 (1.1)	36.76 (1.7
	BayesNet	48.16 (3.3)	49.71 (2.8)	49.87 (4.1)		
	SVM	46.52(2.4)	47.40 (1.7)	46.34 (2.0)		

Tabela B.18: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Macro Precisão Hierárquica no Segundo Nível da Hierarquia

Conjunto de Dados			m Top-Down		1	n One-Shot
Conjunio de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	$13.54 \ (2.6)$	11.42(2.7)	$09.18 \ (0.6)$		
	C4.5	$08.11 \ (0.4)$	07.77 (0.4)	$07.23 \ (0.2)$		
Expr	Ripper	11.62(2.4)	10.33(2.1)	$07.80 \ (0.2)$	06.28 (0.3)	06.97 (0.2)
	BayesNet	09.47 (0.8)	10.52 (1.2)	09.36(1.0)		
	SVM	08.02 (0.6)	$07.53 \ (0.5)$	06.11 (0.4)		
	KNN	10.40 (2.1)	08.70 (0.5)	08.79 (0.6)		
	C4.5	07.77(0.8)	$07.02 \ (0.2)$	$06.83 \ (0.2)$		
CellCycle	Ripper	09.15(1.4)	$09.42 \ (0.9)$	$07.60 \ (0.7)$	05.80 (0.3)	06.08 (0.1
	BayesNet	$08.61 \ (0.4)$	$08.71 \ (0.1)$	08.45 (0.2)		
	SVM	11.22 (1.7)	09.82 (1.2)	09.34 (1.6)		
	KNN	07.88 (0.4)	07.39 (0.5)	07.24 (0.5)		
	C4.5	07.75 (0.5)	$07.16 \ (0.3)$	06.69(0.5)		
Church	Ripper	09.49 (1.6)	05.50(1.3)	06.47(1.6)	04.99 (0.2)	05.13 (0.1
	BayesNet	07.70(0.7)	07.92(0.1)	07.36 (0.4)		
	SVM	08.29(0.8)	$08.23 \ (0.5)$	$07.41 \ (0.6)$		
	KNN	08.56 (0.4)	08.45 (2.0)	07.82 (0.2)		
	C4.5	08.24 (0.4)	06.79 (0.2)	06.56 (0.3)		
Derisi	Ripper	08.70 (0.3)	06.76 (0.8)	07.56 (0.2)	05.56 (0.2)	06.12 (0.7
	BayesNet	07.64 (0.1)	07.54 (0.7)	07.73 (0.2)	, ,	`
	SVM	08.09 (0.4)	07.49 (0.2)	07.43 (0.3)		
	KNN	11.65 (2.5)	10.76 (2.0)	09.70 (1.1)		
Eisen	C4.5	08.05 (1.1)	07.53 (0.2)	06.95 (0.5)		
	Ripper	10.47 (2.7)	10.55 (1.5)	07.94 (0.5)	06.01 (0.4)	06.67 (0.3
	BayesNet	08.82 (0.8)	09.04 (0.4)	09.37 (0.9)	00.01 (0.1)	00.01 (0.0
	SVM	11.70 (2.3)	12.44 (2.9)	09.92 (2.0)		
	KNN	14.66 (3.5)	12.06 (2.5)	09.67 (1.9)		
	C4.5	09.59 (1.3)	08.19 (0.8)	07.16 (0.2)		
Gasch1	Ripper	12.62 (2.0)	09.49 (2.1)	08.13 (0.6)	06.05 (0.2)	06.67 (0.4
Gusciii	BayesNet	08.66 (0.5)	08.66 (0.7)	08.67 (0.2)	00.00 (0.2)	00.01 (0.1
	SVM	18.11 (3.8)	14.94 (3.8)	09.73 (1.1)		
	KNN	11.34 (3.4)	09.07 (1.7)	08.28 (0.5)		
	C4.5	07.93 (0.8)	07.05 (0.2)	06.70 (0.1)		
Gasch2	Ripper	09.52 (1.5)	08.12 (1.5)	07.27 (0.2)	05.61 (0.2)	05.94 (0.1
Gasciiz	BayesNet	07.96 (0.2)	07.60 (0.4)	07.27 (0.2)	05.01 (0.2)	05.54 (0.1
	SVM	09.36 (1.2)	$09.50 \ (0.4)$	07.91 (0.3)		
	KNN	06.81 (1.1)	06.33 (0.6)	06.20 (0.7)		
Dhonotre	C4.5	06.00 (0.5)	07.42 (0.9)	06.91 (0.5)	05.01 (0.0)	OF 10 (0.9
Phenotype	Ripper	06.30 (0.4) 06.67 (0.4)	06.48 (3.1)	06.07 (0.2)	05.21 (0.2)	05.18 (0.3
	BayesNet SVM	08.41 (2.0)	$06.44 \ (0.3)$ $07.00 \ (0.7)$	06.35 (0.4) 06.77 (0.2)		
	KNN	08.17 (0.2)	07.89 (0.3)	07.76 (0.6)		
Common	C4.5	07.51 (0.2)	07.26 (0.4)	$07.03 \ (0.2)$	05 05 (0.3)	06 10 (0.9
Sequence	Ripper	09.20 (2.2)	08.38 (0.6)	07.03 (0.3)	05.85 (0.3)	06.18 (0.3
	BayesNet	08.51 (0.2)	07.88 (0.2)	08.17 (0.08)		
	SVM	28.28 (8.6)	16.93 (5.4)	10.24 (0.4)		
	KNN	07.61 (0.2)	06.95 (0.2)	07.22 (0.3)		
	C4.5	07.48 (0.4)	06.75 (0.2)	06.45 (0.1)		
SPO	Ripper	09.21 (1.5)	09.55 (1.2)	07.92 (0.5)	05.30 (0.1)	05.65 (0.2)
	BayesNet	07.40 (0.5)	07.65 (0.4)	07.67 (0.6)		
	SVM	07.15(0.3)	07.29(0.2)	07.13(0.3)		

Tabela B.19: Comparação das Abordagens Top-Down e One-Shot Utilizando a Medida

Macro Precisão Hierárquica no Terceiro Nível da Hierarquia

Conjunto de Dados			n Top-Down		1	n One-Shot
Jonjunto de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	$04.17\ (0.8)$	03.81(1.7)	02.42(0.1)		
	C4.5	02.61 (0.9)	$02.31 \ (0.5)$	$01.96 \ (0.06)$		
Expr	Ripper	03.76 (0.9)	02.79 (0.5)	$02.11 \ (0.07)$	$01.70 \ (0.07)$	01.88 (0.06
	BayesNet	$03.08 \ (0.8)$	02.92 (0.4)	$02.54 \ (0.30)$		
	SVM	02.15 (0.2)	02.04 (0.1)	01.65 (0.10)		
	KNN	03.23 (0.7)	02.56 (0.5)	02.38 (0.20)		
CellCycle	C4.5	02.31 (0.6)	$02.10 \ (0.4)$	$01.85 \ (0.05)$		
	Ripper	02.73(0.6)	02.58(0.2)	$02.16 \ (0.30)$	01.62 (0.09)	01.75 (0.2
	BayesNet	$02.41 \ (0.2)$	$02.36 \ (0.03)$	$02.27 \ (0.04)$		
	SVM	$03.61\ (1.0)$	03.14(1.0)	02.77(0.70)		
	KNN	02.13 (0.1)	02.00 (0.1)	01.96 (0.1)		
	C4.5	02.10(0.1)	02.15(0.4)	01.81 (0.1)		
Church	Ripper	02.60 (0.4)	01.49 (0.3)	01.75(0.4)	01.35 (0.05)	01.54 (0.3
	BayesNet	02.06 (0.2)	02.14 (0.04)	01.99(0.1)		
	SVM	$02.14\ (0.07)$	02.23(0.1)	02.00(0.1)		
	KNN	02.32 (0.1)	02.29 (0.5)	02.12 (0.05)		
	C4.5	02.33 (0.1)	01.86 (0.1)	01.78 (0.08)		
Derisi	Ripper	02.35 (0.08)	01.83 (0.2)	02.05 (0.06)	01.50 (0.04)	01.61 (0.1
	BayesNet	02.07 (0.03)	02.04 (0.2)	02.09 (0.05)	,	`
	SVM	02.24 (0.2)	02.03 (0.05)	02.01 (0.07)		
	KNN	03.33 (0.6)	03.18 (0.6)	02.75 (0.3)		
Eisen	C4.5	02.56 (0.8)	02.18 (0.1)	01.96 (0.1)		
	Ripper	02.80 (0.6)	03.21 (0.3)	02.25 (0.1)	01.69 (0.1)	01.88 (0.08
	BayesNet	02.57 (0.4)	02.77 (0.5)	02.86 (0.5)	, ,	
	SVM	04.13 (0.8)	04.28 (1.0)	02.75 (0.5)		
	KNN	04.44 (0.9)	03.60 (0.6)	03.00 (0.9)		
	C4.5	02.81 (0.4)	02.44 (0.4)	01.93 (0.05)		
Gasch1	Ripper	03.77 (1.1)	02.57 (0.5)	02.17 (0.2)	01.69 (0.09)	01.80 (0.1
	BayesNet	02.32 (0.08)	02.39 (0.2)	02.31 (0.03)	02100 (0100)	02.00 (0.2
	SVM	06.32 (1.1)	04.52 (1.6)	03.24 (0.5)		
	KNN	03.69 (1.6)	02.79 (0.7)	02.24 (0.1)		
	C4.5	02.12 (0.2)	02.16 (0.4)	01.81 (0.04)		
Gasch2	Ripper	02.64 (0.4)	02.30 (0.3)	01.97 (0.05)	01.52 (0.06)	01.61 (0.04
	BayesNet	02.17 (0.05)	02.07 (0.1)	02.10 (0.08)	(- 1-)	
	SVM	02.84 (0.5)	03.32 (1.1)	02.46 (0.5)		
	KNN	02.48 (1.1)	01.79 (0.1)	01.97 (0.6)		
	C4.5	01.69 (0.1)	02.09 (0.2)	01.95 (0.1)		
Phenotype	Ripper	01.99 (0.5)	01.83 (0.9)	01.71 (0.05)	01.42 (0.03)	01.46 (0.09
Phenotype			()	- ()	()	
r nonety pe	BayesNet		01.82 (0.08)	01.79(0.1)		
r Herioty pe	BayesNet SVM	01.88 (0.1) 02.59 (0.8)	01.82 (0.08) 01.98 (0.2)	01.79 (0.1) 01.91 (0.06)		
	SVM	01.88 (0.1) 02.59 (0.8)	01.98 (0.2)	01.91 (0.06)		
	SVM KNN	01.88 (0.1) 02.59 (0.8) 02.63 (0.6)	01.98 (0.2) 02.14 (0.09)	01.91 (0.06) 02.24 (0.5)		
	KNN C4.5	01.88 (0.1) 02.59 (0.8) 02.63 (0.6) 02.03 (0.05)	01.98 (0.2) 02.14 (0.09) 01.96 (0.1)	01.91 (0.06) 02.24 (0.5) 01.90 (0.06)	01.58 (0.08)	01.88 (0.4
	SVM KNN C4.5 Ripper	01.88 (0.1) 02.59 (0.8) 02.63 (0.6) 02.03 (0.05) 02.49 (0.6)	01.98 (0.2) 02.14 (0.09) 01.96 (0.1) 02.27 (0.1)	01.91 (0.06) 02.24 (0.5) 01.90 (0.06) 01.90 (0.07)	01.58 (0.08)	01.88 (0.4
	SVM KNN C4.5 Ripper BayesNet	01.88 (0.1) 02.59 (0.8) 02.63 (0.6) 02.03 (0.05) 02.49 (0.6) 02.25 (0.07)	01.98 (0.2) 02.14 (0.09) 01.96 (0.1) 02.27 (0.1) 02.13 (0.07)	01.91 (0.06) 02.24 (0.5) 01.90 (0.06) 01.90 (0.07) 02.21 (0.02)	01.58 (0.08)	01.88 (0.4
Sequence	SVM KNN C4.5 Ripper BayesNet SVM	01.88 (0.1) 02.59 (0.8) 02.63 (0.6) 02.03 (0.05) 02.49 (0.6) 02.25 (0.07) 12.18 (3.7)	01.98 (0.2) 02.14 (0.09) 01.96 (0.1) 02.27 (0.1) 02.13 (0.07) 04.58 (1.5)	01.91 (0.06) 02.24 (0.5) 01.90 (0.06) 01.90 (0.07) 02.21 (0.02) 03.39 (0.6)	01.58 (0.08)	01.88 (0.4
	SVM KNN C4.5 Ripper BayesNet SVM KNN	01.88 (0.1) 02.59 (0.8) 02.63 (0.6) 02.03 (0.05) 02.49 (0.6) 02.25 (0.07) 12.18 (3.7) 02.06 (0.07)	01.98 (0.2) 02.14 (0.09) 01.96 (0.1) 02.27 (0.1) 02.13 (0.07) 04.58 (1.5) 01.88 (0.05)	01.91 (0.06) 02.24 (0.5) 01.90 (0.06) 01.90 (0.07) 02.21 (0.02) 03.39 (0.6) 01.95 (0.09)	01.58 (0.08)	01.88 (0.4
Sequence	SVM KNN C4.5 Ripper BayesNet SVM KNN C4.5	01.88 (0.1) 02.59 (0.8) 02.63 (0.6) 02.03 (0.05) 02.49 (0.6) 02.25 (0.07) 12.18 (3.7) 02.06 (0.07) 02.23 (0.3)	01.98 (0.2) 02.14 (0.09) 01.96 (0.1) 02.27 (0.1) 02.13 (0.07) 04.58 (1.5) 01.88 (0.05) 01.83 (0.07)	01.91 (0.06) 02.24 (0.5) 01.90 (0.06) 01.90 (0.07) 02.21 (0.02) 03.39 (0.6) 01.95 (0.09) 01.74 (0.04)		01.88 (0.4
	SVM KNN C4.5 Ripper BayesNet SVM KNN	01.88 (0.1) 02.59 (0.8) 02.63 (0.6) 02.03 (0.05) 02.49 (0.6) 02.25 (0.07) 12.18 (3.7) 02.06 (0.07)	01.98 (0.2) 02.14 (0.09) 01.96 (0.1) 02.27 (0.1) 02.13 (0.07) 04.58 (1.5) 01.88 (0.05)	01.91 (0.06) 02.24 (0.5) 01.90 (0.06) 01.90 (0.07) 02.21 (0.02) 03.39 (0.6) 01.95 (0.09)	01.58 (0.08) 01.43 (0.04)	01.88 (0.4 01.53 (0.07

Tabela B.20: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Macro Precisão Hierárquica no Quarto Nível da Hierarquia

Conjunto de Dados		0	m Top-Down)	n One-Shot
		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	02.73 (0.7)	02.53 (1.1)	01.29 (0.09)		
_	C4.5	01.65 (0.9)	01.34 (0.3)	01.04 (0.03)		
Expr	Ripper	02.35 (0.9)	01.49 (0.3)	01.13 (0.03)	00.90 (0.04)	01.00 (0.03
	BayesNet	02.08 (1.1)	01.69 (0.3)	01.35 (0.1)		
	SVM	01.15 (0.09)	01.20 (0.3)	00.92 (0.1)		
	KNN	$02.54 \ (0.6)$	01.48 (0.5)	01.32 (0.2)		
	C4.5	01.23 (0.3)	01.34 (0.7)	00.98 (0.03)		
CellCycle	Ripper	01.39 (0.2)	01.37 (0.1)	01.09 (0.1)	00.83 (0.04)	00.93 (0.1)
	BayesNet	01.29 (0.1)	01.26 (0.02)	01.21 (0.02)		
	SVM	02.10 (0.8)	01.90 (0.6)	01.62 (0.6)		
	KNN	$01.14\ (0.06)$	$01.07 \ (0.07)$	$01.04 \ (0.07)$		
	C4.5	$01.12 \ (0.07)$	$01.14 \ (0.2)$	$00.96 \ (0.07)$		
Church	Ripper	$01.42\ (0.2)$	00.79 (0.2)	$00.93 \ (0.2)$	00.72 (0.02)	00.82 (0.2)
	BayesNet	$01.10 \ (0.1)$	$01.14 \ (0.02)$	$01.06 \ (0.05)$		
	SVM	$01.40 \ (0.5)$	01.19 (0.08)	$01.07 \ (0.08)$		
	KNN	01.35 (0.2)	01.22(0.3)	$01.13 \ (0.02)$		
Derisi	C4.5	$01.26 \ (0.06)$	00.99 (0.05)	$00.94 \ (0.04)$		
	Ripper	01.29(0.1)	$01.08 \ (0.2)$	$01.09 \ (0.03)$	00.80 (0.02)	00.86 (0.06
	BayesNet	$01.10 \ (0.01)$	01.09(0.1)	$01.11 \ (0.02)$		
	SVM	$01.43\ (0.6)$	$01.08 \ (0.03)$	$01.07 \ (0.04)$		
	KNN	01.78 (0.4)	01.73 (0.3)	01.48 (0.1)		
Eisen	C4.5	01.48(0.4)	$01.18 \; (0.07)$	$01.06 \ (0.07)$		
	Ripper	$01.50 \ (0.3)$	01.76(0.1)	$01.22 \ (0.07)$	00.91 (0.06)	01.02 (0.04
	BayesNet	01.38(0.1)	01.50 (0.3)	01.68 (0.5)		
	SVM	02.90(1.0)	$03.31\ (0.9)$	01.72(0.4)		
	KNN	03.80 (1.7)	02.41 (0.7)	01.71 (0.7)		
	C4.5	01.83(0.4)	01.30(0.2)	$01.03 \ (0.02)$		
Gasch1	Ripper	02.26 (0.8)	01.37(0.3)	01.19(0.2)	00.92 (0.1)	00.96 (0.06
	BayesNet	01.38(0.3)	01.27(0.1)	$01.23 \ (0.02)$		
	SVM	04.90 (0.4)	03.23(1.0)	02.45 (0.9)		
	KNN	02.91 (1.2)	01.76 (0.8)	01.19 (0.08)		
	C4.5	01.15 (0.09)	01.16 (0.2)	00.96 (0.02)		
Gasch2	Ripper	01.50 (0.2)	01.23(0.2)	01.05 (0.02)	00.81 (0.03)	00.94 (0.1
	BayesNet	01.27(0.2)	$01.13 \ (0.08)$	$01.12 \ (0.04)$		
	SVM	01.90 (0.4)	$02.21\ (0.7)$	$01.44 \ (0.5)$		
	KNN	01.49 (0.8)	00.97 (0.09)	01.22 (0.3)		
	C4.5	00.92 (0.08)	01.15 (0.1)	01.07 (0.07)		
Phenotype	Ripper	01.09 (0.3)	01.00 (0.5)	00.93 (0.03)	00.89 (0.2)	01.04 (0.3)
	BayesNet	01.15(0.2)	01.11(0.3)	00.98 (0.06)		
	SVM	$01.89 \ (0.8)$	$01.32 \ (0.2)$	01.05 (0.03)		
	KNN	02.40 (0.4)	01.92 (0.4)	01.30 (0.3)		
	C4.5	01.64 (0.7)	01.38 (0.5)	01.01 (0.03)		
Sequence	Ripper	01.80 (0.4)	01.21 (0.08)	01.01 (0.04)	00.84 (0.04)	01.11 (0.3
	BayesNet	01.65 (0.6)	01.14 (0.03)	01.18 (0.01)		
	SVM	09.03 (2.7)	02.78 (0.8)	02.25 (0.8)		
	KNN	01.10 (0.03)	01.11 (0.2)	01.04 (0.05)		
	C4.5	01.30 (0.4)	01.08 (0.2)	00.93 (0.02)		
SPO	Ripper	01.46 (0.3)	01.48 (0.2)	01.14 (0.07)	00.76 (0.02)	00.92 (0.2
	BayesNet	01.07 (0.07)	01.14 (0.1)	01.10 (0.09)		` '
	SVM	01.14 (0.3)	01.05 (0.03)	01.03 (0.04)		

Tabela B.21: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Macro Revocação Hierárquica no Primeiro Nível da Hierarquia

acro Revocação	Hierarquic			a Hierarquia		
Conjunto de Dados			m Top-Down	TILLE CO	Abordagem	
		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	44.90 (1.0)	44.96 (1.9)	53.91 (1.7)		
D	C4.5	51.76 (1.9)	50.49 (3.4)	72.52 (3.0)	F1 FC (1 F)	EO 77 (1.4
Expr	Ripper	50.50 (3.2)	39.64 (2.1)	71.19 (0.9)	51.76 (1.7)	52.77 (1.4
	BayesNet	68.02 (3.2)	60.47 (2.4)	59.40 (4.4)		
	SVM	31.38 (1.3)	33.81 (1.2)	52.83 (0.5)		
	KNN	45.85 (2.0)	45.22 (1.7)	53.48 (1.9)		
	C4.5	47.89(2.9)	46.18 (2.2)	73.04 (1.7)		
CellCycle	Ripper	43.06 (2.4)	35.72 (2.1)	68.04 (2.8)	51.40 (2.0)	47.57 (2.3)
	BayesNet	54.45 (2.7)	46.87 (2.6)	58.33 (2.2)		
	SVM	51.52 (1.3)	52.83 (3.5)	62.43 (1.8)		
	KNN	35.26(2.8)	35.95(1.6)	50.62 (1.6)		
	C4.5	34.03(2.1)	38.42 (0.7)	57.74 (1.7)		
Church	Ripper	34.38(1.7)	27.19(1.0)	54.97 (1.7)	74.84 (2.8)	66.45 (2.1)
	BayesNet	61.13(1.3)	$36.71\ (1.0)$	48.07 (4.6)		
	SVM	33.93 (1.5)	36.78 (1.4)	56.48 (1.9)		
	KNN	41.00 (1.0)	40.62 (1.3)	50.07 (1.1)		
	C4.5	32.66 (8.0)	43.39(0.7)	67.47 (1.6)		
Derisi	Ripper	38.88(2.2)	30.53(1.6)	62.05 (1.9)	50.39 (1.9)	48.70 (4.2
	BayesNet	58.90 (1.9)	41.84(2.6)	50.18 (3.1)		
	SVM	40.35(1.3)	40.97(1.2)	56.25 (2.8)		
	KNN	50.05 (2.6)	48.81 (1.2)	58.62 (2.3)		
Eisen	C4.5	50.63 (2.8)	48.86 (2.5)	67.98 (3.6)		
	Ripper	47.30 (3.6)	40.40 (2.3)	69.80 (1.9)	53.78 (2.8)	54.81 (2.3
	BayesNet	62.49 (2.9)	50.79 (3.2)	61.24 (3.0)	,	`
	SVM	53.95 (3.2)	54.88 (1.6)	63.81 (3.9)		
	KNN	51.06 (2.3)	50.34 (3.1)	59.15 (1.0)		
	C4.5	54.13 (1.8)	50.61 (2.9)	73.80 (4.5)		
Gasch1	Ripper	46.57 (2.2)	36.43 (1.5)	71.98 (0.7)	52.04 (2.5)	52.88 (2.2
	BayesNet	67.73 (2.2)	55.91 (2.6)	54.41 (2.9)	- (-)	(
	SVM	51.53 (2.4)	51.75 (2.3)	63.25 (2.1)		
	KNN	44.20 (1.5)	43.06 (2.0)	53.48 (2.3)		
	C4.5	41.32 (4.4)	45.80 (2.9)	70.93 (2.2)		
Gasch2	Ripper	39.68 (3.8)	31.98 (1.9)	64.99 (2.5)	48.76 (1.9)	48.65 (1.6
	BayesNet	56.48 (3.3)	42.85 (1.2)	57.05 (2.0)	()	
	SVM	51.60 (1.1)	49.86 (2.1)	59.45 (0.9)		
	KNN	24.77 (3.3)	31.56 (2.3)	45.44 (6.9)		
	C4.5	25.55 (4.2)	34.99 (3.5)	54.85 (7.1)		
Phenotype	Ripper	28.76 (3.8)	26.23 (0.9)	52.65 (1.6)	85.66 (0.9)	72.15 (7.6
	BayesNet	47.94 (4.3)	47.58 (2.2)	55.47 (5.0)	(312)	(
	SVM	31.21 (3.6)	33.10 (2.1)	55.24 (6.7)		
	KNN	43.82 (2.8)	45.08 (2.9)	52.33 (2.7)		
	C4.5	48.61 (2.0)	48.21 (3.4)	69.75 (2.7)		
Sequence	Ripper	42.31 (4.3)	32.69 (1.2)	67.96 (3.5)	50.62 (2.3)	47.99 (2.8
Sequence	BayesNet	62.42 (2.2)	52.09 (1.2) $52.77 (3.7)$	52.58 (4.2)	00.02 (2.0)	±1.00 (2.0
	SVM	26.78 (1.8)	37.21 (2.0)	57.01 (1.6)		
	KNN	36.48 (1.9)	36.61 (1.3)	47.77 (2.2)		
CDO	C4.5	40.77 (9.4)	43.61 (1.9)	66.38 (2.2)	40.40.70.0\	4E EO /1 :
SPO	Ripper	36.36 (3.3)	32.09 (1.9)	61.01 (2.5)	48.48 (2.0)	45.78 (1.4
	BayesNet	54.97 (0.9)	45.33 (2.7)	48.87 (2.4)		
	SVM	42.76(2.5)	43.30 (3.2)	53.44 (1.6)		

Tabela B.22: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Macro Revocação Hierárquica no Segundo Nível da Hierarquia

Conjunto de Dados		_	n Top-Down		Abordagem	
Conjunto de Bados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	08.28(1.1)	08.39(1.6)	08.89 (0.6)		
	C4.5	$08.32 \ (0.5)$	07.76 (0.5)	11.16 (0.4)		
Expr	Ripper	09.62(1.9)	$06.18 \; (0.2)$	10.95 (0.1)	$07.96 \ (0.2)$	08.12 (0.2
	BayesNet	12.56 (2.2)	11.25 (1.1)	09.53 (1.2)		
	SVM	05.11 (0.2)	05.20 (0.2)	08.13 (0.08)		
	KNN	$07.33\ (0.6)$	06.95 (0.2)	08.23 (0.3)		
	C4.5	$07.56 \ (0.6)$	$07.10 \ (0.3)$	$11.24 \ (0.2)$		
CellCycle	Ripper	06.96 (0.4)	05.49(0.3)	$11.24\ (1.5)$	$07.90 \ (0.3)$	07.31 (0.3
	BayesNet	08.46 (0.3)	$07.21\ (0.4)$	09.11 (0.4)		
	SVM	09.97 (1.6)	09.12 (1.9)	10.37 (1.5)		
	KNN	05.42 (0.4)	05.53 (0.2)	07.78 (0.2)		
	C4.5	$05.23 \ (0.3)$	05.91(0.1)	08.88 (0.2)		
Church	Ripper	$06.30 \ (0.8)$	04.18 (0.1)	08.45 (0.2)	$11.51\ (0.4)$	10.29 (0.3
	BayesNet	$09.71 \ (0.5)$	05.65(0.1)	07.39(0.7)		
	SVM	05.37(0.3)	05.66 (0.2)	08.68 (0.3)		
	KNN	06.30 (0.1)	06.34 (0.4)	07.70 (0.1)		
	C4.5	05.02 (1.2)	06.67 (0.1)	10.38 (0.2)		
Derisi	Ripper	05.98 (0.3)	04.72 (0.2)	09.54 (0.3)	07.75(0.3)	08.26 (2.1
	BayesNet	09.06 (0.3)	06.43 (0.4)	07.72 (0.5)	,	`
	SVM	06.20 (0.2)	06.30 (0.2)	08.65 (0.4)		
	KNN	08.40 (0.6)	08.85 (2.0)	09.27 (0.7)		
Eisen	C4.5	08.33 (1.1)	07.51 (0.4)	10.51 (0.6)		
	Ripper	08.17 (2.0)	06.56 (0.7)	11.12 (0.7)	08.27 (0.4)	08.43 (0.3
	BayesNet	09.77 (0.6)	07.81 (0.5)	10.38 (1.8)	00.21 (0.1)	00.10 (0.0
	SVM	09.95 (1.4)	10.39 (2.0)	11.82 (1.6)		
	KNN	10.99 (1.8)	10.16 (1.7)	09.99 (1.7)		
	C4.5	09.56 (0.9)	08.30 (1.1)	11.55 (0.8)		
Gasch1	Ripper	10.17 (2.0)	06.14 (0.7)	11.84 (1.7)	08.00 (0.4)	08.13 (0.3
Gasciii	BayesNet	10.17 (2.0) $11.15 (1.6)$	08.82 (0.5)	08.42 (0.5)	00.00 (0.4)	00.13 (0.6
	SVM	11.06 (1.7)	10.29 (1.6)	10.12 (1.0)		
	KNN	07.34 (0.5)	06.99 (0.9)	08.23 (0.3)		
Gasch2	C4.5	06.42 (0.6)	07.04 (0.4)	11.01 (0.5)	07.50 (0.3)	07.49.70.6
Gasciiz	Ripper BayesNet	06.33 (0.6) $08.69 (0.5)$	04.92 (0.3)	09.99 (0.4)	07.50 (0.5)	07.48 (0.2
	SVM	08.84 (0.9)	06.59 (0.2)	08.78 (0.3) 09.24 (0.2)		
			08.77 (1.5)	` ′		
	KNN	03.95 (0.6)	04.85 (0.3)	06.99 (1.1)		
701	C4.5	03.93 (0.6)	05.38 (0.5)	08.44 (1.1)	o. (o.o)	440/46
Phenotype	Ripper	04.42 (0.6)	04.03 (0.1)	08.10 (0.2)	$15.64 \ (3.2)$	11.10 (1.2
	BayesNet	07.37 (0.6)	07.32 (0.3)	08.53 (0.7)		
	SVM	04.80 (0.5)	05.09 (0.3)	08.50 (1.0)		
	KNN	$06.74 \ (0.4)$	06.93 (0.4)	$08.16 \ (0.5)$		
	C4.5	07.48 (0.3)	07.41 (0.5)	10.81 (0.5)		
Sequence	Ripper	06.78 (1.2)	05.03 (0.2)	10.45 (0.5)	07.79 (0.3)	07.38 (0.4
	BayesNet	09.93 (0.5)	08.12 (0.6)	08.09 (0.6)		
	SVM	06.53 (1.0)	06.18 (0.4)	08.77 (0.2)		
	KNN	05.61 (0.3)	05.63 (0.2)	07.35 (0.3)		-
	C4.5	06.27(1.5)	$06.71 \ (0.3)$	$10.21\ (0.3)$		
SPO	Ripper	05.98(0.9)	04.93 (0.3)	09.38 (0.4)	$07.46 \ (0.3)$	07.04 (0.2
	BayesNet	08.46 (0.1)	06.97(0.4)	07.52 (0.3)		
	SVM	06.58(0.4)	06.66(0.5)	08.22(0.2)		

Tabela B.23: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Macro Revocação Hierárquica no Terceiro Nível da Hierarquia

Conjunto de Dados		_	em Top-Down		Abordagem	One-Shot
Conjunto de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	02.80 (0.7)	02.74 (1.0)	02.31 (0.2)		
	C4.5	02.33(0.1)	$02.31\ (0.5)$	03.02(0.1)		
Expr	Ripper	$03.02 \ (0.6)$	$01.67 \ (0.06)$	$02.96 \ (0.04)$	02.15 (0.07)	02.19 (0.06
	BayesNet	$04.00 \ (0.9)$	$03.14 \ (0.4)$	02.58 (0.3)		
	SVM	01.37 (0.07)	01.40 (0.05)	02.20 (0.02)		
CellCycle	KNN	02.26 (0.4)	02.09 (0.5)	02.23 (0.08)		
	C4.5	02.27(0.6)	$02.13 \ (0.4)$	$03.04\ (0.07)$		
	Ripper	02.17(0.4)	01.49 (0.08)	$03.25 \ (0.5)$	02.35(0.4)	02.19 (0.4)
	BayesNet	02.36 (0.2)	01.95(0.1)	$02.45 \ (0.09)$		
	SVM	02.99(0.7)	02.96(1.2)	03.06 (0.6)		
	KNN	01.47 (0.1)	01.49 (0.06)	02.10 (0.06)		
	C4.5	01.42 (0.08)	01.80 (0.4)	02.40 (0.07)		
Church	Ripper	01.75 (0.2)	01.13 (0.04)	02.29 (0.07)	$03.12\ (0.1)$	02.97 (0.5
	BayesNet	02.59 (0.1)	01.53 (0.04)	02.00 (0.2)	, ,	
	SVM	01.41 (0.06)	01.53 (0.06)	$02.35 \ (0.08)$		
	KNN	01.70 (0.04)	01.72 (0.1)	02.08 (0.04)		
Derisi	C4.5	01.39 (0.4)	01.84 (0.09)	02.81 (0.06)		
	Ripper	01.62 (0.09)	01.28 (0.06)	02.58 (0.08)	02.10 (0.08)	02.03 (0.1
	BayesNet	02.45 (0.08)	01.74 (0.1)	02.09 (0.1)	,	`
	SVM	01.72 (0.1)	01.70 (0.05)	02.34 (0.1)		
	KNN	02.46 (0.2)	02.70 (0.6)	02.84 (0.5)		
Eisen	C4.5	02.49 (0.5)	02.34 (0.5)	02.95 (0.1)		
	Ripper	02.32 (0.5)	02.50 (0.4)	03.32 (0.4)	02.34 (0.1)	02.38 (0.1
	BayesNet	02.91 (0.4)	02.28 (0.2)	03.15 (0.5)	02.01 (0.1)	02.00 (0.1
	SVM	03.25 (0.7)	03.75 (1.3)	03.32 (0.4)		
	KNN	03.78 (1.6)	03.09 (0.4)	03.19 (0.9)		
	C4.5	02.82 (0.5)	02.49 (0.3)	03.07 (0.2)		
Gasch1	Ripper	03.04 (0.6)	$01.70 \ (0.2)$	03.20 (0.5)	02.37 (0.5)	02.20 (0.1
Gasciii	BayesNet	02.95 (0.2)	02.66 (0.2)	02.26 (0.1)	02.51 (0.5)	02.20 (0.1
	SVM	04.30 (1.0)	03.23 (1.0)	03.40 (0.8)		
	KNN		, ,			
	C4.5	02.61 (1.0) $01.73 (0.2)$	02.10 (0.3) 01.99 (0.1)	02.23 (0.09) 02.95 (0.09)		
Gasch2	Ripper	01.73 (0.2)	01.99 (0.1) $01.43 (0.2)$	02.93 (0.09) $02.70 (0.1)$	02.03 (0.07)	02.02 (0.07
Gasciiz	BayesNet	02.40 (0.1)	01.43 (0.2) $01.81 (0.1)$	02.37 (0.08)	02.03 (0.01)	02.02 (0.07
	SVM	02.53 (0.2)	03.13 (0.7)	02.85 (0.6)		
	KNN	01.22 (0.2)	01.37 (0.1)	02.19 (0.4)		
	C4.5	01.22 (0.2) $01.11 (0.2)$	01.57 (0.1) $01.52 (0.1)$	$02.19 \ (0.4)$ $02.38 \ (0.3)$		
Phenotype	Ripper	01.11 (0.2) $01.30 (0.1)$	01.32 (0.1) $01.14 (0.04)$	02.29 (0.07)	03.72 (0.04)	03.13 (0.3
1 henotype	BayesNet	02.08 (0.2)	02.07 (0.09)	02.41 (0.2)	03.72 (0.04)	03.13 (0.3
	SVM	02.08 (0.2) $01.46 (0.3)$	01.44 (0.09)	02.40 (0.3)		
	KNN	01.96 (0.2)	01.88 (0.1)	02.39 (0.3)		
Sequence	C4.5	02.02 (0.08)	02.01 (0.1)	02.90 (0.1)	09 11 (0 1)	00 00 (0 E
bequence	Ripper	01.83 (0.3) 02.60 (0.08)	$01.36 \ (0.05)$	02.83 (0.1)	02.11 (0.1)	02.20 (0.5)
	BayesNet SVM	` ,	02.20 (0.1)	02.19 (0.2)		
		02.69 (0.5)	01.67 (0.1)	03.00 (0.5)		
	KNN	01.52 (0.08)	01.52 (0.05)	01.99 (0.09)		
an o	C4.5	01.80 (0.6)	01.81 (0.08)	02.76 (0.09)	00.00 (5.55)	04.05 (5.00
SPO	Ripper	01.62 (0.2)	01.35 (0.08)	02.54 (0.1)	02.02 (0.08)	01.90 (0.05
	BayesNet	02.29 (0.04)	01.99 (0.2)	02.03 (0.1)		
	SVM	01.78 (0.1)	01.80 (0.1)	02.22 (0.07)		

Tabela B.24: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Macro Revocação Hierárquica no Quarto Nível da Hierarquia

Conjunto de Dados		_	em Top-Down		Abordagem	
Conjunto de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	01.90 (0.6)	$01.86 \ (0.7)$	$01.23 \ (0.09)$		
	C4.5	$01.44 \ (0.4)$	01.34 (0.3)	$01.61 \ (0.06)$		
Expr	Ripper	$01.82 \ (0.6)$	00.89 (0.03)	$01.58 \ (0.02)$	01.15 (0.04)	01.17 (0.03
	BayesNet	$02.65 \ (1.1)$	01.80 (0.4)	01.38 (0.1)		
	SVM	00.73 (0.04)	00.86 (0.2)	01.21 (0.07)		
	KNN	01.69 (0.5)	$01.22 \ (0.5)$	$01.41 \ (0.5)$		
	C4.5	$01.21\ (0.3)$	$01.36 \ (0.7)$	$01.62 \ (0.04)$		
CellCycle	Ripper	$01.14\ (0.2)$	$00.79 \ (0.04)$	01.62 (0.2)	01.14 (0.04)	01.17 (0.2
	BayesNet	01.27 (0.1)	$01.04 \ (0.06)$	$01.30 \ (0.05)$		
	SVM	01.80 (0.6)	01.80 (0.8)	$01.85 \ (0.6)$		
	KNN	00.78 (0.06)	00.79 (0.03)	01.12 (0.03)		
	C4.5	00.75 (0.04)	00.96 (0.2)	$01.28 \ (0.04)$		
Church	Ripper	00.94(0.1)	$00.60 \ (0.02)$	$01.22 \ (0.03)$	01.66 (0.06)	01.58 (0.3
	BayesNet	$01.38 \ (0.08)$	$00.81 \ (0.02)$	01.07(0.1)		
	SVM	$00.80 \ (0.08)$	$00.81 \ (0.03)$	$01.25 \ (0.04)$		
	KNN	00.94 (0.08)	00.91 (0.05)	01.11 (0.02)		
Derisi	C4.5	00.77(0.2)	00.98 (0.05)	$01.49\ (0.03)$		
	Ripper	00.87 (0.03)	00.79 (0.26)	01.38 (0.04)	01.12 (0.04)	01.08 (0.09
	BayesNet	01.30 (0.04)	00.92 (0.06)	01.11 (0.07)	, ,	
	SVM	01.10 (0.4)	00.91 (0.02)	01.25 (0.06)		
	KNN	01.27 (0.08)	01.47 (0.3)	01.42 (0.1)		
Eisen	C4.5	01.56 (0.3)	01.26 (0.2)	01.60 (0.08)		
	Ripper	01.24 (0.3)	01.37 (0.2)	01.79 (0.2)	01.26 (0.06)	01.29 (0.0
	BayesNet	01.58 (0.2)	01.23 (0.1)	01.88 (0.5)		
	SVM	02.12 (0.8)	02.82 (1.2)	02.03 (0.3)		
	KNN	02.88 (1.8)	02.15 (0.4)	01.77 (0.7)		
	C4.5	01.70 (0.4)	01.33 (0.2)	01.64 (0.1)		
Gasch1	Ripper	01.82 (0.6)	00.90 (0.1)	01.82 (0.5)	01.38 (0.5)	01.17 (0.05
Gubelli	BayesNet	01.70 (0.2)	01.42 (0.1)	01.21 (0.06)	01.50 (0.5)	01.11 (0.00
	SVM	03.14 (0.5)	02.29 (0.9)	02.50 (1.0)		
	KNN	01.91 (0.9)	01.33 (0.5)	01.19 (0.05)		
	C4.5	00.94 (0.07)	01.33 (0.3) $01.07 (0.09)$	01.19 (0.05)		
Gasch2	Ripper	01.00 (0.1)	00.76 (0.03)	01.44 (0.05)	01.08 (0.04)	01.30 (0.3
Gasciiz	BayesNet	01.30 (0.1)	01.08 (0.2)	01.27 (0.04)	01.03 (0.04)	01.30 (0.3
	SVM	$01.50 \ (0.1)$ $01.50 \ (0.2)$	02.11 (0.5)	01.27 (0.04) $01.67 (0.4)$		
	KNN	$00.67 (0.1) \\ 00.60 (0.1)$	00.75 (0.05)	01.44 (0.4)		
Dhamatana	C4.5	` ′	00.83 (0.08)	01.30 (0.1)	00.16 (0.0)	01.05 (0.4
Phenotype	Ripper	00.71 (0.07)	00.62 (0.02)	01.25 (0.04)	02.16 (0.2)	01.95 (0.4
	BayesNet SVM	01.26 (0.2) 01.13 (0.4)	01.25 (0.3) 01.02 (0.3)	$01.32 (0.1) \\ 01.31 (0.1)$		
	<u> </u>					
	KNN	01.65 (0.4)	01.65 (0.5)	01.31 (0.2)		
C	C4.5	01.69 (0.9)	01.40 (0.5)	01.55 (0.06)	01.10 (0.05)	01.00 (0.0
Sequence	Ripper	01.38 (0.4)	00.72 (0.02)	01.51 (0.08)	01.12 (0.05)	01.23 (0.2
	BayesNet	01.85 (0.6)	01.17 (0.08)	01.16 (0.09)		
	SVM	02.38 (0.6)	01.22 (0.5)	02.04 (0.8)		
	KNN	$00.81 \ (0.04)$	00.86 (0.1)	$01.06 \ (0.05)$		
	C4.5	$01.07 \ (0.5)$	$01.08 \; (0.2)$	$01.47 \ (0.05)$		
SPO	Ripper	00.87 (0.1)	$00.83 \ (0.2)$	$01.35 \ (0.05)$	01.07 (0.04)	$01.13 \ (0.2$
	BayesNet	$01.22\ (0.02)$	01.06 (0.1)	$01.08 \; (0.05)$		
	SVM	00.98 (0.08)	00.96 (0.07)	01.18 (0.03)		

Tabela B.25: Comparação das Abordagens Top-Down e One-Shot Utilizando a Medida

Micro Precisão Hierárquica Ponderada no Primeiro Nível da Hierarquia

	Abordagem Top-Down		Abordagem One-Shot		
	HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
KNN	57.81 (1.3)	57.38 (3.0)	54.18 (1.0)		
C4.5	53.78(2.4)	52.99(2.0)	49.82(1.7)		
Ripper	60.10(3.2)	57.50(1.9)	51.67(1.7)	44.88 (1.7)	48.13 (1.9
BayesNet	56.78(1.9)	60.06(2.8)	60.26 (2.4)		
SVM	64.11 (4.1)	61.72 (1.9)	48.90 (1.5)		
KNN	58.43 (2.2)	57.69 (2.2)	55.16 (2.1)		
C4.5	52.42(3.2)	49.66(1.5)	47.13(1.1)		
Ripper	57.38 (1.7)	55.50(2.7)	48.93(1.9)	41.43 (1.7)	43.35 (1.0
BayesNet	58.28 (3.4)	59.23(1.7)	56.97 (1.7)		
SVM	59.39(2.7)	$60.72\ (2.8)$	57.76 (1.8)		
KNN	51.22 (1.2)	49.16 (0.7)	49.04 (1.7)		
C4.5	52.98 (2.5)	48.89 (0.8)	47.24 (1.8)		
Ripper	55.27 (1.1)	48.94 (0.6)	46.40 (2.0)	34.16 (1.1)	35.46 (1.4
BayesNet	48.98 (1.1)	52.81 (1.0)	49.23 (2.7)		
SVM	53.53(1.5)	52.48(1.3)	49.15(1.6)		
KNN	51.81 (1.7)	49.88 (1.7)	49.20 (1.3)		
C4.5	54.94 (4.7)	46.36 (1.1)	` '		
Ripper	55.21 (1.6)	50.40 (0.5)	48.00 (1.5)	39.18 (1.1)	41.77 (1.8
BayesNet	49.95 (1.5)	` ′	49.86 (1.7)	, ,	`
SVM	52.53 (2.3)	50.12 (1.5)	49.31 (1.5)		
KNN	58.19 (3.7)	57.41 (1.5)	56.12 (3.0)		
	, ,	` ′	` '		
	` ′	` ′	` '	42.37 (1.8)	45.86 (1.4
	` ,	, ,	, ,	, ,	`
SVM	59.80 (2.8)	62.42 (1.9)	57.61 (3.2)		
KNN	57.90 (2.6)	57.50 (1.4)	55.17 (1.7)		
	` ,	` ′	` ′		
	` ′	` ′		43.03 (1.7)	45.72 (1.9
	` ′	, ,	, ,	-5100 (-11)	
-	` ,	65.99 (1.9)	, ,		
		53 59 (1.3)			
	` ′	, ,	` ′		
	(/			41.00 (1.7)	42.66 (0.9
			` ,		(
SVM	55.92 (1.3)	57.74 (0.7)			
KNN	54.31 (2.6)	50 25 (0.8)			
	, ,	, ,	` ,		
	` ′	` '		33.88 (0.7)	34.93 (1.8
			, ,		
SVM	` ′		47.22 (1.9)		
KNN			48 78 (1.6)		
	` ′	, ,	` ′		
	` ′	, ,	, ,	42.39 (1.6)	45.09 (2.3
		, ,	` '		-5.00 (2.0
-	` ′	, ,	, ,		
	` ′	` '	` ′		
	` ,		, ,	38 83 (0 6)	41 99 (9 t
Ripper	$57.24\ (2.5)$	54.15(2.1)	50.16 (1.5)	38.82 (0.8)	41.32 (2.0
BayesNet	49.87(2.6)	50.81(2.5)	50.01(2.4)		
	C4.5 Ripper BayesNet SVM KNN C4.5 Ripper BayesNet SVM	KNN 57.81 (1.3) C4.5 53.78 (2.4) Ripper 60.10 (3.2) BayesNet 56.78 (1.9) SVM 64.11 (4.1) KNN 58.43 (2.2) C4.5 52.42 (3.2) Ripper 57.38 (1.7) BayesNet 58.28 (3.4) SVM 59.39 (2.7) KNN 51.22 (1.2) C4.5 52.98 (2.5) Ripper 55.27 (1.1) BayesNet 48.98 (1.1) SVM 53.53 (1.5) KNN 51.81 (1.7) C4.5 54.94 (4.7) Ripper 55.21 (1.6) BayesNet 49.95 (1.5) SVM 52.53 (2.3) KNN 58.19 (3.7) C4.5 51.57 (1.8) Ripper 58.55 (3.0) BayesNet 58.28 (3.1) SVM 59.80 (2.8) KNN 57.90 (2.6) C4.5 54.66 (2.4) Ripper 58.94 (4.8) BayesNet 54.74 (1.0)	KNN 57.81 (1.3) 57.38 (3.0) C4.5 53.78 (2.4) 52.99 (2.0) Ripper 60.10 (3.2) 57.50 (1.9) BayesNet 56.78 (1.9) 60.06 (2.8) SVM 64.11 (4.1) 61.72 (1.9) KNN 58.43 (2.2) 57.69 (2.2) C4.5 52.42 (3.2) 49.66 (1.5) Ripper 57.38 (1.7) 55.50 (2.7) BayesNet 58.28 (3.4) 59.23 (1.7) SVM 59.39 (2.7) 60.72 (2.8) KNN 51.22 (1.2) 49.16 (0.7) C4.5 52.98 (2.5) 48.89 (0.8) Ripper 55.27 (1.1) 48.94 (0.6) BayesNet 48.98 (1.1) 52.81 (1.0) SVM 53.53 (1.5) 52.48 (1.3) KNN 51.81 (1.7) 49.88 (1.7) C4.5 54.94 (4.7) 46.36 (1.1) Ripper 55.21 (1.6) 50.40 (0.5) BayesNet 49.95 (1.5) 51.12 (4.8) SVM 52.53 (2.3) 50.12 (1.5) KNN 58.19	KNN 57.81 (1.3) 57.38 (3.0) 54.18 (1.0) C4.5 53.78 (2.4) 52.99 (2.0) 49.82 (1.7) Ripper 60.10 (3.2) 57.50 (1.9) 51.67 (1.7) BayesNet 56.78 (1.9) 60.06 (2.8) 60.26 (2.4) SVM 64.11 (4.1) 61.72 (1.9) 48.90 (1.5) KNN 58.43 (2.2) 57.69 (2.2) 55.16 (2.1) C4.5 52.42 (3.2) 49.66 (1.5) 47.13 (1.1) Ripper 57.38 (1.7) 55.50 (2.7) 48.93 (1.9) BayesNet 58.28 (3.4) 59.23 (1.7) 56.97 (1.7) SVM 59.39 (2.7) 60.72 (2.8) 57.76 (1.8) KNN 51.22 (1.2) 49.16 (0.7) 49.04 (1.7) C4.5 52.98 (2.5) 48.89 (0.8) 47.24 (1.8) Ripper 55.27 (1.1) 48.94 (0.6) 46.40 (2.0) BayesNet 48.98 (1.1) 52.81 (1.0) 49.23 (2.7) SVM 53.53 (1.5) 52.48 (1.3) 49.15 (1.6) KNN 51.81 (1.7) 49.88 (1.7) 49.20 (1.3)<	KNN 57.81 (1.3) 57.38 (3.0) 54.18 (1.0) C4.5 53.78 (2.4) 52.99 (2.0) 49.82 (1.7) Ripper 60.10 (3.2) 57.50 (1.9) 51.67 (1.7) 44.88 (1.7) BayesNet 56.78 (1.9) 60.06 (2.8) 60.26 (2.4) 50.70 48.90 (1.5) KNN 58.43 (2.2) 57.69 (2.2) 55.16 (2.1) 7.70 7.70 C4.5 52.42 (3.2) 49.66 (1.5) 47.13 (1.1) 41.43 (1.7) BayesNet 58.28 (3.4) 59.23 (1.7) 56.97 (1.7) 55.90 (2.7) 48.93 (1.9) 41.43 (1.7) SVM 59.39 (2.7) 60.72 (2.8) 57.76 (1.8) 41.43 (1.7) 49.84 (1.7) 56.97 (1.7) 55.29 (2.5) 48.89 (8.8) 47.24 (1.8) 41.61 (1.1) 41.43 (1.7) 49.84 (1.7) 49.23 (2.7) 57.40 (4.1) 48.94 (0.6) 46.40 (2.0) 34.16 (1.1) 34.16 (1.1) 49.23 (2.7) 57.41 (1.1) 48.94 (0.6) 46.40 (2.0) 34.16 (1.1) 49.24 (2.7) 58.41 (1.1) 49.24 (2.7) 48.94 (1.1) 52.81 (1.0) 49.23 (2.7) 57.41 (1.1)

Tabela B.26: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Micro Precisão Hierárquica Ponderada no Segundo Nível da Hierarquia

Conjunto de Dados		0	n Top-Down		Abordagen	
		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	34.03 (1.1)	31.88 (1.8)	24.07 (1.2)		
	C4.5	28.99 (2.3)	28.38 (1.3)	16.88 (1.3)		
Expr	Ripper	38.24 (3.8)	$37.48\ (1.5)$	20.44 (1.6)	21.27 (0.9)	23.91 (1.2)
	BayesNet	31.58(2.3)	36.55 (3.6)	25.77(1.8)		
	SVM	54.32 (3.4)	40.30 (2.3)	22.96 (1.4)		
	KNN	36.13(2.2)	32.85(2.3)	24.15(1.2)		
	C4.5	28.05(2.0)	26.00(1.4)	$15.31\ (1.1)$		
CellCycle	Ripper	36.15(3.0)	36.08(1.9)	19.76 (1.6)	19.48 (0.8)	22.07 (0.4)
	BayesNet	37.20(3.7)	$35.84\ (1.6)$	25.24 (0.5)		
	SVM	36.12 (2.5)	37.47 (3.8)	26.86 (1.9)		
	KNN	29.12 (1.3)	26.96 (0.5)	20.91 (0.8)		
	C4.5	30.06(1.7)	$27.01\ (0.6)$	17.80(1.1)		
Church	Ripper	33.69(2.7)	33.16 (0.7)	20.51(2.0)	10.23 (0.7)	11.95 (1.
	BayesNet	26.46 (0.4)	28.94(1.1)	19.62(1.1)		
	SVM	30.75(1.6)	28.93(0.8)	19.94(1.6)		
	KNN	31.05 (1.3)	27.33 (1.1)	20.81 (1.1)		
Derisi	C4.5	36.35 (3.3)	24.68 (0.9)	14.57 (1.0)		
	Ripper	34.12 (1.8)	33.42 (0.4)	20.28 (1.3)	18.23 (0.3)	20.85 (0.0
	BayesNet	25.13 (1.8)	28.56 (2.7)	21.07 (0.8)	, ,	`
	SVM	29.07 (1.2)	27.40 (1.4)	19.88 (1.1)		
	KNN	35.81 (2.5)	32.38 (1.4)	24.77 (2.6)		
Eisen	C4.5	27.23 (1.2)	28.03 (1.1)	16.23 (2.3)		
	Ripper	36.80 (3.9)	36.00 (1.8)	21.20 (1.8)	20.47 (1.1)	22.93 (0.3
	BayesNet	33.67 (3.6)	34.49 (2.6)	26.05 (2.1)		
	SVM	37.05 (2.7)	40.05 (1.5)	26.94 (3.2)		
	KNN	35.56 (2.0)	33.24 (0.6)	24.57 (1.8)		
	C4.5	29.98 (1.6)	28.63 (1.1)	16.73 (1.2)		
Gasch1	Ripper	37.92 (4.5)	35.54 (2.2)	20.02 (2.1)	19.71 (0.6)	23.03 (1.0
Gasciii	BayesNet	28.24 (1.5)	31.68 (1.9)	23.09 (0.8)	10.71 (0.0)	20.00 (1.
	SVM	43.98 (2.7)	44.91 (2.4)	29.58 (2.8)		
	KNN	34.07 (2.3)	30.74 (1.6)	24.06 (2.3)		
	C4.5	29.97 (2.2)	26.65 (1.0)	15.34 (0.8)		
Gasch2	Ripper	36.54 (2.2)	35.97 (2.3)	19.74 (0.8) $19.74 (1.4)$	19.26 (0.9)	21.41 (0.
Gasciiz	BayesNet	27.07 (1.9)	30.67 (1.3)	21.55 (0.8)	19.20 (0.9)	21.41 (0.
	SVM	31.31 (1.4)	34.25 (0.5)	22.78 (0.5)		
				. ,		
	KNN	42.11 (4.8) 39.02 (2.5)	30.02 (1.5)	20.64 (2.0)		
Phenotype	C4.5 Ripper	39.87 (3.1)	30.76 (1.8) 36.76 (1.0)	19.27 (0.7) 18.98 (0.7)	10.23 (0.5)	12.41 (1.9
Тпепотуре	BayesNet	28.23 (1.9)	24.45 (2.2)	16.69 (0.9)	10.23 (0.3)	12.41 (1.
	SVM	41.14 (2.2)	32.55 (1.6)	20.13 (1.2)		
	KNN	28.88 (1.7)	28.03 (1.0)	20.22 (0.5)		
Cognonac	C4.5	29.26 (0.4)	28.40 (2.1)	16.83 (1.6)	20.00 (0.0)	99.04.(0.)
Sequence	Ripper	37.93 (1.6)	33.03 (0.9)	18.99 (0.9)	20.88 (0.8)	23.04 (0.8
	BayesNet	34.10 (1.3)	32.11 (1.7)	24.86 (1.3)		
	SVM	63.19 (3.3)	42.23 (2.5)	25.51 (0.7)		
	KNN	29.66 (1.5)	26.74 (0.3)	20.43 (1.1)		
	C4.5	29.82 (2.0)	25.10 (1.1)	$13.64 \ (0.5)$		
SPO	Ripper	36.45 (3.3)	36.33 (1.2)	21.02 (1.3)	18.29 (0.5)	21.31 (0.3
	BayesNet	24.90 (1.5)	28.69 (1.1)	21.64 (1.4)		
	SVM	28.77(1.3)	28.37(1.7)	20.62(1.7)		

Tabela B.27: Comparação das Abordagens Top-Down e One-Shot Utilizando a Medida

Micro Precisão Hierárquica Ponderada no Terceiro Nível da Hierarquia

Abordagem Top-Down			Abordagem One-Shot		
	HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
KNN	25.80(2.3)	23.85(1.7)	11.39 (0.4)		
C4.5	20.63(1.5)	20.17 (0.7)	$06.70 \ (0.5)$		
Ripper	29.87(3.0)	28.05 (1.2)	$08.34 \ (0.6)$	12.86 (0.4)	$15.33 \ (0.8)$
BayesNet	21.78(2.0)	27.21(2.5)	13.22 (1.3)		
SVM	47.80 (4.2)	30.93 (1.7)	14.12 (0.9)		
KNN	28.08 (2.9)	24.82 (1.8)	11.90 (0.6)		
C4.5	19.72(1.3)	18.97(0.6)	$06.03 \ (0.4)$		
Ripper	27.14(2.7)	27.24(1.6)	$08.43 \ (0.8)$	11.47 (0.5)	14.17 (0.2
BayesNet	27.49(3.2)	26.99(1.9)	12.27(0.4)		
SVM	27.54(2.4)	27.93(3.2)	13.20(1.1)		
KNN	21.33 (0.8)	19.92 (0.5)	10.25 (0.6)		
C4.5	21.72 (1.5)	20.21 (0.6)	08.16 (0.5)		
Ripper	, ,	25.95 (0.8)	11.69 (1.2)	03.95 (0.3)	05.17 (0.8
	, ,	, ,	10.20 (0.5)	` ´	`
SVM	22.44 (1.5)	21.29 (0.5)	09.62 (0.6)		
KNN	23 18 (0.9)	20.58 (1.2)	09 75 (0.5)		
	` /	` ,	` ′		
	` ,	, ,	, ,	10.95 (0.3)	13.40 (0.3
	, ,	, ,		10.55 (0.5)	10.40 (0.6
	, ,	` '	` ′		
			` '		
	, ,	, ,	` ′		
	` /	, ,	` ′	10.04 (0.0)	1400 (0
**	, ,	` '	` ′	12.34 (0.9)	14.90 (0.4
*	` ′		, ,		
			, ,		
	, ,	, ,	, ,		
	, ,	` '	` ′		
				11.35 (0.4)	14.92 (1.0
	, ,	, ,			
SVM	36.56 (2.3)	35.10 (2.2)	15.41 (1.6)		
KNN	26.08(2.5)	23.15(0.4)	11.39(1.1)		
C4.5	21.55(2.4)	$19.10 \ (0.6)$	$06.01 \ (0.4)$		
Ripper	28.20 (2.4)	27.65(2.1)	08.72 (0.5)	11.36 (0.6)	13.70 (0.3
BayesNet	18.80(1.7)	22.51(1.2)	$10.40 \ (0.3)$		
SVM	23.02 (0.6)	25.10 (0.9)	10.89 (0.2)		
KNN	36.42 (5.4)	23.06 (1.9)	10.87 (1.8)		
C4.5	32.29(2.8)	22.77(1.3)	09.03(0.7)		
Ripper	31.22(3.5)	29.24(0.8)	10.19 (0.6)	04.22 (0.3)	07.08 (1.9
BayesNet	20.25(1.4)	18.64 (1.1)	09.14 (1.1)		
SVM	33.56(2.1)	23.76(0.8)	09.58 (0.9)		
KNN	20.13 (1.1)	19.39 (0.4)	08.75 (0.2)		
C4.5	` /	, ,	, ,		
	, ,	` '	` ′	12.82 (0.5)	15.63 (0.6
	, ,	` ′			(0.0
SVM	55.76 (4.8)	33.65 (1.6)	15.91 (0.5)		
1 ~ •					
KNIN	21.09 (0.0)	20 17 (0.7)	00 46 (0 4)		
KNN C4.5	21.98 (0.9)	20.17 (0.7)	09.46 (0.4)		
C4.5	21.86 (2.0)	18.42 (0.8)	05.48 (0.2)	10.97 (0.3)	19 57 (0.5
	, ,	, ,	, ,	10.87 (0.3)	13.57 (0.3
	KNN C4.5 Ripper BayesNet SVM	Abordagen HMC-BR KNN	Abordagem Top-Down HMC-BR HMC-LP KNN	Abordagem Top-Down	HMC-BR

Tabela B.28: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Micro Precisão Hierárquica Ponderada no Quarto Nível da Hierarquia

Conjunto de Dados		Abordagem Top-Down			Abordagem One-Shot	
James de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
Expr	KNN	21.75 (1.9)	20.46 (1.7)	07.69 (0.4)		
	C4.5	$17.94\ (1.3)$	17.35 (0.8)	04.28 (0.3)		
	Ripper	26.75(3.0)	23.43 (0.9)	$05.36 \ (0.5)$	$10.15 \ (0.6)$	$12.73 \ (1.3)$
	BayesNet	18.43 (2.3)	23.06(2.3)	09.58 (1.5)		
	SVM	42.18 (3.7)	25.05 (1.5)	10.37 (0.6)		
CellCycle	KNN	23.45(2.3)	21.76(1.8)	07.93 (0.4)		
	C4.5	$16.81 \ (1.2)$	$16.18 \; (0.7)$	$03.80 \ (0.3)$		
	Ripper	23.49(2.2)	22.55(1.5)	$05.32 \ (0.5)$	08.59 (0.3)	11.22 (0.3
	BayesNet	24.37 (3.0)	22.67(1.9)	08.52 (0.1)		
	SVM	24.12 (2.1)	24.18 (2.8)	09.27 (0.8)		
Church	KNN	17.81 (0.9)	16.60 (0.7)	06.83 (0.3)		
	C4.5	17.93(1.5)	17.12(0.5)	05.36(0.2)		
	Ripper	20.97(2.0)	20.80 (0.8)	07.77(0.8)	02.64 (0.2)	03.51 (0.0
	BayesNet	14.59(0.7)	17.37 (0.8)	07.03 (0.5)		
	SVM	18.54 (1.9)	17.32(0.5)	06.49(0.5)		
Derisi	KNN	19.47 (0.8)	17.27 (1.0)	06.27 (0.3)		
	C4.5	24.43 (2.3)	15.67 (0.5)	03.66 (0.2)		
	Ripper	21.97 (2.5)	20.51 (0.5)	06.13 (0.3)	08.25 (0.3)	10.71 (0.
	BayesNet	13.66 (1.0)	17.16 (1.4)	06.49 (0.3)	()	- (-
	SVM	17.50 (0.9)	17.04 (1.0)	06.10 (0.3)		
	KNN	21.79 (1.7)	20.93 (2.4)	07.78 (0.6)		
Eisen	C4.5	16.56 (1.3)	17.29 (0.6)	04.01 (0.5)		
	Ripper	25.10 (3.9)	22.94 (2.1)	04.01 (0.3) 05.53 (0.4)	09.33 (0.8)	12.35 (0.
	BayesNet	21.02 (2.9)	21.91 (2.1)	08.15 (0.4)	09.55 (0.6)	12.55 (0.
	SVM	25.34 (2.6)	26.96 (2.2)	08.13 (0.1)		
Gasch1	KNN					
	C4.5	22.74 (0.9) 17.69 (0.8)	21.70 (1.1) 17.71 (0.8)	07.99 (0.7) $04.18 (0.2)$		
	Ripper	24.97 (3.7)	22.75 (1.6)	` ,	08.59 (0.6)	12.06 (1.0
	BayesNet	14.72 (1.7)	19.19 (1.2)	05.19 (0.6) 08.41 (0.6)	08.59 (0.0)	12.00 (1.0
	SVM	31.86 (2.1)	30.88 (2.1)	10.81 (1.3)		
Gasch2	KNN	22.51 (2.0)	20.03 (0.2)	07.48 (0.9)		
	C4.5	18.65 (2.6)	16.62 (0.4)	03.80 (0.3)	00.01.(0.0)	11.00 (0.
	Ripper	24.80 (2.0)	23.09 (1.7)	05.67 (0.3)	08.61 (0.6)	11.06 (0.0
	BayesNet	16.89 (1.5)	19.26 (0.9)	07.09 (0.5)		
	SVM	20.05 (0.8)	22.07 (1.4)	07.35 (0.2)		
Phenotype	KNN	33.16 (6.2)	19.31 (2.2)	07.24 (1.4)		
	C4.5	28.17 (3.0)	19.04 (1.3)	05.81 (0.4)		
	Ripper	26.90 (2.6)	23.61 (0.8)	06.61 (0.7)	$02.80 \ (0.3)$	05.13 (1.0
	BayesNet	18.48 (1.5)	15.38 (0.8)	06.70 (0.7)		
	SVM	29.15 (2.4)	20.45 (1.2)	06.25 (0.6)		
Sequence	KNN	16.29(1.1)	15.91 (0.3)	05.48(0.1)		
	C4.5	17.38 (0.5)	18.14 (1.5)	$04.19 \ (0.3)$		
	Ripper	24.74 (1.5)	20.66 (1.3)	$04.83 \ (0.3)$	09.90 (0.4)	13.33 (1.0
	BayesNet	21.74(1.7)	21.76 (0.6)	$08.31 \ (0.4)$		
	SVM	52.40 (4.6)	27.56 (1.4)	11.77 (0.5)		
SPO	KNN	18.32 (0.8)	16.75 (0.4)	05.92 (0.2)		
	C4.5	18.34 (1.6)	15.76 (0.5)	03.45 (0.1)		
	Ripper	24.65 (3.0)	22.57 (1.0)	06.09 (0.5)	08.07 (0.4)	10.69 (0.3
	BayesNet	14.26 (0.9)	17.39 (0.5)	07.72 (0.8)	` ′	`
	SVM	17.77 (1.0)	17.75 (0.8)	06.44 (0.3)		

Tabela B.29: Comparação das Abordagens Top-Down e One-Shot Utilizando a Medida

Micro Revocação Hierárquica Ponderada no Primeiro Nível da Hierarquia

	-	-		1	
	HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
KNN	$49.71 \ (0.8)$	48.17(1.6)	60.20(1.6)		
C4.5	53.97(1.2)	53.02(2.5)	75.40(3.3)		
			$76.22 \ (1.1)$	56.59 (1.5)	55.94 (1.6
-	, ,	, ,	58.79(3.9)		
SVM	40.54 (1.6)	43.86 (1.6)	66.94 (0.9)		
KNN	51.94(2.0)	49.94(1.7)	61.17 (1.6)		
C4.5	50.95(1.9)	50.35(2.0)	$76.05 \ (1.9)$		
Ripper	$47.01\ (1.8)$	41.75(1.8)	74.56(2.2)	55.69 (1.6)	51.67 (2.8
BayesNet	55.03(1.9)	49.02(2.2)	59.16(1.7)		
SVM	55.64 (1.6)	56.43 (2.8)	66.58 (1.8)		
KNN	39.79 (2.6)	39.72 (0.8)	59.64 (1.3)		
C4.5	38.24(2.0)	41.81 (0.9)	64.99(2.0)		
Ripper	39.70(2.9)	34.94 (0.7)	66.77(2.1)	78.63 (1.7)	70.95 (1.4
BayesNet	63.56 (0.9)	38.43 (1.5)	49.35(5.1)		
SVM	37.90 (2.2)	40.25 (1.2)	64.70(2.1)		
KNN	44.52 (1.7)	43.09 (1.1)	55.98 (1.6)		
C4.5	33.55 (10.0)	45.87 (1.1)	71.12 (2.4)		
Ripper	41.22 (2.3)	37.13 (0.6)	69.44(2.2)	54.29 (1.1)	51.77 (3.3
BayesNet	59.60 (1.5)	40.69(2.7)	49.43(4.0)		
SVM	43.34 (1.1)	43.84(1.5)	61.57(2.6)		
KNN	53.90 (2.4)	51.23 (1.3)	64.16 (2.6)		
C4.5	, ,	, ,			
Ripper	, ,	, ,		57.52 (1.7)	56.64 (2.4
BayesNet				, ,	`
SVM	56.71 (3.0)	57.53 (1.8)	67.24 (3.4)		
KNN					
	, ,	, ,	, ,		
	, ,	, ,	, ,	56.67 (2.8)	54.10 (2.1
BayesNet	, ,	54.36 (2.4)	51.50 (2.8)	, ,	`
SVM	56.20 (2.1)	56.02 (1.7)	67.08 (2.2)		
KNN	50.02 (1.4)				
		, ,			
	, ,	, ,	` ,	54.57 (1.6)	53.41 (2.2
	, ,				00111 (111
	(/	` '	` ′		
	, ,	` /	, ,		
	, ,	, ,		87.88 (1.7)	74.50 (7.5
	` ′	, ,		01100 (211)	. 1.00 (
-	, ,	, ,	, ,		
	, ,	, ,	, ,		
	` '	, ,		56 18 (2.0)	53.92 (3.3
				00.10 (2.0)	00.02 (0.0
SVM	32.20 (2.1)	45.47 (2.1)	69.75 (1.8)		
	32.20 (2.4)	10.11 (2.1)		<u> </u>	
TZNINI	40 41 (0.0)	41 P1 /1 A			
KNN	42.41 (2.3)	41.51 (1.4)	56.01 (2.8)		
C4.5	44.81 (12.0)	47.66 (1.6)	70.45 (2.2)	F4.94 (9.0)	F1 40 (1 0
		, ,	, ,	54.34 (2.0)	51.46 (1.8
	KNN C4.5 Ripper BayesNet SVM	Abordaged HMC-BR KNN 49.71 (0.8) C4.5 53.97 (1.2) Ripper 52.99 (3.4) BayesNet 68.11 (2.3) SVM 40.54 (1.6) KNN 51.94 (2.0) C4.5 50.95 (1.9) Ripper 47.01 (1.8) BayesNet 55.03 (1.9) SVM 55.64 (1.6) KNN 39.79 (2.6) C4.5 38.24 (2.0) Ripper 39.70 (2.9) BayesNet 63.56 (0.9) SVM 37.90 (2.2) KNN 44.52 (1.7) C4.5 33.55 (10.0) Ripper 41.22 (2.3) BayesNet 59.60 (1.5) SVM 43.34 (1.1) KNN 53.90 (2.4) C4.5 53.59 (2.3) Ripper 50.53 (3.8) BayesNet 62.76 (1.7) SVM 56.71 (3.0) KNN 53.90 (3.4) C4.5 56.15 (1.1) Ripper 49.39 (2.0) BayesNet 67.15 (1.8) SVM 56.20 (2.1) KNN 50.02 (1.4) C4.5 45.96 (5.1) Ripper 44.42 (4.0) BayesNet 58.15 (2.8) SVM 54.72 (1.6) KNN 32.15 (4.3) C4.5 32.97 (5.5) Ripper 36.86 (4.8) BayesNet 49.09 (3.7) SVM 39.02 (4.0) KNN 49.80 (3.6) C4.5 53.27 (2.1) Ripper 48.33 (3.7)	Abordagem Top-Down HMC-BR HMC-LP KNN 49.71 (0.8) 48.17 (1.6) C4.5 53.97 (1.2) 53.02 (2.5) Ripper 52.99 (3.4) 43.93 (1.5) BayesNet 68.11 (2.3) 60.85 (2.3) SVM 40.54 (1.6) 43.86 (1.6) KNN 51.94 (2.0) 49.94 (1.7) C4.5 50.95 (1.9) 50.35 (2.0) Ripper 47.01 (1.8) 41.75 (1.8) BayesNet 55.03 (1.9) 49.02 (2.2) SVM 55.64 (1.6) 56.43 (2.8) KNN 39.79 (2.6) 39.72 (0.8) C4.5 38.24 (2.0) 41.81 (0.9) Ripper 39.70 (2.9) 34.94 (0.7) BayesNet 63.56 (0.9) 38.43 (1.5) SVM 37.90 (2.2) 40.25 (1.2) KNN 44.52 (1.7) 43.09 (1.1) C4.5 33.55 (10.0) 45.87 (1.1) Ripper 41.22 (2.3) 37.13 (0.6) BayesNet 59.60 (1.5) 40.69 (2.7) SVM 43.34 (1.1) 43.84 (1.5) KNN 53.90 (2.4) 51.23 (1.3) C4.5 53.59 (2.3) 52.17 (1.3) Ripper 50.53 (3.8) 44.31 (2.0) BayesNet 62.76 (1.7) 49.81 (2.8) SVM 56.71 (3.0) 57.53 (1.8) KNN 53.90 (3.4) 52.06 (2.2) C4.5 56.15 (1.1) 52.84 (2.0) Ripper 49.39 (2.0) 40.90 (0.9) BayesNet 67.15 (1.8) 54.36 (2.4) SVM 56.20 (2.1) 56.02 (1.7) KNN 50.02 (1.4) 46.66 (1.3) C4.5 45.96 (5.1) 49.87 (2.7) Ripper 44.42 (4.0) 39.43 (1.7) BayesNet 58.15 (2.8) 43.81 (0.8) SVM 54.72 (1.6) 54.01 (1.5) KNN 32.15 (4.3) 39.27 (1.6) C4.5 32.97 (5.5) 41.01 (1.8) Ripper 36.86 (4.8) 37.94 (0.8) BayesNet 49.09 (3.7) 47.47 (1.7) SVM 39.02 (4.0) 40.52 (1.6) KNN 49.80 (3.6) 48.22 (2.3) C4.5 53.27 (2.1) 52.07 (2.9) Ripper 48.33 (3.7) 39.83 (1.1)	Abordagem Top-Down HMC-BR HMC-LP HMC-CT KNN 49.71 (0.8) 48.17 (1.6) 60.20 (1.6) C4.5 53.97 (1.2) 53.02 (2.5) 75.40 (3.3) Ripper 52.99 (3.4) 43.93 (1.5) 76.22 (1.1) BayesNet 68.11 (2.3) 60.85 (2.3) 58.79 (3.9) SVM 40.54 (1.6) 43.86 (1.6) 66.94 (0.9) KNN 51.94 (2.0) 49.94 (1.7) 61.17 (1.6) C4.5 50.95 (1.9) 50.35 (2.0) 76.05 (1.9) Ripper 47.01 (1.8) 41.75 (1.8) 74.56 (2.2) BayesNet 55.03 (1.9) 49.02 (2.2) 59.16 (1.7) SVM 55.64 (1.6) 56.43 (2.8) 66.58 (1.8) KNN 39.79 (2.6) 39.72 (0.8) 59.64 (1.3) C4.5 38.24 (2.0) 41.81 (0.9) 64.99 (2.0) Ripper 39.70 (2.9) 34.94 (0.7) 66.77 (2.1) BayesNet 63.56 (0.9) 38.43 (1.5) 49.35 (5.1) SVM 37.90 (2.2) 40.25 (1.2) 64.70 (2.1) KNN 44.52 (1.7) 43.09 (1.1) 55.98 (1.6) C4.5 33.55 (10.0) 45.87 (1.1) 71.12 (2.4) Ripper 41.22 (2.3) 37.13 (0.6) 69.44 (2.2) BayesNet 59.60 (1.5) 40.69 (2.7) 49.43 (4.0) SVM 43.34 (1.1) 43.84 (1.5) 61.57 (2.6) KNN 53.90 (2.4) 51.23 (1.3) 64.16 (2.6) C4.5 53.59 (2.3) 52.17 (1.3) 73.28 (3.9) Ripper 50.53 (3.8) 44.31 (2.0) 75.49 (2.1) BayesNet 62.76 (1.7) 49.81 (2.8) 60.85 (3.4) SVM 56.71 (3.0) 57.53 (1.8) 67.24 (3.4) KNN 53.90 (3.4) 52.06 (2.2) 63.49 (1.5) C4.5 56.15 (1.1) 52.84 (2.0) 76.35 (4.4) Ripper 49.39 (2.0) 40.90 (0.9) 76.79 (0.4) BayesNet 67.15 (1.8) 54.36 (2.4) 51.50 (2.8) SVM 56.20 (2.1) 56.02 (1.7) 67.08 (2.2) KNN 50.02 (1.4) 46.66 (1.3) 60.41 (2.4) C4.5 45.96 (5.1) 49.87 (2.7) 74.51 (2.9) Ripper 44.42 (4.0) 39.43 (1.7) 73.20 (1.0) BayesNet 58.15 (2.8) 43.81 (0.8) 58.72 (1.7) SVM 54.72 (1.6) 54.01 (1.5) 62.68 (0.9) KNN 32.15 (4.3) 39.27 (1.6) 55.01 (6.5) C4.5 33.97 (5.5) 41.01 (1.8) 63.61 (7.3) Ripper 36.86 (4.8) 37.94 (0.8) 65.14 (1.5) SVM 39.02 (4.0) 40.52 (1.6) 64.20 (6.2) KNN 49.80 (3.6) 48.22 (2.3) 59.75 (3.8) C4.5 53.27 (2.1) 52.07 (2.9) 74.47 (2.3) Ripper 44.33 (3.7) 39.83 (1.1) 76.43 (2.0)	HMC-BR

Tabela B.30: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Micro Revocação Hierárquica Ponderada no Segundo Nível da Hierarquia

Conjunto de Dados		_	m Top-Down		Abordagem	
conjunio de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	26.08 (0.8)	25.53(1.1)	35.86(2.2)		
	C4.5	29.66 (1.3)	28.69(1.5)	52.08(5.6)		
Expr	Ripper	28.93(1.5)	27.02(1.3)	$52.28\ (2.4)$	35.49 (0.8)	33.61 (1.0
	BayesNet	41.55 (2.5)	34.43 (2.8)	33.70(3.4)		
	SVM	21.59 (0.8)	27.02 (1.7)	44.63 (2.4)		
	KNN	27.42(1.9)	26.75(1.6)	36.53 (0.9)		
	C4.5	26.62(1.3)	26.39(1.4)	$52.87\ (2.5)$		
CellCycle	Ripper	25.17(1.3)	26.13(1.1)	49.85(2.3)	36.93 (0.4)	32.03 (1.7)
	BayesNet	31.67(1.4)	26.83(1.6)	$35.61\ (1.6)$		
	SVM	30.63 (1.6)	32.21 (3.1)	41.49 (2.5)		
	KNN	19.80 (1.3)	20.30 (0.6)	37.26 (1.3)		
	C4.5	19.21 (1.5)	21.36(0.7)	43.71 (2.8)		
Church	Ripper	20.22(1.9)	24.50 (0.8)	43.85(2.1)	$52.74\ (1.2)$	45.99 (2.5
	BayesNet	38.01 (1.2)	19.28 (0.8)	28.06 (3.3)		
	SVM	19.14 (1.6)	20.64 (0.9)	40.82(2.0)		
	KNN	22.14 (0.9)	22.02 (0.5)	33.57 (1.4)		
	C4.5	16.82 (5.3)	24.35 (1.0)	48.97 (3.4)		
Derisi	Ripper	20.61 (1.2)	24.95 (0.7)	44.63 (2.9)	37.51 (1.2)	32.64 (1.7
	BayesNet	34.25 (1.9)	21.41 (1.9)	28.39 (2.4)	,	(
	SVM	21.80 (0.6)	22.89 (1.1)	36.53 (1.7)		
	KNN	29.60 (2.5)	27.40 (0.9)	39.08 (4.1)		
Eisen	C4.5	28.46 (1.4)	27.73 (1.0)	50.82 (4.1)		
	Ripper	27.87 (2.6)	25.57 (1.4)	53.02 (1.6)	37.87 (0.9)	34.97 (0.7
	BayesNet	37.21 (3.0)	26.36 (1.7)	36.21 (3.7)	01.01 (0.5)	04.01 (0.1
	SVM	32.11 (2.7)	34.03 (1.7)	43.21 (4.3)		
	KNN C4.5	29.66 (2.8)	28.27 (1.0)	38.28 (1.7)		
Gasch1	Ripper	31.10 (1.5)	28.41 (1.3) 26.47 (1.3)	53.51 (4.6) 52.63 (1.4)	37.14 (1.6)	32.96 (1.5
Gasciii	BayesNet	27.70 (1.8)	` ′	` ′	37.14 (1.0)	32.90 (1.6
	SVM	40.82 (2.2) 33.47 (2.3)	28.36 (1.1) 34.25 (1.9)	28.77 (1.2) 42.59 (2.8)		
	<u> </u>			` ` `		
	KNN	26.04 (1.5)	24.64 (1.2)	36.13 (2.2)		
C 10	C4.5	24.18 (2.8)	26.45 (1.7)	50.67 (3.5)	26.10 (0.7)	22 22 /1 1
Gasch2	Ripper	23.98 (2.7)	25.90 (1.5)	49.48 (1.4)	$36.10 \ (0.7)$	33.22 (1.1
	BayesNet	32.45 (2.0)	23.53 (0.7)	34.05 (1.4)		
	SVM	29.69 (1.1)	30.31 (1.2)	37.60 (1.0)		
	KNN	15.92 (2.8)	21.80 (1.0)	35.89 (4.9)		
	C4.5	15.52 (2.6)	22.20 (1.2)	42.87 (7.8)		
Phenotype	Ripper	17.85 (2.1)	26.79 (0.6)	46.68 (1.3)	65.95 (3.7)	51.82 (7.8
	BayesNet	26.68 (1.7)	24.61 (1.7)	31.84 (2.9)		
	SVM	19.45 (2.2)	23.56 (1.2)	42.71 (5.1)		
	KNN	26.30(2.3)	$24.81\ (1.3)$	36.17 (3.1)		
	C4.5	28.98(1.2)	28.86(2.4)	$52.64\ (1.4)$		
Sequence	Ripper	26.18(2.5)	25.12 (0.8)	53.91 (2.3)	35.59(1.1)	33.27 (1.4
	BayesNet	39.43(3.2)	$30.61\ (1.9)$	36.20(4.4)		
	SVM	19.45 (2.4)	28.13 (1.6)	45.65 (1.2)		
	KNN	21.15 (1.1)	21.32 (0.5)	33.10 (1.6)		
	C4.5	23.17 (6.9)	25.03 (0.9)	47.44 (3.1)		
SPO	Ripper	20.85 (2.0)	25.44 (1.3)	46.60 (1.1)	36.69 (1.7)	32.31 (0.9
	BayesNet	30.64 (0.7)	23.33 (1.1)	26.70 (1.2)	, ,	•
	SVM	25.23 (1.6)	25.33 (2.0)	35.91 (2.0)		

Tabela B.31: Comparação das Abordagens Top-Down e One-Shot Utilizando a Medida

Micro Revocação Hierárquica Ponderada no Terceiro Nível da Hierarquia

no revocação r				iro Nível da		One Shot
Conjunto de Dados		Abordage HMC-BR	m Top-Down HMC-LP	HMC-CT	Abordagem C4.5H	Clus-HMC
	KNN	18.83 (1.4)	18.32 (1.0)	29.18 (1.8)	C4.511	Clus-IIIVIC
	C4.5	21.52 (0.9)	20.60 (0.8)	47.36 (5.3)		
Expr	Ripper	21.06 (1.4)	19.12 (0.8)	43.51 (2.9)	31.03 (1.6)	27.29 (1.4
	BayesNet	30.91 (2.2)	25.32 (2.2)	27.60 (3.6)	(=10)	_,,_, (_,,
	SVM	15.75 (0.9)	19.56 (1.3)	36.58 (1.3)		
	KNN	20.13 (2.2)	19.40 (1.3)	29.85 (1.6)		
	C4.5	19.21 (1.0)	19.02 (0.6)	51.21 (1.7)		
CellCycle	Ripper	17.85 (0.8)	18.84 (0.7)	41.24 (2.2)	34.46 (0.3)	26.46 (1.2
	BayesNet	22.85 (1.4)	19.16 (1.2)	26.98 (1.7)		
	SVM	22.01 (1.3)	23.00 (2.4)	32.74 (1.9)		
	KNN	13.47 (1.1)	14.33 (0.5)	32.32 (1.3)		
	C4.5	13.28 (0.9)	15.06 (0.5)	39.39 (3.4)		
Church	Ripper	14.10 (1.1)	18.88 (0.9)	36.72 (2.4)	48.08 (2.7)	42.78 (3.5
	BayesNet	26.20 (1.3)	14.01 (1.3)	23.28 (3.4)		(-
	SVM	13.50 (1.4)	14.60 (0.6)	35.14 (0.9)		
	KNN	15.28 (0.7)	15.63 (0.3)	29.12 (1.3)		
	C4.5	11.50 (3.6)	17.85 (0.7)	48.24 (3.8)		
Derisi	Ripper	14.06 (0.9)	18.55 (0.7)	36.87 (1.7)	36.28 (1.5)	28.35 (1.3
Bellisi	BayesNet	24.19 (1.5)	15.00 (1.7)	22.45 (1.5)	30.20 (1.0)	20.00 (1.0
	SVM	15.59 (0.4)	16.53 (0.7)	32.95 (1.2)		
	KNN	21.37 (2.1)	19.97 (1.6)	31.78 (3.4)		
Eisen	C4.5	20.88 (1.4)	19.78 (0.6)	47.43 (5.0)		
	Ripper	21.14 (1.9)	18.44 (1.5)	45.46 (2.7)	34.83 (1.2)	28.82 (1.2
	BayesNet	27.77 (2.6)	18.85 (1.6)	29.18 (3.4)	01.00 (1.2)	20102 (111
	SVM	23.92 (2.1)	24.85 (1.5)	34.49 (3.3)		
	KNN	21.22 (1.8)	20.15 (0.6)	31.29 (1.1)		
	C4.5	22.17 (0.8)	19.98 (1.2)	48.74 (5.4)		
Gasch1	Ripper	19.74 (1.6)	19.32 (1.3)	44.10 (2.1)	34.28 (2.1)	27.11 (2.0
	BayesNet	29.98 (2.0)	20.13 (0.6)	24.07 (1.8)	,	
	SVM	25.29 (2.1)	25.41 (1.7)	33.23 (1.2)		
	KNN	18.73 (1.6)	17.59 (0.3)	29.88 (0.7)		
	C4.5	17.09 (1.7)	19.10 (1.1)	48.19 (3.7)		
Gasch2	Ripper	17.13 (2.4)	19.60 (1.2)	40.50 (1.5)	34.11 (1.5)	28.09 (1.6
	BayesNet	22.92 (1.4)	17.15 (0.4)	26.62 (1.6)	, ,	
	SVM	21.31 (0.3)	21.40 (0.8)	31.54 (1.5)		
	KNN	11.11 (1.9)	16.03 (0.7)	31.11 (4.3)		
	C4.5	10.91 (2.0)	16.19 (0.6)	36.84 (7.1)		
Phenotype	Ripper	12.39 (1.3)	21.07 (0.4)	40.91 (1.4)	68.10 (5.2)	47.54 (7.8
V-1	BayesNet	18.06 (1.4)	16.64 (0.8)	24.72 (2.6)	, ,	`
	SVM	13.78 (1.7)	17.16 (0.6)	36.98 (4.8)		
	KNN	18.36 (1.7)	17.62 (0.7)	30.04 (2.7)		
	C4.5	21.09 (1.1)	21.02 (1.9)	49.43 (2.1)		
Sequence	Ripper	18.44 (1.9)	18.12 (0.8)	45.83 (2.6)	30.65 (0.8)	28.83 (1.4
-	BayesNet	28.81 (3.6)	21.98 (1.7)	29.13 (3.5)	` ´	`
	SVM	15.37 (2.2)	21.17 (1.1)	37.73 (1.6)		
	KNN	14.75 (0.6)	15.21 (0.4)	28.59 (1.2)		
	C4.5	16.92 (5.1)	18.21 (0.9)	46.18 (3.0)		
SPO	Ripper	15.03 (1.4)	18.80 (1.1)	37.19 (1.4)	35.47 (2.6)	27.52 (1.3
	BayesNet	22.13 (0.4)	16.16 (1.0)	20.82 (0.7)		- (
	SVM	18.03 (0.9)	18.27 (0.8)	31.17 (0.9)		

Tabela B.32: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Micro Revocação Hierárquica Ponderada no Quarto Nível da Hierarquia

Conjunto de Dados		0	m Top-Down		Abordagem	One-Shot
Conjunto de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	16.75(1.1)	$16.32\ (1.2)$	28.15 (1.8)		
	C4.5	19.06 (0.9)	17.89 (0.6)	$46.37 \ (4.9)$		
Expr	Ripper	18.60 (1.1)	16.35 (0.6)	42.83(3.8)	28.41 (2.3)	24.72 (1.5)
	BayesNet	27.74 (2.2)	22.04 (2.0)	26.13 (4.1)		
	SVM	13.43 (0.6)	16.37 (1.2)	36.96 (2.1)		
	KNN	17.99(1.9)	17.74 (1.5)	29.00(1.3)		
	C4.5	$16.71 \ (0.7)$	$16.31 \ (0.4)$	$50.49\ (1.9)$		
CellCycle	Ripper	15.59 (0.8)	15.91 (0.9)	40.84(3.2)	31.40 (1.0)	23.30 (0.9
	BayesNet	20.24 (0.8)	16.62 (0.8)	26.04(2.1)		
	SVM	19.44 (0.9)	20.79 (2.3)	31.53(2.0)		
	KNN	11.58(1.0)	12.20 (0.4)	$31.23\ (1.3)$		
	C4.5	11.43 (1.0)	12.98(0.5)	38.73(3.7)		
Church	Ripper	12.19(0.9)	15.73(0.7)	36.09(2.8)	47.51 (3.8)	41.27 (4.9
	BayesNet	21.32(1.0)	12.05 (1.0)	21.88(3.5)		
	SVM	11.59 (0.9)	12.69 (0.4)	34.66 (1.3)		
	KNN	13.36 (0.5)	13.43 (0.4)	28.29 (1.3)		
	C4.5	09.85(3.2)	15.41 (0.9)	47.94 (3.5)		
Derisi	Ripper	12.04 (0.7)	15.58 (0.5)	36.54 (1.1)	33.39 (1.8)	25.56 (1.8
	BayesNet	19.99 (1.2)	12.76 (1.6)	20.04 (1.2)		
	SVM	13.45 (0.5)	14.36 (0.7)	31.88 (1.4)		
	KNN	19.16 (2.0)	17.76 (1.7)	30.69 (3.2)		
Eisen	C4.5	18.34 (1.6)	17.15 (0.6)	46.29 (5.1)		
	Ripper	18.45 (1.3)	15.70 (1.2)	44.14 (3.1)	31.16 (1.3)	26.04 (1.5
	BayesNet	24.71 (2.2)	16.34 (1.2)	27.54 (3.3)	,	`
	SVM	21.19 (1.9)	22.27 (1.2)	32.63 (3.5)		
	KNN	19.08 (1.8)	18.00 (0.2)	30.19 (1.1)		
	C4.5	19.66 (0.6)	17.40 (1.3)	47.97 (5.6)		
Gasch1	Ripper	17.24 (1.7)	16.67 (1.1)	43.43 (2.6)	31.38 (2.5)	23.75 (2.4
	BayesNet	25.89 (2.1)	17.17 (0.4)	22.72 (1.5)	(2.0)	
	SVM	22.11 (1.8)	23.48 (1.7)	31.90 (1.3)		
	KNN	16.92 (1.1)	15.54 (0.5)	29.02 (1.1)		
	C4.5	15.01 (1.7)	$16.51 \ (0.3)$	48.02 (3.9)		
Gasch2	Ripper	15.37 (2.5)	16.63 (1.3)	39.38 (2.6)	31.43 (1.4)	24.96 (2.3
Gasciiz	BayesNet	20.77 (1.4)	14.77 (0.4)	24.75 (1.9)	01.10 (1.1)	21.00 (2.0
	SVM	18.63 (0.5)	19.15 (0.7)	29.70 (1.9)		
	KNN	09.40 (1.7)	13.98 (0.8)	29.66 (4.8)		
	C4.5	09.40 (1.7)	14.37 (0.8)	34.84 (6.9)		
Phenotype	Ripper	10.47 (1.2)	17.88 (0.5)	39.75 (3.0)	68.48 (6.2)	44.5 (9.1
1 henotype	BayesNet	14.87 (1.2)	14.57 (0.6)	21.94 (2.7)	00.40 (0.2)	44.0 (3.1
	SVM	11.86 (1.7)	15.27 (0.6)	34.94 (3.9)		
	KNN C4.5	15.96 (1.4)	15.09 (0.6) 18.26 (1.6)	28.29 (2.7)		
Sequence	Ripper	18.79 (1.2) 15.93 (1.7)	18.26 (1.6) 15.28 (0.7)	48.90 (1.6) 45.58 (4.5)	27.22 (0.7)	27.37 (1.9
bequence	BayesNet	24.81 (3.1)	15.28 (0.7) $19.11 (1.4)$	45.58 (4.5) 28.31 (4.0)	21.22 (0.1)	21.31 (1.8
	SVM	24.81 (3.1) 13.80 (1.7)	19.11 (1.4) 18.05 (0.9)	28.31 (4.0) 36.87 (2.2)		
					I	
	KNN	13.07 (0.6)	13.15 (0.3)	27.89 (1.2)		
gp.o	C4.5	14.76 (5.1)	15.66 (0.8)	45.76 (2.8)	99.05 (2.0)	00.00 /3 =
SPO	Ripper	12.73 (1.0)	15.95 (1.0)	36.67 (2.2)	33.25 (2.9)	23.89 (1.2
	BayesNet	18.80 (0.6)	13.96 (0.8)	18.99 (0.5)		
	SVM	15.91 (1.0)	16.01 (0.8)	30.01 (1.4)		

Tabela B.33: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Macro Precisão Hierárquica Ponderada no Primeiro Nível da Hierarquia

Conjunto de Dados		_	n Top-Down		-	n One-Shot
conjunto de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HM0
	KNN	$61.51\ (1.8)$	59.63(4.4)	57.16 (1.7)		
	C4.5	51.06(2.3)	50.55 (2.5)	$47.01\ (1.5)$		
Expr	Ripper	58.48(3.3)	$62.15 \ (4.0)$	$50.74\ (1.6)$	40.86 (1.8)	45.29 (1.4
	BayesNet	55.37(2.2)	57.72(2.9)	59.08(2.7)		
	SVM	49.75 (3.0)	48.97 (3.2)	39.74(2.5)		
	KNN	60.91 (4.7)	56.59 (3.7)	57.17 (4.3)		
	C4.5	49.82(3.9)	45.62(1.4)	44.39(1.3)		
CellCycle	Ripper	56.37(3.7)	61.26 (5.7)	46.92(1.7)	37.71 (1.9)	39.53 (0.7
	BayesNet	55.49(3.3)	$56.64\ (0.9)$	54.30(1.3)		
	SVM	57.68 (2.4)	59.37 (2.7)	55.73 (1.4)		
	KNN	51.22 (2.8)	48.05 (3.5)	47.11 (3.1)		
	C4.5	50.41 (3.2)	46.55 (1.9)	43.52 (3.2)		
Church	Ripper	53.31 (9.3)	35.80 (8.4)	42.09 (10.0)	32.45 (1.2)	33.23 (0.
	BayesNet	47.50 (2.2)	51.50 (0.9)	47.85 (2.6)	, ,	
	SVM	51.38 (1.8)	53.52 (3.7)	48.18 (3.8)		
	KNN	55.66 (2.8)	49.96 (2.2)	50.85 (1.2)		
	C4.5	53.56 (3.1)	44.14 (1.3)	42.67 (2.1)		
Derisi	Ripper	56.57 (1.9)	43.37 (5.9)	49.17 (1.4)	36.13 (1.1)	38.76 (2.
	BayesNet	49.70 (0.7)	49.05 (4.5)	50.24 (1.2)	(=12)	33113 (=1
	SVM	52.61 (3.1)	48.71 (1.3)	48.34 (1.8)		
	KNN	63.22 (3.6)	61.58 (2.6)	60.57 (2.1)		
Eisen	C4.5	48.83 (1.8)	48.94 (1.4)	45.06 (3.1)		
	Ripper	58.04 (4.3)	62.31 (2.6)	51.26 (2.6)	39.09 (2.4)	43.35 (1.
	BayesNet	56.43 (3.2)	58.76 (3.0)	57.98 (2.2)	33.03 (2.4)	40.00 (1.
	SVM	58.79 (2.4)	62.06 (3.1)	55.22 (2.5)		
	KNN	60.15 (3.3)	58.52 (3.2)	56.37 (2.3)		
G 11	C4.5	52.29 (2.0)	51.12 (1.9)	46.39 (1.1)	90 90 (1.0)	49.90.70
Gasch1	Ripper	57.54 (4.1)	53.51 (5.7)	50.36 (1.7)	39.32 (1.2)	43.38 (2.
	BayesNet SVM	54.19 (1.2) 66.19 (3.0)	55.06 (2.6)	55.52 (0.8)		
			64.67 (3.8)	59.51 (3.1)		
	KNN	57.06 (4.3)	52.75 (2.9)	53.85 (3.7)		
	C4.5	49.92 (2.4)	45.87 (1.8)	43.48 (1.0)		
Gasch2	Ripper	54.37 (3.0)	52.77 (9.5)	47.29 (1.2)	36.49 (1.4)	38.64 (1.
	BayesNet	51.75 (1.3)	49.39 (2.7)	50.51 (1.9)		
	SVM	53.76 (1.1)	54.77 (2.1)	50.61 (0.8)		
	KNN	40.14(3.0)	41.14(3.9)	40.32(4.7)		
	C4.5	39.03(3.3)	48.26(5.7)	44.95(3.1)		
Phenotype	Ripper	40.98(2.6)	$42.12\ (20.0)$	39.45(1.3)	32.71 (0.6)	33.68 (2.
	BayesNet	43.40(2.5)	41.87(2.0)	41.30(2.7)		
	SVM	54.67 (13.0)	45.51 (5.0)	44.01 (1.4)		
	KNN	53.13 (1.4)	51.32 (2.2)	48.79 (1.7)		
	C4.5	48.85 (1.3)	$47.21\ (2.5)$	45.60 (1.5)		
Sequence	Ripper	53.67(3.2)	54.48 (3.8)	45.71 (1.8)	38.05 (1.9)	40.2 (1.9
	BayesNet	54.05 (1.8)	51.23(1.7)	53.13 (0.53)		
	SVM	79.99(5.1)	73.08 (3.7)	66.55 (2.7)		
	KNN	49.49 (1.7)	45.22 (1.2)	46.97 (2.1)		
	C4.5	48.65 (2.5)	43.87 (1.7)	41.93 (1.0)		
SPO	Ripper	54.91 (3.2)	62.10 (7.5)	51.47 (3.2)	34.48 (1.1)	36.76 (1.
	BayesNet	48.16 (3.3)	49.71 (2.8)	49.87 (4.1)	` ´	`
	SVM	46.52 (2.4)	47.40 (1.7)	46.34 (2.0)		

Tabela B.34: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Macro Precisão Hierárquica Ponderada no Segundo Nível da Hierarquia

Conjunto de Dados			n Top-Down		Abordagen	
conjunio de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	19.98(2.0)	16.77(3.3)	13.28 (3.6)		
	C4.5	12.20 (1.6)	10.69(2.5)	08.89 (1.1)		
Expr	Ripper	19.50 (3.0)	26.30(5.9)	11.78 (2.7)	06.96 (0.9)	07.58 (0.9)
	BayesNet	14.96 (2.2)	20.14 (4.9)	14.46 (4.0)		
	SVM	17.47 (0.8)	35.60 (2.7)	24.55 (2.1)		
	KNN	16.23(4.1)	$13.51\ (1.8)$	12.76(2.0)		
	C4.5	09.58(1.9)	08.12(1.7)	07.95 (0.6)		
CellCycle	Ripper	16.76 (5.0)	26.84 (3.5)	09.74(1.1)	05.97 (0.6)	06.46 (0.9)
	BayesNet	19.48(3.9)	23.66 (8.7)	18.47(3.4)		
	SVM	20.69 (3.8)	24.83 (6.1)	16.08 (2.9)		
	KNN	11.99 (3.0)	08.20 (1.4)	09.98 (2.0)		
	C4.5	09.64(2.6)	07.81 (0.4)	07.77(0.8)		
Church	Ripper	16.31 (4.4)	28.47(3.3)	21.91(3.2)	06.83 (1.9)	05.70 (0.3
	BayesNet	12.01(3.3)	$13.40 \ (4.5)$	11.08(2.3)		
	SVM	10.45 (2.7)	10.22(2.3)	09.46 (0.9)		
	KNN	12.46 (5.0)	10.96 (2.9)	09.88 (1.3)		
Derisi	C4.5	16.26 (5.5)	08.22 (1.8)	07.13 (0.7)		
	Ripper	19.68 (4.1)	29.72 (5.1)	18.73 (4.4)	05.56 (0.2)	07.54 (1.6
	BayesNet	12.99 (4.2)	18.20 (5.5)	15.14 (7.7)	, ,	`
	SVM	09.31 (0.6)	10.27 (2.0)	08.30 (0.8)		
	KNN	18.76 (4.9)	19.21 (4.3)	16.29 (2.0)		
Eisen	C4.5	12.83 (2.7)	09.21 (1.5)	08.60 (1.2)		
	Ripper	19.87 (4.4)	21.03 (5.4)	13.31 (2.2)	07.12 (1.6)	07.16 (0.7
	BayesNet	18.93 (3.2)	19.64 (4.7)	17.66 (3.9)		(
	SVM	23.07 (7.0)	28.29 (7.2)	17.87 (3.0)		
	KNN	27.92 (3.9)	22.40 (4.1)	15.82 (5.9)		
	C4.5	14.05 (3.4)	11.99 (2.9)	09.17 (1.0)		
Gasch1	Ripper	22.14 (4.1)	26.31 (6.0)	11.11 (1.0)	06.89 (0.9)	06.95 (0.6
Gubelli	BayesNet	14.89 (2.6)	17.99 (3.2)	14.20 (5.3)	00.00 (0.0)	00.00 (0.0
	SVM	36.13 (4.5)	32.53 (9.9)	23.44 (3.1)		
	KNN	17.64 (3.7)	14.17 (3.4)	13.36 (1.5)		
	C4.5	11.39 (4.8)	08.50 (0.9)	07.77 (0.2)		
Gasch2	Ripper	19.17 (2.7)	25.98 (5.1)	10.99 (2.0)	05.91 (0.5)	06.47 (0.8
Gasciiz	BayesNet	11.55 (1.8)	16.86 (3.3)	11.62 (1.6)	05.51 (0.5)	00.47 (0.0
	SVM	16.41 (3.9)	15.91 (2.5)	10.19 (1.2)		
				, ,		
	KNN C4.5	13.27 (5.0) 10.98 (4.5)	16.75 (5.8)	13.49 (2.9)		
Phenotype	Ripper	$10.98 \ (4.3)$ $12.23 \ (2.4)$	13.78 (4.2) 32.37 (3.6)	10.02 (2.2)	06.48 (0.5)	05.94 (0.8
r nenotype	BayesNet	13.09 (5.6)	13.50 (2.0)	25.21 (0.5) 12.34 (7.7)	00.48 (0.5)	05.94 (0.6
	SVM	16.14 (5.1)	15.30 (2.0) $15.31 (2.3)$	09.53 (1.5)		
	KNN	12.09 (4.3)	11.24 (2.6)	09.50 (1.6)		
Common on	C4.5	11.91 (2.9)	10.41 (3.5)	08.81 (1.0)	06.07.(0.7)	07 40 /1 9
Sequence	Ripper	19.60 (3.5)	19.01 (2.7)	09.17 (0.8)	06.07 (0.7)	07.48 (1.3
	BayesNet SVM	16.80 (3.9)	14.00 (2.6)	12.81 (2.8)		
		37.85 (11.0)	37.46 (1.2)	28.51 (1.5)		
	KNN	07.61 (0.2)	07.89 (1.0)	08.11 (0.5)		
an a	C4.5	09.22 (1.6)	07.20 (0.8)	06.70 (0.3)		
SPO	Ripper	16.13 (5.9)	30.65 (2.3)	13.92 (5.4)	06.23 (1.9)	$05.71 \ (0.2$
	BayesNet	11.49 (2.9)	13.83 (2.8)	14.77 (3.8)		
	SVM	10.39(2.5)	08.73(1.3)	08.84(1.7)		

Tabela B.35: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Macro Precisão Hierárquica Ponderada no Terceiro Nível da Hierarquia

Conjunto de Dados			m Top-Down		Abordagen	
Conjunito de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	12.50(2.2)	09.97(1.7)	06.85 (1.9)		
	C4.5	07.19(1.7)	07.33(2.3)	03.12 (1.1)		
Expr	Ripper	12.69(2.0)	20.36 (4.3)	$04.10 \ (0.8)$	$04.24\ (1.6)$	04.72 (1.1
	BayesNet	09.22(1.6)	11.53 (3.7)	06.86 (2.8)		
	SVM	12.56 (2.4)	$35.50 \ (4.4)$	27.10 (4.1)		
	KNN	10.59(2.9)	09.17(3.1)	06.69 (1.6)		
	C4.5	06.38(2.1)	05.59(1.3)	02.55 (0.4)		
CellCycle	Ripper	09.23(1.2)	24.12(4.6)	04.98(1.1)	02.99(0.7)	04.02 (0.6
	BayesNet	08.43(2.2)	13.36(4.6)	11.66 (2.4)		
	SVM	13.55 (2.7)	16.20 (2.3)	08.65 (1.4)		
	KNN	07.06 (1.6)	05.49 (0.8)	05.40 (1.5)		
	C4.5	07.37(2.4)	04.60(1.3)	03.09 (0.6)		
Church	Ripper	09.08(2.7)	34.75(1.7)	14.66 (1.6)	02.70(1.3)	02.65 (0.8
	BayesNet	05.20(2.1)	10.61(2.9)	12.45 (3.1)		
	SVM	07.29(2.2)	06.73(1.9)	03.97(1.4)		
	KNN	06.43 (3.0)	04.85 (2.5)	05.32 (1.3)		
Derisi	C4.5	08.40 (4.2)	05.52 (2.1)	01.95 (0.2)		
	Ripper	08.98 (2.0)	31.28 (4.6)	09.63 (0.5)	02.49 (0.9)	04.54 (0.9
	BayesNet	07.86 (1.5)	10.93 (4.7)	13.05 (3.6)	, ,	`
	SVM	07.02 (1.8)	05.42 (1.3)	03.03 (1.2)		
	KNN	10.53 (3.4)	11.98 (1.6)	08.25 (1.8)		
Eisen	C4.5	06.04 (0.6)	06.23 (2.1)	03.27 (1.2)		
	Ripper	13.37 (4.2)	17.26 (7.2)	05.94 (2.0)	04.10 (1.7)	04.90 (1.8
	BayesNet	10.51 (2.0)	11.87 (3.4)	13.00 (4.1)	(-)	(
	SVM	14.54 (4.5)	17.71 (4.7)	07.86 (1.1)		
	KNN	14.88 (1.3)	12.39 (3.3)	09.09 (2.3)		
	C4.5	07.22 (1.9)	06.54 (0.9)	02.63 (0.7)		
Gasch1	Ripper	12.43 (4.6)	23.53 (5.0)	04.42 (1.0)	03.02 (0.6)	04.21 (1.0
Gubelli	BayesNet	09.13 (1.3)	08.61 (2.2)	11.14 (2.2)	00.02 (0.0)	01.21 (1.
	SVM	22.13 (2.9)	26.38 (6.7)	17.15 (3.1)		
	KNN	11.00 (5.5)	10.37 (1.4)	06.97 (3.5)		
	C4.5	08.09 (3.0)	04.67 (0.5)	00.97 (3.3) $02.24 (0.5)$		
Gasch2	Ripper	11.67 (4.0)	30.37 (1.4)	02.24 (0.5)	03.08 (0.8)	04.50 (1.9
Gasciiz	BayesNet	08.93 (2.7)	08.64 (1.6)	08.43 (2.6)	03.00 (0.0)	04.00 (1
	SVM	09.02 (1.4)	12.44 (2.3)	05.73 (1.2)		
				` '		
	KNN C4.5	07.53 (2.6) 05.23 (0.9)	14.21 (3.2)	08.35 (4.5) 06.22 (3.1)		
Phenotype	Ripper	$05.23 \ (0.9)$ $07.22 \ (3.3)$	12.60 (1.0) 36.68 (4.0)	23.68 (4.0)	02.12 (0.3)	02.95 (1.
r nenotype	BayesNet	06.04 (2.8)	05.27 (1.9)	03.97 (1.4)	02.12 (0.3)	02.95 (1.
	SVM	08.39 (3.3)	13.35 (2.6)	03.97 (1.4) $03.85 (0.7)$		
	KNN	06.20 (1.6)	05.98 (3.1)	04.08 (1.1)		
Cognopae	C4.5	06.11 (2.9)	07.07 (2.8)	02.86 (0.4)	09 99 /1 1\	05.00.70.4
Sequence	Ripper	11.85 (2.0)	20.81 (3.0)	02.87 (0.6)	03.23 (1.1)	05.00 (0.6
	BayesNet	10.34 (1.7)	11.25 (1.3)	07.72 (2.0)		
	SVM	21.83 (7.3)	37.52 (3.0)	25.34 (1.4)		
	KNN	04.69 (2.4)	05.53 (2.2)	04.08 (1.1)		
an 0	C4.5	06.80 (1.5)	03.98 (1.3)	02.07 (0.4)		
SPO	Ripper	09.86 (2.9)	29.41 (3.9)	07.59 (2.1)	03.36 (1.8)	03.43 (1.6
	BayesNet	06.76 (1.1)	06.57 (2.2)	07.74 (1.5)		
	SVM	06.71(1.3)	07.18(1.4)	03.83 (0.9)		

Tabela B.36: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Macro Precisão Hierárquica Ponderada no Quarto Nível da Hierarquia

Conjunto de Dados		_	m Top-Down		Abordagen	
Jan-10 do Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	$12.17\ (1.8)$	13.13 (2.8)	07.39(1.6)		
_	C4.5	08.91 (3.5)	09.47 (2.0)	02.49 (0.5)		
Expr	Ripper	14.44 (1.8)	22.11(4.5)	03.38 (0.6)	05.76 (1.8)	08.32 (2.1
	BayesNet	09.64 (2.5)	12.76 (2.7)	07.99 (4.2)		
	SVM	09.67 (1.2)	39.13 (2.6)	29.37 (2.1)		
	KNN	12.34 (3.2)	11.74(4.2)	06.73(1.2)		
	C4.5	07.28(2.0)	06.73(1.4)	$01.70 \ (0.3)$		
CellCycle	Ripper	09.39(1.7)	27.74 (6.1)	04.90 (0.3)	$03.42 \ (0.8)$	05.87 (0.6)
	BayesNet	07.56 (1.8)	16.06 (5.0)	13.78(2.0)		
	SVM	14.86 (2.6)	19.73 (3.5)	08.66 (2.2)		
	KNN	07.60 (1.1)	07.68 (2.0)	05.37 (1.1)		
	C4.5	07.84(2.0)	06.30(1.4)	$02.80 \ (0.8)$		
Church	Ripper	08.38(2.0)	38.52 (2.4)	15.97(2.8)	02.25 (0.9)	02.82 (0.8
	BayesNet	06.22(3.4)	14.82(2.7)	15.26 (1.5)		
	SVM	07.51(2.4)	08.34(1.5)	03.84(1.3)		
	KNN	06.73 (2.3)	07.28 (1.2)	05.21 (1.1)		
Derisi	C4.5	07.89(2.9)	06.28 (2.1)	01.40 (0.3)		
	Ripper	07.73 (1.2)	33.89(4.2)	08.91 (1.0)	02.51 (0.7)	05.51 (1.8
	BayesNet	08.21 (1.1)	12.21(2.5)	12.61(3.5)		
	SVM	07.52(2.3)	05.94(0.9)	02.42(1.5)		
	KNN	08.47 (2.7)	12.31 (1.0)	07.55 (2.1)		
Eisen	C4.5	07.62 (0.7)	08.74 (3.0)	02.13 (0.9)		
	Ripper	13.62 (2.9)	18.97 (8.3)	05.11 (1.8)	03.72 (1.9)	07.32 (2.9
	BayesNet	10.04 (1.4)	14.77 (3.0)	12.65 (2.2)	, ,	`
	SVM	16.37 (4.2)	19.88 (4.0)	08.51 (0.8)		
	KNN	16.33 (3.1)	15.34 (4.2)	08.99 (2.2)		
	C4.5	07.92 (1.7)	08.42 (0.5)	02.59 (0.2)		
Gasch1	Ripper	12.52 (3.9)	27.15 (5.1)	04.87 (1.1)	03.67 (0.8)	05.84 (1.3
	BayesNet	08.44 (3.3)	10.21 (1.5)	12.98 (2.7)	,	`
	SVM	20.94 (1.4)	30.78 (4.9)	18.42 (3.6)		
	KNN	10.99 (2.5)	14.66 (2.4)	06.20 (2.7)		
	C4.5	08.73 (1.8)	06.48 (1.8)	01.64 (0.4)		
Gasch2	Ripper	09.27 (1.8)	31.63 (1.0)	07.13 (2.0)	03.80 (1.2)	05.77 (2.2
	BayesNet	08.27 (2.7)	09.83 (2.2)	10.40 (1.5)	,	
	SVM	09.06 (2.0)	14.37 (1.3)	05.42 (1.7)		
	KNN	05.55 (2.5)	14.25 (3.7)	09.80 (4.8)		
	C4.5	04.43 (1.3)	15.73 (1.8)	06.34 (3.0)		
Phenotype	Ripper	05.90 (1.6)	38.08 (4.5)	20.96 (4.2)	02.06 (1.1)	02.66 (1.2
	BayesNet	06.82 (3.7)	06.46 (1.1)	03.38 (1.3)	02:00 (2:2)	02.00 (
	SVM	07.61 (2.8)	15.18 (2.6)	03.35 (0.9)		
	KNN	07.71 (2.0)	09.27 (2.1)	03.77 (1.2)		
	C4.5	07.71 (2.0) 07.26 (3.1)	09.27 (2.1)	03.77 (1.2) $02.13 (0.3)$		
Soguence	Ripper	12.89 (3.3)	23.29 (4.0)	02.13 (0.5) $02.20 (0.5)$	04.29 (1.1)	08.25 (1.4
equence	BayesNet	12.75 (1.6)	12.52 (1.8)	07.89 (2.2)	J.1.20 (1.11)	00.20 (1.5
bequence		12.10 (1.0)	, ,	, ,		
Sequence	-	16.66 (3.9)	42.34 (3.2)	28,68 (3.8)		
Sequence	SVM	16.66 (3.9)	42.34 (3.2)	28.68 (3.8)		
Bequence	SVM KNN	05.20 (2.5)	05.89 (1.1)	03.20 (1.3)		
	SVM KNN C4.5	05.20 (2.5) 07.24 (1.7)	05.89 (1.1) 05.51 (2.2)	03.20 (1.3) 01.54 (0.4)	02 62 (1.2)	04 22 /1 5
SPO	SVM KNN	05.20 (2.5)	05.89 (1.1)	03.20 (1.3)	03.62 (1.2)	04.33 (1.3

Tabela B.37: Comparação das Abordagens Top-Down e One-Shot Utilizando a Medida

Macro Revocação Hierárquica Ponderada no Primeiro Nível da Hierarquia

0 : 4 1 5 1	Abordagem Top-Down				Abordagem One-Shot		
Conjunto de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC	
	KNN	44.90 (1.0)	44.96 (1.9)	53.91 (1.7)			
	C4.5	51.76(1.9)	50.49(3.4)	72.52 (3.0)			
Expr	Ripper	50.50(3.2)	39.64(2.1)	71.19(0.9)	51.76 (1.7)	52.77 (1.4)	
	BayesNet	68.02(3.2)	60.47(2.4)	59.40 (4.4)			
	SVM	31.38 (1.3)	33.81 (1.2)	52.83 (0.5)			
	KNN	45.85 (2.0)	45.22 (1.7)	53.48 (1.9)			
	C4.5	47.89(2.9)	46.18(2.2)	$73.04\ (1.7)$			
CellCycle	Ripper	43.06(2.4)	35.72(2.1)	68.04(2.8)	51.40(2.0)	47.57(2.3)	
	BayesNet	54.45(2.7)	46.87(2.6)	58.33(2.2)			
	SVM	51.52 (1.3)	52.83 (3.5)	62.43 (1.8)			
	KNN	35.26 (2.8)	35.95 (1.6)	50.62 (1.6)			
	C4.5	34.03(2.1)	38.42 (0.7)	57.74 (1.7)			
Church	Ripper	34.38(1.7)	27.19(1.0)	54.97(1.7)	74.84 (2.8)	66.45 (2.1	
	BayesNet	61.13(1.3)	$36.71\ (1.0)$	48.07(4.6)			
	SVM	33.93(1.5)	36.78 (1.4)	56.48 (1.9)			
	KNN	41.00 (1.0)	40.62 (1.3)	50.07 (1.1)			
	C4.5	32.66 (8.0)	43.39 (0.7)	67.47 (1.6)			
Derisi	Ripper	38.88(2.2)	30.53(1.6)	62.05(1.9)	50.39 (1.9)	48.70 (4.2	
	BayesNet	58.90 (1.9)	41.84(2.6)	50.18 (3.1)			
	SVM	40.35(1.3)	40.97(1.2)	56.25 (2.8)			
	KNN	50.05 (2.6)	48.81 (1.2)	58.62 (2.3)			
	C4.5	50.63 (2.8)	48.86 (2.5)	67.98 (3.6)			
Eisen	Ripper	47.30 (3.6)	40.40 (2.3)	69.80 (1.9)	53.78 (2.8)	54.81 (2.3	
	BayesNet	62.49 (2.9)	50.79 (3.2)	61.24 (3.0)	, ,		
	SVM	53.95 (3.2)	54.88 (1.6)	63.81 (3.9)			
	KNN	51.06 (2.3)	50.34 (3.1)	59.15 (1.0)			
	C4.5	54.13 (1.8)	50.61 (2.9)	73.80 (4.5)			
Gasch1	Ripper	46.57 (2.2)	36.43 (1.5)	71.98 (0.7)	52.04 (2.5)	52.88 (2.2	
	BayesNet	67.73 (2.2)	55.91 (2.6)	54.41 (2.9)			
	SVM	51.53(2.4)	51.75(2.3)	63.25(2.1)			
	KNN	44.20 (1.5)	43.06 (2.0)	53.48 (2.3)			
	C4.5	41.32 (4.4)	45.80 (2.9)	70.93 (2.2)			
Gasch2	Ripper	39.68 (3.8)	31.98 (1.9)	64.99 (2.5)	48.76 (1.9)	48.65 (1.6	
	BayesNet	56.48 (3.3)	42.85 (1.2)	57.05 (2.0)			
	SVM	51.60 (1.1)	49.86(2.1)	59.45 (0.9)			
	KNN	24.77 (3.3)	31.56 (2.3)	45.44 (6.9)			
	C4.5	25.55 (4.2)	34.99 (3.5)	54.85 (7.1)			
Phenotype	Ripper	28.76 (3.8)	26.23 (0.9)	52.65 (1.6)	85.66 (0.9)	72.15 (7.6	
	BayesNet	47.94 (4.3)	47.58 (2.2)	55.47 (5.0)			
	SVM	31.21 (3.6)	33.10(2.1)	55.24(6.7)			
	KNN	43.82 (2.8)	45.08 (2.9)	52.33 (2.7)			
	C4.5	48.61 (2.0)	48.21 (3.4)	69.75 (2.7)			
Sequence	Ripper	42.31 (4.3)	32.69 (1.2)	67.96 (3.5)	50.62 (2.3)	47.99 (2.8	
	BayesNet	62.42 (2.2)	52.77 (3.7)	52.58 (4.2)		•	
	SVM	26.78 (1.8)	37.21 (2.0)	57.01 (1.6)			
	KNN	36.48 (1.9)	36.61 (1.3)	47.77 (2.2)			
	C4.5	40.77 (9.4)	43.61 (1.9)	66.38 (2.2)			
SPO	Ripper	36.36 (3.3)	32.09 (1.9)	61.01 (2.5)	48.48 (2.0)	45.78 (1.4	
	BayesNet	54.97 (0.9)	45.33 (2.7)	48.87 (2.4)	(- /		
	SVM	42.76 (2.5)	43.30 (3.2)	53.44 (1.6)			

Tabela B.38: Comparação das Abordagens Top-Down e One-Shot Utilizando a Medida

Macro Revocação Hierárquica Ponderada no Segundo Nível da Hierarquia

Conjunto de Dados		_	m Top-Down		Abordagem	
Conjunio de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	$11.11\ (1.3)$	11.13(2.1)	13.36 (2.6)		
	C4.5	$13.21\ (2.2)$	10.40(2.2)	17.29(4.6)		
Expr	Ripper	13.77(3.7)	15.24(2.7)	21.79 (4.0)	09.66 (2.3)	09.57 (2.0)
	BayesNet	19.71 (4.4)	18.15 (3.6)	14.87 (3.9)		
	SVM	07.05 (0.4)	14.76 (3.2)	20.60 (2.1)		
	KNN	09.88(1.2)	10.04 (0.9)	13.94(2.1)		
	C4.5	09.19(1.7)	08.14(1.6)	16.32 (3.4)		
CellCycle	Ripper	09.99(1.2)	14.57(1.9)	19.78(2.6)	$08.11 \ (0.6)$	07.78 (1.1
	BayesNet	$14.71 \ (1.5)$	12.33(2.6)	14.87(2.9)		
	SVM	14.57 (2.1)	16.71 (3.9)	$19.13 \ (2.0)$		
	KNN	05.97 (0.8)	06.04 (0.7)	11.28 (1.2)		
	C4.5	05.59(0.7)	06.19(0.1)	12.38 (2.8)		
Church	Ripper	$07.91\ (1.5)$	$15.50 \ (0.7)$	18.27 (3.4)	17.15(2.8)	13.11 (1.6
	BayesNet	13.26(1.1)	$06.36 \ (0.4)$	08.51 (1.4)		
	SVM	06.03 (0.3)	06.57 (0.7)	11.70 (1.5)		
	KNN	06.92 (0.9)	07.30 (1.2)	10.33 (0.9)		
	C4.5	07.66 (1.7)	07.90 (1.4)	12.76(2.1)		
Derisi	Ripper	09.16 (1.6)	15.79(1.7)	$17.02 \ (1.9)$	07.75(0.3)	09.71 (2.0
	BayesNet	13.36(2.5)	11.18(2.6)	10.91(3.2)		
	SVM	07.06 (0.3)	08.24 (0.9)	10.87(2.0)		
	KNN	12.53 (3.0)	12.76 (2.2)	16.40 (5.8)		
Eisen	C4.5	11.28 (1.2)	09.55 (1.4)	16.60 (4.6)		
	Ripper	12.80 (2.7)	13.66 (2.5)	25.00 (2.0)	09.98 (2.1)	10.15 (2.1
	BayesNet	20.25 (3.3)	14.63 (2.7)	18.41 (4.0)	` '	`
	SVM	15.62 (3.3)	21.71 (3.6)	21.15 (4.2)		
	KNN	15.53 (3.4)	14.48 (2.0)	16.37 (3.0)		
	C4.5	13.59 (3.4)	11.30 (2.2)	19.31 (5.1)		
Gasch1	Ripper	15.43 (2.8)	15.64 (2.9)	21.15 (2.6)	10.55 (1.8)	08.59 (0.7
	BayesNet	17.50 (2.7)	15.37 (2.5)	12.76 (2.9)	,	`
	SVM	18.55 (3.1)	19.40 (2.9)	24.12 (4.3)		
	KNN	09.32 (1.3)	09.39 (2.0)	13.55 (2.8)		
	C4.5	07.67 (1.1)	08.41 (1.2)	15.31 (1.8)		
Gasch2	Ripper	09.51 (1.7)	14.94 (1.5)	18.84 (2.0)	07.74 (0.3)	07.87 (0.5
	BayesNet	11.23 (1.3)	10.57 (0.8)	11.92 (1.8)	,	`
	SVM	12.85 (1.3)	13.77 (1.9)	13.60 (1.6)		
	KNN	05.17 (1.5)	07.65 (1.5)	13.57 (3.6)		
	C4.5	04.46 (0.7)	08.71 (2.8)	15.42 (5.1)		
Phenotype	Ripper	05.22 (0.6)	15.91 (0.5)	21.65 (2.2)	27.86 (5.3)	14.81 (5.4
V 1	BayesNet	10.32 (2.2)	10.08 (2.1)	13.55 (5.8)	,	`
	SVM	06.24 (1.2)	10.54 (1.3)	14.26 (3.7)		
	KNN	07.64 (1.0)	07.95 (0.5)	10.67 (1.5)		
	C4.5	10.22 (1.3)	10.48 (3.4)	18.53 (2.7)		
Sequence	Ripper	10.17 (1.6)	$13.14 \ (0.4)$	20.27 (3.2)	08.62 (2.0)	08.49 (0.9
1	BayesNet	15.86 (2.5)	12.23 (1.3)	15.99 (4.1)	(- ···)	(0.0
	SVM	07.90 (1.3)	15.40 (1.3)	22.64 (2.2)		
	KNN	05.61 (0.3)	06.15 (0.5)	09.25 (1.3)		
	C4.5	$05.01 \ (0.3)$ $07.28 \ (2.2)$	07.06 (0.4)	09.23 (1.3) $11.19 (0.9)$		
SPO	Ripper	07.28 (2.2)	14.15 (2.0)	16.45 (2.3)	08.25 (1.3)	07.18 (0.3
DI 0	BayesNet	09.78 (1.3)	08.85 (1.0)	12.63 (1.8)	00.20 (1.9)	01.10 (0.6
	SVM	09.08 (2.4)	07.92 (1.3)	10.93 (2.7)		

Tabela B.39: Comparação das Abordagens Top-Down e One-Shot Utilizando a Medida

Macro Revocação Hierárquica Ponderada no Terceiro Nível da Hierarquia

cro Revocação	Hierarquic			ro Nivel da .		
Conjunto de Dados			em Top-Down		Abordagen	
conjunito de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	05.38 (0.9)	05.79(1.4)	$07.25\ (1.6)$		
	C4.5	06.80 (1.4)	$06.64\ (1.6)$	05.92(2.0)		
Expr	Ripper	07.60 (1.2)	08.85 (1.6)	08.17 (1.5)	06.38 (2.3)	06.12 (1.7)
	BayesNet	10.04 (2.7)	08.89 (2.1)	10.19 (3.4)		
	SVM	03.11 (0.6)	09.79 (1.6)	16.99 (3.6)		
	KNN	04.57(1.1)	05.93(0.9)	06.90(1.6)		
	C4.5	$05.51\ (1.9)$	04.42 (0.9)	$05.74 \ (0.5)$		
CellCycle	Ripper	05.07(1.1)	08.87(1.7)	$07.40 \ (0.8)$	04.76 (1.1)	05.08 (1.6
	BayesNet	05.77(1.3)	05.35 (0.9)	$07.54\ (2.0)$		
	SVM	07.27 (0.9)	08.87 (1.2)	08.92 (2.5)		
	KNN	$02.41\ (0.6)$	$02.83 \ (0.4)$	05.73(1.9)		
	C4.5	$02.44 \ (0.6)$	03.35 (0.6)	04.73(1.2)		
Church	Ripper	02.90 (0.5)	11.87 (1.2)	11.11 (1.9)	06.80 (1.8)	05.47 (0.6)
	BayesNet	$04.10 \ (0.7)$	$03.16 \ (0.6)$	$07.17\ (1.6)$		
	SVM	03.77 (1.1)	04.73 (1.4)	04.61 (1.2)		
	KNN	02.57 (0.6)	03.28 (0.7)	05.01 (0.8)		
	C4.5	02.83(1.1)	04.43(1.3)	03.94 (0.8)		
Derisi	Ripper	03.00 (0.6)	$10.08\ (1.6)$	07.19(1.8)	04.35 (2.0)	06.37 (0.9
	BayesNet	04.65 (0.8)	04.19(1.2)	$07.94\ (1.8)$		
	SVM	04.93(1.6)	$04.61\ (1.7)$	$03.84\ (1.6)$		
	KNN	05.14 (1.3)	06.44 (1.6)	08.16 (3.0)		
Eisen	C4.5	04.98 (0.6)	05.96 (1.5)	06.68 (1.4)		
	Ripper	07.38(1.9)	08.22(2.9)	09.46 (1.2)	06.04 (2.9)	06.81 (1.6
	BayesNet	08.37(1.5)	06.50(1.0)	10.36(2.2)		
	SVM	08.43(2.3)	$11.57\ (2.9)$	08.42(1.0)		
	KNN	07.51 (1.7)	07.00 (0.5)	09.14 (2.0)		
	C4.5	07.08 (1.3)	05.41 (1.2)	05.48 (1.4)		
Gasch1	Ripper	07.49 (3.2)	09.42 (1.8)	07.73 (1.2)	05.49 (1.2)	05.44 (1.3
	BayesNet	07.43(1.2)	05.79(0.3)	11.62 (2.1)		
	SVM	09.63(1.8)	12.08(2.3)	$15.02 \ (3.0)$		
	KNN	05.09 (1.6)	05.46 (0.9)	07.79 (3.6)		
	C4.5	05.33 (1.7)	03.82 (0.5)	04.33 (0.7)		
Gasch2	Ripper	05.46 (2.0)	10.83 (0.9)	09.14 (1.8)	04.50 (1.4)	05.59 (2.5
	BayesNet	05.54 (1.2)	05.08 (0.4)	06.91 (1.4)		`
	SVM	06.78 (1.3)	08.05 (1.5)	07.60 (1.4)		
	KNN	01.93 (0.6)	04.88 (1.2)	08.62 (5.7)		
	C4.5	01.95 (0.3)	05.41 (1.1)	06.67 (2.1)		
Phenotype	Ripper	02.06 (0.4)	12.35 (1.0)	$19.74 \ (2.9)$	09.56 (2.2)	06.01 (2.3
V 1	BayesNet	03.30 (1.3)	03.10 (0.4)	03.42 (0.8)	, ,	`
	SVM	03.62 (1.2)	06.26 (0.9)	05.93 (1.3)		
	KNN	03.01 (0.5)	03.63 (1.1)	04.22 (0.8)	·	
	C4.5	04.94 (0.9)	07.00 (1.5)	07.09 (1.5)		
Sequence	Ripper	04.97 (1.3)	08.12 (0.9)	05.91 (1.2)	04.75 (1.7)	06.73 (0.8
• • • •	BayesNet	08.37 (2.1)	06.69 (1.2)	10.87 (2.2)		(0.0
	SVM	04.15 (0.8)	11.16 (1.3)	17.89 (2.3)		
	KNN	02.37 (0.7)	03.26 (1.0)	04.03 (1.3)		
	C4.5	03.96 (1.2)	03.20 (1.0) $03.58 (1.1)$	03.78 (0.9)		
			10.39 (1.1)	06.01 (1.2)	04.86 (2.8)	03.51 (1.0
SPO	Rinner					
SPO	Ripper BayesNet	03.97 (0.5) $04.63 (0.8)$	03.11 (0.7)	05.63 (1.3)	04.00 (2.0)	00.01 (1.0

Tabela B.40: Comparação das Abordagens $\mathit{Top\text{-}Down}$ e $\mathit{One\text{-}Shot}$ Utilizando a Medida

Macro Revocação Hierárquica Ponderada no Quarto Nível da Hierarquia

Conjunto de Dados		_	em Top-Down		Abordagen	
Conjunto de Dados		HMC-BR	HMC-LP	HMC-CT	C4.5H	Clus-HMC
	KNN	04.53 (0.6)	05.75(1.4)	07.67 (1.6)		
	C4.5	08.34(2.3)	$07.24\ (1.4)$	05.22(2.0)		
Expr	Ripper	$07.61 \ (0.7)$	07.95 (1.5)	06.75 (1.0)	09.25 (2.3)	10.15 (2.5)
	BayesNet	09.70(2.2)	$07.38\ (1.8)$	12.46 (4.8)		
	SVM	02.11 (0.3)	07.31 (1.3)	24.75 (3.8)		
	KNN	04.68(1.3)	06.35(2.0)	$07.32 \ (0.6)$		
	C4.5	06.60 (1.3)	04.48(1.0)	04.43(1.1)		
CellCycle	Ripper	$04.01\ (1.1)$	07.69(1.9)	$07.21\ (0.9)$	06.15(1.3)	07.71 (1.4)
	BayesNet	$04.45 \ (0.5)$	04.57(1.0)	09.41(2.3)		
	SVM	07.48 (1.1)	10.28 (1.7)	09.54(2.3)		
	KNN	02.37 (0.5)	$03.50 \ (0.8)$	05.85 (2.5)		
	C4.5	$02.66 \ (0.6)$	04.41(1.1)	05.28(1.5)		
Church	Ripper	$02.36 \ (0.3)$	$08.70 \ (0.7)$	13.79(2.0)	07.14(2.6)	06.19 (0.8
	BayesNet	$02.81\ (0.9)$	$03.81\ (1.0)$	08.74(1.4)		
	SVM	03.82 (1.9)	05.71(2.0)	04.67 (1.0)		
	KNN	02.08 (0.5)	03.55 (0.7)	05.34 (1.1)		
	C4.5	02.65 (1.5)	$04.51\ (1.7)$	02.89 (0.9)		
Derisi	Ripper	$02.11\ (0.5)$	$07.68 \ (0.8)$	07.79(1.0)	04.50 (1.2)	07.42 (1.3
	BayesNet	03.33(1.1)	03.28(1.2)	08.49(2.4)		
	SVM	05.22(2.1)	04.76(2.3)	03.35(1.9)		
	KNN	04.13 (1.4)	05.38 (0.9)	08.81 (2.7)		
Eisen	C4.5	06.56 (1.5)	06.31 (2.7)	04.34 (1.2)		
	Ripper	07.81 (1.5)	06.84 (2.4)	07.81 (1.9)	05.73 (3.2)	09.96 (3.4
	BayesNet	06.60 (1.6)	$05.04\ (0.9)$	11.75 (1.3)		
	SVM	08.65(2.2)	11.74(2.9)	09.80 (1.8)		
	KNN	06.94 (2.1)	07.05 (1.4)	09.12 (1.2)		
	C4.5	07.80 (2.1)	06.42 (2.0)	05.26 (0.3)		
Gasch1	Ripper	06.50 (2.4)	08.67 (1.4)	07.29 (1.6)	08.36 (1.9)	07.59 (0.9
	BayesNet	06.60 (2.7)	04.31 (0.5)	13.17 (1.9)	, ,	,
	SVM	08.28 (1.2)	13.05(2.4)	17.36 (3.5)		
	KNN	04.62 (1.4)	06.63 (0.8)	07.10 (3.1)		
	C4.5	05.86 (0.9)	04.07 (0.7)	03.20 (1.0)		
Gasch2	Ripper	04.92 (1.3)	08.76 (0.9)	09.25 (2.6)	06.76 (1.5)	07.19 (3.1
	BayesNet	04.77 (1.9)	03.63 (0.5)	08.63 (1.9)		
	SVM	06.11 (1.9)	08.87 (1.6)	07.63(2.3)		
	KNN	01.41 (0.5)	04.54 (1.1)	09.06 (4.8)		
	C4.5	01.28 (0.3)	05.88 (0.8)	06.35 (1.9)		
Phenotype	Ripper	01.36 (0.3)	08.65(0.7)	19.42 (4.4)	07.70 (3.1)	04.95 (2.3
* -	BayesNet	02.83 (1.1)	03.52 (0.8)	03.03 (1.0)	` ,	`
	SVM	03.91 (1.3)	06.79 (1.5)	04.79 (1.5)		
	KNN	03.40 (1.2)	04.11 (1.6)	04.33 (0.8)		
	C4.5	06.29 (1.1)	08.03 (1.3)	05.82 (1.9)		
Sequence	Ripper	05.45 (2.1)	06.86 (0.6)	04.53 (0.9)	06.23 (1.8)	11.77 (1.3
•	BayesNet	07.29 (2.4)	06.57 (1.3)	11.94 (2.3)	,	
	SVM	03.56 (0.9)	09.48 (0.9)	22.79 (3.2)		
	KNN	02.15 (1.1)	02.95 (0.9)	03.59 (0.9)		
	C4.5	03.47 (0.9)	03.38 (1.1)	02.89 (0.6)		
SPO	Ripper	03.39 (0.7)	08.74 (1.4)	06.90 (1.0)	05.29 (1.4)	04.95 (1.7
0	BayesNet	03.59 (0.7)	03.01 (0.6)	05.08 (1.7)	00.20 (1.4)	5 2.00 (1.1
				()		

Neste apêndice são apresentados os resultados dos testes estatísticos realizados para os experimentos do Capítulo 6. Os testes foram realizados com a finalidade de verificar a significância estatística das diferenças encontradas na comparação dos resultados das diferentes técnicas de classificação. Os testes estatísticos escolhidos foram o teste de Friedman (Friedman, 1937, 1940) e Nemenyi (Nemenyi, 1963)¹. Conforme explicado no Capítulo 6, estes testes são mais adequados para comparações envolvendo vários algoritmos e vários conjuntos de dados (Demšar, 2006).

Os resultados dos testes estão organizados em seis seções, sendo que cada seção apresenta os resultados dos testes estatísticos para uma medida de avaliação específica. Em cada uma das medidas de avaliação, os resultados dos testes são apresentados nível a nível da hierarquia. Para todos os testes, foram utilizados dois níveis de significância, 5% ($\alpha = 0.05$) e 10% ($\alpha = 0.1$).

Para facilitar o entendimento das análises estatísticas realizadas, os resultados são apresentados em tabelas. Para a interpretação das tabelas, são utilizados os símbolos apresentados na Tabela C.1. Esses símbolos representam as relações entre os resultados dos algoritmos localizados nas linhas das tabelas com os algoritmos localizados nas colunas das tabelas.

¹As implementações dos testes estatísticos foram feitas por Thiago Covões, do Laboratório de Computação Bioinspirada do ICMC-USP de São Carlos.

	Tabela C.1. Legenda para os resultados dos restes Estatisticos
Símbolo	Significado
	Indica que o algoritmo localizado na coluna da tabela obteve resultados
A	estatisticamente melhores do que algoritmo localizado na linha da tabela,
	com nível de significância igual a 5%
	Indica que o algoritmo localizado na coluna da tabela obteve resultados
Δ	estatisticamente melhores do que algoritmo localizado na linha da tabela,
	com nível de significância igual a 10%
	Indica que o algoritmo localizado na linha da tabela obteve resultados es-
∇	tatisticamente melhores do que algoritmo localizado na coluna da tabela,
	com nível de significância igual a 10%
	Indica que o algoritmo localizado na linha da tabela obteve resultados es-
▼	tatisticamente melhores do que algoritmo localizado na coluna da tabela,
	com nível de significância igual a 5%

Tabela C.1: Legenda para os Resultados dos Testes Estatísticos

${\it C.1}$ Análise dos Resultados da Medida ${\it Hierarchical~Loss~Func-tion}$

Conforme pode ser observado na Tabela C.2, no primeiro nível das hierarquias, houve bons resultados de todas as técnicas da abordagem *Top-Down*, porém destacaram-se as técnicas HMC-BR e HMC-LP. Nessas duas técnicas, os testes estatísticos mostraram diferenças estatisticamente significantes na comparação com a maioria dos algoritmos da técnica HMC-CT, e também na comparação com os resultados das técnicas C4.5H e Clus-HMC.

Os resultados no segundo, terceiro e quarto níveis (Tabelas C.3, C.4 e C.5), mostram que a técnica HMC-LP destacou-se, principalmente nos níveis três e quatro, em comparação com a técnica HMC-CT. Os testes estatísticos apresentaram diferenças estatisticamente significantes na comparação dos resultados da maioria dos algoritmos das técnicas HMC-BR e HMC-CT, bem como das técnicas HMC-LP e HMC-CT. Também houve diferenças estatisticamente significantes quando comparados alguns algoritmos das técnicas HMC-BR e HMC-LP com as técnicas C4.5H e Clus-HMC. Praticamente não houve diferenças estatisticamente significantes na comparação dos resultados das técnicas HMC-BR e HMC-LP.

Tabela C.2: Análise dos Resultados da Medida *Hierarchical Loss Function* no Primeiro

Nível das Hierarquias

				HMC-					HMC-					HMC-		
		KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM
	KNN															
	C4.5															
HMC-BR																
	BayesNet															
	SVM															
	KNN															
	C4.5			A		A					A					
HMC-LP	Ripper															
	BayesNet															
	SVM															
	KNN					A										
	C4.5	A	Δ	A	Δ	A			Δ	A	A					Δ
HMC-CT		Δ		A		A					A					
	BayesNet															
	SVM															
	.5H	A	A	A	A	A	Δ		A	A	A				A	A
Clus-	HMC	A		A		A				Δ	A					

Tabela C.3: Análise dos Resultados da Medida *Hierarchical Loss Function* no Segundo

Nível das Hierarquias

				HMC-					HMC-					HMC-		
		KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM
	KNN															
	C4.5															
HMC-BR	Ripper															
	BayesNet															
	SVM															
	KNN															
	C4.5															
HMC-LP	Ripper															
	BayesNet															
	SVM															
	KNN	A		A		A	A			A	Δ					
	C4.5	A	A	A	Δ	A	A		A	A	A					
HMC-CT	Ripper	A	A	A		A	A		A	A	A					
	BayesNet	A		A												
	SVM	A	Δ	A		A	A		Δ	A	A					
C4.	.5H	A		A		A	A			A						
Clus-	HMC	A		Δ												

Tabela C.4: Análise dos Resultados da Medida *Hierarchical Loss Function* no Terceiro

Nível das Hierarquias

I VIVCI G		1		HMC-	BR				HMC-	·LP				HMC-	CT	
		KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM
	KNN															
	C4.5															
HMC-BR																
	BayesNet															
	SVM															
	KNN															
	C4.5															
HMC-LP	Ripper															
	BayesNet															
	SVM															
	KNN	A		Δ		Δ	A		A	A	A					
	C4.5	A	A	A		A	A	A	A	A	A					
HMC-CT		A	Δ	A		A	A	A	A	A	A					
	BayesNet	A		Δ			A		A	A	A					
	SVM	A		A		Δ	A		A	A	A					
	.5H	Δ					A		A	A	A					
Clus-	HMC								A	Δ						

Tabela C.5:	Análise	${\rm dos}$	Resultados	da	Medida	Hierarchical	Loss	Function n	10	Quarto
Nível das Hi	erarquias									

				HMC-					HMC-					HMC-		
		KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM
	KNN															
	C4.5															
HMC-BR																
	BayesNet								A							
	SVM															
	KNN															
	C4.5															
HMC-LP	Ripper															
	BayesNet															
	SVM															
	KNN	Δ		A			A	A	A	A	A					
	C4.5	A	A	A		A	A	A	A	A	A					
HMC-CT	Ripper	A	A	A		A	A	A	A	A	A					
	BayesNet			Δ			A	Δ	A	A	A					
	SVM			A			A	A	A	A	A					
C4.	.5H			Δ			A	Δ	A	A	A					
Clus-	HMC						Δ		A	A						

C.2 Análise dos Resultados da Medida F1 Hierárquica

De acordo com os resultados mostrados nas tabelas C.6 e C.7, no primeiro e segundo níveis das hierarquias, as técnicas HMC-BR e HMC-CT obtiveram os melhores resultados, apresentando diferenças estatisticamente significantes na comparação com alguns resultados da técnica HMC-LP, e na comparação com os resultados das técnicas C4.5H e Clus-HMC.

Nos níveis três e quatro, os melhores resultados foram obtidos pelas técnicas HMC-BR e HMC-LP (Tabelas C.8 e C.9), com destaque para a técnica HMC-BR, que apresentou diferenças estatisticamente significantes na comparação com a maioria dos resultados das outras técnicas.

Tabela C.6: Análise dos Resultados da Medida F1 Hierárquica no Primeiro Nível das Hierarquias

				HMC-					HMC-					HMC-		
		KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM
	KNN													A		A
	C4.5				A								A	A		A
HMC-BR														A		A
	BayesNet															
	SVM													A		Δ
	KNN				A							Δ	A	A		A
	C4.5				Δ								A	A		A
HMC-LP	Ripper				A						A	A	A	A	A	A
	BayesNet													A		A
	SVM															
	KNN															
	C4.5															
HMC-CT	Ripper															
	BayesNet															
	SVM															
C4	.5H				A							A	A	A		A
Clus-	HMC				A							Δ	A	A		A

Tabela C.7: Análise dos Resultados da Medida F1 Hierárquica no Segundo Nível das Hierarquias

				HMC-					HMC-					HMC-		
		KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM
	KNN															
	C4.5				A											A
HMC-BR	Ripper															
	BayesNet															
	SVM															
	KNN				A											A
	C4.5				A									Δ		A
$\operatorname{HMC-LP}$	Ripper	A		A	A	A					A	A	Δ	A	A	A
	BayesNet				A											Δ
	SVM															
	KNN															
	C4.5				Δ											
HMC-CT	Ripper															
	BayesNet															
	SVM															
C4	.5H			Δ	A	A					Δ	A		A	A	A
Clus-	HMC				A							A		A	A	A

Tabela C.8: Análise dos Resultados da Medida F1 Hierárquica no Terceiro Nível das Hierarquias

				HMC-					HMC-					HMC-		
		KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM
	KNN									-						
	C4.5															
HMC-BR	Ripper															
	BayesNet															
	SVM															
	KNN															
	C4.5				A											
HMC-LP	Ripper	A	A	A	A	A	A			A	A				Δ	A
	BayesNet															
	SVM															
	KNN				A											
	C4.5	A		A	A					Δ	Δ					A
HMC-CT					A											
	BayesNet				Δ											
	SVM															
C4	.5H	A		A	A					A	Δ					A
Clus-	HMC	A		A	A											

Tabela C.9: Análise dos Resultados da Medida F1 Hierárquica no Quarto Nível das Hierarquias

				HMC-					HMC-					HMC-		
		KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM
	KNN															
	C4.5															
HMC-BR	Ripper															
	BayesNet															
	SVM															
	KNN															
	C4.5				Δ											
HMC-LP		A	A	A	A	A	A	Δ		A	A					
	BayesNet															
	SVM															
	KNN	Δ		A	A	Δ										
	C4.5	A	A	A	A	A	A			A	A					
HMC-CT		A	Δ	A	A	A	Δ			A	A					
	BayesNet				A											
	SVM				A											
C4.	.5H	A	Δ	A	A	A				A	Δ					
Clus-	HMC	A		A	A	A										

C.3 Análise dos Resultados da Medida Micro F1 Hierárquica

No primeiro nível das hierarquias, conforme pode ser observado na Tabela C.10, os melhores resultados foram obtidos pela técnica HMC-CT. Os resultados dos testes estatísticos mostraram que houve diferenças estatisticamente significantes na comparação dos resultados dos algoritmos da técnica HMC-CT com a maioria dos algoritmos utilizados nas técnicas HMC-BR e HMC-CT, e também na comparação com os resultados das técnicas C4.5H e Clus-HMC.

A partir do segundo nível das hierarquias (Tabelas C.11, C.12 e C.13), ganham destaque os desempenhos das técnicas HMC-BR e HMC-LP, ficando o desempenho da técnica HMC-CT inferior. Nos níveis três e quatro, inclusive, as técnicas C4.5H e Clus-HMC obtiveram melhores resultados que alguns algoritmos utilizados na técnica HMC-CT, mostrando diferenças estatisticamente significantes. Algoritmos utilizados nas técnicas HMC-BR e HMC-LP também apresentaram resultados estatisticamente superiores aos resultados da técnica HMC-CT.

Tabela C.10: Análise dos Resultados da Medida Micro F1 Hierárquica no Primeiro Nível das Hierarquias

				HMC-					HMC-					HMC-		
		KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM
	KNN													A		A
	C4.5				A								A	A		A
HMC-BR														A		A
	BayesNet															
	SVM													A		Δ
	KNN				A							Δ	A	A		A
	C4.5				Δ								A	A		A
HMC-LP					A						A	A	A	A	A	A
	BayesNet													A		A
	SVM															
	KNN															
	C4.5															
HMC-CT																
	BayesNet															
	SVM															
C4.	5H				A							A	A	A		A
Clus-	HMC				A							Δ	A	A		A

Tabela C.11: Análise dos Resultados da Medida Micro F1 Hierárquica no Segundo Nível das Hierarquias

				HMC-					HMC-					HMC-		
		KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM
	KNN															
	C4.5							Δ								
HMC-BR								A								
	BayesNet															
	SVM															
	KNN															
	C4.5															
HMC-LP		Δ			A	A	A	A			A	Δ				A
	BayesNet															
	SVM															
	KNN															
	C4.5				A	A	A	A			A					A
HMC-CT					A	Δ	A	A			A					
	BayesNet				A			A			A					
	SVM															
C4.	5H				Δ			A			A					
Clus-	HMC															

Tabela C.12: Análise dos Resultados da Medida Micro F1 Hierárquica no Terceiro Nível das Hierarquias

				HMC-					HMC-					HMC-		
		KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM
	KNN															
	C4.5															
HMC-BR	Ripper															
	BayesNet															
	SVM															
	KNN															
	C4.5															
HMC-LP					Δ	Δ	A	A			A					
	BayesNet															
	SVM															
	KNN				Δ	Δ	A	A			A					
	C4.5	A	A	Δ	A	A	A	A		A	A					
HMC-CT		A	A		A	A	A	A		A	A					
	BayesNet	Δ			A	A	A	A		Δ	A					
	SVM						A	A			A					
C4.	.5H						Δ	A			Δ					
Clus-	HMC												▼	∇		

Tabela C.13: Análise dos Resultados da Medida Micro F1 Hierárquica no Quarto Nível das Hierarquias

				HMC-	BR				HMC-	·LP				HMC-	CT	
		KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM
	KNN															
	C4.5															
HMC-BR	Ripper															
	BayesNet															
	SVM															
	KNN															
	C4.5															
HMC-LP							A	A			A					
	BayesNet															
	SVM															
	KNN				A	A	A	A			A					
	C4.5	A		A	A											
HMC-CT		A	A	Δ	A	A	A	A		A	A					
	BayesNet	Δ			A	A	A	A			A					
	SVM					Δ	A	A			A					
C4.	.5H						A	A			A					
Clus-	HMC												▼	∇		

C.4 Análise dos Resultados da Medida Micro F1 Hierárquica Ponderada

Nos resultados da medida micro F1 hierárquica ponderada, foram obtidos resultados semelhantes aos obtidos pela medida micro F1 hierárquica. No primeiro nível das hierarquias (Tabela C.14), a técnica HMC-CT obteve melhores resultados, sendo que a comparação com as outras técnicas mostrou resultados estatisticamente significantes.

No segundo, terceiro e quarto níveis hierárquicos (Tabelas C.15, C.16 e C.17), os desempenhos das técnicas HMC-BR e HMC-LP foram superiores aos da técnica HMC-CT. Esses melhores desempenhos destacaram-se principalmente nos níveis três e quatro das hierarquias. Na maioria dos casos, houve diferenças estatisticamente significantes na comparação das técnicas.

Tabela C.14: Análise dos Resultados da Medida Micro F1 Hierárquica Ponderada no Primeiro Nível das Hierarquias

				HMC-					HMC-					HMC-		
		KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM
	KNN													A		A
	C4.5				A								A	A		A
HMC-BR	Ripper													A		A
	BayesNet															
	SVM													A		Δ
	KNN				A							Δ	A	A		A
	C4.5				Δ								A	A		A
$\operatorname{HMC-LP}$					A						A	A	A	A	A	A
	BayesNet													A		A
	SVM															
	KNN															
	C4.5															
HMC-CT	Ripper															
	BayesNet															
	SVM															
C4	.5H				A							A	A	A		A
Clus-	HMC				A							Δ	A	A		A

Tabela C.15: Análise dos Resultados da Medida Micro F1 Hierárquica Ponderada no Segundo Nível das Hierarquias

				HMC-	·BR				HMC-	-LP				HMC-	CT	
		KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM
	KNN															
	C4.5				A						A					
HMC-BR																
	BayesNet															
	SVM															
	KNN				A						A					
	C4.5				A						A					
$\operatorname{HMC-LP}$																
	BayesNet															
	SVM															
	KNN															
	C4.5			A	A	A			A		A			Δ		A
HMC-CT																
	BayesNet				A				Δ		A					Δ
	SVM															
C4.	5H	Δ		A	A	A			A		A			Δ		A
Clus-	HMC				A	A			A		A					A

Tabela C.16: Análise dos Resultados da Medida Micro F1 Hierárquica Ponderada no

Terceiro Nível das Hierarquias

rercent	7 111101	Table 1	11101	HMC-			Ι		HMC-	LD		1		HMC-	CT	
		*****	~			OTT	T / N / N /	~			OT TO E	*****	1012			LOX 73. f
		KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM
	KNN															
	C4.5															
HMC-BR	Ripper															
	BayesNet															
	SVM															
	KNN															
	C4.5															
HMC-LP	Ripper															
	BayesNet															
	SVM															
	KNN	Δ		A	A	A			A	Δ	A					
	C4.5	A		A	A	A	A	A	A	A	A					
HMC-CT	Ripper			A	A	A			A		A					
	BayesNet	A		A	A	A			A	A	A					
	SVM				A	Δ			A		A					
C4.	5H			A	A	A			A		A					
Clus-	HMC				A				Δ		A					

Tabela C.17: Análise dos Resultados da Medida Micro F1 Hierárquica Ponderada no

Quarto Nível das Hierarquias

				HMC-					HMC-					HMC-		
		KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM
	KNN															
	C4.5															
HMC-BR	Ripper															
	BayesNet															
	SVM															
	KNN															
	C4.5															
$\operatorname{HMC-LP}$	Ripper															
	BayesNet															
	SVM															
	KNN	Δ		A	A	A			A	Δ	A					
	C4.5	A	Δ	A												
HMC-CT	Ripper	A		A	A	A			A	A	A					
	BayesNet	A		A	A	A			A	Δ	A					
	SVM			Δ	A	A			A		A					
C4.	5H	A	A	A			A		A							
Clus-	HMC				A	Δ					•					

C.5 Análise dos Resultados da Medida Macro F1 Hierárquica

No primeiro nível das hierarquias (Tabela C.18), os melhores resultados foram obtidos pela técnica HMC-CT, mostrando diferenças estatisticamente superiores na comparação com as outras técnicas, principalmente na comparação com as técnicas C4.5H e Clus-HMC.

No segundo, terceiro e quarto níveis das hierarquias (Tabelas C.19, C.20 e C.21), não foram detectadas muitas diferenças estatisticamente significantes nos resultados das técnicas. Apenas alguns resultados das técnicas HMC-BR e HMC-CT apresentaram melhores resultados, com diferenças estatisticamente significantes, na comparação com os resultados das outras técnicas.

Tabela C.18: Análise dos Resultados da Medida Macro F1 Hierárquica no Primeiro Nível das Hierarquias

				HMC-					HMC-					HMC-		
		KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM
	KNN				Δ											
	C4.5				A								A	A	Δ	A
HMC-BR	* *				A									A		A
	BayesNet															
	SVM				Δ											
	KNN				A									A		A
	C4.5				A								A	A	Δ	A
HMC-LP					A					Δ		A	A	A	A	A
	BayesNet															
	SVM															
	KNN															
	C4.5															
HMC-CT																
	BayesNet															
	SVM															
C4.	C4.5H				A							A	A	A	A	A
Clus-	HMC				A							Δ	A	A	A	A

Tabela C.19: Análise dos Resultados da Medida Macro F1 Hierárquica no Segundo Nível das Hierarquias

445 1110	1			HMC-	DD				HMC-	I D				HMC-	CT	
		*****	1012			OX 73. 6	*****				OX 13.6	*****				lov va
		KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM
	KNN															
	C4.5				A											A
HMC-BR																
	BayesNet															
	SVM															
	KNN															
	C4.5				A									Δ		A
$\operatorname{HMC-LP}$	Ripper				A	A					A	A	A	A	A	A
	BayesNet															
	SVM															
	KNN															
	C4.5															
HMC-CT	Ripper															
	BayesNet															
	SVM															
C4.	5H				A								Δ	A		A
Clus-	HMC				A									Δ		A

Tabela C.20: Análise dos Resultados da Medida Macro F1 Hierárquica no Terceiro Nível

das Hierarquias

				HMC-					HMC-					HMC-		
		KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVN
	KNN															
	C4.5				Δ											Δ
HMC-BR																
	BayesNet															
	SVM															
	KNN															
	C4.5				Δ											A
HMC-LP	Ripper	Δ			A	A					Δ	Δ	Δ	A	Δ	A
	BayesNet															
	SVM															
	KNN															
	C4.5															
HMC-CT																
	BayesNet															
	SVM															
C4.	5H				A	Δ						Δ		A		A
Clus-	HMC				A											A

Tabela C.21: Análise dos Resultados da Medida Macro F1 Hierárquica no Quarto Nível das Hierarquias

				HMC-					HMC-					HMC-		
		KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM
	KNN															
	C4.5															
HMC-BR	Ripper															
	BayesNet															
	SVM															
	KNN															
	C4.5															
HMC-LP	Ripper	Δ			A	A										A
	BayesNet															
	SVM															
	KNN															
	C4.5															
HMC-CT	Ripper															
	BayesNet															
	SVM															
C4.	.5H	Δ			A	A										A
Clus-	HMC					Δ										Δ

C.6 Análise dos Resultados da Medida Macro F1 Hierárquica Ponderada

De acordo com os resultados no primeiro nível das hierarquias (Tabela C.22), o melhor desempenho foi obtido pela técnica HMC-CT. Os testes estatísticos apresentaram diferenças estatisticamente significantes na comparação com alguns dos algoritmos utilizados nas técnicas HMC-BR e HMC-LP. Na comparação com as técnicas C4.5H e Clus-HMC, todos os algoritmos da técnica HMC-CT apresentaram resultados superiores, com diferenças estatisticamente significantes.

A partir do segundo nível das hierarquias (Tabelas C.23, C.24 e C.25), começam a ganhar destaque os desempenhos das técnicas HMC-BR e HMC-LP, principalmente da técnica HMC-LP. Os testes estatísticos, porém, detectaram diferenças estatisticamente significantes em apenas alguns resultados na comparação das técnicas, principalmente no terceiro nível das hierarquias.

Tabela C.22: Análise dos Resultados da Medida Macro F1 Hierárquica Ponderada no Primeiro Nível das Hierarquias

				HMC-					HMC-					HMC-		
		KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM
	KNN				Δ											
	C4.5				A								A	A	Δ	A
HMC-BR	Ripper				A									A		A
	BayesNet															
	SVM				Δ											
	KNN				A									A		A
	C4.5				A								A	A	Δ	A
$\operatorname{HMC-LP}$	Ripper				A					Δ		A	A	A	A	A
	BayesNet															
	SVM															
	KNN															
	C4.5															
HMC-CT	Ripper															
	BayesNet															
	SVM															
C4.	.5H				A							A	A	A	A	A
Clus-	HMC				A							Δ	A	A	A	A

Tabela C.23: Análise dos Resultados da Medida Macro F1 Hierárquica Ponderada no

Segundo Nível das Hierarquias

				HMC-					HMC-					HMC-		
		KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM
	KNN								A							
	C4.5								A		A			Δ		Δ
HMC-BR																
	BayesNet															
	SVM															
	KNN								A		Δ					
	C4.5				A				A	Δ	A			A	Δ	A
HMC-LP																
	BayesNet															
	SVM															
	KNN															
	C4.5								A							
HMC-CT																
	BayesNet															
	SVM															
C4.				Δ	A				A	A	A			A	A	A
Clus-	HMC			Δ	A				A	A	A	Δ		A	A	A

Tabela C.24: Análise dos Resultados da Medida Macro F1 Hierárquica Ponderada no

Terceiro Nível das Hierarquias

	111101	HMC-BR						HMC-LP						HMC-CT					
		KNN	C4.5			SVM	KNN	C4.5		BayesNet	SVM	KNN	C4.5			SVM			
HMC-BR	KNN								A		A				Δ				
	C4.5								A		A				Δ				
	Ripper																		
	BayesNet																		
	SVM																		
HMC-LP	KNN								Δ										
	C4.5								A		A								
	Ripper																		
	BayesNet																		
	SVM																		
HMC-CT	KNN																		
	C4.5				A	A			A	Δ	A			Δ	A	A			
	Ripper																		
	BayesNet																		
	SVM																		
C4.5H					Δ	A			A		A			Δ	A	Δ			
Clus-HMC									A		A				A				

Tabela C.25: Análise dos Resultados da Medida Macro F1 Hierárquica Ponderada no

Quarto Nível das Hierarquias

		HMC-BR						HMC-LP					HMC-CT					
		KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM	KNN	C4.5	Ripper	BayesNet	SVM		
HMC-BR	KNN								A		A				A			
	C4.5								A		Δ							
	Ripper								Δ									
	BayesNet								Δ									
	SVM																	
HMC-LP	KNN																	
	C4.5								A		Δ							
	BayesNet																	
	SVM																	
HMC-CT	KNN																	
	C4.5					A			A		A				A	A		
	Ripper																	
	BayesNet																	
	SVM																	
C4.5H						Δ			A		A				A			
Clus-HMC																		