

**SIDNEI NICOLI**

**Método adaptativo para a interpretação de medições oriundas de sistemas  
elétricos**

São Paulo

2017

**SIDNEI NICOLI**

**Método adaptativo para a interpretação de medições oriundas de sistemas elétricos**

Tese apresentada à Escola Politécnica da  
Universidade de São Paulo para a obtenção do  
título de Doutor em Ciências

São Paulo

2017

**SIDNEI NICOLI**

**Método adaptativo para a interpretação de medições oriundas de sistemas elétricos**

Tese apresentada à Escola Politécnica da  
Universidade de São Paulo para a obtenção do  
título de Doutor em Ciências

Área de concentração: Sistemas de Potência

Orientador: Prof. Dr. José Antonio Jardini

São Paulo

2017

## FICHA CATALOGRÁFICA

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, \_\_\_\_\_ de \_\_\_\_\_ de \_\_\_\_\_

Assinatura do autor: \_\_\_\_\_

Assinatura do orientador: \_\_\_\_\_

Nicoli, Sidnei

Método Adaptativo para a Identificação de Medições Oriundas de Sistemas Elétricos / S. Nicoli – versão corr. -- São Paulo, 2017. 228 p.

Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Energia e Automação Elétrica.

1.Análise de séries temporais 2.Monitoramento ambiental (Influencias) 3.Monitoramento ambiental (Medição) 4.Novidades (Medição) I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Energia e Automação Elétricas

Dedico este trabalho a minha família.

## AGRADECIMENTOS

Agradeço ao meu orientador, Prof. José Antonio Jardini, pela oportunidade, e a Luiz Carlos Magrini pelas contribuições a este trabalho.

Aos meus pais, José e Josefa Nicoli, presentes em minha memória, pelo incentivo e o sacrifício que dedicaram a minha formação e a minha vida.

Ao meu amigo, Rômulo Gonçalves Lins, pelo seu apoio e a sua participação em artigos que envolveram os conhecimentos desta tese.

Aos amigos de trabalho, Alex Lopes, Milana dos Santos, Ricardo Leon, Fabiana Silva, Marcos Bassini, Gerson Saiki, Diogo Alves, Luciano Ogiboski, Osvaldo Rein Jr, Ronaldo Casolari, Maurício Jardini, Ferdinando Crispino, Paula Kayano, Igor Matsuo, Thales Sousa, Patrícia Silveira, Patrícia Albertini e Monica Brasão, com os quais compartilhei dias muito agradáveis.

Agradeço à Ana Maria Badiali por seu apoio às questões relacionadas à formatação deste texto.

Agradeço à EDP e a CESP, que financiaram os projetos de pesquisas relacionados a esta tese.

Agradeço aos amigos da Estácio, Edson Russo, Ana Carolina Russo e Walter Sarantini, pelo apoio recebido durante o desenvolvimento da tese.

À ANEEL que, por meio de seu programa de P&D, permitiu os projetos de pesquisas que contribuíram para este trabalho.

A todos que, direta ou indiretamente, colaboraram com este trabalho, mas não puderam ser citados.

Agradeço a Deus acima de tudo, que está acima de todas as coisas, que permite que eu caminhe com minhas próprias pernas quando tudo está fácil, e que me carrega quando tudo está difícil.

## RESUMO

Esta pesquisa aborda um método adaptativo baseado em regras e aplicado a sistemas de informação não supervisionados de fluxo contínuo com a utilização de técnicas inteligentes, utilizando séries históricas de medições. O objetivo principal é identificar medições que podem ser categorizadas de diversas formas, tais como novidades, *outliers*, anomalias, dentre outras, apesar do método apresentado abordar somente monitoramento e autonomia, o mesmo pode ser utilizado para controle de ações desde que esteja integrado aos dispositivos de campo necessários para atuação de um processo. Para tanto, utiliza uma representação do conhecimento adequada, critérios de busca inteligente, inferência e critérios para a aprendizagem, possibilitando um processo de melhoria contínua. Os estudos realizados com medições possibilitaram ao estabelecimento de processos que conduziram ao surgimento de um Método Adaptativo para a Interpretação de Medições aplicável a sistemas inteligentes e do Método Complementar, capaz de auxiliar na interpretação de resultados obtidos por outros métodos já estabelecidos para identificação de anomalias. Para esse fim, foram consideradas: uma técnica adaptativa, a importância do ambiente de influência relativo ao ponto de medição e a utilização das novidades como referência às mudanças que ocorrem em um ambiente.

**Palavras-chave:** Interpretação de medições. Séries temporais. Método adaptativo. Ambiente de influência. Novidades (Medição). Anomalia.

## ABSTRACT

This research addresses a rules-based adaptive method applied to unsupervised continuous flow information systems using intelligent techniques using historical series of measurements. The main goal is to identify measurements that can be categorized in different ways, such as novelties, outliers, anomalies, among others, although the presented method only approach monitoring and autonomy, it can be used to control actions since it is integrated with the devices required to perform a process control. To do so, it uses a representation of adequate knowledge, intelligent search criteria, inference and criteria for learning, enabling a process of continuous improvement. The research conducted with real measurements allowed the development of optimized computational routines that guided the development of a new method based on adaptive techniques for measurements interpretation and the Complementary Method, able to assist the interpretation of results obtained by other methods already established to identify non-standard. To this end, we considered: an adaptive technique, the importance of the influence environment relative to the measurement point and the use of the novelties as a reference to the changes that occur in an environment.

**Keywords:** Measurements Interpretation. Time Series. Adaptive method. Environmental influence. Novelties (Measurement). Anomalies.



## LISTA DE FIGURAS

Figura 1 - Ilustração do Método MAIM de forma simplificada.....	2
Figura 2 - Ilustração do Método Complementar de forma simplificada .....	3
Figura 3 - Formas de representação visual das tabelas de decisão tradicional.....	18
Figura 4 - Estrutura básica de uma Tabela de Decisões Adaptativa .....	23
Figura 5- Estrutura simplificada de uma Tabela de Decisões Adaptativa.....	24
Figura 6 - Sensor de pressão piezoresistivo .....	35
Figura 7 - Sensor MENS: (a) não encapsulado, (b) encapsulado.....	36
Figura 8 - Sensor tipo MENS código MPXV7002.....	37
Figura 9 – Medidor de Vazão instalado.....	38
Figura 10 – Medidor de vazão e seus componentes .....	38
Figura 11 - Exemplo de parte adaptativa de uma TDA.....	44
Figura 12 - Exemplo de uma TDA utilizada .....	45
Figura 13- Curva da Banheira .....	54
Figura 14 - Curva tempo resposta .....	55

Figura 15 - Curva com ambientes reconhecidos separados por linhas tracejadas (valores em mm).....	56
Figura 16 - Medições realizadas entre 2001 e 2010, em períodos de quinze dias em um eixo	59
Figura 17 – Quantidade de ocorrências x Valor Medições individuais em ( $10^{-4}$ mm).....	60
Figura 18 - Medições em ( $10^{-9}$ mm) obtidas de 01/08/2014 às 13h50min à 01/10/2014 às 10h04min, em períodos de 15min para o medidor de vazão. ....	62
Figura 19 - Quantidade de ocorrências x Valor Medições individuais em ( $10^{-9}$ mm). ....	63
Figura 20 - Medições realizadas de 29/08/2014 às 16h07min à 01/10/2014 às 10h04min,.....	64
Figura 21 - Quantidade de ocorrências x Valor Medições individuais (em $10^{-6}$ m). ....	65
Figura 22 – Resumo das validações .....	68
Figura 23 - Arquitetura do sistema computacional utilizado pelo método.....	70
Figura 24 – Modelo de dados para Tabela de Decisões adaptativa e convencional.....	73
Figura 25 - Tabela de decisões simples .....	76
Figura 26 - Informações existentes no módulo adaptativo.....	77
Figura 27 - Interface envolvendo Funções, parâmetros, variáveis e geradores.....	78
Figura 28 - Autômato referente a uma calculadora simples.....	78

Figura 29 – Aumento do número de operações da calculadora simples. ....	79
Figura 30 - Exemplo de Operação da tabela de decisões para a Calculadora simples.....	80
Figura 31 - Fluxo de dados do Algoritmo proposto .....	82
Figura 32- Exemplo esquemático de aplicação com a criação de uma nova regra .....	85
Figura 33 - Esquema básico do Método Complementar .....	87
Figura 34 - Exemplo de uma representação entre dois nodos e sua informação mínima.....	89
Figura 35– Diagrama Ishikawa (Espinha de Peixe / Causa-Efeito). ....	90
Figura 36 - Exemplo de diagrama de Causa Efeito simplificado para leitura não conforme...	90
Figura 37 – Interação simplificada do Método principal com o Complementar.....	92
Figura 38 - Trecho de um mapa tendo como centro o instrumento de medição .....	95
Figura 39 - Modelo de dados desenvolvido para o Método Complementar. ....	98
Figura 40 – Esquema da instalação utilizada.....	103
Figura 41- Serie temporal obtida pelo Piezômetro.....	104
Figura 42- Pontos estudados nesta aplicação no piezômetro .....	105
Figura 43- Histograma com os dados da série temporal .....	106

Figura 44 – Parte de um mapa conceitual tendo como elemento central o sensor .....	108
Figura 45 – Serie Temporal do Medidor Triortogonal MT-V2-3 eixo Um.....	113
Figura 46 - Informações utilizadas na Tabela de decisões .....	114
Figura 47 - Tabela de decisões da aplicação com as funções relacionadas às ações .....	116
Figura 48 - Série temporal de um eixo do sensor triortogonal instalado no vertedouro .....	120
Figura 49 - Tabela de decisões da aplicação .....	121
Figura 50 – Série Temporal do sensor Piezométrico PZ1762 .....	125
Figura 51 - T.D simplificada aplicada ao sensor Piezométrico com funções relacionadas às ações .....	127
Figura 52 - Tabela de Dados Subjacente .....	132
Figura 53 – Série histórica do Extensômetro de Haste, com medições em mm .....	133
Figura 54- Registros das sequencias analisadas do Extensômetro de Haste .....	134
Figura 55 - Série histórica do Medidor Triortogonal em um de seus eixos, com medições em $10^{-2}$ mm.....	135
Figura 56 - Registros das sequencias analisadas de um eixo do Medidor Triortogonal .....	135
Figura 57 - Série histórica do Medidor de Vazão Triangular, com medições em mm.....	136

Figura 58 - Registros das sequencias analisadas do Medidor de vazão triangular.....	136
Figura 59 - Série histórica do Piezômetro, em ( $10^{-3}$ m) .....	137
Figura 60 - Registros das sequencias analisadas do Piezomêtro.....	138
Figura 61- Árvore de decisão entre duas alternativas de altura de Ensecadeira.....	158
Figura 62 – Árvore de decisões .....	159
Figura 63 - Exemplo de representação de uma árvore de decisão adaptativa .....	161
Figura 64 – Arquitetura de Inferência Fuzzy.....	163
Figura 65 - Árvore de decisão gerada com o método apresentado por Maletzke .....	166
Figura 66 - Transformações de uma série temporal em uma palavra.....	176
Figura 67 - Fluxograma com os passos utilizados no SAX.....	176
Figura 68 - Representação gráfica da série original (valores adimensionais) .....	177
Figura 69 - Representação gráfica com a série normalizada (valores adimensionais).....	178
Figura 70 - Representação gráfica da aplicação com o SAX, (valores adimensionais) .....	178
Figura 71 - Estrutura obtida do exemplo .....	179
Figura 72- Medições referentes ao Extensômetro de Haste (em mm) e ao Medidor Triortogonal ( $10^{-2}$ mm).....	185

Figura 73 - Medições do Medidor de Vazões Triangular (em mm) e Piezômetro (em $10^{-3}$ m) .....	186
Figura 74 – Fluxograma básico do método de Basha e Ameen. ....	190
Figura 75 - Dados do exemplo, considerando uma série experimental (valores adimensionais). .....	190
Figura 76 - Dados do exemplo normalizados para cada grupo (valores adimensionais) .....	191
Figura 77 - Resultados do calculo da distância Euclidiana .....	191
Figura 78 - Esquema do Sistema computacional OLINDDA .....	199

## LISTA DE TABELAS

Tabela 1- Relação Quantidade de grupos formados x Arranjos possíveis .....	59
Tabela 2 - Distribuição de todas as novidades considerando todos os agrupamentos, .....	61
Tabela 3 - Relação Quantidade de grupos formados x Arranjos possíveis .....	62
Tabela 4- Distribuição de todas as novidades considerando todos os agrupamentos, .....	64
Tabela 5 - Relação Quantidade de grupos formados x Arranjos possíveis .....	65
Tabela 6 - Distribuição de todas as novidades considerando todos os agrupamentos, .....	66
Tabela 7- Exemplo de Trecho da tabela de nebulosidade .....	109
Tabela 8 – Exemplo de Trecho da tabela de Precipitação .....	109
Tabela 9 - Trecho da série temporal com os registros das sequencias 181, 182 e 183. ....	117
Tabela 10 - Registros da série Histórica com a indicação da mudança da identificação do ambiente 3 para o 4.....	117
Tabela 11 – Cinco últimos registros medidos após as sequencia 181, 182 e 183. ....	118
Tabela 12 Grupos registrados durante cada sequencia.....	118
Tabela 13 – Registros da série temporal das sequencias 863, 864 e 865. ....	121
Tabela 14 - Registros dos 5 últimos valores após as sequencias 863, 864 e 865.....	122

Tabela 15 - Grupos registrados durante cada sequencia.....	122
Tabela 16 - Dados lidos pelo piezômetro para processamento .....	126
Tabela 17 - Distribuição gaussiana em regiões equiprováveis.....	180
Tabela 18 - Distâncias entre símbolos do alfabeto .....	181
Tabela 19 - Resultados do cálculo simplificado da distância entre os segmentos .....	182
Tabela 20 - segmentos correspondentes a maior distância por sensor. ....	187



## LISTA DE QUADROS

Quadro 1 Características marcantes .....	31
Quadro 2- Categoria de causas e falhas.....	33
Quadro 3 - Materiais defeituosos e Características observáveis .....	33
Quadro 4 - Ocorrências genéricas e Características observáveis .....	34
Quadro 5 - Análise enfocando a capacidade de adaptação dos métodos.....	42
Quadro 6 – Exemplos de possíveis regras .....	74
Quadro 7 - Categorias de conhecimentos relacionadas ao instrumento de medição.....	93
Quadro 8 – Relação sequencial das regras utilizadas .....	110
Quadro 9 – Resultados após o processamento das sequencias de 181 a 183 .....	119
Quadro 10 - Resultados após o processamento de cada sequencia de 863 a 865.....	123
Quadro 11- Resultado do Processamento por sequencia.....	128
Quadro 12 – Resultados obtidos nos testes com HOT SAX modificado.....	132
Quadro 13 – Resumo das indicações de anomalias.....	138
Quadro 14- Exemplos de técnicas computacionais x Inspiração na natureza.....	156
Quadro 15 - Característica da Lógica Fuzzy: vantagens e desvantagens .....	162

Quadro 16 – Exemplo de transformação de entrada analógica em saída digital.....	171
---	-----

## LISTA DE ALGORITMOS

Algoritmo 1 - Etapas do pseudocódigo para o método MAIM (considerando um ciclo) .....	83
Algoritmo 2 Pseudocódigo envolvendo procedimentos TDA.....	84
Algoritmo 3 - Passos operacionais utilizados pelo Método Complementar.....	91

## LISTA DE ABREVIATURAS E SIGLAS

AA	Autômatos Adaptativos
ADA	Árvore de Decisões Adaptativa
AF	Autômato Finito
ERP	<i>Enterprise Resource Planning</i>
ETX	<i>Expected Transmission Count</i>
IAD	Indução por Árvore de Decisão
IOT	<i>Internet of Things</i>
ITIL	<i>Information Technology Infrastructure Library</i>
MAIM	Método Adaptativo para Interpretação de Medições
MEMS	<i>Micro Eletrical Mechanical Systems</i>
MDA	Mineração de Dados Adaptativa
PAA	<i>Piecewise Aggregate Approximation</i>
PMBOK	<i>Project Management Body of Knowledge</i>
POC	<i>Proof of Concept</i>
PSD	<i>Position Sensitive Detector</i>
RSSF	Rede de Sensores Sem Fio
SAX	<i>Symbolic Aggregate Approximation</i>
SCADA	<i>Supervisory Control and Data Acquisition</i>
SQL	<i>Structured Query Language</i>
TD	Tabela de Decisão
TDA	Tabela de Decisão Adaptativa
VBA	<i>Visual Basic for Applications</i>
WAT	<i>Wavelet and Augmented Trie</i>

## SUMÁRIO

<b>1</b>	<b>CONSIDERAÇÕES GERAIS .....</b>	<b>1</b>
1.1	INTRODUÇÃO.....	1
1.2	MÉTODO ADAPTATIVO PARA INTERPRETAÇÃO DE MEDIÇÕES (MAIM) .....	1
1.3	MÉTODO COMPLEMENTAR.....	2
1.4	CONTEXTUALIZAÇÃO .....	4
1.5	OBJETIVOS.....	6
1.6	METODOLOGIA.....	7
1.7	ORGANIZAÇÃO DO TEXTO .....	9
<b>2</b>	<b>ESTADO DA ARTE .....</b>	<b>11</b>
2.1	SÉRIES HISTÓRICAS .....	11
2.2	MÉTODOS ADAPTATIVOS.....	13
2.2.1	Dispositivo ou processo computacional .....	14
2.2.2	Dispositivo não adaptativo guiado por regras .....	15
2.2.3	Dispositivo adaptativo guiado por regras .....	16
2.3	TÉCNICAS UTILIZADAS EM MÉTODOS INTELIGENTES .....	17
2.3.1	Tabela de decisões.....	18
2.3.2	Tabela de Decisões Adaptativa .....	21
2.3.3	Regras de Associação.....	24
2.3.4	Sistemas Especialistas .....	27
2.4	MONITORAMENTO DE BARRAGENS.....	29
2.4.1	Exemplo de informações para Base de Conhecimentos .....	32

2.4.2	Instrumentos analisados nesta pesquisa.....	34
2.4.2.1	Piezômetro de tubo fechado.....	35
2.4.2.2	Piezômetro de tubo aberto .....	36
2.4.2.3	Medidor de Vazão .....	36
2.4.2.4	Medidor de Junta Triortogonal .....	39
2.4.2.5	Extensômetro de Haste .....	39
2.5	CONSIDERAÇÕES FINAIS .....	40
<b>3</b>	<b>ANÁLISE DOS ESTUDOS E DESENVOLVIMENTOS REALIZADOS.....</b>	<b>41</b>
3.1	ESTUDOS QUE NÃO ENVOLVEM TDA.....	41
3.2	ESTUDOS DE ARTIGOS QUE ENVOLVEM TDA.....	42
3.2.1	Uso de métricas de roteamento através de Tabela de Decisão Adaptativa para redes de sensores sem fio	43
3.2.2	Aplicação de Tabelas de Decisão Adaptativas em sistemas de controle de crescimento de plantas ..	44
3.2.3	Operação sustentável de sistemas de ar condicionado usando Tecnologia Adaptativa .....	45
3.2.4	Modelo de um sistema de conservação de alimentos baseado na IoT-A e seleção de elementos para Tabela de Decisão Adaptativa.....	47
3.2.5	Mecanização da aprendizagem com dispositivos adaptativos: conceitos e aplicação .....	48
3.2.6	Descoberta de padrões em bases de dados utilizando Técnicas Adaptativas .....	49
3.3	CONSIDERAÇÕES FINAIS .....	50
<b>4</b>	<b>MÉTODOS DESENVOLVIDOS .....</b>	<b>51</b>
4.1	MÉTODO ADAPTATIVO PARA A INTERPRETAÇÃO DE MEDIÇÕES - MAIM	51
4.1.1	Conceitos utilizados de Novidade e Ambiente .....	52
4.1.1.1	Novidade.....	52
4.1.1.2	Ambiente .....	53

4.1.2	Prova de conceitos envolvendo agrupamentos de medições e novidades .....	56
4.1.2.1	Provas de conceitos com as medições de um eixo Triortogonal.....	58
4.1.2.2	Provas de conceitos com o medidor de vazão .....	61
4.1.2.3	Prova de conceito com Piezômetro com alimentação de energia por painel solar.....	64
4.1.2.4	Considerações finais sobre as provas de conceitos sobre agrupamentos de medições e novidades 66	
4.1.3	Provas de conceitos envolvendo estabilidade e ambiente .....	67
4.1.3.1	Considerações sobre as provas de conceitos envolvendo estabilidade e ambiente.....	68
4.1.4	Requisitos .....	69
4.1.5	Arquitetura do método MAIM .....	70
4.1.6	Considerações sobre as regras .....	71
4.1.6.1	Resumos de informações .....	72
4.1.6.2	Tabelas de decisões .....	72
4.1.6.3	Exemplos de possíveis regras .....	74
4.1.7	Prova de conceito com relação ao módulo adaptativo.....	75
4.1.7.1	Exemplo de prova de conceito com tabela de decisões simples .....	76
4.1.7.2	Exemplo de Prova de conceitos com tabela de decisões adaptativa .....	77
4.1.8	Fluxo de dados e pseudocódigo.....	80
4.2	<b>MÉTODO COMPLEMENTAR</b> .....	85
4.2.1	Considerações teóricas .....	88
4.2.1.1	Mapas Conceituais.....	88
4.2.1.2	Diagrama Causa-Efeito.....	89
4.2.2	Método Complementar propriamente dito .....	91
4.2.2.1	Levantamento dos dados .....	92

4.2.2.2	Categorização das informações obtidas .....	93
4.2.2.3	Mapeamento das influências e pertinências.....	94
4.2.2.4	Regras .....	96
4.2.2.5	Receber Mensagens e Alertas dos Métodos Básicos .....	97
4.2.2.6	Base de Conhecimento .....	97
4.2.2.7	Recepção das Mensagens e Alertas dos Métodos Básicos .....	98
4.2.2.8	Análise e Inferência.....	99
4.2.2.9	Emissão de novas Mensagens e Alertas .....	99
4.2.3	Estudos de Casos .....	99
4.2.3.1	Estudo de caso Método HOT SAX Adaptado .....	99
4.2.3.2	Estudo de caso envolvendo a pesquisa de Schwabacher, Oza e Matthews.....	101
4.3	EXEMPLO DE APLICAÇÃO .....	102
4.4	CONSIDERAÇÕES FINAIS SOBRE O MÉTODO COMPLEMENTAR .....	110
<b>5</b>	<b>APLICAÇÕES DO MÉTODO MAIM .....</b>	<b>112</b>
5.1	INFORMAÇÕES GERAIS DA APLICAÇÃO QUE ENVOLVE O SENSOR TRIORTOGONAL.....	112
5.1.1	Informações sobre o processamento e análise.....	114
5.1.2	Verificação se o Primeiro ponto, o pico, corresponde a uma Anomalia.....	116
5.1.3	Verificação se o Segundo ponto, corresponde a uma anomalia.....	120
5.1.4	Considerações finais sobre os pontos analisados.....	123
5.2	INFORMAÇÕES GERAIS DA APLICAÇÃO QUE ENVOLVE O SENSOR PIEZÔMETRO .....	124
5.2.1	Informações sobre o processamento e análise .....	127
5.2.2	Verificação se o ponto corresponde a uma Anomalia .....	128



5.2.3	Consideração final sobre os pontos analisados.....	129
5.3	CONSIDERAÇÃO FINAL .....	130
<b>6</b>	<b>COMPARAÇÃO DOS RESULTADOS OBTIDOS PELO MAIM E HOT SAX MODIFICADO .....</b>	<b>131</b>
6.1	TESTE REALIZADO COM O EXTENSÔMETRO .....	133
6.2	TESTE REALIZADO COM O MEDIDOR TRIORTOGONAL EM UM DE SEUS EIXOS	134
6.3	TESTE REALIZADO COM O MEDIDOR DE VAZÃO TRIANGULAR .....	136
6.4	TESTE REALIZADO COM O PIEZÔMETRO .....	137
6.5	CONSIDERAÇÕES FINAIS .....	138
<b>7</b>	<b>CONCLUSÃO .....</b>	<b>140</b>
7.1	ASPECTOS GERAIS DO MÉTODO.....	140
7.2	CONTRIBUIÇÕES .....	141
7.3	TRABALHOS FUTUROS PARA A EVOLUÇÃO DESTA PESQUISA .....	143
	<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>144</b>
	<b>APÊNDICE A – COMPLEMENTAÇÃO DO ESTADO DA ARTE.....</b>	<b>153</b>
A 1	BASE DE CONHECIMENTO .....	153
A 2	MÉTODOS ESTATÍSTICOS .....	154
A 3	MÉTODOS INTELIGENTES .....	155
A 3 1	Inteligência Artificial .....	155
A 3 2	Inteligência Computacional .....	156
A 3 3	Mineração de Dados.....	156
A 4	TÉCNICAS UTILIZADAS EM MÉTODOS INTELIGENTES.....	157

A 4 1 Árvores de Decisões.....	157
A 4 2 Árvore de decisão adaptativa - ADA .....	159
A 4 3 Lógica Fuzzy.....	161
A 4 4 Redes Neurais .....	164
<b>A.5 CONSIDERAÇÕES SOBRE TÉCNICAS DIVERSAS UTILIZADAS NO TRATAMENTO DE SÉRIES TEMPORAIS .....</b>	<b>165</b>
<b>A 6 CONSIDERAÇÕES SOBRE MEDIÇÕES .....</b>	<b>168</b>
A 6 1 Medidas e sinais analógicos .....	169
A 6 2 Sinais analógicos.....	170
A 6 2 1 Amostragem.....	171
A 6 2 2 Retenção.....	172
A 6 2 3 Quantificação .....	172
A 6 2 4 Codificação .....	172
A 6 2 4 Eliminação de alguns tipos de erros na comunicação .....	173
<b>APÊNDICE B – DETALHES DAS ANÁLISES E DESENVOLVIMENTOS .....</b>	<b>174</b>
B 1 MÉTODO HOT SAX ORIGINAL.....	174
B 2 MÉTODO HOT SAX MODIFICADO.....	183
B 2 1 Testes realizados .....	185
B 2 2 Considerações sobre o método HOT SAX Modificado .....	188
B 3 MÉTODO DE BASHA E AMEEN .....	188
B 3 1 Considerações sobre o método Basha e Ameen .....	192
B 4 ANÁLISE SINTETIZADA DE ALGUNS MÉTODOS INVESTIGADOS .....	192
B 4 1 Método HOT SAX .....	193

B 4 2 Método WAT .....	193
B 4 3 Método de Fu pré WAT .....	194
B 4 4 Método de Toshniwal e Yadav.....	195
B 4 5 Método de Knorn e Leith .....	195
B 4 6 Método K-Means + ID3 .....	196
B 4 7 Método de Mahoney e Chan .....	196
B 4 8 Método de Koushanfa e Sangiovanni-Vincentelli.....	197
B 4 9 Método de Basha e Ammeen.....	198
B 4 10 Detecção de Novidades com aplicação a fluxos contínuos de dados .....	198
<b>APÊNDICE C - CONFIGURAÇÃO DE <i>HARDWARE</i> E <i>SOFTWARE</i> UTILIZADOS</b>	<b>200</b>
C 1 <b>HARDWARE</b> .....	200
C 2 <b>SOFTWARE</b> .....	200

# 1 CONSIDERAÇÕES GERAIS

## 1.1 INTRODUÇÃO

A pesquisa realizada nesta tese culminou com o desenvolvimento de dois métodos: o Método Adaptativo para Interpretação de Medições (MAIM) e o Método Complementar. Esses métodos interpretam resultados de medições enviando, quando necessário, mensagens e alertas para tomadores de decisões. Esses métodos são independentes, mas podem trabalhar em conjunto. Nesse caso, é o Método Complementar que possui a capacidade de integração com outros métodos quaisquer a partir do recebimento de seus resultados e melhorando-os quando possível.

A motivação para o desenvolvimento desses métodos surgiu na realização em projetos relacionados a Usinas Hidroelétricas, nas quais foram estudados, instalados e monitorados instrumentos aplicados em estruturas de responsabilidade da engenharia civil para a coleta de informações. Durante esses projetos foi percebido que as pesquisas realizadas podiam ser estendidas para outras áreas de supervisão e monitoração envolvendo coleta de informações e sua interpretação.

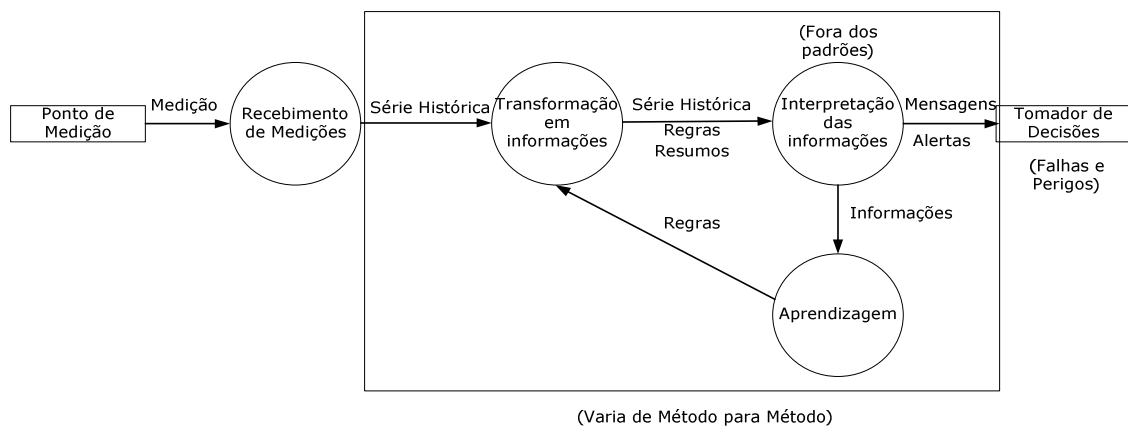
## 1.2 MÉTODO ADAPTATIVO PARA INTERPRETAÇÃO DE MEDIÇÕES (MAIM)

O método MAIM, conforme ilustrado simplificado na Figura 1, pretende atender de forma não supervisionada a necessidade de tomadores de decisão de receberem mensagens e alertas que retratem a realidade com maior assertividade. Nele, as medições recebidas de um ponto de medição são integradas a uma série histórica e em uma base de conhecimento (conforme apêndice A item A1) que, quando interpretadas, geram mensagens e alertas que podem ser enviadas aos tomadores de decisão. Além disso, o monitoramento realizado pelo MAIM pode ser integrado a sistemas corporativos, tais como: *Enterprise Resource Planning* (ERP), *Supervisory Control and Data Acquisition* (SCADA), dentre outros, ajudando com suas informações para a tomada de decisões.

A Base de Conhecimento do MAIM é composta por um complemento de informações incorporado na série histórica por resumos de informações, baseadas em conjunto de medições, e por regras pertinentes, assim, ela auxilia um dos módulos a decidir sobre a geração de mensagens e alertas.

O módulo trabalha com uma Tabela de Decisões Adaptativa, que é modificada toda vez que a série histórica deixa de ser estável, gerando novas regras para auxiliar juntamente com resumos e regras na identificação de possíveis situações de perigo e/ou alerta. Neste caso, um módulo de aprendizagem recebe parte das informações transformando-as em novas regras, conforme apresentado no item 2.3.2 que aborda Tabela de Decisões Adaptativas.

Figura 1 - Ilustração do Método MAIM de forma simplificada



Fonte: Autor

### 1.3 MÉTODO COMPLEMENTAR

O objetivo do Método Complementar é receber os resultados das interpretações do método MAIM ou de outros métodos que lidam com séries históricas (ou temporais) para indicar anomalias e reinterpretá-las confirmando ou não os seus resultados. Além disso, incluir justificativas, e orientar melhor as ações necessárias para aumentar a assertividade do resultado.

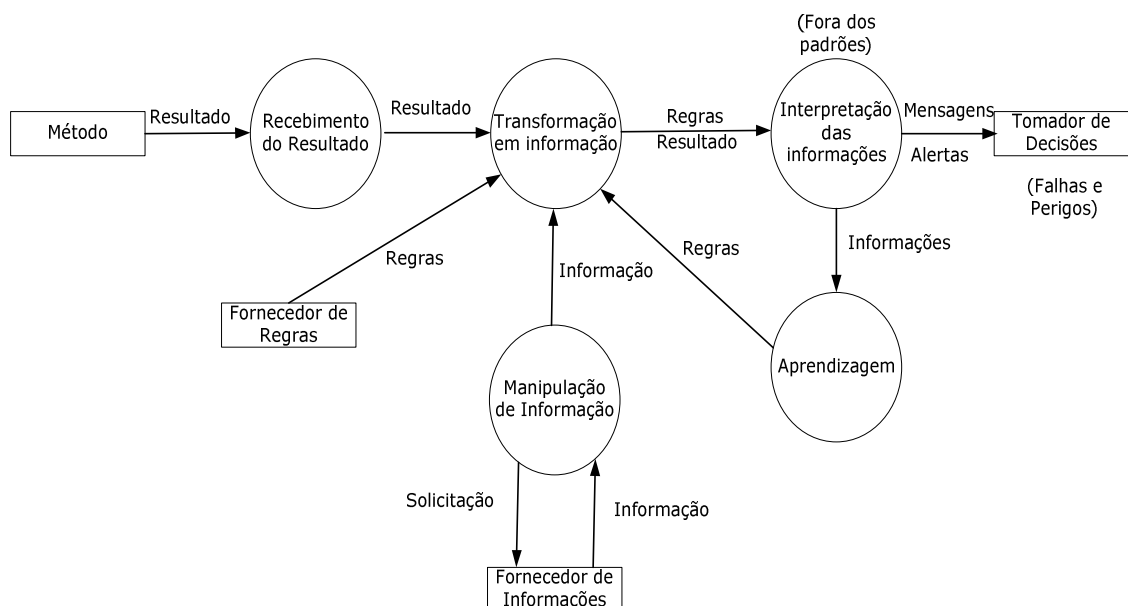
O Método Complementar, conforme ilustrado de forma simplificada na figura 2, recebe resultados obtidos por outros métodos e melhora quando possível a assertividade desses resultados ajudando os Tomadores de Decisão com novas mensagens e alertas.

No Método Complementar as informações trabalhadas são recebidas de diversas fontes, sendo a principal o método que fornece o resultado a ser analisado. As outras fontes fornecem suas informações a qualquer tempo, podendo ter origem em: especialistas, outros sistemas computacionais, outras bases de conhecimento (corporativas e utilidade pública), dentre outras. Existe um módulo específico de manipulação de informações que é capaz de solicitar informações a outras fontes específicas.

A base de conhecimento é composta por regras geradas em diversas origens compondo todo o conhecimento possível em regras com a finalidade de aumentar a assertividade dos resultados. As regras existentes são utilizadas por um módulo inteligente que as interpreta gerando as mensagens e regras pertinentes.

A principal vantagem do Método Complementar está em sua proposta de trabalhar com outros métodos complementando seus resultados e enviando mensagens pertinentes mais elaboradas, auxiliando de forma significativa os tomadores de decisão.

Figura 2 - Ilustração do Método Complementar de forma simplificada



Fonte: Autor

## 1.4 CONTEXTUALIZAÇÃO

A obtenção de medições automáticas e sua inclusão em séries históricas ditas temporais estão presentes há muito tempo nas mais diversas áreas do conhecimento técnico com o objetivo de auxiliar no controle e na tomada de decisões. O simples monitoramento das medições pode identificar situações não esperadas, identificar padrões e tendências. O monitoramento pode evoluir para um controle que, com suas comparações, pode gerar eventos dos mais diversos tipos, como atuação de dispositivos, geração de mensagens e alertas, dentre outras ações. Existem também monitoramentos que dificilmente evoluem para o controle, como alguns aplicados na área elétrica, que não implicam em ações em dispositivos.

Atualmente, o tempo entre as medições automáticas está cada vez menor e o volume de informações cada vez maior, demandando ações além da capacidade humana e gerando uma automatização capaz de incluir qualquer tipo de acompanhamento, com o auxílio de métodos para interpretação e análise de todas as ocorrências, de medições a eventos. Além disso, a comunicação e integração entre sistemas digitais, incluídos os de medições, estão transformando as relações entre eles ampliando as possibilidades de interpretação e análises dos fatos de uma forma tal que é possível construir perfis de objetos, pessoas e fatos integrados.

Acompanhando esses acontecimentos, alguns autores estão ampliando conceitos existentes, tais como os apresentados em (PORTER; HEPPELMENN, 2014) que consideram a existência de estágios mais evoluídos e mais inteligentes com relação aos produtos (incluídos instrumentos de medição) e serviços (que podem incluir geração, transmissão e distribuição de energia), que são: monitoramento, controle, otimização e autonomia. Para eles, o controle contém operações de monitoramento, a otimização contém as operações de controle e, finalmente, a autonomia contém as operações de otimização. Um instrumento ou um grupo de instrumentos podem fazer parte de qualquer um dos estágios apresentados. Assim, com base em Porter; Heppelmenn (2014), tem-se:

- Monitoramento envolvendo sensores e fontes de informação, permitindo saber a condição de um produto, seu ambiente externo, o uso e sua operação;
- Controle com procedimentos ou *software* embutido no produto ou de forma remota, possibilitando a manipulação das funções do produto e personalização da experiência do usuário;
- Otimização, permite que, com as capacidades do monitoramento e do controle, sejam utilizados algoritmos para a otimização do uso e operação de um produto, melhorando desempenho, possibilitando diagnóstico, manutenção e reparos preventivos;
- Autonomia, tem a combinação do monitoramento, do controle e da otimização, permitindo a operação autônoma de um produto, sua autocoordenação de operação com outros produtos e sistemas, melhorias e personalização autônoma do produto além do autodiagnóstico e autorreparação.

Considerando Porter; Heppelmenn (2014), este trabalho apresenta um método, desenvolvido inicialmente para monitoramento, com a capacidade de estar integrado aos estágios de controle, otimização e autonomia, em parte devido ao desenho do banco de dados e da base de conhecimento. Dessa forma, são possíveis a comunicação e a integração com um sistema organizacional existente em uma empresa (que pode envolver módulos de administração, manutenção, dentre outros) e mundial, envolvendo as questões relacionadas aos termos atualmente em voga como: *Smart Grid*, *Smart City*, Internet das Coisas (IoT), redes inteligentes, computação embarcada dentre outras.

Ressalta-se que os métodos desenvolvidos (MAIM e Complementar) estão aderentes ao monitoramento que pode evoluir ou não para um controle, apesar de que os tipos de aplicações envolvidas nesta tese dificilmente sinalizam para uma evolução no sentido do controle, estando restritas apenas na identificação de medições que podem gerar mensagens e alertas. Mesmo assim, sem evoluir para utilização do controle (mas sim em seu projeto), é possível evoluir para a otimização e autonomia. Desse modo, os métodos



foram incluídos em sistemas computacionais que buscam mais flexibilidade em sua operacionalidade.

## 1.5 OBJETIVOS

O objetivo principal corresponde ao estabelecimento de um método de avaliação de medições automáticas no monitoramento, com possível integração nos outros estágios (controle, otimização e autonomia), conforme (PORTER; HEPPELMENN, 2014), que podem conduzir às ideias de flexibilidade e adaptabilidade. Destarte, neste trabalho é considerado o chamado ambiente de influência com relação ao ponto de medição que considera e relaciona as informações envolvendo o ambiente físico, o instrumento de medição, o ambiente computacional, dentre outros.

Outro objetivo importante está relacionado à aprendizagem, sendo assim, um método inteligente e estar considerando em seu algoritmo as características citadas por autores da área de Inteligência Artificial como em (ROUSSELL; NORVIG, 2010), (LUGER, 2008, 2013), e (RICH; KNIGHT, 1993), envolvendo a Representação do Conhecimento, a Busca, a Inferência e o Aprendizado, além de questões envolvendo agentes inteligentes. Por isso, devido a uma necessidade específica é realizada uma busca de informações registradas segundo uma determinada estrutura de representações do conhecimento, obtendo resultados aos quais é aplicada a inferência (“raciocínio”) que dará alguma resposta ou solução a necessidade de forma a aprender com ela.

Devido às considerações sobre as várias etapas, alguns objetivos secundários são considerados para atingir suas proposições. Dentre os quais são citados os seguintes:

- Ser um método não supervisionado;
- Considerar um fluxo contínuo de medições formando séries históricas;
- Em seu algoritmo registrar, interpretar e analisar cada medição recebida;

- Devido ao critério de flexibilidade identificar: novidades (a primeira aparição de uma medição), padrões, anomalias, discórdias, divergências, não conformidades, e *outliers* em geral;
- Identificar mudanças do ambiente relacionado ao ponto de medição;
- Identificar previsões próximas;
- Em seu algoritmo deve estar prevista a realização de diagnósticos;
- Aprender o comportamento da grandeza com a série histórica, com os registros auxiliares e as interpretações realizadas;
- Emitir no mínimo mensagens e alarmes aos diversos interessados do monitoramento e estar preparado para disparar a ativação de diversos tipos de ações;
- Realizar melhoria contínua de forma a agregar novas informações, aprender e descartar aprendizado que não seja mais pertinente.

Um aspecto importante deste estudo é o fato de ele ter sido patrocinado pelas concessionárias EDP e CESP. Desse modo, parte dos testes considera instalações existentes em uma de suas Usinas Hidroelétricas e utilizam medições oriundas de instrumentos civis (utilizados na engenharia civil) tais como os medidores do tipo: Triortogonal, Extensômetro de Haste, Medidor de Vazão Triangular, Piezômetro de tubo, dentre outros. Mas um desdobramento do estudo permite aumentar sua abrangência para outras áreas inclusive a elétrica e mais especificamente as condições relacionadas a sistemas computacionais inteligentes.

## 1.6 METODOLOGIA

Cabe um esclarecimento sobre a organização da tese. A metodologia utilizada para os estudos consistiu nos seguintes critérios:

- Considerar a necessidade da criação de um método para a área de usinas hidroelétricas capaz de trabalhar em um aplicativo que receba medições em

uma série histórica de forma não supervisionada, que aprenda também com essas medições, identifique anomalias e envie alertas aos interessados;

- Em seguida, estabelecer e verificar para a série histórica se é possível identificar medições consideradas novidades (que correspondem a primeira ocorrência) e, indiretamente, identificar modificações do ambiente no entorno de um instrumento de medição (a partir das medições, ambiente de influência com relação ao ponto de medição que considera e relaciona as informações envolvendo: o ambiente físico, o instrumento de medição, o ambiente computacional, dentre outras) para realizar análise e identificação de candidatas à anomalia e às anomalias de forma adaptativa;
- Realizar o levantamento do estado da arte para a obtenção de anomalias em duas classificações, sendo uma para questões adaptativas e outra para questões gerais;
- Realizar o levantamento de métodos, passíveis de testes, para a identificação de anomalias na literatura e internet;
- Realizar testes em métodos e/ou em aplicativos para saber como são identificadas as anomalias, utilizando aplicativos disponíveis no mercado/academia, a serem identificados, ou desenvolvendo segundo as especificações dos métodos estudados;
- Definir os requisitos necessários para o desenvolvimento de aplicativos para métodos adaptativos considerando principalmente o uso de regras, tabela de decisões adaptativa, novidades, ambientes, agrupamentos de medições e não supervisão;
- Realizar provas de conceito (testes de validações de ideias) para estudar a viabilidade dos conceitos e teorias propostas desta tese;
- Desenvolver protótipos, considerando monitoração das medições e prevendo no projeto a evolução da estrutura do aplicativo para controle, otimização e autonomia;
- Testar a metodologia;
- Estabelecer o método final.

O texto desta tese é uma consequência direta da metodologia desenvolvida, embora não seja a sequência descritiva dela.

## 1.7 ORGANIZAÇÃO DO TEXTO

O Capítulo **Erro! Fonte de referência não encontrada.** apresenta a introdução do trabalho, os objetivos, a metodologia aplicada e a organização do texto.

O Capítulo 2, em conjunto com o Apêndice A, apresentam o estado da arte que envolve conceitos, métodos e técnicas existentes.

O Capítulo 3 apresenta uma síntese de pesquisas realizadas e apresentadas no Apêndice B que envolve desenvolvimentos e experimentos incluindo estudos e testes com o método denominado HOT SAX e uma adaptação realizada para o patrocinador EDP e publicada em (NICOLI et al., 2013). Ainda estão apresentados testes com o método de (BASHA; AMEEM, 2007), no apêndice B. Resumos de vários estudos complementares também estão mostrados, realizados com outros métodos além dos relacionados a artigos envolvendo aplicações de Tabela de Decisões Adaptativas.

O capítulo 4 apresenta os dois métodos desenvolvidos junto com provas de conceitos e validações. O primeiro método corresponde ao tema proposto neste trabalho utilizando critérios adaptativos para a análise de pontos candidatos a anomalias, com enfoque inicial em monitoramento, de forma a ficar aderente a outras áreas. Também no primeiro método há considerações sobre novidades utilizadas para a identificação de ambientes operacionais, nos quais está o ponto de medição. O segundo método complementa os resultados obtidos por outros métodos, podendo até recusá-los caso não estejam aderentes a questões técnicas ou às necessidades dos interessados.

O capítulo 5 apresenta aplicações envolvendo o método MAIM, foco principal desta tese.

O capítulo 6 apresenta a comparação do método MAIM com o método HOT SAX modificado, considerando o fato de os dois terem o conjunto básico de requisitos semelhantes.

O Capítulo 7 apresenta as conclusões advindas desta pesquisa, além de propostas para continuação deste trabalho.

## 2 ESTADO DA ARTE

Este capítulo apresenta áreas do conhecimento, métodos e técnicas, considerados no desenvolvimento do método proposto nesta tese. Essas informações constataam que existem diversas abordagens aplicadas à análise de séries históricas que, ao serem agrupadas, prevalecem três grupos caracterizados pelo uso de técnicas estatísticas, técnicas inteligentes e técnicas híbridas, esta envolvendo as outras duas.

Na sequência, serão apresentados conceitos em função de sua consideração no método desenvolvido, tais como: séries históricas de medições, métodos adaptativos e técnicas utilizadas em métodos inteligentes. Em complementação ao exposto neste capítulo, mais informações são apresentadas no apêndice A, estão exibidos métodos estatísticos, métodos inteligentes (inteligência artificial, inteligência computacional e mineração de dados), além dessas outras técnicas: Árvores de Decisão, Árvores de Decisão Adaptativa, Lógica Fuzzy e Redes Neurais.

Cabe ressaltar que, dentre os conteúdos do capítulo, estão: a formalização, o enfoque dado ao uso de regras por algumas técnicas e a abordagem dada aos conceitos adaptativos. Ressaltam-se também as referências apresentadas em (FACELI et al., 2011), que forneceram os conceitos básicos das Regras de Associação, além das referências, segundo Neto (2001, 2007, 2009), Tchemra (2009), Catae; Rocha (2011), e Baraúna (2010) que contribuíram com os conceitos adaptativos em sua formalização e a tabela de decisão adaptativa.

### 2.1 SÉRIES HISTÓRICAS

Segundo Barros (2009), uma Série Temporal é um conjunto de observações ordenadas no tempo (não necessariamente igualmente espaçadas) e que apresentam dependência serial (isto é, dependência entre instantes de tempo), que, para o autor, corresponde a um processo estocástico. Além disso, as séries temporais podem ser um modelo Univariado quando é explicada (prevista) apenas por seus valores passados ou um modelo Multivariado, quando explicada pelos seus valores passados e também pelos

valores passados de outras variáveis. As séries temporais podem apresentar tendências, ciclos e sazonalidades, que são facilmente observáveis em gráficos traçados envolvendo as variáveis: tempo e valor.

Ainda com relação à definição de série temporal (SILVA et al., 2007), “uma série temporal é um conjunto de observações discretas, realizadas em períodos equidistantes e que apresentam uma dependência serial entre essas observações”. Também pode ser considerada como em (MORETTI; TOLOI, 2006) como qualquer conjunto de observações ordenadas no tempo.

Com base nas definições e objetivos para este trabalho, a série histórica de medições compõe um conjunto de registros das ocorrências de medições realizadas por um instrumento em um ponto físico específico correspondendo a uma série temporal. Os registros das medições normalmente possuem uma informação mínima que inclui o valor medido e o instante da leitura realizada.

Cabe destacar que, com relação às séries temporais, uma consideração utilizada por várias técnicas está relacionada às variáveis que envolvem valor e instante também na realização de agrupamentos com elas. Ressalta-se que quanto menor é a unidade de medida utilizada em uma variável (por exemplo, uso de grama, ao invés de quilograma), maior serão as possibilidades de agrupamentos, sendo que essa ideia corresponde à Granularidade. Em outras palavras, a maior Granularidade equivale ao menor detalhamento e a menor granularidade está relacionada ao maior detalhamento. Assim, por exemplo:

- Se a unidade de medida for dia, é possível agrupar por dia, semana, mês, bimestre, trimestre, semestre, ano e assim por diante;
- Se a unidade de medida for por segundo, é possível agrupar por segundos, minutos, horas, dias e assim por diante.

Desse modo, a menor Granularidade possibilita mais flexibilidade em agrupamentos, técnica utilizada em vários métodos nas áreas de mineração de dados, estatística, inteligência artificial, dentre outras.

Outras técnicas são aplicadas às séries históricas e algumas delas são comentadas no item A.2.2 do apêndice A, que trata das considerações às técnicas diversas utilizadas no tratamento de séries temporais.

## 2.2 MÉTODOS ADAPTATIVOS

Existem vários enfoques com relação à adaptação, dentre os quais está a transdisciplinaridade, considerada na Teoria Geral dos Sistemas e citada por (MARTINELLI et al., 2012) no âmbito da Cibernética, na qual os sistemas se adaptam para poderem manter o controle mesmo diante de grandes mudanças ambientais. O autor completa dizendo que: “em sua manifestação mais simples, por meio de algum processo de aprendizagem, o sistema consegue redefinir as referências para seus mecanismos internos de regulação”. Também comenta que a evolução é quando a adaptação envolve, além da aprendizagem, mutações que correspondem a mudanças estruturais e/ou funcionais permanentes no sistema, assim, o sistema é capaz de vencer novas ameaças e aproveitar novas oportunidades que se manifestam no ambiente externo.

A linha de pesquisa desta tese segue os mesmos conceitos dos diversos trabalhos de Neto (2001, 2007, 2009) e Ramos; Neto; Vega (2009). Dessa forma, foi considerada a definição de (NETO, 2009) em que Adaptatividade é “a propriedade que apresenta um sistema, dispositivo ou processo computacional, que lhe permite sem a interferência de agentes externos (mesmo o próprio operador) tomarem decisões de modificar dinamicamente, de forma autônoma, seu próprio comportamento, em resposta à sua configuração corrente e ao estímulo de entrada recebido”.

Este trabalho também considera, segundo Carhuanina; Neto (2016), que em termos gerais, Adaptabilidade define que uma entidade pode ser adaptada por um agente externo que tem o poder de decisão sobre as mudanças a serem feitas, as quais são aplicadas pela entidade a ser adaptada ou, num caso particular, ela mesma pode governar o processo de decisão e aplicação das mudanças em sua própria estrutura. Também, segundo Carhuanina;



Neto (2016), a Adaptatividade, de maneira geral, pode ser definida como a capacidade que possui alguma coisa para efetuar uma mudança a fim de se adequar a uma nova situação.

### 2.2.1 Dispositivo ou processo computacional

O termo dispositivo é:

Qualquer artefato abstrato que, a partir de uma configuração inicial conhecida e fixa, opera migrando sucessivamente de uma configuração para outra, em resposta a sucessivos estímulos de entrada recebidos, e de acordo com um conjunto de regras que define o seu comportamento. (NETO, 2009).

Assim um programa, um processo que receber um estímulo X na configuração Y, passa a ter uma nova configuração Z. Sendo que, configuração, é a representação, em um determinado momento, do conteúdo de todos os elementos variáveis relevantes à operação do dispositivo.

Existem diversos dispositivos conforme sua forma de atuação. Podem ser destacados os seguintes por (TCHEMRA, 2009) e (NETO, 2007):

- Dispositivos de reconhecimento: como os autômatos convencionais, que se caracterizam por conjuntos estáticos e finitos de estados e de regras de transição;
- Dispositivos de geração: no caso das gramáticas livres de contexto que têm como base as regras de produção da gramática;
- Dispositivos para a representação de sistemas assíncronos: como o formalismo clássico dos *Statecharts* utilizados para o sincronismo de eventos;
- Dispositivos estocásticos: são dispositivos que apresentam transições associadas às probabilidades, tal como a rede de Markov que tem a capacidade de representar fenômenos aleatórios;
- Dispositivos de processamento: representados pelos códigos de programação;

- Dispositivos de auxílio à tomada de decisão: como as tabelas de decisão e árvores de decisão que são constituídas por conjuntos de regras.

As diversas técnicas adaptativas existentes utilizam os dispositivos definidos conforme Neto (2009). Dentre as técnicas adaptativas aplicáveis ao tipo de problema em foco, destacam-se as Tabelas de Decisão Adaptativas (TDA), as Árvores de Decisões Adaptativas (ADA) e os Autômatos Adaptativos (AA).

Os itens a seguir apresentam: dispositivos não adaptativos guiados por regras e dispositivos adaptativos guiados por regras. Mais adiante estão discutidas questões relativas a regras com enfoque adaptativo.

### 2.2.2 Dispositivo não adaptativo guiado por regras

Conforme Neto (2009), um dispositivo não adaptativo guiado por regras tem seu comportamento descrito por um conjunto finito e fixo de regras condicionais na forma clássica “se - então” também ditas regras de produção. Esse tipo de dispositivo não permite alteração dinâmica em seus componentes estruturais ou funcionais.

Para Neto (2009) um dispositivo  $D$ , não adaptativo, guiado por regras, pode ser descrito como uma quintupla, conforme a formulação seguinte:  $D = (C, R, S, C_0, A)$  na qual:

- $C$  é o conjunto de todas as possíveis configurações de  $D$ ;
- $S$  é o conjunto de todos os estímulos válidos de entrada;
- $C_0 \in C$  é a única configuração inicial de  $D$ ;
- $A \subseteq C$  é o conjunto de todas as configurações de aceitação de  $D$ ;
- $R$  é uma relação de mudança de configuração para  $D$ :  $R \subseteq C \times (S \cup \{\varepsilon\}) \times C$ , cujos elementos constituem as regras que definem o dispositivo  $D$ :  $r = (c_j, s, c_j') \in R$  com  $c_j, c_j' \in C$ ;  $s \in S$ , as quais podem ser denotadas na

forma  $(c_j, s) \rightarrow c_j'$ , realçando ser  $s$  o estímulo que, aplicado ao dispositivo  $D$  na configuração  $c_j$ , leva-o a uma nova configuração  $c_j'$ ,  $c_j \Rightarrow c_j'$ ;

- A linguagem definida por  $D$  é representada por  $L(D)$ :  $L(D) = \{w \in S^* \mid c \Rightarrow w^* c, c \in A\}$ , onde  $x^*$  denota o fecho de Kleene de um conjunto (por exemplo,  $S$  e  $C$ ) ou operador  $x$  (por exemplo,  $w$  e  $c$ ).

Conforme NETO (2009), “as regras, denotadas como triplas  $(c, s, c')$ , podem ser facilmente transcritas como cláusulas condicionais, da forma: se recebido for  $s$ , então o dispositivo  $D$  poderá alterar sua configuração para  $c'$  .

Cabe ressaltar que ocorre uma alteração (movimento) no dispositivo toda vez que uma dessas regras for aplicada. É importante observar que, segundo Neto (2009), na prática as condições podem levar a ações que por sua vez podem ser funções ou procedimentos a serem executados dentro de um aplicativo computacional.

A tabela de decisões é um dispositivo não adaptativo guiado por regras.

### 2.2.3 Dispositivo adaptativo guiado por regras

Segundo Neto (2009), “é todo dispositivo cujo comportamento incorpore o conceito da adaptatividade”. Ao receber um estímulo  $X$  na configuração  $Y$ , a nova configuração passa a ser  $Z$ , e também o conjunto de regras se altera pela remoção, substituição e inserção de novas regras. Para Neto (2001), o dispositivo adaptativo é um sistema cuja operação é realizada por um conjunto finito de regras, que se modifica dinamicamente.

O dispositivo adaptativo formalizado por (NETO, 2001) é composto por um dispositivo subjacente não adaptativo e um mecanismo adaptativo.

De acordo com as definições apresentadas por Tchemra (2009) com base no descrito em Neto (2001), um dispositivo adaptativo é constituído pela dupla  $DA = (DN, CA)$  onde:

- DN: o núcleo composto por um dispositivo subjacente não adaptativo, cujas operações são definidas por um conjunto finito de regras R;
- CA: o mecanismo adaptativo, que pode agir sobre o núcleo e possui os recursos necessários à mudança da configuração (estrutura) e do comportamento do dispositivo subjacente, enquanto está em operação.

Assim, a partir de um dispositivo adaptativo guiado por regra é possível desenvolver outro dispositivo adaptativo como, por exemplo, a tabela de decisões adaptativa.

Uma característica interessante com relação às funções adaptativas é que elas podem estar relacionadas a mais de uma ação adaptativa, desse modo, antes da inclusão de uma regra a ação de consulta e a exclusão podem ser executadas em outras regras.

Com relação às regras é possível executar uma ação adaptativa antes de executar uma regra e também executar após a regra. O ideal é operar com essas duas possibilidades, devido a uma flexibilidade maior.

### 2.3 TÉCNICAS UTILIZADAS EM MÉTODOS INTELIGENTES

Cabe ressaltar neste item o uso do termo “técnica”, que em algumas situações foi utilizada para verdadeiros métodos. A razão disso é com referência às abordagens inteligentes que contêm métodos que, por sua vez contêm técnicas. Os métodos inteligentes em geral possuem uma série de procedimentos, sendo que alguns são embutidos em ferramentas computacionais específicas, dentre as quais algoritmos e/ou heurísticas. Neles são aplicadas técnicas desenvolvidas a partir de uma ou mais abordagens conceituais, algumas dessas técnicas, por sua vez, são utilizadas em métodos de abordagens diferentes.

### 2.3.1 Tabela de decisões

Numa tabela de decisões tradicional, uma regra é composta por um conjunto de condições e um conjunto de ações. Na realidade, o conjunto de condições existentes determina a execução de um conjunto de ações. Uma tabela de decisões é composta por um conjunto de regras, que pode ser considerada uma base de conhecimento. Dessa forma, é possível realizar uma busca na base de conhecimento para encontrar no mínimo uma determinada regra que satisfaça uma determinada consulta.

Na Figura 3, duas formas de representação visual de uma tabela de decisões tradicional são apresentadas, nela são denotadas regras ( $R_i$ ), condições ( $C_i$ ) e ações ( $A_i$ ), nas quais o índice  $i$  é um número inteiro onde  $1 \leq i \leq n$ . A principal diferença em relação às tabelas da Figura 3 é a troca de linhas e colunas.

Figura 3 - Formas de representação visual das tabelas de decisão tradicional

		REGRAS				
		R1	R2	R3		Rn
C1		SIM		SIM		
C2			SIM	SIM		
C3				SIM		
						SIM
Cn				SIM		
A1		X				
A2			X			
A3		X				
						X
An				X		

		CONDIÇÕES			AÇÕES				
		C1	C2		Cn	A1	A2		An
	R1		SIM				X		
	R2	SIM				X			X
	R3		SIM						
	R4								
REGRAS									
	Rn								X

Fonte: Autor

Os conteúdos das células que, neste caso são o encontro das regras com as condições ou ações variam um pouco conforme o modelo de tabela de decisões. Assim nas células relacionadas com as condições são encontrados: (Sim, Não, “-“ ou nada), (Verdadeiro, Falso, “-“ ou nada), valores numéricos, intervalos numéricos, até textos alfanuméricos, dentre outros. Já nas células relativas às ações, são encontradas marcas

indicando que devem ser aplicadas. Do mesmo modo são encontrados os nomes de funções e procedimentos a serem utilizados e, em alguns casos também aparece uma indicação de Verdadeiro ou Falso.

Outrossim, é importante acrescentar que no cabeçalho das regras pode ser encontrado o nome da regra ou uma sequência de numeração. No caso do cabeçalho de condições em geral, são encontradas descrições das condições e proposições. Já no caso do cabeçalho das ações, são encontradas descrições das ações e, em alguns casos, os próprios procedimentos ou funções.

Nota-se que a grande variedade de conteúdos das células e cabeçalhos (que podem possuir mais de uma linha ou coluna) leva à inferência da existência de uma variedade de procedimentos possíveis para a operação de uma tabela de decisões. Mas toda essa variedade possui um mínimo em comum.

Normalmente, a operação de uma tabela de decisões tem como base um conjunto de condições de entrada que é utilizado em uma busca na base de conhecimento. Caso seja encontrado um conjunto de condições iguais às condições de entrada, as ações da regra são utilizadas. Mas, em alguns casos, podem existir regras com as condições de entrada coincidentes. Quando existe uma única coincidência com as condições de uma regra, tem-se uma solução determinística, mas, quando existem mais de uma coincidência tem-se uma solução não determinística. No caso de diversas soluções, geralmente o usuário da tabela de decisões acaba completando a decisão final. Cabe ressaltar que a ocorrência de nenhuma coincidência é possível e, nessa situação, foi criada uma variação de tabela de decisões que possui uma coluna denominada de “Senão” que executa uma ou mais ações.

Há também a operacionalização por meio do encadeamento de tabelas de decisões que, conforme Tchemra (2009), em geral são separadas por partes lógicas, com base nas ações. Existem dois tipos: a Tabela de decisões Aberta, em que o encadeamento lógico é sem retorno à tabela de decisões anterior, e existe a Tabela de decisões Fechada, na qual ocorre encadeamento lógico para a tabela de decisões anterior.

Conforme apresentado por Tchemra (2009) formalmente, uma tabela de decisão pode ser definida por  $TD = (C, A, R)$ , onde:

- $C$  é o conjunto de condições, que representam as variáveis do problema:

$$C = \{C_i, 1 \leq i \leq n\};$$

- $A$  é a relação de ações, que podem ser executadas para solucionar o problema:

$$A = \{A_j, 1 \leq j \leq m\};$$

- $R$  é o conjunto de regras, composto pela combinação dos valores das variáveis, que levam à execução de ações possíveis:  $R = \{R_k, 1 \leq k \leq p\}$  com a regra  $R_k = (c_{i,k}, a_{j,k})$  sendo:
  - $c_{i,k}$  um valor que a condição  $C_i$  pode assumir na regra;
  - $a_{j,k}$  uma indicação de execução da ação  $A_j$  na regra.

Assim segundo Tchemra (2009), qualquer regra  $R_k$  da tabela pode ser interpretada como uma série de sentenças “se X então Y”, onde X é o conjunto de condições a serem verificadas e Y, as ações a serem executadas. Geralmente, as condições são compostas por “e”s ou “ou”s, ou seja, X é uma sentença da forma: “ $C_1$  e  $C_2$  e ... e  $C_n$ ”, ou uma composição do tipo “ $C_1$  ou  $C_2$  ou ... ou  $C_n$ ”, ou ainda, uma combinação de “e”s, “ou”s e “não”s (negação).

Muitas melhorias foram propostas e implementadas para as tabelas de decisões como, por exemplo, o uso de pesos e de classificações para as diversas soluções encontradas e o aprendizado com interação com o usuário para quando nenhuma solução for encontrada. Mas a mais avançada forma de uma tabela de decisões é a inclusão da adaptatividade, como será apresentado no próximo item.

### 2.3.2 Tabela de Decisões Adaptativa

A Tabela de Decisões Adaptativa (TDA) foi definida em (NETO, 2001) como um dispositivo subjacente composto por uma tabela de decisão convencional, na qual é acrescentada a camada adaptativa composta por um conjunto de linhas para a definição das funções. Quando a ação adaptativa é executada, geralmente, o conjunto de regras é modificado e, em consequência, a quantidade de colunas da tabela. A quantidade de linhas, porém, permanece fixa, pois não há alteração nestas.

Assim, a TDA é uma dupla:  $TDA = (TDN, CA)$ , na qual TDN é a tabela de decisão convencional e CA o mecanismo adaptativo.

A tabela de decisão não adaptativa em (TCHEMRA, 2009) é definida por uma sêxtupla.

$TDN = (CT, NRT, CV, t_0, AT, CRA)$ , em que:

- CT é o conjunto de todas as configurações possíveis da tabela de decisão;
- $NRT \subseteq CT \times CV \times CT \times CRA$  é o conjunto de regras de decisão da tabela;
- CV é o conjunto finito de valores válidos das condições do problema inserido na tabela, incluindo (vazio);
- $t_0 \in CT$  é a configuração inicial da tabela de decisão, o que inclui as condições, ações e regras existentes;
- CRA é o conjunto finito dos possíveis resultados obtidos pela execução das ações quando da aplicação das regras NRT; com  $\mathcal{E} \in CRA$ ;
- Cada regra  $r_t \in NRT$  é da forma  $r_t = (t_i, cv, t_j, r)$ , onde com a entrada de valores das condições  $cv \in CV$ , a regra  $r_t$  gera a resposta  $r \in CRA$  e muda a configuração da tabela corrente  $t_i$  para a tabela  $t_j$ .
- $AT \subseteq CT$  é o conjunto de configurações da tabela de decisão aceitas durante o processo; logo, o subconjunto  $FC = CT - AT$  é composto pelas configurações rejeitadas.



O mecanismo adaptativo CA da TDA é definido pelo conjunto das funções adaptativas FA, cada uma delas declarada por uma sétupla:

FA = (nome, P, V, G, BA, AD, AA), em que:

- Nome é uma identificação usada para referenciar a função adaptativa;
- P é uma lista de parâmetros formais que são utilizados para referenciar valores passados para a função adaptativa quando é chamada;
- V é uma lista de variáveis (é opcional), que são utilizadas para armazenar valores resultantes da aplicação da função adaptativa;
- G é uma lista de geradores (é opcional), utilizados para referenciar novos valores resultantes da aplicação da função adaptativa;
- BA (é opcional) indica uma chamada de ação adaptativa anterior;
- AD representa o corpo da função adaptativa e é composta por um conjunto de ações adaptativas elementares de consulta, inclusão e exclusão que modificam o conjunto de regras corrente da tabela de decisão;
- AA (é opcional) e indica uma chamada de ação adaptativa posterior.

Desta forma, o mecanismo adaptativo CA da TDA é determinado por:

$CA \subseteq BA \times NRT \times AA$ , em que BA e AA são conjuntos de ações adaptativas a serem executadas, respectivamente, antes e depois das regras não adaptativas da tabela, e ambos contêm a ação nula  $\mathcal{E}$ , ou seja,  $BA \cap AA$ .

A Figura 4 apresenta os componentes básicos de uma Tabela de Decisões Adaptativa com base em Tchemra (2009, 2010), Ramos; Neto; Vega (2009) e na definição formal apresentada acima. Este modelo tem como base a função adaptativa, que pode ser utilizada antes ou após o dispositivo subjacente.

Figura 4 - Estrutura básica de uma Tabela de Decisões Adaptativa

		Cabeçalho das colunas da declaração das funções adaptativas que podem conter colunas H, +, -, ?	Cabeçalhos das colunas com Regras que podem conter S, R, E (o número de colunas é variável)
Conjunto de linhas de condições	Nomes / descrições das Condições		Celulas com os conteúdos das condições
Conjunto de linhas de ações	Nomes / descrições das ações	Declaração das funções adaptativas (protótipo)	Celulas indicando a aplicação da ação
Conjunto de linhas das funções adaptativas	Nome das funções adaptativas	As colunas com "?" indica consulta a com "-" exclusão e a com "+" inclusão de regras	Células indicando a execução da função adaptativa
Conjunto de linhas com os parâmetros das funções adaptativas	Nome dos Parâmetros das funções	Na coluna com H em suas células são indicadas se uma função é executada antes (B) ou depois (A)	Células indicando conteúdos dos parâmetros
Conjunto de linhas com as variáveis das funções adaptativas	Nome das variáveis das variáveis	e com (P) os parâmetros, e com (V) as variáveis e com (G) os geradores.	Células indicando conteúdo das variáveis
Conjunto de linhas com os geradores das funções adaptativas	Nome dos geradores		Células indicando conteúdo dos geradores

Fonte: Autor.

Cabe ressaltar que, segundo Ramos; Neto; Vega (2009), uma Tabela de Decisões Adaptativa pode ser estruturada minimamente conforme o apresentado na figura 5, que ilustra que, para as mesmas condições que são utilizadas na Tabela de Decisões subjacente (a uma TDA) para uma determinada regra, podem ser definidas ações adaptativas antes e/ou depois dela. Dessa maneira, as ações adaptativas afetam a TD alterando o seu comportamento. Ao utilizar as abstrações do tipo funções adaptativas e as suas configurações, chega-se à representação da figura 4, com a inclusão formal de Parâmetros, Variáveis, e Geradores, em um corpo de função adaptativo considerando também as ações adaptativas de consulta, exclusão e inclusão de regras.

Nota-se que é provável ocorrer por meio de uma ação adaptativa anterior uma regra eliminar a si mesma, não sendo mais possível executar a ação normal nem a ação adaptativa posterior, devendo-se proceder à execução, se existir, de uma próxima regra. Nesse caso é possível utilizar diretamente uma ação adaptativa (através de uma função).

Figura 5- Estrutura simplificada de uma Tabela de Decisões Adaptativa

		Regras				
		1	2	...	...	n
Condições	C1	Conteudo1_1	Conteudo2_1	...	...	ConteudoN_1
	C2	...	Conteudo2_2	...	...	...
	...	...	...	...	...	...
	...	...	...	...	...	...
	CN	...	...	...	...	ConteudoN_n
Ações	A1	xxx	...	...	...	...
	A2	xxx	...	...	...	...
	...	...	...	...	...	...
	...	...	...	...	...	...
	AN	...	xxx	...	...	...
Ações Adaptativas Antes	AA1	...	...	...	...	...
	AA2	xxx	...	...	...	...
	...	...	...	...	...	...
	...	...	...	...	...	...
	AAN	...	...	...	...	...
Ações Adaptativas depois	AD1	...	...	...	...	...
	AD2	xxx	xxx	...	...	...
	...	...	...	...	...	...
	...	...	...	...	...	...
	ADN	...	...	...	...	...

Fonte: Autor.

### 2.3.3 Regras de Associação

As Regras de Associação têm origem na área de mineração de dados com proposta em Agrawal; Imieliski; Swami (1993) e consolidada nos trabalhos apresentados por Agrawal; Srikant (1994, 1995, 1996), possibilitando um grande número de aplicações. As regras de associação são obtidas na análise de um conjunto de dados históricos (que pode ser uma série temporal) que correspondem a um determinado contexto, como por exemplo: nas compras realizadas por clientes em um supermercado envolvendo todos os itens; e como no consumo diário de energia por hora em uma unidade consumidora envolvendo o dia, a hora, o minuto e o valor do consumo, dentre outras possibilidades.

Geralmente, na análise de Regras de Associação é considerada a existência da transação de um registro que corresponde à ocorrência de um subconjunto de itens em um determinado momento, assim uma transação pode ser considerada como: uma compra com vários itens, uma medição da energia com diversos componentes, um conjunto de medições, por exemplo.

As transações armazenadas possibilitam a identificação de padrões que ajudam na inferência com relação às próximas transações, devido à criação de regras de associação, que comumente são apresentadas na seguinte forma: (Se X então Y), na qual X é chamado de antecedente e corresponde a um conjunto de valores que poderão conduzir a um conjunto de valores Y chamado de conseqüente. O conjunto de valores X também podem ser fatos correspondendo a leituras e Y a conseqüência desses fatos. Na realidade a ocorrência de X está vinculada a ocorrência de Y.

Com base em Faceli et al. (2011, p.181-184), Agrawal ; Imieliski; Swami (1993), Agrawal et al. (1994), e Wu et al. (2007), uma regra de associação é apresentada da seguinte forma:

Seja:  $I = \{i_1, \dots, i_n\}$  um conjunto de itens distintos,

$T = \{t_1, \dots, t_n\}$  um conjunto de transações composta por um conjunto de itens,

em que:  $T \subseteq I$ .

Sejam X e Y conjuntos específicos de itens tal que  $X \subseteq T$  e  $Y \subseteq T$ .

Uma regra de associação (Se X ENTÃO Y), também denotada por  $X \Rightarrow Y$ , e que  $X \subset I$  e  $Y \subset I$  e  $X \cap Y = \emptyset$ .

Considera-se a existência de algumas métricas como suporte, confiança e interesse, sendo que:

Suporte indica a porcentagem de ocorrência de X e Y em todas as transações existentes em T, assim quanto maior este, melhor valor. O suporte é dado por  $\text{Suporte}(X \Rightarrow Y) = (\text{quantidade de registros com X e Y}) / (\text{quantidade total de registros})$ ;

Confiança indica a frequência em que a ocorrência do conjunto de itens X implica na ocorrência do conjunto Y. Ou melhor, mede a frequência de itens em Y que aparece nas transações que contem X.  $\text{Confiança}(X \Rightarrow Y) = P(X \cup Y) / P(X) = (\text{quantidade de registros com X e Y}) / (\text{quantidade de registros com X})$ ;

Interesse ou *Lift* indica que quanto mais frequente se torna Y, quando X ocorre, pode ir de zero a infinito. Interesse ( $X \Rightarrow Y$ ) =  $P(Y/X)$  ou  $P(XY) / P(X)P(Y)$ . Em outras palavras, é um fator pelo qual a confiança excede a confiança esperada. Ele é determinado dividindo a confiança da regra pelo suporte do item pretendido. Corresponde a confiança  $(X \Rightarrow Y)/\text{suporte}(Y)$ ;

Convicção em uma regra mede o quanto ela é convincente. Corresponde ao quociente da frequência esperada de X ocorrer sem Y (é a frequência de erro da regra), é como se X e Y fossem independentes divididos pela frequência de previsões incorretas, conforme a seguir:

$$\text{Convicção}(X \Rightarrow Y) = (1 - \text{suporte}(Y)) / (1 - \text{CONFIANÇA}(X \Rightarrow Y));$$

Pode-se afirmar que quando a regra pertence a uma tabela de dados D com confiança c, se c% dos registros que contêm X também contêm Y;

Pode-se afirmar que quando a regra tem suporte s em uma tabela de dados D se s% dos registros em D contêm  $X \cup Y$ ;

$$\text{Nota-se que a confiança}(X \Rightarrow Y) = P(X \cup Y) / P(A) = \text{suporte}(X \cup Y) / \text{suporte}(A).$$

Com a finalidade de esclarecer o conhecimento formal é apresentado o exemplo a seguir.

Seja uma loja que vende somente seis itens {P, M, L, C, F, I}, e que ultimamente realizou cinco vendas: [{P, M, L}, {P, C}, {P, L}, {C, F, I}, {P, I, L}], na análise das vendas dos itens P e L, com base nas cinco ultimas vendas, e seja X e Y, respectivamente P e L, tem-se que:

$$\text{O (suporte para P e L)} = (\text{Registros com X e Y}) / (\text{Total de Registros}) = 3/5$$

e

$$\text{A confiança (Registros com X e Y) / (Registros com X)} = 3/4$$

Assim é obtida uma regra que:  $P \Rightarrow L$  com suporte = 0,6 e confiança = 0,75.

Afirma-se que, normalmente, as regras de associação visam encontrar os itens que possuem suporte maior ou igual ao suporte mínimo estabelecido e definido por um usuário e encontrar o conjunto de regras de associação com o maior valor de confiança que um mínimo estabelecido por aquele (FACELLI et al., 2011). Outro enfoque afirma que a obtenção das Regras de Associação consiste em: “encontrar conjuntos de itens frequentes, com frequência maior ou igual à especificada pelo usuário como sendo o suporte mínimo, não é trivial, devido à explosão combinatória ocorrida ao gerar os subconjuntos de itens. Mas, uma vez que os conjuntos de itens frequentes são obtidos, é muito simples gerar regras de associação com confiança maior ou igual à especificada pelo usuário como sendo o valor mínimo” (WU et al., 2007).

Para Wu et al. (2007), o algoritmo mais popular para a geração de Regras de Associação é o APRIORI, introduzido em (AGRAWAL; SRIKANT, 1994), no qual é empregada a busca em profundidade possibilitando ao algoritmo gerar conjuntos de itens candidatos (reconhecidos como o padrão) com k elementos a partir de conjuntos de itens de k-1 elementos. A varredura só termina no último elemento da base de dados e os padrões não frequentes forem descartados.

A ideia das Regras de Associação é utilizada com a finalidade de estabelecer uma base de conhecimento possibilitando originalmente o reconhecimento de padrões, podendo também ser utilizada para investigar não padrões, auxiliando no reconhecimento de anomalias.

#### **2.3.4 Sistemas Especialistas**

Os Sistemas especialistas utilizam o conhecimento de um domínio oriundo de pessoas e da documentação existentes em geral. Assim o esforço para obtê-los depende da disposição dos especialistas e de pesquisas na literatura. Com relação à quantidade de conhecimento necessário depende da aplicação, já uma das principais formas de validação do conhecimento é a comprovação em testes e a comparação com o conhecimento de

outros especialistas. O conhecimento é em geral organizado na forma de fatos e regras. São aplicados em diagnósticos, planejamento, previsão, controle, instrução, e interpretação.

Originalmente para Bittencourt (2006), as regras são pares de expressões consistindo em uma condição e uma ação, apresentadas por Emil Post em 1936, e utilizadas no chamado Sistema de Post, (POST, 1943). Mais tarde foi chamado de Sistemas de Produção, redescoberto na década de 70 do século XX para a modelagem do comportamento humano, passando também a ter uma vertente chamada de Sistemas Especialistas, com uma arquitetura contendo inicialmente três módulos: uma base de regras, uma memória de trabalho e um motor de inferência, baseado em lógica de primeira ordem. Outras versões apresentam aplicações com arquiteturas diferenciadas, envolvendo várias técnicas, inclusive com a combinação de métodos diferentes denominadas geralmente de híbridas.

Nas aplicações a representação de regra mais encontrada é do tipo “SE X ENTÃO Y”, e ela pode ser identificada por um nome (também dito identificador) e a parte “SE X” chamada de premissa ou antecedente; e se for verdadeira será acionada a parte “ENTÃO Y” chamada consequente. Cabe ressaltar que a parte consequente pode ser o antecedente para outra regra. Também é possível encontrar as regras da seguinte forma “Se X ENTÃO Y SENÃO Z”.

As regras são principalmente baseadas no conhecimento tácito de especialistas, que é definido como pessoal em Polanyi (1966) e utilizadas por Nonaka; Takeuchi (1995), pois proveem das experiências pessoais, que ocorrem devido à prática, à habilidade pessoal e as aptidões profissionais, e implica também em fatores comportamentais como postura frente a problemas, entre outros. Também podem ser baseadas no conhecimento explícito que está registrado nos mais variados tipos de mídias em publicações que envolvem normas, leis, regras organizacionais, histórias, dentre outras. Alguns dos conteúdos publicados também correspondem em alguns casos a conhecimentos do tipo tácito.

Bittencourt (2006) enfatiza que os Sistemas Especialistas, existentes além de utilizarem regras e realizarem buscas, incluem a justificativa (explicação) para a solução dada. Outrossim, atualmente, considera-se a incerteza com a utilização do grau de verdade

(pertinência ou confiança) tanto do lado do antecedente como do lado do consequente, além da informação do grau de verdade em tempo de execução pelo usuário. Desse modo, com os processamentos são realizados cálculos na composição de diversas regras de forma a obter um novo grau de confiança.

As regras obtidas de especialistas quando bem aplicadas facilitam a solução de problemas, podendo corresponder a correlações difíceis de serem percebidas imediatamente e de serem obtidas estatisticamente de forma imediata. Assim o uso de regras de especialistas podem economizar tempo e recursos quando comparadas com outras formas de obtenção de regras.

Ao processar um Sistema Especialista, os fatos e as regras geram uma inferência que imita um especialista. Em seu algoritmo pode existir o encadeamento para frente (*forward chaining*) ou para trás (*backward chaining*). Segundo Bitemcourt (2006), no encadeamento para frente, a premissa é comparada à descrição da situação atual; as regras nessa situação são selecionadas e a consequente é executada (introdução de novos fatos). Já no encadeamento para trás, o sistema é controlado por uma lista de objetivos; as regras que contêm esse objetivo na consequente possuem a premissa adicionada à lista de objetivos correntes. É comum utilizar encadeamento para frente em planejamento, projeto e classificação, enquanto que para diagnóstico é utilizado encadeamento para trás.

Sistemas Especialistas trabalham com o motor de inferência que, segundo Bitemcourt (2006), consiste em três fases: seleção das regras correspondentes às informações dadas, resolução de conflitos (separa as regras que realmente devem ser utilizadas, considerando critérios como: grau de pertinência, número de condições, dentre outros) e a execução propriamente dita das regras.

## 2.4 MONITORAMENTO DE BARRAGENS

A importância do uso de instrumentação em Barragens pode ser ilustrada conforme Silveira (2014), que analisou a ruptura da usina americana Teton em 1976, com base no conhecimento adquirido na instrumentação da Barragem brasileira de Água Vermelha. O



autor discute as possíveis posições de sensores e os fenômenos envolvidos que acabaram implicando na ruptura da Barragem de Teton. O autor cita que analistas americanos concluíram que:

Eram necessários instrumentos superficiais para a medição dos deslocamentos verticais e horizontais, medidores de recalque e inclinômetros para a medição dos deslocamentos internos do aterro, assim como piezômetros para a medição das pressões neutras no interior do aterro e da fundação. Além desses instrumentos mencionaram a importância da instalação de medidores de vazão, poços para a observação do nível d'água nas proximidades do reservatório, e instrumentos tais como acelerômetros para a realização de registros sísmicos. (SILVEIRA, 2014).

Os analistas citados também ressaltam a necessidade de manual de instruções de operação para inspeções de rotina e treinamento para ajudar na interpretação de condições potencialmente adversas, objetivando relatar prontamente qualquer condição anômala.

Fazendo um balanço das barragens existentes atualmente, percebe-se uma grande quantidade dessas instalações, e boa parte delas corresponde a barragens de hidroelétricas. Além disso, existem registros históricos de vários acidentes com barragens, como as de Teton. Os acidentes levaram ao surgimento de exigências técnicas e normas para a construção, manutenção e acompanhamento.

No caso específico de monitoramento em barragens é recomendado o posicionamento de diversos tipos de sensores no corpo da barragem, podendo até se constituírem em uma rede inteligente. Cada sensor gera sua própria série histórica de medições, que pode ser interpretada individualmente e/ou em composição com outros sensores.

Em seu estado normal as medições obtidas apresentam características próprias que podem sofrer modificações por interferências oriundas de diversos tipos de fenômenos que, em algumas situações podem representar perigo. As interferências podem ser oriundas de:

- Acomodação de terreno;
- Abalos sísmicos;

- Aumento na pressão da coluna d' água;
- Nível d' água;
- Nível pluviométrico;
- Erosão interna (*Piping*);
- Envelhecimento da barragem;
- Mudanças estruturais;
- Operação de equipamentos, dentre outros.

Há várias possibilidades de análise das medições obtidas, dentre elas existe as que envolvem características marcantes, conforme Fong; Nannan (2011), que são utilizadas na previsão de terremotos e sua influência em barragens, descritas no quadro 1.

Quadro 1 Características marcantes

Características Marcantes	Descrição
Nível	Valor médio da série temporal (variância).
Tendência	Através da varredura da série temporal, em longo prazo, ou análise de modificações dos dados.
Sazonalidade	Consistem na repetição de um comportamento, que pode ser observado várias vezes.
Ruído	Varição randômica
Casos isolados	Valores pendentes conhecidos como ruído branco ( <i>White noise</i> )
Autocorrelação	Consiste na correlação cruzada de uma série temporal consigo mesma, medida em autocovariância.
Estacionaridade ( <i>Stationarity</i> )	Medição de um processo estocástico pela troca com a distribuição probabilística conjunta.
Itinerário aleatório	Um fenômeno que troca uma série temporal (histórica) por uma distribuição probabilística onde as modificações nos dados são independentes uns dos outros. Neste caso uma modificação de dados passada de um conjunto de dados não consegue prever um movimento futuro. Os pontos de dados possuem uma forma aleatória e um caminho imprevisível

Fonte: Fong; Nannan (2011)

Dentre outras formas de análise, estão as que consideram sistemas inteligentes computacionais que utilizam o conhecimento existente organizado de forma conveniente para a realização de diagnósticos e de sugestões para a manutenção a partir de informações colhidas em inspeções/vistorias e em instrumentos de medição. Esses conhecimentos em geral estão relacionados entre si. Essa organização de conhecimentos constitui a chamada Base de Conhecimento. No item seguinte estão alguns exemplos de conhecimentos que podem ser utilizados.

#### **2.4.1 Exemplo de informações para Base de Conhecimentos**

A seguir são apresentados diversos conhecimentos nos quadros 2, 3 e 4, oriundos de problemas existentes em barragens hidroelétricas, do ponto de vista civil. Com base nessas informações, é possível criar regras, que correlacionam medições com anomalias. Essas regras podem ser utilizadas pelo Método Complementar que é capaz de justificar e correlacionar fenômenos. Nota-se que é possível que alguns fenômenos podem estar relacionados ao tipo de medição obtida a partir de um instrumento de medida.

Os conhecimentos relacionados envolvem: causas e falhas, materiais defeituosos e suas características observáveis, além de ocorrências genéricas e suas características observáveis.

No quadro 2 estão conhecimentos específicos para barragens, obtidos em PROÁGUA (2002) que contém informações macros relacionadas à causa e efeito organizadas em três colunas: “Falhas”, “Decorrem ou estão associadas”, e “causas”. Assim, dada uma causa, obtêm-se a falha e sua decorrência além da associação com outros fenômenos.

No quadro 3, adaptado de informações obtidas em PROÁGUA (2002), estão relacionadas categorias de causas e falhas. No quadro está relacionada a falha, que é a consequência de uma causa decorrente ou associada a outras situações, por exemplo a deterioração do concreto, que tem como causas a reação de álcalis/agregados, ou ao congelamento, ou ao degelo, ou ainda a lixiviação, o que pode ser decorrente de materiais defeituosos, agregados reativos, ou ainda a agregados de baixa resistência.

Quadro 2- Categoria de causas e falhas

<b>Falha</b>	<b>Decorrem ou estão associadas</b>	<b>Causa</b>
Deterioração da fundação	Qualidade e/ou tratamento das fundações. Apresentam rachaduras visíveis; afundamento localizado; retirada de materiais.	Remoção de matérias sólidas e solúveis; retirada de rochas e erosão.
Instabilidade da fundação	Materiais solúveis; xistos argilosos ou argilas dispersivas que reagem com a água.	Liquefação; deslizamentos; afundamentos e deslocamento de falhas.
Vertedouros defeituosos	Cheia de projeto; adequação de vertedouro; histórico de operação do vertedouro e do descarregador; obstruções; condição a jusante; crescimento da vegetação; fissuras e/ou rachaduras nas estruturas de concreto; equipamentos de má-condição de uso.	Obstruções; revestimentos fraturados; evidencia de sobrecarga da capacidade disponível e comportas e guinchos disponíveis.
Deterioração do concreto	Materiais defeituosos; agregados reativos; agregados de baixa resistência.	Reação álcalis/agregados; congelamento; degelo e lixiviação.
Defeitos de barragens de concreto		Alta subpressão; distribuição imprevista de supressão; deslocamentos e deflexões diferenciais e; sobrecargas.
Defeitos de barragem de terra e/ou enrocamentos	Estabilidade e sanidade das rochas do enrocamento; fraturamento hidráulico; rachaduras do solo; solos de baixa densidade.	Potencial de liquefação; instabilidade de taludes; vazamento excessivo; remoção dos materiais sólidos e solúveis e erosão do talude.
Defeitos de margens do reservatório	Erosões, deslocamentos de falhas; rupturas.	Permeabilidade; instabilidade e fragilidade inerente das barreiras naturais.

Fonte: Adaptado do PROÁGUA (2002)

No quadro 3, adaptado de informações obtidas em PROÁGUA (2002) estão relacionados materiais defeituosos.

Quadro 3 - Materiais defeituosos e Características observáveis

<b>Material Defeituoso</b>	<b>Características Observáveis</b>
Concreto	Reação agregado-álcali, aspectos estranhos e rachaduras; Lixiviação; Ação da geada; Abrasão; Lascamento; Deterioração geral; Perda de resistência.
Rocha	Desintegração; Amolecimento; Dissolução.
Solos	Degradação; Dissolução; Perda de plasticidade; Perda de resistência; Alteração mineralógica.
Solo-cimento	Perda de cimentação; Fragmentação.
Metais	Eletrólise; Corrosão; Corrosão sobtensão; Fadiga; Corte e ruptura; Esfoliamento.
Madeira	Apodrecimento; Encolhimento; Combustão; Ataque por organismos.
Tecidos de revestimento	Perfurações; Separação de uniões; Deterioração pela luz; Desintegração das vedações-limites; Perda de plasticidade e flexibilidade.
Borracha e elastômeros	Endurecimento; Perda de elasticidade; Deterioração pelo calor; Degradação química.
Vedações de juntas	Perda de plasticidade; Encolhimento; Derretimento.

Fonte: Adaptado do PROÁGUA (2002)

No quadro 4, adaptado de PROÁGUA (2002) estão relacionadas ocorrências genéricas quanto a suas características, localização e tempo de existência. Essas

ocorrências são de natureza universal, a despeito do tipo de estrutura ou classe de fundação.

Quadro 4 - Ocorrências genéricas e Características observáveis

Ocorrência	Características observáveis	
Geral	Percolação e vazamento; Relação descarga-nível; Aumentando ou diminuindo; Turvação e erosão interna ( <i>piping</i> ); Cor; Sólidos dissolvidos; Localização e formato; Temperatura; Gosto; Evidência de pressão; Bolhas; Tempo de existência e duração.	
Drenagem	Obstruções; Precipitados químicos e depósitos; Queda desimpedida; Disponibilidade de bomba de poço; Crescimento de bactérias.	
Cavitação	Picotamento de superfície; Evidência sonora; Implosões; Bolsas de vapor.	
Tensões e deformações – evidências e indícios	Concreto	Rachaduras; Esmagamentos; Deslocamentos; Desvios; Cisalhamentos; Fluência.
	Aço	Rachaduras; Estiramentos; Contrações; Dobramentos; Flambagens.
	Madeira	Esmagamento; Flambagem; Dobramento; Cisalhamentos; Alongamentos; Compressões.
	Rocha e nos solos	Rachaduras; Deslocamentos; Recalque; Consolidação; Afundamento; Compressão; Zonas de alongamento e compressão.

Fonte: Adaptado de PROÁGUA (2002)

#### 2.4.2 Instrumentos analisados nesta pesquisa

Para a obtenção de medições realizadas no monitoramento em hidroelétricas, existem vários tipos de instrumentos de medidas civis, tais como: triortogonal, piezômetro de tubo aberto, piezômetro de tubo fechado, medidor de vazão, extensômetro de haste, termômetro, dentre outros.

Nas barragens estudadas são considerados instrumentos manuais e automáticos microprocessados. Dentre as características consideradas nos instrumentos automáticos estão:

- Capacidade de integração à rede de comunicação e enlace para longa distância, com cabos protegidos de roedores, pisadas e umidades;
- Os sinais de saída dos sensores, que são convertidos para digitais e posteriormente, para medições, utilizando um conversor analógico digital robusto que suporta as condições operacionais da barragem (usina);
- As medições, que são enviadas a um computador central (servidor);

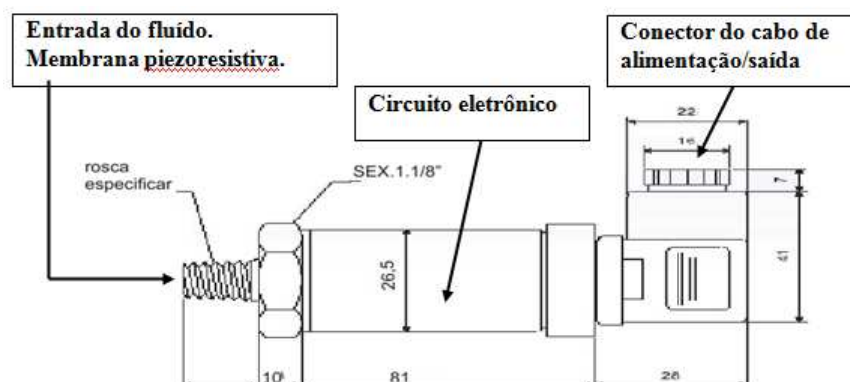
- Robustez para suportar as condições de extrema umidade e calor nas galerias da Usina e a incidência de radiação solar intensa na área externa;
- Proteção contra possíveis interferências eletromagnéticas provocadas pelos geradores;
- Baixo custo.

A seguir serão descritas as características básicas dos instrumentos de medição utilizados nesta pesquisa.

#### 2.4.2.1 Piezômetro de tubo fechado

Neste instrumento, é utilizado um sensor piezoresistivo e um condicionador/transmissor de sinais, conforme os catálogos da WÄRME (2005, 2016). A figura 6 apresenta características baseadas na mudança da resistividade de um material provocada pela compressão ou tração deste. Na prática é instalado em uma derivação de um tubo fechado com manômetro, alimentado por uma corrente de 24 Vdc. O sensor corresponde às faixas de operação, mantendo a sensibilidade e exatidão (de até 0,0001 kgf/cm<sup>2</sup>). Em geral nesse tipo de instrumento a cada local de medição da barragem é associada uma pressão média, um desvio padrão, um valor máximo e mínimo.

Figura 6 - Sensor de pressão piezoresistivo



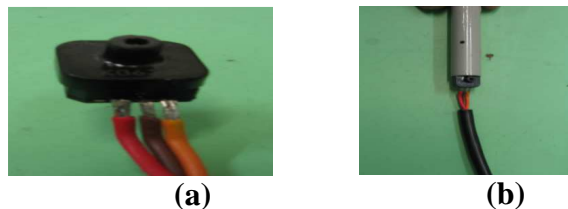
Fonte: WÄRME (2005, 2016)

Normalmente, o material mais utilizado no sensor de pressão por suas propriedades piezoresistivas é o silício, com dopagem adequada em certas direções cristalográficas. Cabe ressaltar a existência de outros fornecedores desse tipo sensor.

#### 2.4.2.2 Piezômetro de tubo aberto

Neste caso, também são utilizados um sensor piezoresistivo e um condicionador/transmissor de sinais. O sensor é baseado na tecnologia *Micro Eletrical Mechanical Systems* (MEMS) assentada na piezoresistividade, apresentando dimensões reduzidas, portanto, pode ser instalado em um tubo com diâmetro também reduzido (a figura 7a). Para ser instalado, o sensor MEMS é encapsulado de forma a prover a rigidez mecânica necessária, se houver aumento significativo nas suas dimensões (figura 7b).

Figura 7 - Sensor MENS: (a) não encapsulado, (b) encapsulado.



Fonte: Autor

#### 2.4.2.3 Medidor de Vazão

Utiliza as seguintes tecnologias para medição: sensor de pressão MEMS, sensor ultrassônico e sensor de boia.

O sensor MEMS, de pressão relativa, é semelhante ao dos piezômetros de tubo aberto, porém com uma maior precisão e menor faixa de medição. Mede a coluna d'água acima do sensor, possuindo duas tomadas de pressão e uma faixa de variação de pressão

bem menor do que o sensor MEMS de pressão absoluta. Dessa forma, pode-se apresentar mais exatidão na medida de coluna d'água.

O sensor MEMS de pressão relativa mede a pressão da coluna d'água (tomada de pressão P1) em relação à pressão atmosférica (tomada de pressão P2), ou seja, o sensor apresenta em sua saída a pressão exercida pela coluna d'água em relação à pressão atmosférica.

A figura 8 apresenta detalhes do sensor MENS código MPXV7002 e de sua montagem objetivando no medidor de vazão a coleta de dados da distância entre sua base e uma superfície plana. Na figura 9 está o sensor instalado para medição da vazão.

Figura 8 - Sensor tipo MENS código MPXV7002



Fonte: Autor; Freescale (2015)

Cabe ressaltar a existência de outros fornecedores, desse tipo de sensor, além do citado.



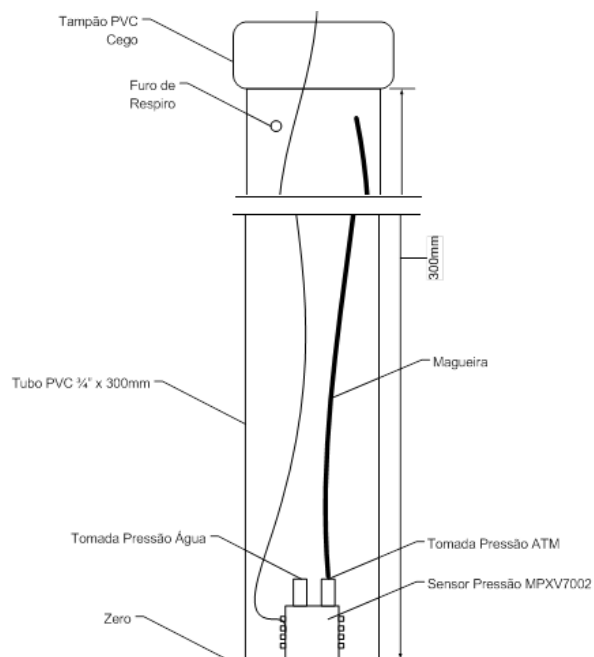
Figura 9 – Medidor de Vazão instalado



Fonte: Autor

A figura 10 apresenta detalhes do tubo de PVC e dos componentes do medidor de vazão, a saber: Tampão de PVC, furo de respiro, mangueira, tomadas de pressão e sensor MEMS.

Figura 10 – Medidor de vazão e seus componentes



Fonte: Autor

#### 2.4.2.4 Medidor de Junta Triortogonal

Os medidores triortogonais mesuram os deslocamentos em juntas entre dois blocos de concreto. Os deslocamentos são medidos em três dimensões triortogonais que representam:

- Abertura / Fechamento;
- Recalque / Elevação;
- Avanço / Recuo.

Os deslocamentos são da ordem de frações de milímetros, assim, sua leitura exige uma precisão de 0,01mm.

O sensor instalado na barragem da usina hidroelétrica de Lajeado utiliza o Efeito Hall, medindo o campo magnético de um ímã permanente, relacionando-o à distância do sensor ao ímã. O princípio de efeito Hall se baseia no surgimento de uma diferença de potencial gerado pela intensidade de um campo magnético. Desta maneira, como a intensidade de um campo magnético varia conforme a distância, o sinal de saída do sensor Hall é um valor de tensão elétrica proporcional à distância medida, com precisão aproximada de 3,5 microm/mV.

#### 2.4.2.5 Extensômetro de Haste

Utilizado para medir o deslocamento da barragem de concreto com referencia a uma base fixa com exatidão de 0,01 mm (10 micrometros). Na realidade mede o deslocamento relativo entre o maciço rochoso (no qual suas hastes estão ancoradas) e a fundação da barragem (onde se encontra instalado o sensor).

Tecnologias consideradas para os sensores são:

- Sensor de Efeito Hall – princípios descritos no medidor de junta triortogonal;

- Sensor Óptico – Este sistema é baseado no uso de um sensor de distância do tipo PSD (*Position Sensitive Detector*), o qual realiza a medição de distância entre o ponto de instalação do sensor infravermelho e uma superfície plana. O sensor é composto por um diodo emissor de luz infravermelha e um sistema de foto detecção constituído por uma lente e um foto transistor. Mas para as instalações não tem uma exatidão necessária.

Sensor de deslocamento Resistivo – baseia-se em um potenciômetro linear de precisão. Este potenciômetro possui uma haste atuada por mola e tem sua resistência variada linearmente com o deslocamento da haste, ou seja, aplicando-se uma tensão ao potenciômetro e obtendo-se uma tensão de saída proporcional à resistência, conseqüentemente ao deslocamento da haste.

## 2.5 CONSIDERAÇÕES FINAIS

Este capítulo contém os conceitos adotados na tese. De uma maneira geral, é dado enfoque à aplicação de regras de diversas formas, enfatizando as aplicadas nas tabelas de decisões e em sua forma adaptativa a TDA. Também enfoca as Regras de associação aplicadas em mineração de dados e as Regras de Produção aplicadas em diversos métodos inteligentes como os Sistemas Especialistas. Além disso, estão apresentadas no item de monitoramento de barragem informações que podem compor parte de uma base de conhecimento aplicada principalmente no Método Complementar.

### 3 ANÁLISE DOS ESTUDOS E DESENVOLVIMENTOS REALIZADOS

Este capítulo apresenta os estudos de métodos e artigos que possuem alguma relação com a pesquisa realizada. Tais estudos estão agrupados em dois itens, sendo que o primeiro apresenta testes e análises realizadas em diversos métodos e o segundo item, a síntese de artigos relacionados à linha de adaptatividade, seguida por esta tese, conforme apresentado no item que trata de Tabela de Decisões Adaptativa.

#### 3.1 ESTUDOS QUE NÃO ENVOLVEM TDA

Durante a pesquisa, vários métodos foram replicados, adaptados e em um caso alterado para atingir o objetivo deste trabalho. Os detalhes da análise de cada método estão apresentados no Apêndice B.

A pesquisa realizada considera um processamento *on-line* não supervisionado, com medições recebidas em tempo real de forma a constituir uma série histórica de dados oriundos de um instrumento de medição para a aplicação dos métodos com a finalidade de reconhecer um valor não conforme com o comportamento histórico dos dados.

Na pesquisa, também é destacada a capacidade de adaptação de cada método, resultando o resumo apresentado no quadro 5 e contendo: os métodos, características básicas, objetivo e se é ou não adaptativo.

Dentre os métodos apresentados no quadro 5, dois deles se destacam e se mostram promissores devido ao fato de reconhecerem os pontos extremos durante as simulações: os métodos HOT SAX e o de (BASHA; AMMEEN, 2007), sendo detalhados no Apêndice B. Cabe ressaltar que nos métodos pesquisados, é possível perceber que nem sempre os valores não conformes são reconhecidos adequadamente. Para tanto, alterações estão implantadas no método HOT SAX Modificado apresentado no Apêndice B.

Quadro 5 - Análise enfocando a capacidade de adaptação dos métodos

<b>Método</b>	<b>Características</b>	<b>Objetivo</b>	<b>Adaptativo</b>
Hot Sax	Estatística (tem normalização) e velocidade do algoritmo	Discórdia, Anomalias	Não, somente encontra uma discórdia.
WAT	Estatística (herança do HOT SAX) e <i>Wavelet</i> .	Discórdia, Anomalias	Sim, encontra o top-k discórdias.
Fu pré WAT	Estatística (herança do HOT SAX) e <i>Wavelet</i> .	Discórdia, Anomalias	Não, só encontra uma discórdia.
Toshniwal e Yadav	Estatística (herança do HOT SAX, além de <i>quartis</i> , <i>Outliers</i> ).	<i>Outlier</i>	Sim
Knorn e Leith	Estatística (filtro Kalman)	Anomalias	Sim
K-Means+ID3	Mineração de dados (K-Means) e Inteligência artificial (árvore de decisões).	Aprendizado, Anomalias	Não
Mahoney e Chan	Estatística (filtro Kalman)	Anomalias	Não
Koushanfa e Sangiovanni-Vincentelli	Estatística (do tipo não paramétrica)	Testes <i>on-line</i> em redes de sensores, Falhas.	Não
Basha e Ammeen	Estatística (teste-t, ki-quadrado).	Subseqüências não usuais, discórdias em séries com periodicidade.	Não

Fonte: Autor

### 3.2 ESTUDOS DE ARTIGOS QUE ENVOLVEM TDA

Os títulos dos subitens a seguir coincidem com o nome dos artigos sintetizados. Neles é observada a utilização de um dispositivo subjacente em conjunto com uma TDA e o uso de termos como a Internet das Coisas Adaptativa, além da Mineração de Dados Adaptativa.

Nos itens estão apresentadas aplicações e/ou suas propostas que envolvem TDA em diversas áreas do conhecimento técnico, reforçando que esta tese enfoca uma nova área para a utilização da TDA.

### **3.2.1 Uso de métricas de roteamento através de Tabela de Decisão Adaptativa para redes de sensores sem fio**

A proposta apresentada por Miguel; Oliveira (2017) visa à administração de energia de uma Rede de Sensores Sem Fio (RSSF). A rede é composta por nós que possuem pequeno ou nenhum poder de processamento, com memória insuficiente para comunicação com a internet e/ou IoT.

Normalmente uma das possibilidades de controle de energia é ligar e desligar periféricos, sensores, rádios e até o processador (se existente no nó). Também é possível equilibrar o roteamento priorizando o menor caminho, o que corresponde a um menor consumo de energia. Esse controle evita que um nó esgote a energia antes de outro e minimize a existência de nós e regiões inacessíveis, sem comunicação.

Dentre as aplicações da RSSF estão: ambientes industriais, acompanhamento médico, atividades de detecção, dentre outras. Esse conjunto de nós pode ser integrado de alguma forma com a Internet e em fase avançada com a IoT.

Na proposta apresentada e ainda não aplicada por Miguel; Oliveira(2017) considera uma fórmula que utiliza a energia média  $M$  da rede, e a energia mínima  $\min M$  e a máxima  $\max M$  encontradas em algum nó da rede (não necessariamente no mesmo nó). Também é utilizada uma TDA com base em funções adaptativas (com variáveis, parâmetros, e geradores), além de considerar as métricas ETX inicialmente. A figura 11 corresponde um exemplo da parte adaptativa da TDA utilizada, com definição de funções adaptativas.

Figura 11 - Exemplo de parte adaptativa de uma TDA

	H	..	..	..	..+	..+	..+	H	..	..	..	..+	..+	..+
State -		1	1	1	1	1	1		1	1	1	1	1	1
Carga <= 18		N	N	S	N	N	S		N	N	S	N	N	S
Carga > 18		S	S	N	S	S	N		S	S	N	S	S	N
Tempo > 300s		N	S	N	N	S	N		N	S	N	N	S	N
State :-		1	1	0	1	1	0		1	1	0	1	1	0
Tempo :-		0	0	0	0	0	0		0	0	0	0	0	0
ETX :-		80%	60%	100%	60%	40%	100%		60%	40%	100%	40%	20%	100%
Energia :-		20%	40%	0%	40%	60%	0%		40%	60%	0%	60%	80%	0%
K	A		V											
L						V		A		V				
M													V	
N														
O				V			V				V			V
p1				p1		60%	60%				p1		40%	40%
p2				p2		40%	40%				p2		60%	60%
p3				p3		40%	40%				p3		20%	20%
p4				p4		60%	60%				p4		80%	80%

Fonte: Miguel; Oliveira (2017)

### 3.2.2 Aplicação de Tabelas de Decisão Adaptativas em sistemas de controle de crescimento de plantas

A proposta apresentada em Moreira; Barretto (2017) tem como foco a automação agrícola no estudo de um sistema para o crescimento de plantas em um ambiente controlado (em malha fechada, *plant factory*), com a finalidade de obter alta produtividade e baixo consumo de recursos. Sua aplicação está prevista para uma instalação com Aeropônia (semelhante à hidropônia, só que com irrigação sem solo; no caso também é utilizada a aeração com água) com alto nível de controle. Um sistema desse tipo considera energia, água e insumos, e pode ter subsistemas (como: iluminação artificial, irrigação, visão) além de subsistemas de manejo de soluções nutritivas na irrigação e aeração, inicialmente coletiva e evoluindo para uso individual para cada planta. O estágio mais avançado de um sistema desse tipo é a autonomia.

A aplicação de Moreira; Barretto (2017) consiste em um modelo inicial para o manejo de patógenos limitantes da cultura da batata, especificamente para o manejo preventivo de requeima (*phytophthora infestans*) em uma planta com Aeropônia, envolvendo fungicidas (3 tipos e suas combinações). Para tanto utiliza uma TD para uma escolha possível utilizando um modelo de TDA para ações adaptativas, antes de utilizar as

ações normais, considerando três estágios com conteúdo verdadeiro (V), falso (F), indiferente (---), e ação a ser executada (X) semelhante à apresentada na figura 12, adaptada de. Moreira; Barretto (2017).

Figura 12 - Exemplo de uma TDA utilizada

				Regras			
				R1	R2	R3	R4
TD	Condições	c1	VS > 24	---	V	---	---
		c2	DAT > 65	V	---	---	---
		c3	Infestação emergencial?	---	F	V	F
		c4	Falha no Isolamento?	---	F	F	V
	Ações normais	a1	Manejo fungicida A		X		
		a2	Manejo fungicida B			X	X
TDA	Ações adaptativas	ba2	Ação preventiva intermediária		X	X	X
	Anteriores	ba3	Ação preventiva final	X			

Fonte: Adaptado de Moreira; Barretto (2017)

Ressalta-se que Moreira; Barretto (2017) indicam à existência na a Coreia da utilização da Iot (*Internet of Things*, internet das coisas) em uma *Smart Farm* com sensoriamento expandido.

### 3.2.3 Operação sustentável de sistemas de ar condicionado usando Tecnologia Adaptativa

A aplicação apresentada em (MARÈ et al., 2015) utiliza Árvores de Decisão Adaptativas (ADA) para a criação de um classificador das condições do ambiente interior relativas ao conforto térmico e à qualidade do ar, para auxiliar na definição de estratégias de operação sustentáveis de sistemas de ar condicionado central, minimizando-se operações manuais e intervenção humana. No método são utilizadas ferramentas IAD (Indução por Árvore de Decisão) como única função adaptativa de uma ADA, mostrando-se como uma solução simples, de fácil implementação e eficaz. Porém, segundo os autores, esta solução consome mais recursos computacionais e aumenta o tempo de execução, comparativamente a uma AD típica.

Entre os conceitos utilizados estão: a síndrome do edifício doente, influência do ambiente exterior, alterações nas funções de controle, conjunto de pontos que definem a



função de controle. A IAD é aplicada em diagnóstico de falhas, detecção de eventos, e explicação e avaliação de fenômenos ambientais.

São realizadas diversas análises e atividades, por exemplo:

- Verificação da adequação das condições ambientais frente aos parâmetros vigentes em normas;
- Comparação dos sistemas de ar condicionado a partir da eficiência na promoção de temperatura, umidade relativa do ar e teor de CO<sub>2</sub> conformes;
- Identificação de padrões de variação das grandezas monitoradas em cada ambiente, ao longo do tempo, em função da ocupação e das estratégias de operação adotadas para cada sistema; análise cruzada de informações de qualidade do ambiente interior e consumo de energia para cada sistema de ar condicionado.
- Monitoração dos consumos de energia de Fan coils e Chiller.

Em sua operação ocorre a coleta contínua de dados de uma rede de sensores nos ambientes internos e externos de uma sala. As leituras são realizadas a cada 30s, compondo uma massa considerável de dados. O conteúdo utilizado para operacionalizar o sistema apresentado por (MAREÈ et al., 2015) é determinado a partir do treinamento de uma ADA com base em uma IAD, sendo suas regras inseridas em uma TDA. Em (WEKA, 2015) há bibliotecas com algoritmos de aprendizagem, que podem ser utilizados. Sempre que os resultados não forem compatíveis com o conforto esperado pelos ocupantes, uma nova ADA é desenvolvida, gerando novas classificações. Como exemplo o artigo apresenta a aplicação em duas salas. Na prática é utilizada uma TD (Tabela de Decisões) convencional com uma camada adaptativa contendo uma IAD para gerar Tabelas Adaptativas. Nas características do classificador para a qualidade do ambiente interior (usadas na construção da árvore) está a temperatura do ar e teor de CO<sub>2</sub>.

### **3.2.4 Modelo de um sistema de conservação de alimentos baseado na IoT-A e seleção de elementos para Tabela de Decisão Adaptativa**

A aplicação apresentada em (KAWANO et al., 2015) envolve o monitoramento agrícola, com o acompanhamento e a observação contínua de uma área de plantio, objetivando avaliar as mudanças ocorridas nesse ambiente. Na aplicação, é utilizada a Internet das Coisas com sensores de radiofrequência. Como esta utiliza em sua monitoração dispositivos adaptativos, ela é dita adaptativa, IoT-A. Nesse monitoramento é importante para o processo de tomada de decisão e auxilia na solução de problemas, como ataques de pragas e doenças, correção do solo, aplicação de insumos e mudanças climáticas que podem prejudicar a produtividade da plantação.

A aplicação considera que a modificação de dados está relacionada ao intervalo dinâmico de amostragem de dados. Assim, utiliza um intervalo longo entre os dados quando ocorre uma baixa variação da grandeza monitorada e um intervalo curto entre os dados quando eles representarem fenômenos que merecem maior atenção (ex: frio ou calor excessivo ou variações repentinas nos dados monitorados). Essas considerações podem representar economia de energia (pois o tempo de comunicação será menor) em alguns tipos de sensores. Em (KAWANO et al., 2015) é enfatizado que no caso dos dados variarem lentamente, não há necessidade de registrá-los quando há alta frequência, economizando energia e evitando registro redundante de dados.

A aplicação desenvolvida pelos autores utiliza duas estratégias. Na Estratégia 1, é considerado um “padrão de normalidade” (valores esperados), que faz com que a rede de sensores adote intervalos de amostragem de dados longos enquanto as informações obtidas correspondem ao padrão. Quando ocorre a leitura de dados fora do padrão de normalidade, os nós da rede reduzem o intervalo de amostragem, de modo a registrar os dados mais frequentemente. A Estratégia 2 é considerada quando ocorrem variações consistentes nos dados, reduzindo o intervalo e aumentando a frequência dos registros de dados, enquanto que baixas variações determinam intervalos maiores. A grandeza da variação dos dados influenciará na determinação do intervalo de amostragem.

Para controlar a dinamicidade dos intervalos de registro de dados são utilizados Autômatos Adaptativos (AA). Cada nó de uma rede de sensores sem fios tem seu AA. Durante uma transição adaptativa, um AA sofre alguma mudança em sua configuração (eliminando, acrescentando, ou simplesmente modificando estados e transições). Isto faz com que uma nova máquina de estados apareça, no lugar da anterior, caracterizando a execução de um passo adicional. Uma TDA representa as regras de uma função adaptativa e, conseqüentemente, toda a lógica do AA.

Um exemplo de uso de um AA envolve o que ocorre na Estratégia 1, na qual existem as seguintes condições:

- Caso o dado coletado pelo nó sensor pertence ao padrão de normalidade (que pode ser um intervalo), o AA transitará para o estado correspondente ao maior intervalo de amostragem de dados;
- Caso o dado coletado se mantenha estável, então o AA permanecerá no estado corrente;
- Caso o dado coletado pelo nó sensor extrapola os limites do padrão de normalidade, o AA transitará para um estado que corresponde a um intervalo de amostragem de dados menor, por meio de uma função adaptativa;
- Caso o valor do dado coletado retorne ao padrão de normalidade, o AA transita de volta para o estado que corresponde ao maior intervalo de amostragem de dados.

Na prática, são utilizados AA com suas regras representadas por uma TDA.

### **3.2.5 Mecanização da aprendizagem com dispositivos adaptativos: conceitos e aplicação**

No trabalho apresentado em (STANGE; NETO, 2013) os autores investigam questões relacionadas à utilização da adaptatividade no processo de aprendizagem de

máquina, tais como: mecanização da aprendizagem, representação do conhecimento, inferência e tomada de decisão. Para isso, é proposta a utilização de dispositivos adaptativos para representar o conhecimento adquirido através da aprendizagem incremental, no jogo da velha.

Para os autores, a aprendizagem de máquina utilizando tecnologia adaptativa pode ser considerada uma técnica inspirada na aprendizagem indutiva supervisionada com o objetivo de melhorar o entendimento das regras, mas principalmente para cumprir exigências de eficiência. Eles enfatizam ainda que, na técnica, uma base de regras é utilizada para representar o comportamento inteligente de um jogador em função de experiências passadas.

Em seu conteúdo, são utilizados: Autômatos Adaptativos AA (na representação de regras aprendidas), com AF (Autômato finito com aceita/rejeita, lógica binária), Máquina de Mealy e, por fim, a TDA para a representação da memória do mecanismo de aprendizagem.

Na aplicação, o conjunto de transições (jogadas) resultante de cada partida é inserido em uma Tabela de Decisão Adaptativa, porém, é relevante destacar que o próprio autômato adaptativo pode representar o conjunto de regras adquiridas. Apesar disso, diferentes dispositivos adaptativos podem apresentar uma solução mais aderente a um problema em particular. Desse modo, a TDA representa a memória do mecanismo de aprendizagem. Outra motivação para a representação utilizando uma TDA está na maior flexibilidade de forma a permitir consulta às regras, além de inclusão e exclusão de regras durante a operação do dispositivo.

### **3.2.6 Descoberta de padrões em bases de dados utilizando Técnicas Adaptativas**

Tchemra; Carmargo (2009) apresenta uma aplicação de Mineração de Dados Adaptativa com o uso de TDA juntamente a regras de associação, na qual o processo consiste na descoberta de padrões e no uso de técnicas de mineração de dados agregadas ao conceito de adaptatividade. Para tanto, utiliza uma base de dados organizada de forma

multidimensional. Nesta aplicação, a tomada de decisão utiliza os padrões descobertos, que representam os critérios de um problema de decisão, como entrada numa tabela de decisão adaptativa baseada em métodos multicritério e técnicas adaptativas.

O modelo de Mineração de Dados Adaptativo (MDA) proposto é definido formalmente com o emprego da sêxtupla:  $MDA = (BD, TC, TDM, PD, BC, FA)$ , cujos elementos são descritos a seguir: BD é uma base de dados composta pelos registros de interesse para a mineração de dados; TC é o tipo de conhecimento a ser descoberto; TDM é a técnica de mineração de dados a ser utilizada; PD é o conjunto de padrões descoberto; BC é a base de conhecimento resultante; FA representa o conjunto de funções adaptativas que forma a camada adaptativa do modelo.

### 3.3 CONSIDERAÇÕES FINAIS

Apesar de alterar um método de forma a torná-lo operacional para os requisitos estudados, percebeu-se que as condições estabelecidas nos métodos originais não identificavam todas as possibilidades de não conformidades. Além disso, as possibilidades de aplicação da Tabela de Decisões Adaptativa podem ser como um mecanismo adaptativo principal e/ou como uma base de dados de regras.

Desse modo, a pesquisa se desdobrou no desenvolvimento de dois métodos. Sendo que o segundo consiste em um Método Complementar para melhorar os resultados apresentados por outros métodos e o primeiro, em um método capaz de se adequar de forma inovadora e original aos requisitos básicos utilizados, constituindo-se no MAIM - Método Adaptativo de Interpretação de Medições.

## 4 MÉTODOS DESENVOLVIDOS

O presente capítulo apresenta o objetivo principal desta tese que é o Método Adaptativo para a Interpretação de Medições (MAIM) além do Método Complementar que é uma consequência das pesquisas realizadas com os resultados oriundos de outros métodos.

O ambiente de desenvolvimento utilizado está apresentado no apêndice C nos itens item C1 para *hardware* e C2 para *software*.

Para chegar aos protótipos com as considerações dos métodos desta tese, são utilizadas provas de conceitos, que correspondem às “evidências que estabelecem que uma ideia, invenção, processo ou modelo de negócio é viável” (BUSINESSDICTIONARY, 2017). Na realidade, o termo também é muito empregado com características próprias na Tecnologia de Informação, Medicina, Licitações, Escritórios de Patentes, dentre outras áreas do conhecimento.

Dessa forma, as provas de conceito são aplicadas para ajudar na investigação de alguns requisitos para os modelos desenvolvidos e demonstrar a viabilidade de alguns conceitos que culminam com a aceitabilidade e aplicabilidade deles, pois podem ser ajustados durante os testes.

As provas de conceito estão utilizadas para Agrupamentos e Novidades (item 4.1.2), Estabilidade e Ambiente (item 4.1.3), e Modulo adaptativo (Item 4.1.7).

Neste capítulo, inicialmente está apresentado o método MAIM e, posteriormente, o Método Complementar.

### 4.1 MÉTODO ADAPTATIVO PARA A INTERPRETAÇÃO DE MEDIÇÕES - MAIM

O MAIM foi desenvolvido com a finalidade de identificar medições em séries históricas, que podem ser categorizadas como candidatas a anomalias ou anomalias (não conformidades). Após o reconhecimento da medição e sua categoria, caso necessário,

pode-se interpretar e enviar mensagens ou alertas a um grupo de pessoas interessadas ou ao Método Complementar, que pode justificar a mensagem e correlacioná-la a outros dados, dando nova interpretação ou recusá-la. Para atingir esse objetivo básico de forma original e prática, foi estabelecida uma base de conhecimento montada em um banco de dados relacional.

A base de conhecimento é composta: pela própria série histórica, regras, grupos de conteúdos consecutivos e sua quantificação, dentre outros. Com relação às regras, parte delas está organizada de forma semelhante aos princípios apresentados no item 2.4.4 que trata de tabelas adaptativas. A principal diferença é que os conceitos estão distribuídos em um modelo de dados relacional operado segundo os conceitos apresentados de uma TDA, contendo todas as regras, incluindo suas condições, ações, e componentes adaptativos, tais como funções, parâmetros, variáveis e geradores de regras.

Para tanto são necessários os conceitos de Novidade (primeira ocorrência de um valor) e Ambiente evidenciados de forma original com considerações relativas a um sistema adaptativo.

#### **4.1.1 Conceitos utilizados de Novidade e Ambiente**

Os conceitos de Novidade e Ambiente constituem uma nova abordagem e são considerados de forma original no método desenvolvido nesta Tese conjuntamente às questões que envolvem sistemas adaptativos.

##### **4.1.1.1 Novidade**

Para Ferreira (2015), novidade, em um de seus significados, corresponde a uma alteração inesperada no andamento regular das coisas.

Segundo Paim (2007), as novidades ou anomalias em uma série temporal podem ser vistas como valores ou como uma sequência de valores inesperados quando comparados a um conjunto de valores considerados normais.

Keogh; Lin; Fu (2005) utiliza o HOT SAX para obter discórdias em séries temporais, sendo que, para os autores, discórdia é definida como a subsequência da série que mais se difere do restante das subsequências, dada uma métrica de distância. Com relação à novidade (KEOGH; LIN; FU, 2005), eles utilizaram seu algoritmo para detectar novidades em séries temporais, definindo como novidades as “n” subsequências com maior distância.

Para este trabalho, o termo novidade corresponde: a uma informação que ocorreu uma única vez em um ambiente ou a primeira vez da ocorrência de um valor. Assim, a primeira ocorrência equivale a uma novidade e, se houver mais ocorrências dessa informação, ela deixa de ser novidade.

#### 4.1.1.2 Ambiente

A utilização das considerações a seguir foi motivada pelas observações em gráficos de algumas séries históricas em que as medições ao longo do tempo ficavam dentro de um conjunto valores e, repentinamente, mudavam para um novo conjunto de valores. Para a ocorrência dessa mudança, existiam várias possibilidades. Um exemplo corresponde ao caso que ocorreu (em usinas estudadas nesta pesquisa) com a mudança da estrutura de apoio de um instrumento de medida, que alterou a faixa de medidas obtidas. Dessa forma, foi considerado que a mudança em um ambiente deve ser avaliada para a interpretação de resultados. Para tanto, foram investigadas outras considerações com o ambiente, conforme segue.

Na visão da realidade estabelecida pela Teoria Geral dos Sistemas, Martinelli et al. (2012) apresentam que tudo que extrapola a fronteira de um sistema fará parte do ambiente, mas deve ser considerado somente aquilo que tem alguma chance de interferir no sistema e de lhe impor restrições, estabelecendo assim o contexto do sistema.



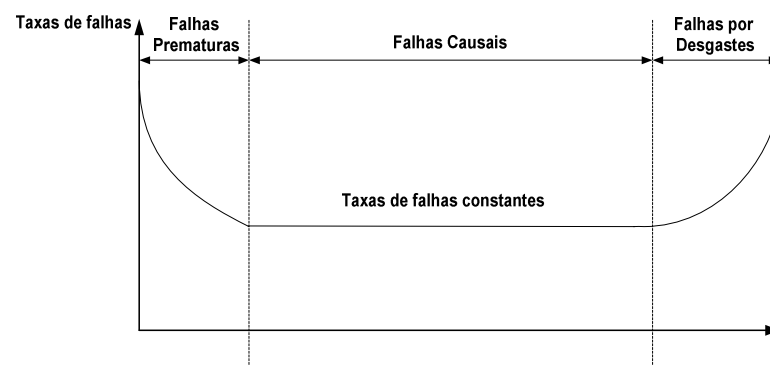
Para Ferreira (2015) ambiente é “aquilo que cerca ou envolve os seres vivos, ou as coisas, por todos os lados”. Para este trabalho, o que envolve um ponto de medição é constituído pelo ambiente físico propriamente dito, pelo conjunto de medições realizadas, pelos conceitos pertinentes e pelo ambiente virtual, que pode ser registrado em uma base de conhecimentos, organizada em um conjunto de tabelas de um banco de dados relacional. Sobre essa base de conhecimento é concebível realizar buscas de informações e realizar inferências de forma a realizar um aprendizado.

Esse ambiente é variável, pois as modificações ocorridas conduzem a novas situações, especialmente com relação às medições realizadas, podendo acontecer inferências diferentes. Assim, para efeito de estudo, são considerados os seguintes estados de um ambiente: estável, em evolução para uma estabilidade, e novo (que passa a existir após a quebra da estabilidade existente anteriormente). Esses três estados de um ambiente podem ser percebidos visualmente em algumas séries históricas como a seguir.

As figuras 13, 14, e 15 apresentam gráficos que indicam mudanças de ambiente, segundo o critério adotado.

Na figura 13, está Curva da Banheira, conforme Lafraia (2014), nela estão representadas as falhas que podem ocorrer na vida de equipamentos, componentes, *software*, dentre outros.

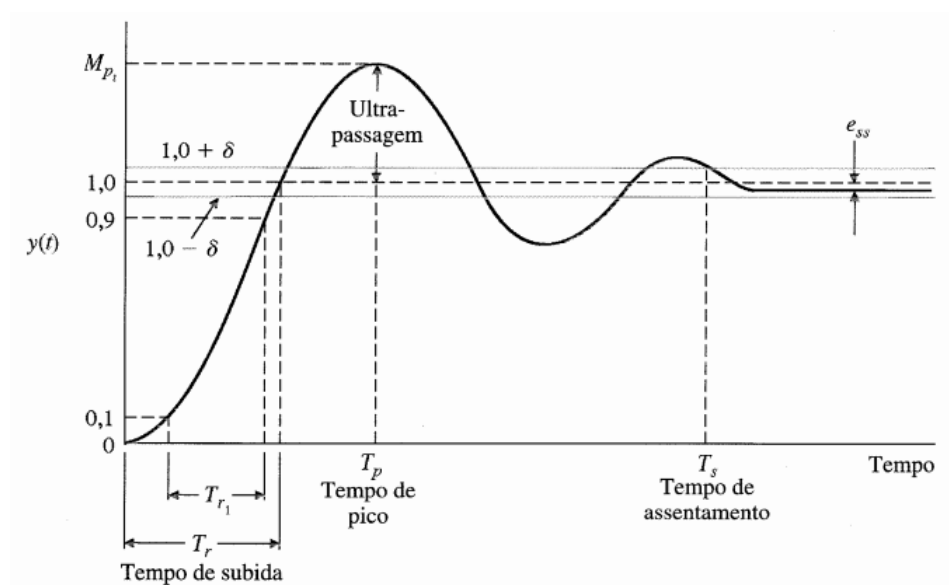
Figura 13- Curva da Banheira



Fonte: Lafraia (2014)

Na figura 14, está representada a curva típica de resposta no tempo e seus parâmetros que, na grande maioria das vezes, envolve valores de tempo pequenos. Na figura 14, está indicada uma faixa percentual de  $\pm\delta$  que, após o tempo de assentamento, os valores de resposta devem permanecer. Na realidade indica uma situação de estabilidade.

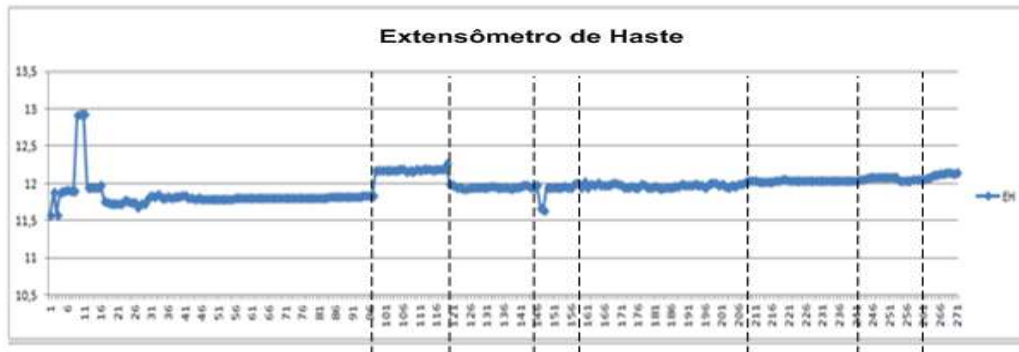
Figura 14 - Curva tempo resposta



Fonte: Dorf; Bishop (2013)

A figura 15 corresponde a uma curva referente a um extensômetro de haste, instalado em uma hidroelétrica. As medições (em mm) foram obtidas de 13/08/2001 á 26/08/2010 em períodos de 15 dias, numerados sequencialmente de 1 a 271. Ainda na figura 15, está uma divisão visual, segundo critério adotado neste estudo, apresentando os ambientes reconhecidos, entre linhas tracejadas. O primeiro ambiente contém a acomodação de terreno (que pode ocorrer no início das operações de uma barragem) tendo em sua maior expressão um patamar com os maiores valores do gráfico. Essa primeira divisão, está entre as sequencias 1 e 96, que correspondem a 13/08/2001 e 08/07/2003. Após a ocorrência de várias novidades (ambiente em evolução), verifica-se uma estabilidade na qual as medições são repetições.

Figura 15 - Curva com ambientes reconhecidos separados por linhas tracejadas (valores em mm)



Fonte: Autor

Os próximos itens apresentam testes envolvendo agrupamentos, novidades, estabilidade e ambiente.

#### 4.1.2 Prova de conceitos envolvendo agrupamentos de medições e novidades

No método MAIM ocorre o recebimento das medições e a identificação de novidades bem como o agrupamento dessas medições.

As novidades são utilizadas para a identificação dos estados de um Ambiente (novo, em evolução, e estável) e o agrupamento de medições estão relacionados às sequencias de entradas das medições, podendo ser utilizados no registro das novidades e quando possível, na previsão de sequencias segundo suas ocorrências.

Nos agrupamentos podem ser registradas sequencias de medições que contêm de uma até infinitas medições, desse modo, torna-se necessário saber como ocorre a ocupação do espaço computacional desses dados.

Para tanto, considera-se a recepção das medições periodicamente, registrando suas ocorrências. Assim, cada medição tem sua ocorrência atualizada, de modo que, na primeira ocorrência seu registro é criado com o valor numérico 1 indicando novidade, e nas

próximas ocorrências é aumentada a contagem de 1, em seu registro histórico, deixando de ser novidade.

Durante o processamento, as medições são reunidas em grupos contendo de uma até cinco medições (poderiam ser mais), seu registro é feito, como o explicado, para a frequência de uma medição. Um grupo pode ser considerado novidade quando seu valor for 1. Os grupos são criados e registrados na sequência de entrada das medições.

Para efeito de identificação e explicação, os agrupamentos serão denominados por um código composto pelo prefixo “G” seguido pela quantidade de elementos do grupo indo de 1 até N, sendo limitados nestes testes para 5. Portanto os grupos considerados são G1, G2, G3, G4 e G5.

Considere, como exemplo, que já existam cinco medições registradas historicamente com os seguintes valores (coincidindo com a ordem de chegada): 1, 2, 3, 4, 5. Seus agrupamentos, registrados na base de conhecimento, serão: G1 {1}, G2 {2,1}, G3 {3, 2, 1}, G4 {4, 3, 2, 1}, G5 {5, 4, 3, 2, 1}. Em um registro histórico está indicado o agrupamento, o número de ocorrências e a data/hora, por exemplo, o registro fictício G3 {293, 278, 341} com 37 ocorrências até o dia 17/05/2010 às 16h00 min.

Assim, neste item, serão investigadas a obtenção de novidades e o agrupamento de medições, utilizando séries históricas reais. Para tanto, duas validações são utilizadas.

A primeira validação consiste em saber se a quantidade de agrupamentos obtidos é muito menor que a quantidade de agrupamentos possíveis indicando que não há necessidade de armazenar todos os registros possíveis, utilizando espaço para registros e tempo de processamento desnecessariamente.

A segunda validação visa verificar a diminuição da ocorrência de novidades ao longo do tempo. Além disso, visa a responder, se a melhor forma de novidade deve considerar somente uma nova medição ou combinações de uma a cinco medições ou mais.

Na prova de conceito, foram utilizadas as séries históricas relativas à leitura de sensores utilizados na monitoração de uma barragem hidroelétrica, a saber: Triortogonal,

Medidor de Vazão, e o Piezômetro com alimentação de energia por painel solar. As medições existentes nas séries históricas foram fornecidas para o processamento computacional uma de cada vez, da mais antiga para a mais recente, simulando o recebimento de informações e seu registro em tempo real.

Nas provas de conceito são apresentadas a curva da leitura (figuras 16, 18, 20), e uma tabela contendo a indicação do tipo de grupo (tabelas 1, 3, 5), a quantidade de elementos nesse tipo de grupo, quantidade de grupos formados e a quantidade de arranjos possíveis para a formação de grupos. Essas tabelas são obtidas em consultas do tipo SQL realizadas no gerenciador de banco de dados. A quantidade de novidades obtidas coincide com o número de grupos formados, de modo que sempre um grupo será novidade em seu primeiro registro. Para o cálculo dos arranjos possíveis, foi considerada a fórmula geral para arranjos simples que corresponde a  $A_{n,p} = n!/(n-p)!$ . Na fórmula: “A” corresponde ao arranjo, “n” ao número de elementos, “p” ao número de elementos arranjados.

Também está retratada a correspondência entre o traçado gráfico da série histórica e um histograma com a frequência dos valores lidos. Outra investigação identifica as novidades ao longo do tempo, com o uso de consultas SQL, em função da data de surgimento da novidade.

Enfatizando que as provas de conceito dos próximos itens verificam principalmente a correspondência entre a quantidade grupos de arranjos formados e a quantidade de grupos de arranjos possíveis. Além disso, constata a correspondência dos gráficos do tipo histograma com o gráfico dos Valores x tempo.

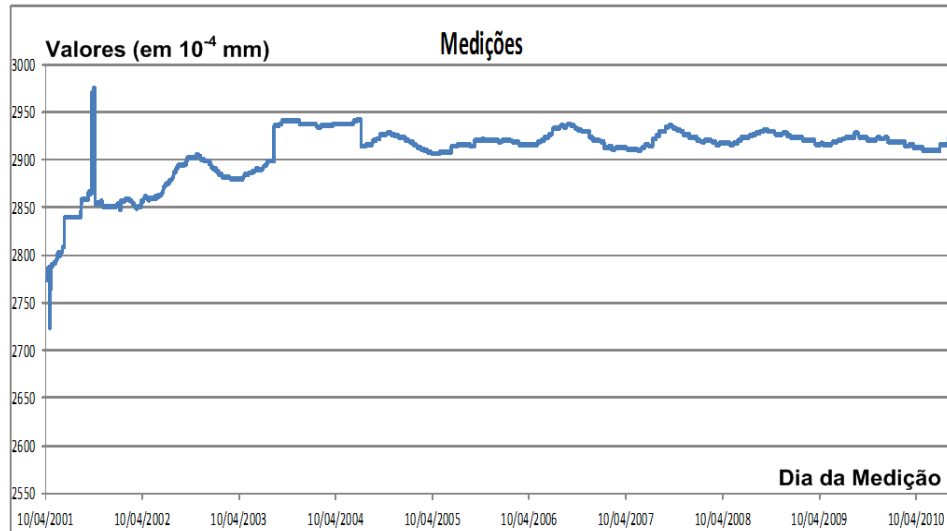
Cabe ressaltar que, os dados obtidos para análise foram utilizados em um aplicativo capaz de receber as medições e registrá-las em agrupamentos do tipo G1 à G5.

#### 4.1.2.1 Provas de conceitos com as medições de um eixo Triortogonal

Nesta prova de conceito estão envolvidas 3426 medições manuais em ( $10^{-4}$  mm) realizadas manualmente entre 2001 e 2010 em períodos de quinze dias, traçadas no gráfico

da figura 16, sendo que uma de suas características é a existência de patamares (com valores constantes), além de regiões que oscilam em torno de um valor médio, em seu desenho.

Figura 16 - Medições realizadas entre 2001 e 2010, em períodos de quinze dias em um eixo do medidor triortogonal.



Fonte: Autor

O balanço do processamento está apresentado na tabela 1, na qual a quantidade de grupos formados é muito menor que os arranjos possíveis. A quantidade de novidades que aparece no período foi de 2513.

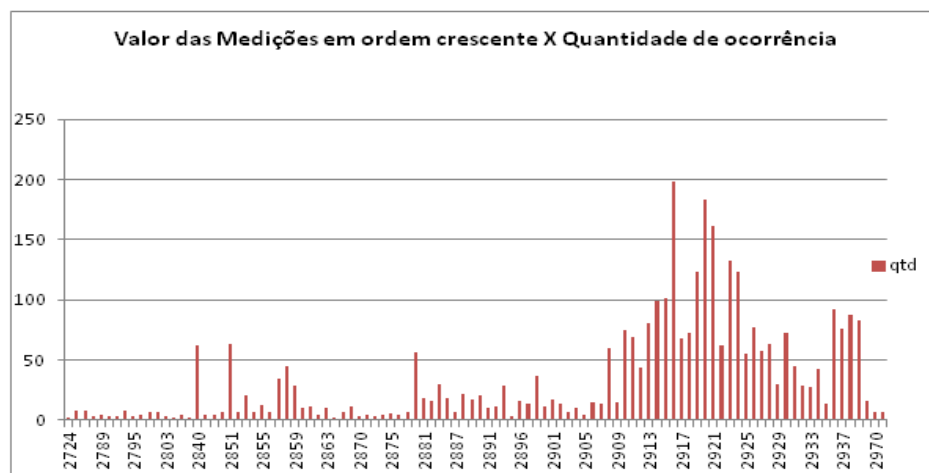
Tabela 1- Relação Quantidade de grupos formados x Arranjos possíveis

Agrupamentos	Elementos por grupo	Quantidade de grupos formados	Arranjos possíveis
G1	1	102	102
G2	2	310	10302
G3	3	514	1030200
G4	4	709	101989800
G5	5	878	9995000400
Total		2513	10098030804

Fonte: Autor

Na figura 17, é mostrado um histograma com os valores individuais de medição indicados de forma crescente e com a frequência de cada valor; a distribuição corresponde a 102 valores.

Figura 17 – Quantidade de ocorrências x Valor Medições individuais em ( $10^{-4}$  mm).



Fonte: Autor

Na análise dos gráficos apresentados nas figuras 16 e 17, percebe-se que o principal elemento de ligação entre eles são os valores de medição. O histograma da figura 17 contabiliza o número de vezes em que os valores indicados aparecem em um determinado período, além disso, a distribuição, mais evidente, resultante dos valores está após aproximadamente o valor 2909. No caso da figura 16, a percepção visual indica que depois de um determinado instante no tempo, as medições oscilam entre os valores de 2900 e 2950 em ( $10^{-4}$  mm).

Na tabela 2 estão apresentadas duas relações: a primeira resume a quantidade de todas as 2513 novidades que aparecem de 2001 a 2010, envolvendo os grupos do tipo G1 à G5, que possuem de 1 a 5 elementos, na segunda, estão as novidades que envolvem somente um valor, do grupo tipo G1. Note-se que, em ambas as relações, ocorrem a

diminuição da frequência de novidades, ao longo dos anos, mas fica em evidência na segunda tabela o ano de 2001, quando as medições iniciam em abril e, em 2010, quando terminam em 18 de novembro. A segunda relação evidencia o ano de 2009 sem novidades, mas os grupos do tipo G2 à G5 continuam a combinação de medições gerando novidades. Do ponto de vista da segunda existe um período estável sem a geração de novos valores de leitura.

Tabela 2 - Distribuição de todas as novidades considerando todos os agrupamentos, de G1 até G5, e somente o agrupamento G1.

<b>Ano</b>	<b>Quantidade Novidades grupos de G1 à G5</b>		<b>Ano</b>	<b>Quantidade Novidades de G1</b>
2001	395		2001	29
2002	673		2002	32
2003	270		2003	10
2004	235		2004	13
2005	215		2005	9
2006	225		2006	5
2007	200		2007	2
2008	115		2008	1
2009	110		2009	0
2010	75		2010	1
Total	2513			102

Fonte: Autor

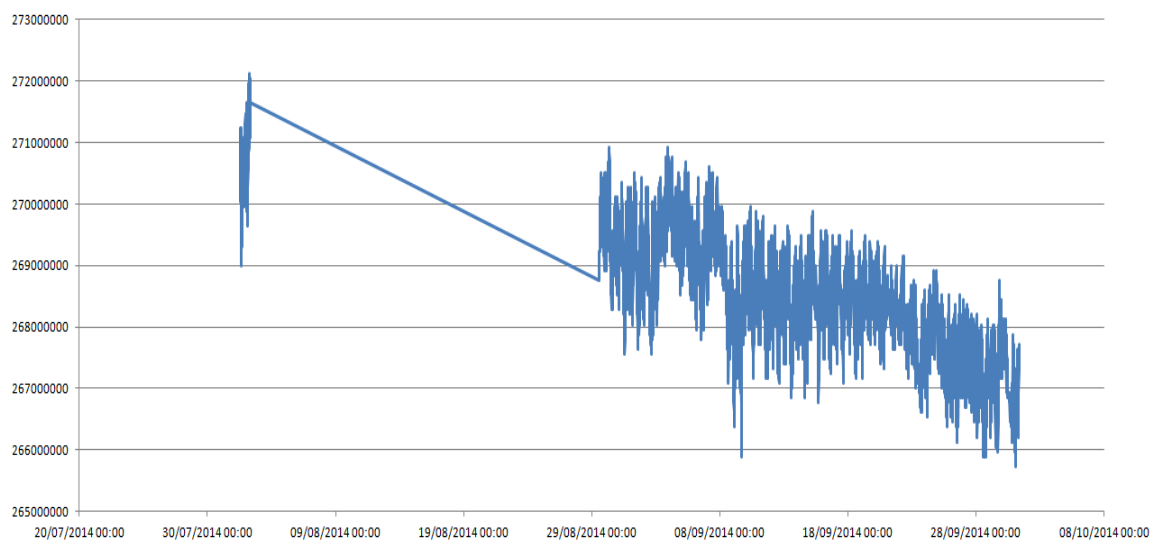
#### 4.1.2.2 Provas de conceitos com o medidor de vazão

Nesta prova de conceito estão envolvidas 3426 medições realizadas em 2014, obtidas em períodos de 15 minutos, de 01/08/2014 13h50min a 01/10/2014 10h04min, traçadas no gráfico da figura 18. No gráfico apesar de todos os pontos estarem unidos, existe um período, no qual, não ocorreu leitura, que é um indicio de anomalia.

O balanço do processamento está apresentado na tabela 3 mostrando que a quantidade de grupos formados é muito menor que os arranjos possíveis. A quantidade de novidades que apareceram no período foi de 10514.



Figura 18 - Medições em ( $10^{-9}$  mm) obtidas de 01/08/2014 às 13h50min à 01/10/2014 às 10h04min, em períodos de 15min para o medidor de vazão.



Fonte: Autor

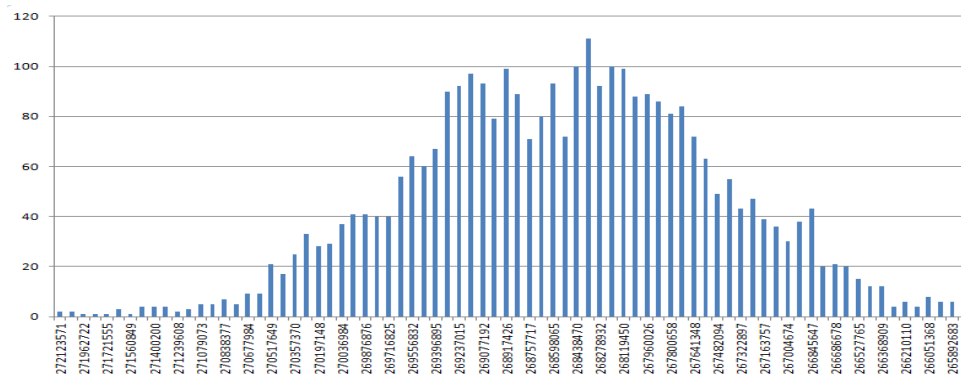
Tabela 3 - Relação Quantidade de grupos formados x Arranjos possíveis

Agrupamentos	Elementos por grupo	Quantidade de grupos formados	Arranjos possíveis
G1	1	78	78
G2	2	1171	6006
G3	3	2901	456456
G4	4	3170	34234200
G5	5	3194	2533330800
Total		10514	2568027540

Fonte: Autor

Na figura 19, está um histograma com os valores individuais de medição indicados de forma crescente e a frequência da ocorrência de cada valor; a distribuição corresponde a 78 valores.

Figura 19 - Quantidade de ocorrências x Valor Medições individuais em ( $10^{-9}$  mm).



Fonte: Autor

Os dados analisados no histograma da figura 19 correspondem à análise visual da figura 18, contendo as medições ao longo do tempo. Vale observar que as medições estão em eixos diferentes.

Na análise dos gráficos apresentados nas figuras 19 e 18, percebe-se que o principal elemento de ligação entre eles são os valores de medição. O histograma da figura 19 contabiliza o número de vezes em que os valores indicados aparecem em um determinado período, além disso, a distribuição resultante está aproximadamente centralizada. No caso da figura 18, a percepção visual indica que a vazão sofre uma diminuição ao longo do tempo.

Na tabela 4, estão apresentadas duas relações semanais indicando o surgimento de novidades: na primeira está apresentada a relação de todos os 10514 grupos, desde os com um elemento até os que possuem 5 elementos, isto é, dos tipos G1 até G5. Na segunda relação estão os grupos que envolvem um elemento, grupo do tipo G1. Também fica evidente a diminuição da quantidade de novidades, apesar de um período sem medições (a identificação 38), mas não a indicação de um período significativo sem novidades.

Tabela 4- Distribuição de todas as novidades considerando todos os agrupamentos, de G1 até G5, e somente o agrupamento G1

Identificação da Semana	Quantidade Novidades, grupos de G1 à G5	Identificação da Semana	Quantidade de Novidades, grupo G1
31	323	31	32
35	511	35	13
36	2365	36	11
37	2213	37	16
38	2044	39	2
39	2049	40	4
40	1009		

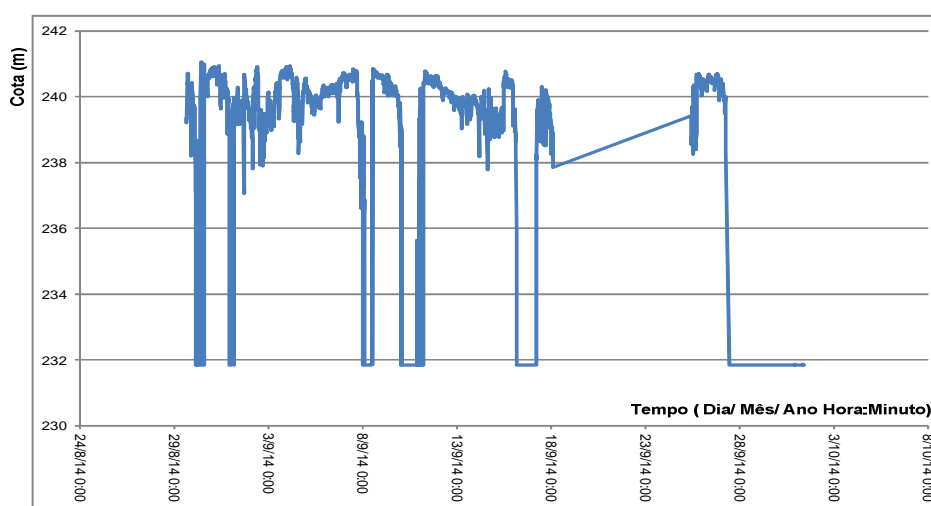
Fonte: Autor

#### 4.1.2.3 Prova de conceito com Piezômetro com alimentação de energia por painel solar

Nesta prova de conceito estão envolvidas 2289 medições em metro de coluna de água e representadas a partir de uma cota de referência da usina hidroelétrica. Os valores foram obtidos no ano de 2014 em períodos de 15 minutos, de 29/08/2014 16h07min à 01/10/2014 10h04min, traçadas no gráfico da figura 20.

Observa-se que na parte inferior do traçado do gráfico da figura 20, existem patamares que correspondem a leituras de uma coluna sem água ou a medições erradas relacionadas à diminuição do fornecimento de energia para o instrumento.

Figura 20 - Medições realizadas de 29/08/2014 às 16h07min à 01/10/2014 às 10h04min, em períodos de 15min.



Fonte: Autor

O balanço do processamento está apresentado na tabela 5, no qual a quantidade de grupos formados é muito menor que os arranjos possíveis. A quantidade de novidades que aparecem no período foi de 9716.

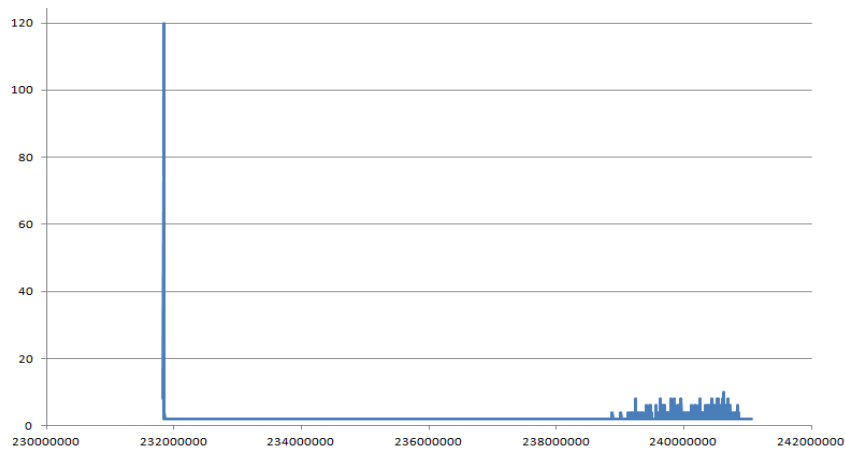
Tabela 5 - Relação Quantidade de grupos formados x Arranjos possíveis

Agrupamentos	Elementos por grupo	Quantidade de grupos formados	Arranjos possíveis
G1	1	1445	1445
G2	2	1917	2086580
G3	3	2039	3010934940
G4	4	2127	4341768183480
G5	5	2188	6256487952394680
Total		9716	6,26083E+15

Fonte: Autor

Na figura 21 está um histograma com os valores individuais de medição, indicados de forma crescente e a frequência de cada valor, a distribuição corresponde a 1445 valores.

Figura 21 - Quantidade de ocorrências x Valor Medições individuais (em  $10^{-6}$  m).



Fonte: Autor

Na análise dos gráficos apresentados nas figuras 20 e 21, percebe-se que o principal elemento de ligação entre eles são os valores de medição. O histograma da figura 21 contabiliza o número de vezes em que os valores indicados aparecem em um determinado período, além disso, a distribuição resultante está concentrada ao lado direito, existe um ponto à esquerda que possui uma frequência alta e corresponde ao valor de fundo do sensor na diminuição de fornecimento de energia. Nesses gráficos é evidente a existência de anomalias que devem ser investigadas.

Na tabela 6 estão apresentadas duas relações semanais indicando o surgimento de novidades: na primeira, a relação de todos os 9716 grupos, desde os com um elemento até os que possuem 5 elementos, grupos G1 à G5. Na segunda, os grupos que envolvem um elemento, grupo G1. Também fica evidente a diminuição de novidades. Mas só a segunda relação indica a possibilidade de um futuro período sem novidades.

Tabela 6 - Distribuição de todas as novidades considerando todos os agrupamentos, de G1 até G5, e somente o agrupamento G1

<b>Identificação da Semana</b>	<b>Quantidade Novidades, grupos de G1 à G5</b>		<b>Identificação da Semana</b>	<b>Quantidade Novidades, grupos G1</b>
35	600		35	110
36	3282		36	595
37	2921		37	407
38	1529		38	222
39	869		39	108
40	515		40	3

Fonte: Autor

#### 4.1.2.4 Considerações finais sobre as provas de conceitos sobre agrupamentos de medições e novidades

A realização das provas de conceitos comprova que:

- A quantidade de grupos formados corresponde a uma quantidade pequena quando comparados à quantidade de arranjos possíveis;
- Evidencia a correspondência entre um gráfico traçado de uma série temporal com o histograma;
- Ocorre a diminuição da ocorrência de novidades ao longo do tempo;
- Com relação à estabilidade. Nos testes do medidor de vazão não é possível a indicação de estabilidade, no caso do triortogonal, ocorre uma estabilidade, e no Piezômetro existe a eminência da obtenção de uma estabilidade, envolvendo medidas de fim de escala que constituem anomalias relacionadas ao sensor;
- Pode-se inferir que o aparecimento de uma novidade do tipo G1 envolve o aparecimento ao mesmo tempo de novidades para todos os agrupamentos utilizados G2, G3, G4 e G5, não impedindo o aparecimento posterior de novidades do tipo G3, G4, G5 ou superior;
- Como observado nas tabelas 1, 3 e 5, quanto maior o número de medições (em um grupo), para uma mesma série temporal, menor é a probabilidade de seu arranjo ser repetido, desconsiderando as sequencias com um e dois elementos. Assim é recomendado para previsões o uso de sequencias a partir de 3, mas não muito maior (quanto maior a sequencia considerada maior a quantidade de possíveis registros). Nesta tese o método está utilizando agrupamentos de cinco medições, note que no uso de número impar não ocorre empate.

#### **4.1.3 Provas de conceitos envolvendo estabilidade e ambiente**

Neste item estão as provas de conceitos com o medidor Triortogonal, considerando novidades, estabilidade e ambiente, visando identificar a configuração mais vantajosa para eles. Com relação ao ambiente o método estabelece os seguintes estados: novo, em evolução e estável. Sendo que o novo corresponde à inicialização de um ambiente, estável a um ambiente em que não surgem mais novidades após um determinado número de

medições e, em evolução, a um ambiente que não alcançou a estabilidade. Um ambiente deixa de ser estável quando uma novidade aparece levando ao surgimento de um novo ambiente.

A validação considera a mesma série histórica do medidor triortogonal do item anterior, além dos mesmos agrupamentos e novidades. A diferença está relacionada com as detecções das mudanças de ambiente, assim define que um ambiente se torne estável após uma quantidade determinada de medições sem novidades, a saber: 10, 15, 20 e 25 medições.

Neste item estão apresentadas oito simulações, quatro considerando todas as novidades envolvidas nos agrupamentos indistintamente com uma até cinco medições, e quatro utilizando somente a novidade envolvendo os valores, que corresponde ao grupo com uma única medição. Na figura 22 são apresentados os resumos dos testes envolvendo 3426 medições.

Figura 22 – Resumo das validações

Considerando os 5 tipos de agrupamentos G1 à G5		Considerando somente o agrupamento G1	
medições	3426	medições	3426
QTD p/ estabilidade	QTD ambientes obtidos	QTD p/ estabilidade	QTD ambientes obtidos
10	96	10	52
15	46	15	34
20	41	20	28
25	30	25	24

Fonte: Autor

#### 4.1.3.1 Considerações sobre as provas de conceitos envolvendo estabilidade e ambiente

Os resumos apresentados na figura 22 conduzem a ideia de que o número de ambientes com relação ao medidor triortogonal aumenta quando são considerados todos os agrupamentos e diminui, quando utilizada somente a novidade composta de uma medição. Outra conclusão indica que quanto maior a quantidade de medições sem novidades para

um ambiente passar do estado em evolução para o estado estável, menor é o número de ambientes criados.

Para o método, uma novidade é candidata à anomalia, dessa forma quando relacionada ao início de um novo ambiente sua candidatura é potencializada. Assim, neste momento será considerado pelo método o agrupamento de uma medição e a quantidade de dez medições para passar do estado em evolução para o estado estável (no método esse valor é parametrizado, podendo ser decidido a partir de testes e considerando o intervalo entre as medições).

#### **4.1.4 Requisitos**

O método proposto considera parte das proposições apresentadas no item 1.2, além de outras necessidades. Assim atende aos seguintes requisitos:

- Os procedimentos do método relacionados ao processamento devem ser não supervisionados;
- Utilização de uma base de conhecimento híbrida capaz de comportar parâmetros, regras e tabelas convencionais;
- Considerar um fluxo contínuo de medições formando séries históricas;
- Operar sem necessidade de treinamento, com uma parte das medições;
- Em seus procedimentos registrar, e analisar cada medição recebida;
- Identificar novidades para auxiliar na identificação de ambientes;
- Identificar mudanças do ambiente relacionado ao ponto de medição;
- Utilizar as novidades e o ambiente para auxiliar na inferência;
- Devido ao critério de flexibilidade identificar: padrões, e não padrões (anomalias, *outliers* em geral, não conformidades, dentre outras);
- Realizar inferência;
- Aprender com a série histórica, os registros auxiliares e as interpretações realizadas, adaptando-se ao longo do tempo;
- Identificar previsões próximas;

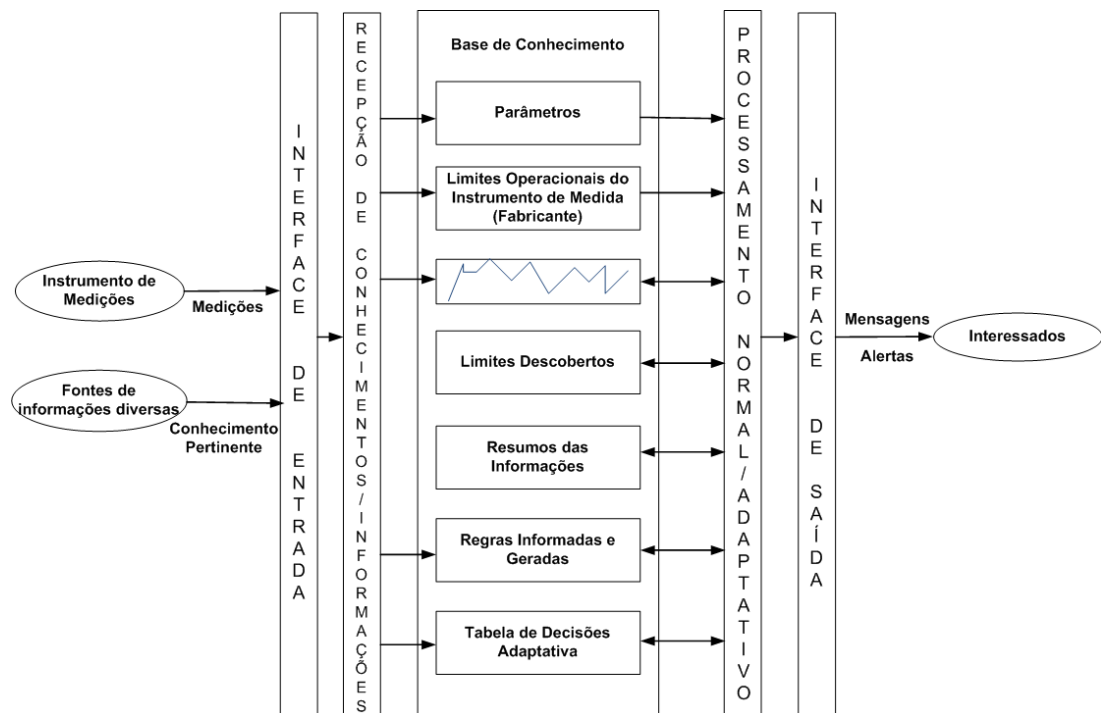


- Emitir no mínimo mensagens e alarmes aos diversos interessados do processo medido e estar preparado para disparar a ativação de diversos tipos de ações;
- Realizar um processo de melhoria continua de forma a agregar novas informações, aprender e descartar aprendizado que não seja mais pertinente.

#### 4.1.5 Arquitetura do método MAIM

A arquitetura, figura 23, envolve a recepção de informações oriundas do instrumento de medida e de Fontes Diversas, tais como: especialistas e do fabricante do instrumento de medidas.

Figura 23 - Arquitetura do sistema computacional utilizado pelo método



Fonte: Autor

Cabe ressaltar, que toda a informação recebida, incluindo parâmetros para auxiliar no processamento, medições, dentre outras, são preparados para ajudar na realização de um processamento adaptativo.

As medições são armazenadas diretamente como uma série temporal, e através de um procedimento de conversão em informações sumarizadas (agrupamentos de medidas) e em limites descobertos (limites de operação, valor máximo, valor mínimo, dentre outras).

As fontes diversas fornecem conhecimentos pertinentes na forma de parâmetros (como tolerância, precisão, dentre outras), de regras para a TDA, em limites operacionais (limites de valores suportados pelo instrumento, informados pelo fabricante, temperatura de operação, dentre outras). Cabe acrescentar que a ideia básica é trabalhar com as informações possíveis, caso alguns limites e intervalos não existam, eles não devem ser informados para a base de conhecimento.

A recepção das informações cria uma base de conhecimento que é processada pelo módulo de processamento (normal e adaptativo), seguindo uma ordem determinada (na sequência: parâmetros, limites operacionais do instrumento de medida, série temporal, limites descobertos, cálculos, resumos das Informações, e regras Informadas e geradas, Tabela de Decisões Adaptativa, e novamente regras) gerando interpretações, mensagens e regras.

As regras existentes em uma TDA ajudam a identificar se uma medida corresponde a um não padrão, caso isso ocorra são enviadas mensagens e/ou alertas aos interessados. O processamento da TDA atualiza informações na série temporal indicando se aquele ponto gerou mensagem, alerta, e se ele corresponde a uma anomalia. No caso da ocorrência de uma nova combinação de condições o processamento da TDA decide: se o valor em uma regra anterior deve ser alterado ou se uma nova regra deve ser criada.

#### **4.1.6 Considerações sobre as regras**

Na arquitetura apresentada as regras estão presentes nas seguintes formas:

- Resumos de informações, constituindo condições como as existentes nas regras de associação permitindo alguma previsão em casos evidentes;
- Limites descobertos, para cada ambiente identificado, pelo processamento das séries históricas, dentre eles estão o maior e o menor valor medido;
- Limites operacionais, não obrigatórios, que podem ser informados por operadores, dentre os quais podem estar o maior e o menor valor possível de ser medido pelo instrumento de medida, fundo de escala;
- Integradas a um modelo de dados desenvolvido para tabelas de decisões adaptativas ou não.

#### 4.1.6.1 Resumos de informações

Durante o processamento são realizados agrupamentos, que podem ser compostos por uma quantidade determinada de medições, o qual vai de uma ao limite de cinco medições nos exemplos desta tese (podendo ser modificado nos parâmetros de processamento), sendo suas ocorrências quantificadas para cada ambiente. Assim, são obtidos índices que podem indicar a possibilidade do próximo valor a ser lido. Por exemplo: em um conjunto de medições sequencialmente oriundas de um instrumento de medida {a, b, c, d, e, f, g} pode ocorrer a combinação {a, e, g} em 80%; das combinações de medições; se forem lidas {a, e}, é provável que próxima leitura seja {g}, para constituir {a, e, g}.

#### 4.1.6.2 Tabelas de decisões

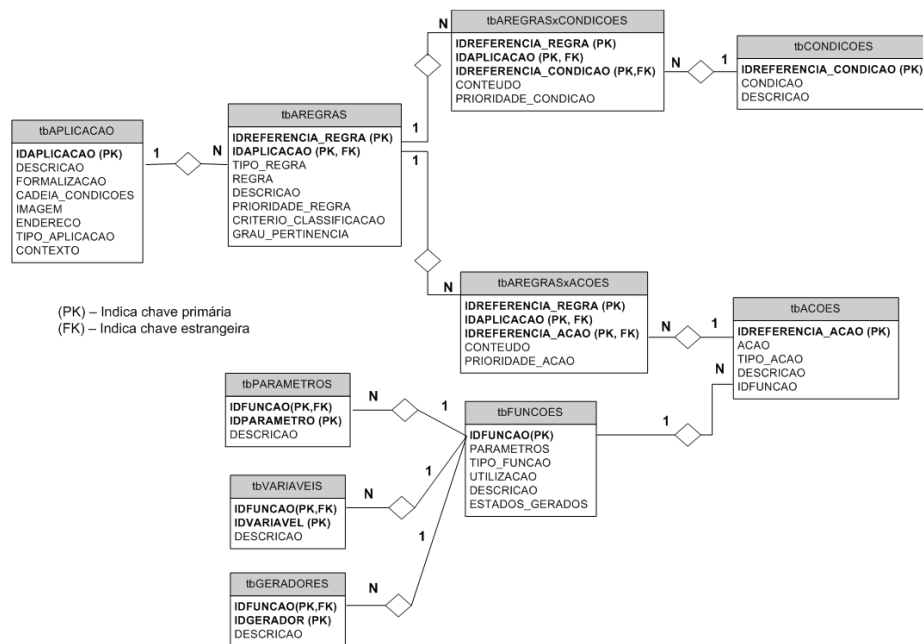
A utilização de Tabela de decisões seja ela adaptativa ou não foi idealizada de forma a ser independente da aplicação. Desse modo, ela pode ser utilizada em monitoramento em diversas áreas. Seu modelo de dados está preparado para aplicações de monitoramento, controle (desde que a velocidade de processamento seja adequada), otimização e autonomia, pois pode conter regras específicas para cada uma delas. Assim,

em sua concepção existe a indicação da aplicação. Mas neste trabalho está restrito a monitoramentos comuns na área elétrica.

O modelo de dados desenvolvido inclui os conceitos existentes nas tabelas adaptativas. Dessa maneira, são consideradas as condições, ações, funções adaptativas com seus parâmetros, variáveis e geradores. Nada impede de serem consideradas somente as condições e ações presentes nas tabelas de decisões convencionais. As funções, parâmetros, variáveis e geradores são ativados quando ocorre a indicação do uso da adaptatividade, possibilitando operações de consulta, exclusão e adição de regras.

O modelo de dados para decisões adaptativas está representado na figura 24.

Figura 24 – Modelo de dados para Tabela de Decisões adaptativa e convencional



Fonte: Autor

Nas tabelas apresentadas na figura 24, existem atributos relacionados a prioridades de regras, condições e ações, que caso sejam aplicados orientam na pesquisa das regras existentes nas tabelas. Cabe destacar que junto às regras também pode ocorrer o acréscimo

de um protótipo da função adaptativa e dos modelos para consulta, exclusão e acréscimo de regras usadas para alterar a base de regras existente. Um ponto interessante incluído no modelo de dados é o grau de pertinência (que indica o grau de verdade, podendo variar de 0 à 1) para regras, que pode ajudar em inferências como as utilizadas nas regras de Associação, na lógica Fuzzy e em alguns Sistemas Especialistas. Também está presente um campo que pode auxiliar na classificação de regras.

Cabe Ressaltar que o modelo de dados contempla duas situações possíveis de Tabela de Dados Adaptativa, apresentadas no item 2.3.2 nas figuras 4 e 5, ambas com um núcleo subjacente contendo uma Tabela de Decisões.

#### 4.1.6.3 Exemplos de possíveis regras

No quadro 6 são relacionadas algumas possíveis regras que podem ser consideradas em alguma aplicação.

Quadro 6 – Exemplos de possíveis regras

<b>Tipo de regra</b>	<b>Condições</b>	<b>Ações</b>
Verificação dos intervalos de tempos	Diferença de tempo (na espera de medidas) > 2 períodos	Mensagem, anomalia de falta de medição.
Regras de previsão para o ambiente	Identificação do Ambiente, agrupamentos variando de uma a 5 medições (podendo ser mais através de parâmetro informado) contendo a última medição.	Indicação da probabilidade para uma próxima medição
Regras de previsão para o ambiente	Identificação do Ambiente, e inclinação de uma quantidade determinada de pontos. Se inclinações só positivas Se inclinações só negativas Se misturadas positivas e negativas	Indicação da próxima inclinação (isso pode ser realizado através do agrupamento de inclinações)..
Intervalo de valores de medidas aceitos pelo instrumento de medição, informado pelo fabricante.	Se menor que o intervalo informado Se maior que o intervalo informado	Indicação de alarme, anomalia.
Inclinação	Se muito acentuada positiva e negativa	O que pode indicar um valor repentino que pode ser candidato a anomalia.
Ambientes	Se Estado do Ambiente estável e aparece uma novidade	Estado do Ambiente novo indica candidato a anomalia. Iniciar a montagem de agrupamentos para o novo ambiente. Nova contagem de medições sem novidades. <b>Mudar regras associadas ao ambiente.</b>
Ambiente	Se indicação anterior do estado do ambiente é novo	Mudar para estado do ambiente em evolução
Ambiente	Se Estado do Ambiente é estável e não aparece anomalia por uma quantidade determinada de medições (parametrizada no sistema)	Manter o estado do ambiente estável

Continuação do Quadro 6		
Novidade	Identificação de novidade obtida com a as atualizações dos grupos	Candidata anomalia
Intervalo obtido pelas medições do ambiente	Se menor que o intervalo informado Se maior que o intervalo informado	Candidato à anomalia
Limites existentes da natureza dos fenômenos medidos	Se maior ou ,menor aos limites informados	Candidato a anomalia. Um exemplo é o caso da concentração de ozônio na atmosfera.
Regra prática hipotética com informação do operador	Medição em intervalo determinado pelo operador	Verificar instrumento vizinho (devido a proximidade, entre as medições, verificando se estão com a mesma tendência)
Regra prática hipotética com informação do operador	Inclinações alternam entre positivo e negativo por um determinado período	Indicação de um caminhão passando perto (situação hipotética para algumas áreas de aplicação)
Regra prática hipotética com informação do operador	Inclinações alternam entre positivo e negativo por um determinado período e diferenças entre medições são maiores que um determinado valor	Indicação de vibrações em turbina próxima (situação hipotética, com correlação pessoal com outros fenômenos)
Regra prática hipotética com informação do operador	Identificação de leitura de fundo	Indicação de falha no instrumento de medição

Fonte: Autor

Possíveis regras podem ser estabelecidas, relacionando os diversos tipos de limites existentes, dentre os quais: do equipamento (LE), da natureza do fenômeno (LF), e do limite operacional existente (LO).

Por exemplo: eventualmente esses limites podem estar distribuídos da seguinte forma, como em uma escala, (LE+) (LF+)(LO+) - (LO-) (LF-)(LE-), na qual, + e - representam valores superiores e inferiores. Dessa forma podem ser estabelecidas as seguintes regras:

- Se ocorrer a superação do LE (maior que LF e LO) também ocorre superação do LF e LO, portanto é uma anomalia;
- Se ocorrer a superação do LF (maior que o LO) e isso, for crítico, é uma anomalia;
- Se ocorrer a superação do LO e não superação do LF e LE somente é um caso a ser investigado, candidato a uma anomalia.

#### 4.1.7 Prova de conceito com relação ao módulo adaptativo

As provas de conceitos deste item visam apresentar a adequação do modelo de dados e de rotinas desenvolvidas às aplicações que consideram tabelas de decisões

tradicionais e adaptativas. No caso das tabelas de decisão Adaptativas são consideradas uma Tabela de Decisões convencional e um mecanismo adaptativo.

Para as provas de conceitos, as interfaces são simples para facilitar a manipulação de informações, com operações de inclusão, exclusão e alteração. As provas de conceitos estão aplicadas à uma tabela de decisões simples e a uma tabela adaptativa simulando uma calculadora (é apresentada neste item devido à possibilidade de visualização do ocorrido). O modelo de dados genérico está apresentado na figura 24.

#### 4.1.7.1 Exemplo de prova de conceito com tabela de decisões simples

Para a prova de conceito, foi utilizada uma tabela simples, como a mostrada na figura 25. Nota-se que, na tabela, todas as possibilidades estão contempladas, num total de 8 regras. As células relativas às condições estão preenchidas por “S” ou “N”, correspondendo respectivamente a sim e não. As células relativas às ações podem ser preenchidas por “X” indicando a utilização da ação. Neste caso, a tabela de decisões simples corresponde a uma base de conhecimento, pois nela estão contidos os conhecimentos para a tomada de decisões.

A partir das informações fornecidas relacionadas às condições, é realizada uma busca da regra que coincide com elas e, ao encontrar essa regra a ação correspondente é executada.

Figura 25 - Tabela de decisões simples

		REGRAS							
		1	2	3	4	5	6	7	8
CONDIÇÕES	Ocupa cargo chefia	S	S	S	S	N	N	N	N
	Idade maior que 40	S	S	N	N	S	S	N	N
	Mais de 2 anos no cargo	S	N	S	N	S	N	S	N
AÇÕES	Exame especial	X	X	X		X	X		
	Exame normal				X			X	X

Fonte: Autor

A figura 26 mostra a tabela de decisões da figura 25 existente no sistema computacional indicando as condições, as ações, às funções relacionadas às ações e às

prioridades para a realização de busca. Observa-se que é previsto o uso de prioridades. Na figura 26, a coluna “REFERENCIA\_REGRA” é o código, de identificação da regra, gerado na sua criação.

Figura 26 - Informações existentes no módulo adaptativo

REGRA	REFERENCIA_REGRA	PRIORIDADE	TIPO_REGRA	Indica_se_v	Indica_se_c	Indica_se_c	Enviar_mensagem_para_exame_v	Enviar_mensagem_para_exame_n
-1			PRIORIDADES	1	2	3	1	2
0			FUNÇÕES				mensagem_exame_especial	mensagem_exame_normal
1	48	1	NORMAL	S	S	S	X	
2	49	1	NORMAL	S	S	N	X	
3	50	1	NORMAL	S	N	S	X	
4	51	1	NORMAL	S	N	N		X
5	52	1	NORMAL	N	S	S	X	
6	53	1	NORMAL	N	S	N	X	
7	54	1	NORMAL	N	N	S		X
8	55	1	NORMAL	N	N	N		X

Fonte: Autor

Com relação ao processamento, todas as provas de conceitos indicam e executam ações correspondentes ao envio de mensagens.

#### 4.1.7.2 Exemplo de Prova de conceitos com tabela de decisões adaptativa

A prova de conceito com relação a um dispositivo utilizando a tabela de decisões adaptativa é um pouco mais complexa. Ele envolve funções adaptativas contemplando a possibilidade de criar, consultar e excluir regras. Junto às regras também ocorre o acréscimo do protótipo da função adaptativa e dos modelos para consulta, exclusão e acréscimo de regras usadas para alterar a base de regras existente. Também há interface para a entrada de informações com relação às funções adaptativas, parâmetros, variáveis e geradores.

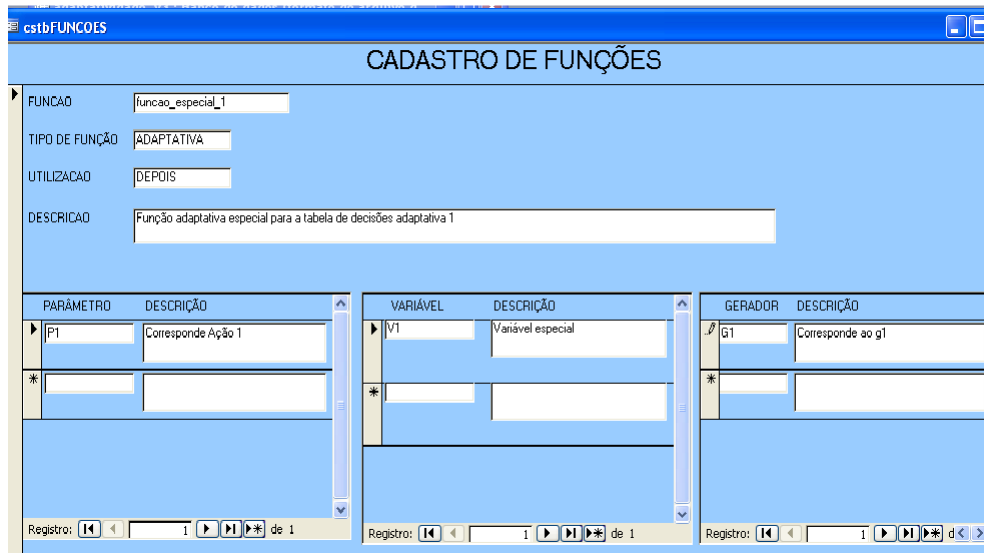
Na figura 27, encontra-se a interface de cadastramento de funções, com os correspondentes parâmetros, variáveis e geradores de regras.

Para realizar a prova de conceitos com o sistema computacional desenvolvido, é utilizado o autômato da figura 28 correspondendo a uma calculadora. Nele estão os arcos



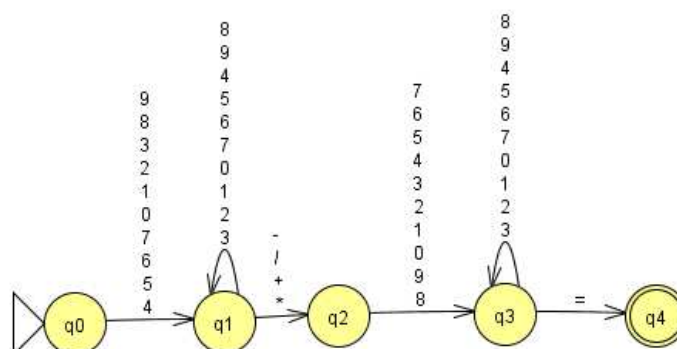
com as possibilidades numéricas e nodos indicados por q, o nodo com círculos concêntricos corresponde ao final do autômato. Ele pode ser adaptado de forma a ampliar a quantidade de números e operações matemáticas a serem realizadas, como as possibilidades apresentadas na figura 29.

Figura 27 - Interface envolvendo Funções, parâmetros, variáveis e geradores.



Fonte: Autor

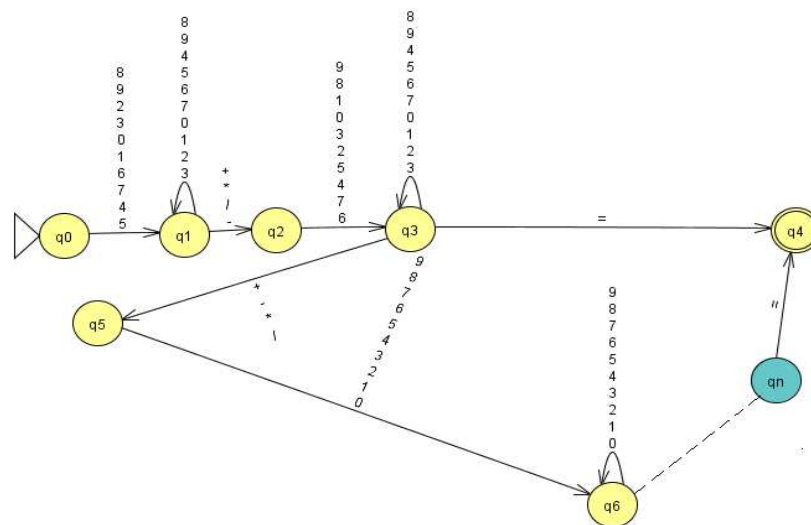
Figura 28 - Autômato referente a uma calculadora simples



Fonte: Autor

Na figura 29 representa um autômato adaptativo que inclui arcos e nós como os seguintes: arco entre q1 e q2, nó q2, arco entre q2 e q3, nó q3. Por exemplo, está criado o trecho: arco entre q3 e q5, nó q5, arco entre q5 e q6, nó q6, e o arco entre q6 e q7, só que na representação outros trechos já estão criados e o q7 não aparece. No caso o autômato ideal é o não adaptativo, mas para dar sequencia ao estudo está simulado nos teste o autômato adaptativo da calculadora simples.

Figura 29 – Aumento do número de operações da calculadora simples.



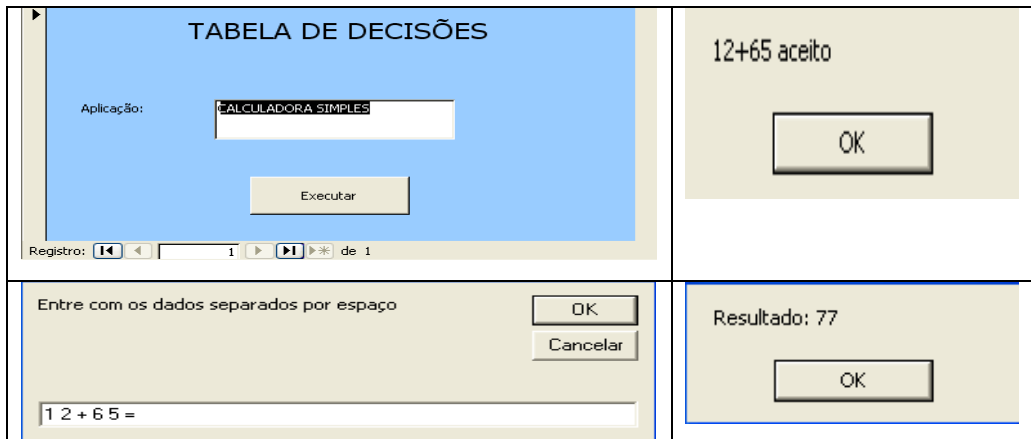
Fonte: Autor

Durante as operações o número de nodos e arcos é modificado para compor novas operações. Dessa forma, em tela semelhante à da Figura 30, que apresenta um exemplo simples, é possível informar dados (com caracteres separados por espaço) e calcular as seguintes expressões mais complexas, como:

$2 + 3 4 =$  com a expressão  $2+34$  com resultado de 36;

$2 + 3 4 - 2 * 3 + 4 / 2 =$  com expressão  $2+34-2*3+4/2$  com resultado de 32.

Figura 30 - Exemplo de Operação da tabela de decisões para a Calculadora simples



Fonte: Autor

As provas de conceitos realizadas comprovaram que o autômato para uma calculadora simples se adapta a novas combinações de cálculo.

#### 4.1.8 Fluxo de dados e pseudocódigo

Os procedimentos de inferência, assim como o método, consideram novidade o valor atual de uma medição que não possui registro na base de conhecimento. Também considera que um ambiente é estável quando a quantidade de novidades está decrescendo, tendendo a zero, em uma sequência crescente de períodos iguais.

Diversas mudanças nos ambientes físico e virtual, de operação de um instrumento de medida, podem ocorrer refletindo nas medições recebidas pelo sistema computacional. A troca ou calibração de um instrumento em um ponto de medição é uma mudança local, já o aumento do ritmo das chuvas em uma região consiste em uma mudança que pode alterar o nível da represa e refletir em algum instrumento de medida.

A base de conhecimento corresponde ao registro de grupos de elementos, resumos, regras e parâmetros (como tolerâncias, precisão, dentre outros critérios).

O fluxograma apresentado na figura 31 e o pseudocódigo apresentado como Algoritmo 1, complementado pelo Algoritmo 2, estão aderentes à arquitetura ilustrada na

figura 23 e envolve operações realizadas na interface de entrada, na conversão, na inferência e interface de saída. Algumas dessas operações implicam a manipulação da base de dados, gerando a criação de novas informações, consultas, alterações e, em casos específicos, a exclusão de informações.

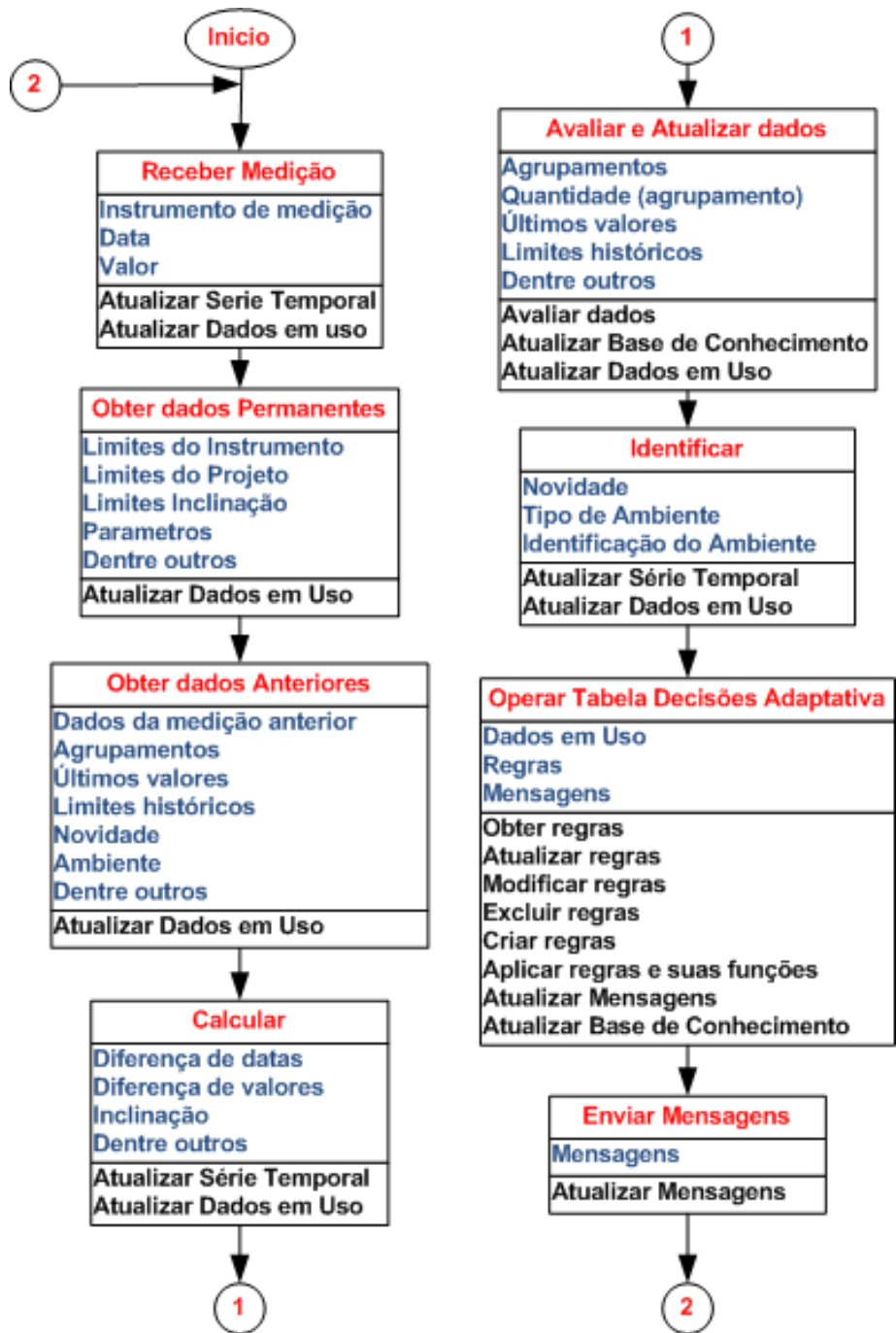
Todo o sistema computacional criado segue os princípios para ter uma representação de conhecimentos, um mecanismo de busca, a realização da inferência e aprendizagem com a inclusão, exclusão e modificações de regras seguindo os conceitos já apresentados sobre TDA.

Assim na Figura 31, estão apresentados os principais elementos e operações realizadas pelo sistema proposto. Em cada operação indicada é mostrada a principal atividade do bloco, os dados e as funções envolvidas em cada um deles. Cabe ressaltar que, apesar de usar símbolo semelhante ao existente em UML o fluxograma não utiliza esse padrão e nem é voltado a objeto.

Basicamente o fluxograma da figura 31 indica que o sistema recebe a medição, seu valor e quando foi realizada (data/hora/etc.), atualizando o banco de dados e iniciando um conjunto de dados, que serão completados antes da utilização da Tabela de Decisões Adaptativa.

Com a informação do instrumento de medição são reunidas características fixas com dados e parâmetros, além de dados que são relativos à medição anterior e resumos relacionados aos agrupamentos, suas ocorrências, limites históricos, últimos valores, dentre outros. Também são realizados cálculos entre a medição anterior e a atual (como diferença entre datas, valores, Inclinações). Na sequência, são identificadas a existência de novidade e as informações relacionadas ao ambiente (tipo e identificação numérica). Em seguida, é aplicada a Tabela de Decisões Adaptativa (em cujos procedimentos está a capacidade de adaptar regras existentes e aplicá-las, podendo também excluir, criar e em situações especiais alterar o conteúdo, modificar diretamente um regra e obter regras especiais em outros sistemas).

Figura 31 - Fluxo de dados do Algoritmo proposto



Fonte: Autor

A seguir estão apresentados os pseudocódigos 1 e 2 com mais detalhes que o fluxograma. Dando um enfoque à utilização da Tabela de Decisões Adaptativa que pode

aplicar ações adaptativas antes e depois da aplicação das ações previstas diretamente por ela, mudando regras quando necessário.

Algoritmo 1 - Etapas do pseudocódigo para o método MAIM (considerando um ciclo)

---

**Entrada:** valor de medição obtida a partir de um sensor;

**Saída:** Mensagem / Alarmes, em geral indicando medição candidata à anomalia.

---

1. Utilizar a última medição realizada (uma por vez);
2. Registrar na série temporal instrumento de medição, o valor da medição e a data;
3. Obter da base de conhecimento os Parâmetros relacionados ao instrumento de medição;
4. Obter limites do instrumento de medição (inferior e superior, se existirem);
5. Obter limites de projeto (inferior e superior, se existirem);
6. Obter limites operacionais adotados de inclinação (inferior e superior, se existirem), dentre outros;
7. Obter dados da medição anterior e dos últimos de agrupamentos, últimos valores, limites históricos, novidade, tipo de ambiente, identificação do ambiente, dentre outros
8. Calcular a diferença das datas entre a medição atual e a anterior;
9. Calcular a diferença entre os valores atual e anterior;
10. Calcular a inclinação entre valor atual com o valor anterior, dentre outros;
11. Registrar na série temporal (atual com referência ao anterior): diferença entre datas, diferença entre valores, e inclinação entre os valores;
12. Avaliar e Atualizar: agrupamento de dados com o valor atual (para 1, 2 3, 4 e 5; quantidade somando um, quando existir o grupo e criar quando não inexistente); grau de confiança e suporte dos agrupamentos (relacionados a regras de associação). Identificar novidade (que corresponde a um agrupamento que aparece pela primeira vez, e pode ter de um a 5 elementos), Tipo de Ambiente (novo, em evolução, estável), e Identificação do Ambiente (numérica);
13. Atualizar os últimos valores, modificando quando necessário os limites operacionais (inferior e superior);
14. **Procedimentos Envolvendo TDA;**
15. Enviar mensagens e alertas aos interessados, caso necessário;
16. Retornar para etapa 1.

---

Fonte: Autor

---

**Procedimentos Envolvendo TDA**

**Entrada:** Regras (do tipo Se condições então Ações), e conteúdos necessários para utilização das regras;

**Saída:** Mensagens / Alarmes

Pode ser executada mais de uma regra aumentando o tamanho da Mensagem.

---

1. Busca das Regras Pertinentes;
2. Executar enquanto existir regras Pertinentes;
  - a. Executar o lado das condições;
    - i. Verificar se a regra possui as variáveis disponíveis para a operação
      - i.1 - Se sim;
        - i.a - Executar o lado das condições;
          - i.i - Se verdadeiro;
            - i.i.1 - Verificar se existem ações adaptativas a serem executadas antes;
              - i.i.1.a - Executar as ações e funções adaptativas antes;
            - i.i.2 - Verificar se existem ações normais a serem executadas;
              - i.i.2.a - Executar as ações e funções normais;
              - i.i.2.b - Compor mensagem
            - i.i.3 - Verificar se existem ações adaptativas a serem executadas depois;
              - i.i.3.a - Executar as ações e funções adaptativas depois
          - i.ii - Se Falso
            1. Não executar ações
    - b. Continuar Executar enquanto existir regras Pertinentes
  3. Atualizar a base de conhecimento
  4. Ir para próxima operação (no caso é enviar mensagem).

---

Fonte: Autor

Cabe ressaltar que, nos dois modelos apresentados de Tabela de Decisões Adaptativa no item 2.3.2 e nas figuras 4 e 5, para as mesmas condições podem existir ações adaptativas que podem ser aplicadas antes (anterior) e após (posterior) a execução das ações da Tabela de Decisões Subjacente (interior a TDA).

Um exemplo de ação criando uma nova regra antes da execução é apresentado esquematicamente na figura 32, nela existia a regra 3 com o conteúdo da condição C\_4 como “Dentro” com uma ação adaptativa anterior que verifica a existência da condição C\_4 como “Fora” e, caso não exista, cria com sua ação AA\_1 uma nova regra que, no

caso, é R9 que praticamente é uma cópia da R3 só que no conteúdo C\_4 com “Fora”, e uma nova ação A\_5 utilizada pela regra R9.

A configuração inicial da TDA, no exemplo é modificada para outra, como indicado na figura 32, por meio de uma ação anterior e, neste caso a execução da tabela continua com a nova regra com os dados e informações atuais. No caso de uma regra ser criada ou excluída em ação posterior, a execução da TDA seria aplicada aos dados e informação da próxima sequência da série histórica.

Figura 32- Exemplo esquemático de aplicação com a criação de uma nova regra

Regras	.....	R3	.....	=====>	Regras	.....	R3	.....	R9
Prioridades	.....	1	.....	=====>	Prioridades	.....	1	.....	1
Condições:				=====>	Condições:				
C_1	.....	Não	.....	=====>	C_1	.....	Não	.....	Não
C_2	.....	Estável	.....	=====>	C_2	.....	Estável	.....	Estável
C_3	.....	Menor	.....	=====>	C_3	.....	Menor	.....	Menor
C_4	.....	Dentro	.....	=====>	C_4	.....	Dentro	.....	Fora
	.....		.....	=====>		.....		.....	
Ações				=====>	Ações				
A_1	.....		.....	=====>	A_1	.....		.....	
A_2	.....		.....	=====>	A_2	.....		.....	
A_3	.....		.....	=====>	A_3	.....		.....	
A_4	.....	xxx	.....	=====>	A_4	.....	xxx	.....	
Ações Antes				=====>	A_5	.....		.....	xxx
AA_1	.....	xxx	.....	=====>	Ações Antes				
AA_2	.....		.....	=====>	AA_1	.....	xxx	.....	xxx
	.....		.....	=====>	AA_2	.....		.....	
Ações Depois				=====>	Ações Depois				
AD_1	.....		.....	=====>	AD_1	.....		.....	
AD_2	.....		.....	=====>	AD_2	.....		.....	

Fonte: Autor

## 4.2 MÉTODO COMPLEMENTAR

O Método Complementar é uma consequência dos estudos apresentados nos itens anteriores, nos quais alguns métodos identificam em séries temporais valores fora dos padrões históricos existentes e, devido aos seus processos, podem ser enviadas mensagens



e alertas, sendo que algumas delas são imprecisas e/ou podem ser descartadas devido à importância dada pelos interessados. Desse modo, o Método Complementar pretende aumentar a assertividade de algumas mensagens ou alertas, descartar as desnecessárias, e incluir justificativas e explicações, além de indicar correlações.

O Método Complementar é utilizado após a análise e interpretação inicial das séries temporais realizada por um ou mais métodos, que serão denominados de Métodos Básicos, pode se inclusive o método MAIM um deles.

As informações necessárias para o Método Complementar podem ser oriundas: de correlações não consideradas nos Métodos Básicos, da experiência de pessoas envolvidas, dos ambientes físico e virtual, da abrangência do conhecimento utilizado, dentre outras fontes. Com relação ao ambiente virtual, estão incluídas: a Internet, informações oriundas da rede de comunicação interna, inclusive informações de outros pontos de medição e de seus instrumentos de medição.

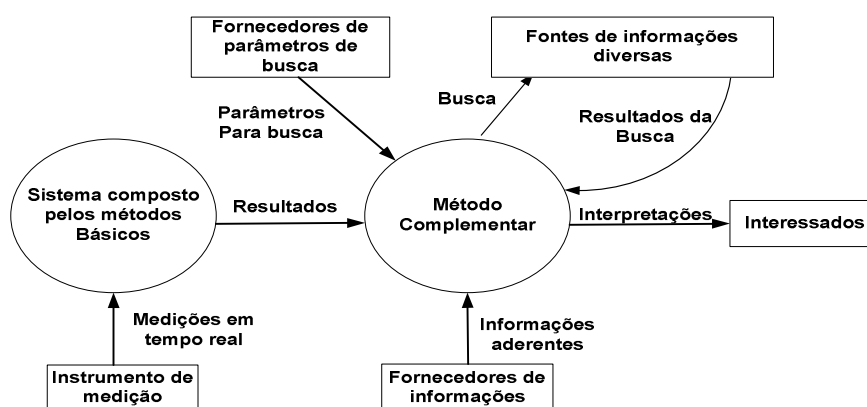
Como parte das informações, envolve o instante (portanto subentendendo data, hora, dentre outros indicadores temporais), as correlações devem considerar que os eventos envolvidos estejam relacionados no tempo. Dessa forma, o valor medido para ser considerado deve estar dentro de um intervalo relacionado a um evento externo, por exemplo, uma medição foi realizada às 10 horas, e no mesmo instante ocorreu um abalo sísmico, que influenciou o seu valor. A correlação irá buscar o intervalo de tempo em que ocorreu o abalo sísmico.

Para tanto, as principais questões envolvidas na aplicação, deste método, estão relacionadas à coleta de informações, sua relevância, sua representação, na busca para realizar interpretações e análise, e na complementação das mensagens oriundas do método.

A utilização do Método Complementar, não garante uma resposta 100% correta, assim é necessário entender suas aplicações e analisar os impedimentos existentes para obter uma assertividade maior. Logo, um conjunto de informações deve ser reunido em uma base de conhecimento para serem manipulados pelo Método Complementar de forma inteligente.

O esquema básico do Método Complementar está apresentado na figura 33. Neste, é possível observar a recepção dos resultados do Sistema composto pelos Métodos Básicos, das informações aderentes à aplicação e dos parâmetros de busca. Além disso, o Método Complementar realiza buscas em fontes de informações diversas para incorporar seus resultados, como por exemplo, na Internet. As informações recebidas compõe uma base de conhecimento. Cabe ressaltar que as informações fornecidas ao Método Complementar são organizadas na forma de regras, mapas e tabelas que ao serem computadas, podem gerar interpretações na forma de mensagens e alertas para os interessados.

Figura 33 - Esquema básico do Método Complementar



Fonte: Autor

No caso das informações provenientes das fontes de informações diversas, existem algumas categorias disponíveis em órgãos governamentais, nos sites de busca públicos, dentre os quais estão as que envolvem questões ambientais, climáticas, geológicas, financeiras (bolsa de valores), sísmicas, marés, dentre outras. Apesar da existência de uma gama enorme de informações, somente devem ser utilizadas as pertinentes com a aplicação. Assim, por exemplo, sensores instalados em uma boia marítima fornecem dados sobre uma anomalia ocorrida ao Método Complementar, que por sua vez, poderia ser

correlacionada com informações obtidas em sites específicos da internet contendo dados climáticos, e sobre as marés referentes à posição da boia.

#### **4.2.1 Considerações teóricas**

Parte dos conceitos já foi apresentada no item 2.3 e no apêndice A e envolvem as séries temporais, regras de produção, regras de associação, as regras originárias de árvores de decisão e da aplicação da lógica Fuzzy, dentre outros conceitos e técnicas.

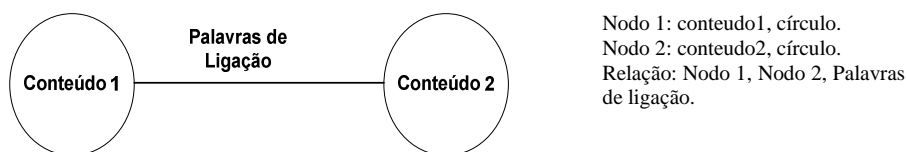
Alguns novos conceitos serão apresentados, como os relativos: aos Mapas Conceituais por permitirem o mapeamento das influências de um ambiente e ao Diagrama Causa-Efeito por permitir o estabelecimento de regras.

##### **4.2.1.1 Mapas Conceituais**

Um mapa conceitual consiste em uma técnica descrita em (NOVAK, 1998) e seu objetivo é relacionar conceitos (conhecimentos) originalmente de forma hierárquica com a finalidade de facilitar o aprendizado. Nos conceitos, podem ser utilizados nomes, substantivos, verbos, adjetivos e expressões matemáticas. Em sua representação na forma de grafos, é comum encontrar nodos (que podem ser representados por figuras) com rótulos indicando conceitos e ligados por arestas com rótulos, contendo as chamadas palavras de ligações (não obrigatórias, que em geral são explicativas para o relacionamento entre os nodos).

Basicamente é possível armazenar o mapa conceitual contendo no mínimo a indicação de um nodo, seu conteúdo, a figura referente a ele, e no caso da relação de dois nodos, pode existir a indicação do sentido (em geral através de uma seta), além das palavras de ligação entre eles, como indicado na figura 34.

Figura 34 - Exemplo de uma representação entre dois nodos e sua informação mínima.



Fonte: Autor

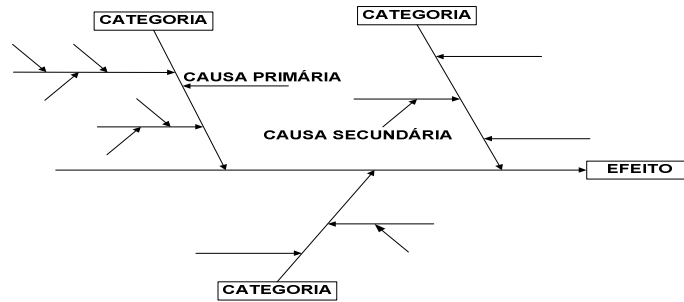
Diversas técnicas foram criadas contendo essas ideias básicas para a representação do conhecimento, com diferenças na aplicação, interpretação das relações e visualização dos mapas. Desse modo, diversos produtos podem ser encontrados, como: Decision Explorer, Inspiration, CMAP, SemNet, Axon Idea Processor, VisiMap and InfoMap, Activity Map, Text Vision, SMARTIdeas, MindMapper, dentre outros.

Assim é possível construir mapas conceituais de forma manual, envolvendo diretamente pessoas ou de forma automática, a partir de fontes, tais como: textuais em arquivos, bases de conhecimentos ou internet.

#### 4.2.1.2 Diagrama Causa-Efeito

Segundo Groover (2011), o diagrama de Causa e Efeito também é conhecido como Espinha de Peixe ou Ishikawa, não constitui uma ferramenta estatística, geralmente é aplicado por equipes de qualidade com a finalidade de identificar as causas de um determinado efeito. Foi proposto inicialmente em 1943 por Kaoru Ishikawa, e posteriormente aperfeiçoado, é utilizado no controle de qualidade e tem sua forma básica como indicado na figura 35, na qual se observa sua organização em: Categorias de causas, causas primárias e secundárias referente a um efeito.

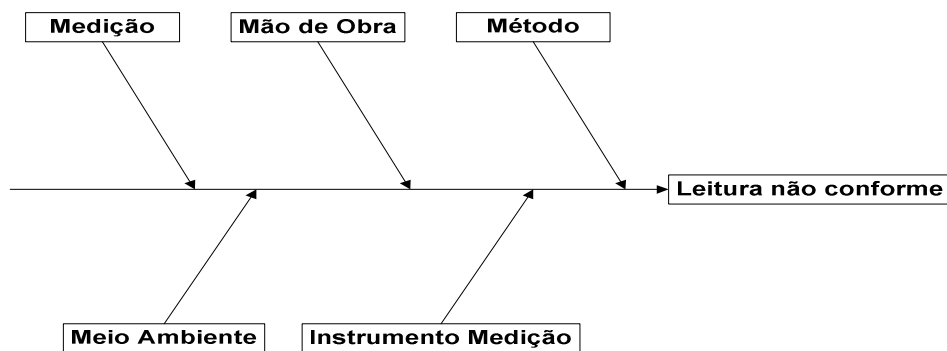
Figura 35– Diagrama Ishikawa (Espinha de Peixe / Causa-Efeito).



Fonte: Groover (2011)

Na figura 36, há um diagrama envolvendo cinco categorias de causas influenciando a obtenção da leitura não conforme que são: medição, mão de obra, método de trabalho utilizado, meio ambiente, e instrumento de medição. Em outras palavras estão compreendidos o valor da medição, as pessoas envolvidas diretamente com o instrumento e o processo de trabalho, além das influências do meio ambiente e as próprias características do instrumento e seus limites.

Figura 36 - Exemplo de diagrama de Causa Efeito simplificado para leitura não conforme.



Fonte: Autor

Esse diagrama pode ser armazenado indicando o efeito e os ramos e seus índices relacionados a ele, por exemplo: Efeito, Categoria 1, Causa primária 1.1, Causa secundária 1.1.1, Causa secundária 1.1.2, Categoria 2, Causa primária 2.1. O diagrama causa efeito

também poderia evoluir para regras. Outra possibilidade é armazenar o diagrama de causa efeito como um mapa conceitual.

#### 4.2.2 Método Complementar propriamente dito

O Método Complementar, cujos passos operacionais estão apresentados no algoritmo 3, visa a partir das mensagens ou alertas enviados pelos Métodos Básicos, ajudar com soluções mais assertivas e quando não pertinentes descartá-las. Para tanto propõe realizar a inferência utilizando uma base de conhecimento desenvolvida e atualizada a partir mapas conceituais relacionando as influências do ambiente e de regras obtidas de uma série de conhecimentos pertinentes ao um ponto de medição.

Algoritmo 3 - Passos operacionais utilizados pelo Método Complementar

---

**Entradas:** Conhecimentos, Medições.

**Saídas:** Alarmes e Mensagens

---

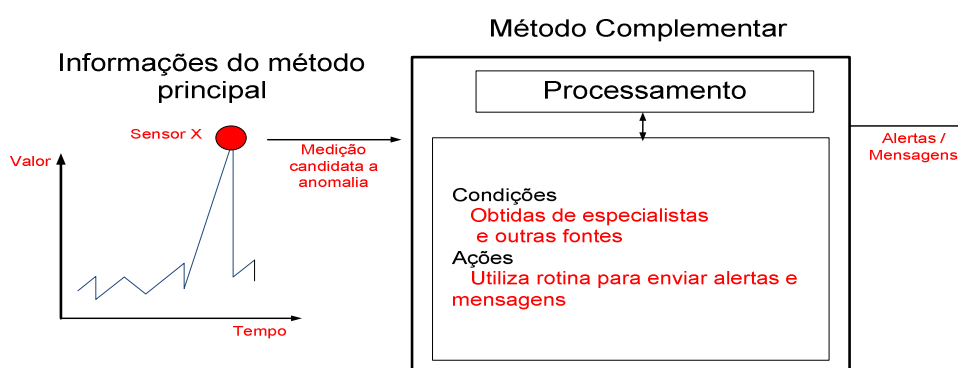
1. Levantamento inicial do conhecimento pertinente ao ponto de medição;
2. Categorizar esse conhecimento;
3. Organizar o conhecimento de forma a obter mapeamentos, regras e tabelas;
4. Inserir o conhecimento na “base de conhecimento”;
5. Receber medição (conteúdo), interpretações, mensagens e alertas dos Métodos Básicos;
6. Se necessário buscar informações em bases de dados parametrizadas, e prepará-las na forma de regras;
7. Realizar a análise utilizando inferência considerando a base de conhecimento;
8. Verificar se medição é candidata à anomalia;
9. Caso verdade emitir novas mensagens e alertas para os interessados, e caso contrário descartar as interpretações, mensagens e alertas recebidos não enviando novas mensagens e alertas;
10. Num processo de melhoria continua realizar novos levantamentos pertinentes ao ponto de medição;
11. Repetir novamente os passos a partir do segundo

---

Fonte: Autor

De forma simplificada, a figura 37, indica o processamento do método principal entregando sua interpretação na forma de medição candidata à anomalia, recebida pelo método complementar que, após o seu processamento utilizando condições oriunda de especialistas e/ou outras fontes pode enviar mensagens e/ou alertas.

Figura 37 – Interação simplificada do Método principal com o Complementar



Fonte: Autor

A seguir são discutidos os passos operacionais de 1 a 8.

#### 4.2.2.1 Levantamento dos dados

As informações levantadas são constituídas de dados qualitativos e quantitativos e de conhecimentos tácitos e explícitos referentes ao ponto de medição e originários de:

- Pessoas envolvidas utilizando técnicas de entrevista;
- Da literatura, envolvendo livros, artigos, manuais, dentre outros;
- Da utilização de robôs de busca para obtenção de conhecimentos específicos existentes na Internet ou Intranet de uma organização;
- De outras fontes existentes na rede corporativa.

#### 4.2.2.2 Categorização das informações obtidas

O método Complementar facilita a categorização das informações, e possibilita a inclusão de novas categorias a qualquer momento. No quadro 7 estão as categorias sugeridas, dentre várias possibilidades, relacionadas a um instrumento de medição.

Quadro 7 - Categorias de conhecimentos relacionadas ao instrumento de medição

Categoria	Características	Subcategoria	Conhecimentos Possíveis
Pessoal	Experiência específica na mesma instalação, e/ou experiência com várias instalações semelhantes.	Fornecedor	Conhecimentos em geral sobre os produtos
		Vendedor	Experiência com diversos compradores
		Comprador	Informações de diversos fornecedores
		Manutenção	Experiência com instrumento específico
		Operador de supervisão	Experiência na monitoração e controle
		Operador de campo	Experiência em campo com vários instrumentos
		Coletor de dados	Experiência em campo com vários instrumentos
		Interessados pelos resultados em geral	Experiência com mensagens e alertas
		Reguladores (órgãos oficiais por ex: ONS)	Experiências semelhantes em outras instalações
Temporal	Organização em períodos de tempo	Qualquer período tradicional como: anual, mensal, semanal. Além de períodos determinados especificamente.	Correlação entre eventos periódicos e períodos equivalentes utilizados nos métodos básicos.
Contextual	Correlação com variáveis	Ambiente Físico	Pó, Umidade, Vibração, seres vivos agressivos ao equipamento, corrosão, etc.
		Outras variáveis	Pressão, temperatura, etc.
		Outras séries temporais	Valores
Instrumento	Propriedades do instrumento		Informações diversas como: dimensões, limites, dentre outros.
Outros instrumentos	Correlação com instrumentos iguais e/ou diferentes		Valores e resultados no mesmo instante. Correlação direta de variáveis
Manutenção	Relacionada ao instrumento e ao seu ambiente físico		Falhas, árvores de falhas, manutenção preventiva, preditiva, e corretiva.
Conhecimentos diversos	Quaisquer	Diversas	Relações diretas e indiretas

Fonte: Autor

Cuidados devem ser tomados com relação à pertinência entre os conceitos, pois num mundo complexo existem coincidências, no entanto a relações causais devem ser analisadas com cuidado. Se forem utilizados dados oriundos da estatística, existem ferramentas para analisar correlações, estabelecendo um grau de relação, por exemplo, o valor 1 equivale a correlação total, se 0, não existe correlação e os dados são independentes, se -1 a correlação é perfeitamente inversa. Outra variável estatística



interessante é a significância que verifica a chance de um resultado ser consequência de um mero acaso.

#### 4.2.2.3 Mapeamento das influências e pertinências

O mapeamento das influências existentes ao ponto de medição e suas pertinências, na forma de mapa conceitual, visa relacionar hierarquicamente um conceito central a outros conceitos. A quantidade de elementos pertinentes utilizados depende das categorizações das informações utilizadas e do nível de detalhes possíveis e permitidos pela organização proprietária e/ou usuária do instrumento. O Método Complementar considera que os elementos próximos ao elemento central do mapa estão diretamente relacionados a ele, e esses elementos próximos, por sua vez, podem possuir outras relações diretas a eles e não relacionados ao elemento central, podendo existir várias camadas de relacionamentos entre os conceitos.

Cabe ressaltar que é possível construir mapas mais elaborados com mais detalhes, como arestas com vários significados, como por exemplo, conteúdos de tipos diferentes, com vários significados, como hierarquias, dependências, relações constantes ou eventuais, dentre outros.

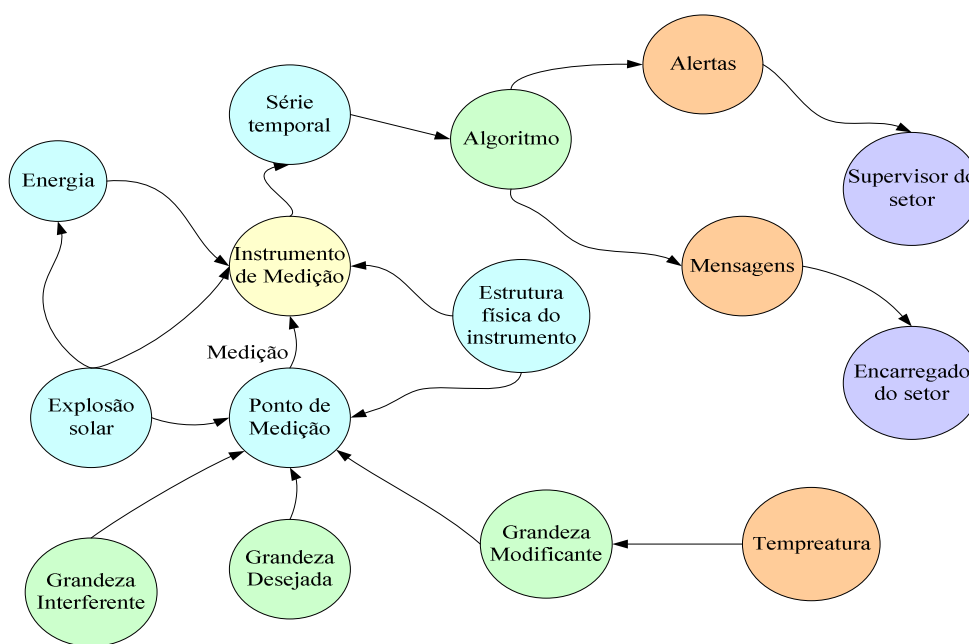
Na figura 38 está representado um exemplo de um trecho de um mapa qualitativo com relações simples indicando os elementos (nodos) e o sentido de sua influência.

A figura 38 tem como ponto central um instrumento de medição. No mapa podem ser percebidas algumas camadas de relacionamentos, assim:

- Na primeira camada, encontram-se energia, indução eletromagnética, ponto de medição, e estrutura física do instrumento, série temporal;
- Na segunda camada, são encontrados os termos: algoritmo, grandeza desejada, interferente e modificante;
- Na terceira camada, são encontradas: alertas, mensagens e temperatura;

- Na quarta camada são encontradas duas categorias de interessados: o supervisor do setor e o encarregado do setor.

Figura 38 - Trecho de um mapa tendo como centro o instrumento de medição



Fonte: Autor

Ainda na figura 38, é possível perceber que suas arestas são direcionadas indicando o sentido da influência de um nó em outro, por exemplo, a temperatura está relacionada à grandeza modificante, o algoritmo que influencia os alertas que, por sua vez, influenciam o supervisor do setor.

A partir da figura 38 é possível inferir sobre um trecho do mapa em que as grandezas desejada, interferente e modificante estão relacionadas ao ponto de medição, e a temperatura é relacionada à grandeza modificante. O instrumento de medição alimenta a série temporal interpretada por um algoritmo que emite alerta para o supervisor do setor e emite mensagens para o encarregado do setor. Note que na aresta do ponto de medição com o instrumento existe uma palavra de ligação indicando medição, que pode ser interpretada que a medição vai do ponto até o instrumento.

O mapa conceitual contém possíveis causas para os valores gerados em um ponto de medição e capturados pelo instrumento de medição que por sua vez envia para registros na série temporal. Nesse mapa estão registrados todos os possíveis caminhos que influenciam o elemento central. Embora seja visível no mapa um caminho completo ele pode ter somente uma parte ativada assim se não existirem informações sobre a grandeza interferente sua relação pode ser desconsiderada. Se por exemplo o ponto de medição for retirado, todos os caminhos que chegam a ele são desconsiderados. No caso de um nodo intermediário ser desconsiderado, só será considerado o caminho após esse nodo e no sentido do elemento central. Para que isso ocorra cada nodo deve possuir atributos que podem gerar restrições ou considerações.

#### 4.2.2.4 Regras

As regras, no Método Complementar constituem a base de conhecimento, sendo originárias de várias fontes inclusive dos mapas conceituais devido a informações pertinentes e/ou das influencias ambientais. Assim como nos mapas conceituais, as regras podem ser originárias das diversas categorias de conhecimentos, só que elas relacionam explicitamente as causas aos efeitos, sendo que um efeito em uma regra, pode ser causa em outra regra.

As regras podem ser apresentadas da seguinte forma no método:

Causa 1 e causa 2 e ..... e causa n implicam no efeito 1 e efeito 2 e ... e efeito m

O método entende que cada causa e cada efeito pode possuir seu grau de pertinência.

As regras podem ser obtidas de forma automática a partir de árvore de decisões, árvores de falha, mineração de dados obtendo regras de associação, dentre outras possibilidades.

#### 4.2.2.5 Receber Mensagens e Alertas dos Métodos Básicos

As informações fornecidas pelos Métodos Básicos devem formar uma fila FIFO por instrumento de medição para serem processados pelo Método Complementar. Cada instrumento de medição tem sua própria base de conhecimento.

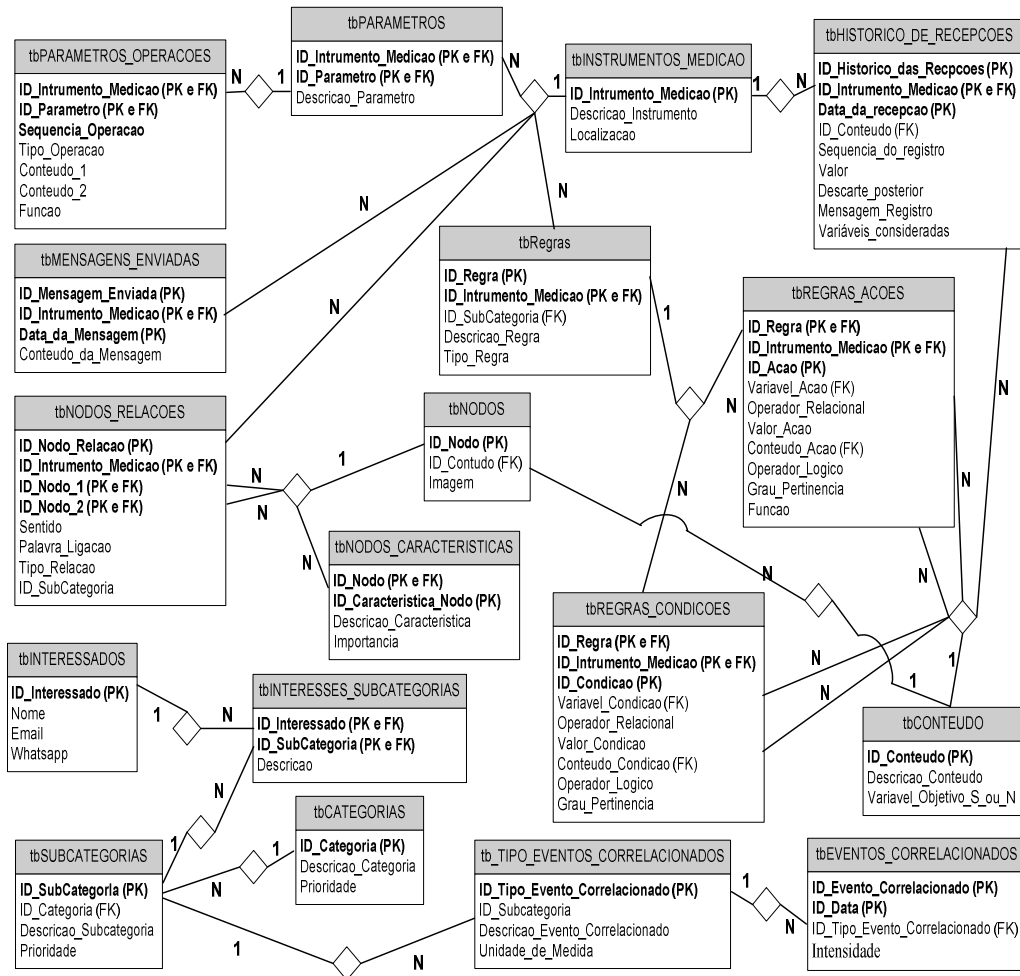
O Método Complementar obtém as informações do instrumento e realiza a inferência utilizando a base de conhecimento.

#### 4.2.2.6 Base de Conhecimento

A base de conhecimento foi desenvolvida para ser utilizada em um banco de dados relacional de forma a permitir o processamento relacionado a correlações temporais, ao mapeamento existente e às regras existentes. Na figura 39 está a estrutura do modelo de dados desenvolvido.

Na estrutura do modelo, estão representadas tabelas com as regras oriundas de informações diretas, de mapas conceituais e de diagramas de causa efeito. Além disso, estão registrados parâmetros utilizados nas operações e regras. Também estão contemplados dados dos instrumentos de medições, mensagens e seus interessados divididos em categorias, e histórico de recepções.

Figura 39 - Modelo de dados desenvolvido para o Método Complementar.



Fonte: Autor

#### 4.2.2.7 Recepção das Mensagens e Alertas dos Métodos Básicos

As mensagens e alertas resultantes do processamento dos Métodos Básicos são gravadas em uma tabela específica que orienta o processamento para cada registro em ordem de chegada à base de conhecimento do Método Complementar.

#### 4.2.2.8 Análise e Inferência

A análise tem início com o enquadramento dos conhecimentos temporais pertinentes com a data oriunda da mensagem ou alerta emitido pelo processamento dos Métodos Básicos. Na sequência ocorre o processamento segundo uma prioridade definida para as categorias iniciando pelas tabelas temporais, passando pelas regras existentes e finalmente utilizando as informações hierárquicas do mapa.

#### 4.2.2.9 Emissão de novas Mensagens e Alertas

A emissão de novas mensagens e alertas é realizada em função dos interessados segundo as categorias declaradas como pertinentes a eles. Para efeito de registro elas são armazenadas conforme a sua criação, permitindo consultas por instrumento, período, conteúdo e se foi descartada ou não pelo Método Complementar.

### 4.2.3 Estudos de Casos

Neste item são apresentadas considerações para o uso do Método Complementar após a aplicação de métodos em séries temporais, apresentando os seus requisitos originais, o ambiente da aplicação e as possíveis regras aplicadas em seus resultados.

#### 4.2.3.1 Estudo de caso Método HOT SAX Adaptado

Este estudo de caso envolve os itens 3.1 e 3.2, que respectivamente explicam o método HOT SAX e o método HOT SAX Adaptado.

Durante o desenvolvimento do método Hot Sax adaptado, foi acrescentado que as mensagens e alarmes fossem emitidos e enviados o mínimo possível aos interessados, devido à falta de tempo destes. Também, devido à existência desse pouco tempo

disponível, foi descartado o levantamento de correlações práticas e de correlações a partir de outros sistemas de coletas de dados. Desse modo, praticamente seria utilizado o próprio método Hot Sax adaptado, diminuindo as possibilidades de acerto com relação as não conformidades.

Mesmo assim as medições históricas possuíam registros que envolviam, com relação às medições, os valores obtidos e a data e hora de sua obtenção.

Desse modo, o enquadramento da aplicação do método na instalação específica da empresa está comentado no item a seguir.

**a) Proposta para a aplicação**

Devido à ausência de registros de informações pertinentes e à dificuldade de tomar tempo do pessoal da instalação, para a aquisição de informações e identificação de correlações, a proposta é mínima e corresponde à obtenção de dados externos de domínio público, na obtenção de informações das séries temporais relacionadas ao nível de água da represa, e das séries temporais de outros instrumentos de medição próximos.

Essa proposta não utiliza o pessoal interno da empresa, e atualizaria as informações fornecidas de forma automática, possibilitando o uso pelo Método Complementar. Algumas das informações envolveriam a correlação do instrumento de medição, sua proteção, fenômenos provenientes de induções eletromagnéticas e de abalos sísmicos com reflexo locais. Além disso, envolveria séries temporais com registros locais, no caso nível d'água. Basicamente seriam utilizados robôs de busca na internet e nos registros locais correspondentes aos níveis da água na represa, essas informações seriam capturadas e incorporadas na base de conhecimento para correlações temporais.

A partir dessas informações, seria possível afirmar, por exemplo, em uma mensagem ou alarme, que a ocorrência foi influenciada por um abalo sísmico cujo período coincide com a medição indicada como não conforme pelo Método Básico, deixando claro que essa mensagem terá um atraso em função da obtenção da informação do abalo e sua correlação, mas existirá uma informação mais assertiva sobre a ocorrência desse evento.

As informações obtidas dessa forma fariam uma correlação podendo descartas ou não uma mensagem ou alarme. No caso de emitir nova mensagem ou alarme, nova informação será acrescentada com o fenômeno correlacionado e de sua influência nas medições indicadas.

#### b) Considerações com relação ao estudo de caso

Esse estudo de caso mostra que é possível evoluir o método com a busca de informações externas na Internet ou outras bases de conhecimento, como no caso de níveis de água disponíveis na organização. Essas informações seriam obtidas com o uso de robôs de busca específicos direcionados para as fontes de conhecimentos também específicas. Cabe ressaltar que informações oriundas de outros instrumentos de medição com relação à mesma data seriam importantes.

No caso as possíveis correlações poderiam eliminar a emissão de alarmes desnecessários. Outras correlações poderiam ser climáticas e estar relacionadas a enchentes, volumes de chuvas, dentre outros.

#### 4.2.3.2 Estudo de caso envolvendo a pesquisa de Schwabacher, Oza e Matthews

O estudo apresentado em Schwabacher; Oza; Matthews (2009) envolve o *Space Shuttle* (Ônibus Espacial) considerando quatro algoritmos não supervisionados e nove anomalias.

Para Schwabacher; Oza; Matthews (2009), a existência de diferentes definições de anomalia, os algoritmos comparados detectaram diferentes anomalias; desse modo, os autores concluíram que é útil utilizar vários algoritmos.

Com relação a especialistas, Schwabacher; Oza; Matthews (2009) enfatizam que, a abordagem utilizada na detecção de anomalias em dados do sensor era utilizar um grande número de especialistas humanos, como os controladores de voo que observavam os dados em tempo quase real durante cada voo, e os engenheiros, que estudavam os dados após



cada voo. Além disso, verificações de limite sinalizavam quando uma determinada variável ficava fora de um intervalo pré-determinado.

Segundo Schwabacher; Oza; Matthews (2009) os seres humanos podem não ser capazes de reconhecer as falhas que envolvem relações entre grande número de variáveis. Além disso, algumas falhas potenciais poderiam acontecer rápido demais para que os seres humanos pudessem detectá-las e reagir antes que se tornassem catastróficas.

Mas, em seus testes, tiveram que considerar a validação subjetiva das anomalias por especialistas de domínio. Sinalizam que nos resultados apresentados no artigo, os algoritmos de detecção de anomalias não fazem uso de qualquer conhecimento especializado. Assim, eles indicaram a pretensão de explorar formas de usar o conhecimento de fundo no contexto automatizado de detecção de anomalia. Uma possível fonte de conhecimento é a determinação feita por especialistas do domínio de que certas anomalias detectadas pelos algoritmos não são significativas. Uma maneira de fazer o uso deste conhecimento é a utilização de algoritmos semissupervisionados de aprendizagem e proporcionar-lhes com exemplos etiquetados como nominal para cada anomalia candidata que os peritos julgarem como nominal. O algoritmo, então, pode evitar sinalização incorreta de uma anomalia no futuro, quando padrões similares aparecessem nos dados.

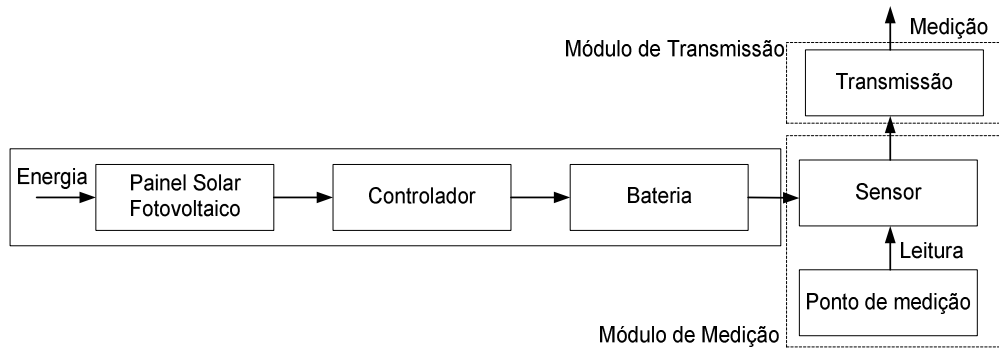
Na proposta para a utilização do Método Complementar basta capturar o resultado dos métodos e utilizá-las com regras fornecidas por especialistas de forma a evitar as interpretações incorretas, permitindo uma validação mais assertiva.

#### 4.3 EXEMPLO DE APLICAÇÃO

A aplicação estudada é composta por um conjunto isolado de dispositivos instalados no entorno de uma barragem hidroelétrica. Nela, os sinais provenientes de um piezômetro (responsável para medir a pressão estática da barragem) são transmitidos para serem utilizados em um computador central. O instrumento é composto por três módulos: suprimento de energia, medição e transmissão de dados. A bateria é carregada por um

painel solar fotovoltaico responsável pela geração de energia elétrica. A figura 40 mostra o esquema da instalação.

Figura 40 – Esquema da instalação utilizada



Fonte: Autor

A partir das medidas obtidas no ponto de medição (de agosto a outubro de 2014) pelo piezômetro instalado, uma série temporal contendo 2289 amostras foi armazenada e traçada.

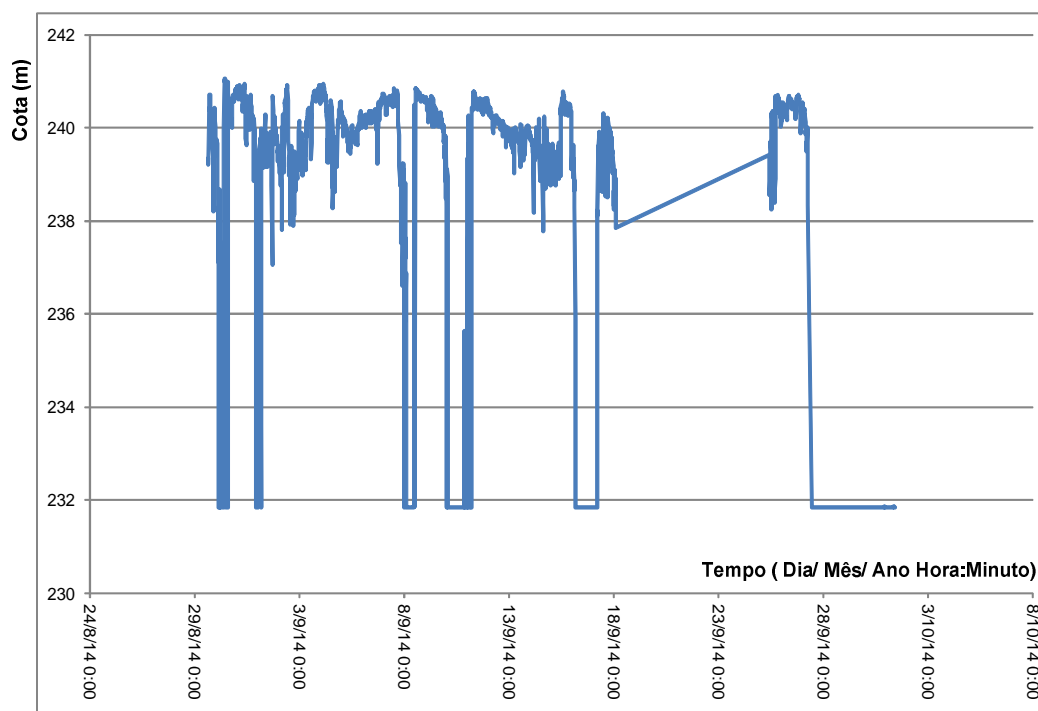
As medições utilizadas correspondem ao piezômetro PZ1762, instalado em uma usina hidroelétrica pertencente a CESP na barragem de terra. A leitura é feita em volts pelo sensor MPXH6250A podendo estar no intervalo de 0,3 à 4,9 V, segundo o fabricante também correspondendo ao intervalo de 20 à 250 kPa, ou de 3 a 36 psi, com erro 1,5% e trabalhando bem de 0 à 85 °C).

No sistema, os valores de leitura estão em volts e transformados em metro de coluna d'água e, em seguida transformados em cota em função de uma cota de referência da usina utilizados pelo sistema de monitoramento. Desse modo, as medições são utilizadas em cotas, por exemplo, 210 m no gráfico e utilizado computacionalmente como 210.000.000 ( $10^{-6}$  m).

O período de aquisição de sinal é 15 minutos de 29/08/2014, 16h07min à 01/10/2014 10h04min. A figura 41 mostra a série temporal que representa a pressão medida pelo piezômetro instalado, transformada em coluna de água e, posteriormente em

cota de referência. Assim, o eixo Y representa os valores medidos em metro de coluna de água e transformados em função de uma cota de referência da usina e o eixo x representa o tempo.

Figura 41- Serie temporal obtida pelo Piezômetro



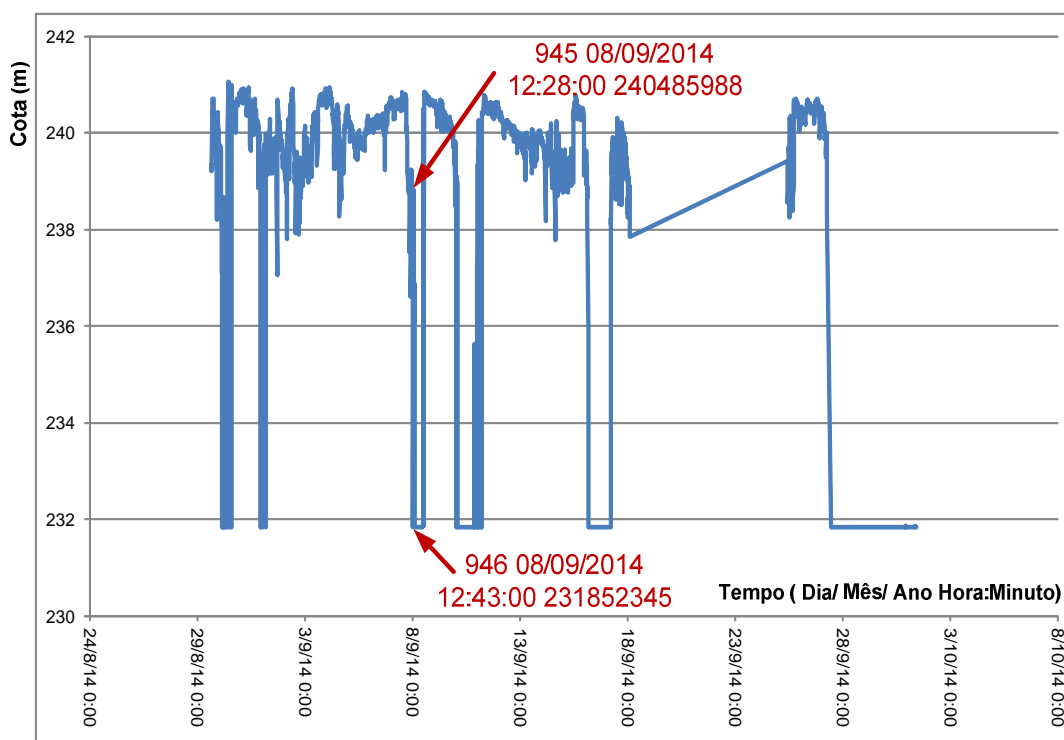
Fonte: Autor

No caso do piezômetro PZ1762, o valor de referência é aproximadamente a cota de 240,520004 metros, cabendo ressaltar que, se a leitura do instrumento tender a zero volt e também mais próximo de zero, estará a coluna de água (sem água), o que corresponde a cota de instalação que está por volta da cota de 231 metros. Dessa forma, na figura 41, percebe-se que os valores inferiores (nos patamares) podem indicar falta de água ou falta total de energia na leitura dando a leitura de fundo.

Também, na figura 41, deve-se atentar para o fato de existir uma sequência das medições da série (armazenada nos registros), a data/hora e o valor inteiro (utilizados

assim para facilitar o processamento), o valor real (em cota em metro) corresponde ao valor inteiro dividido por 1000000, sendo que um valor inteiro da medição (no computador), que é 231852345, corresponde a um valor real de 231,852345 metros, que, por sua vez, tem um valor correspondente em coluna de água e em volt. Com a finalidade de facilitar o entendimento do processamento realizado nesta aplicação estão apresentados no gráfico da figura 42 dois pontos utilizados nesta aplicação: 945 e 946.

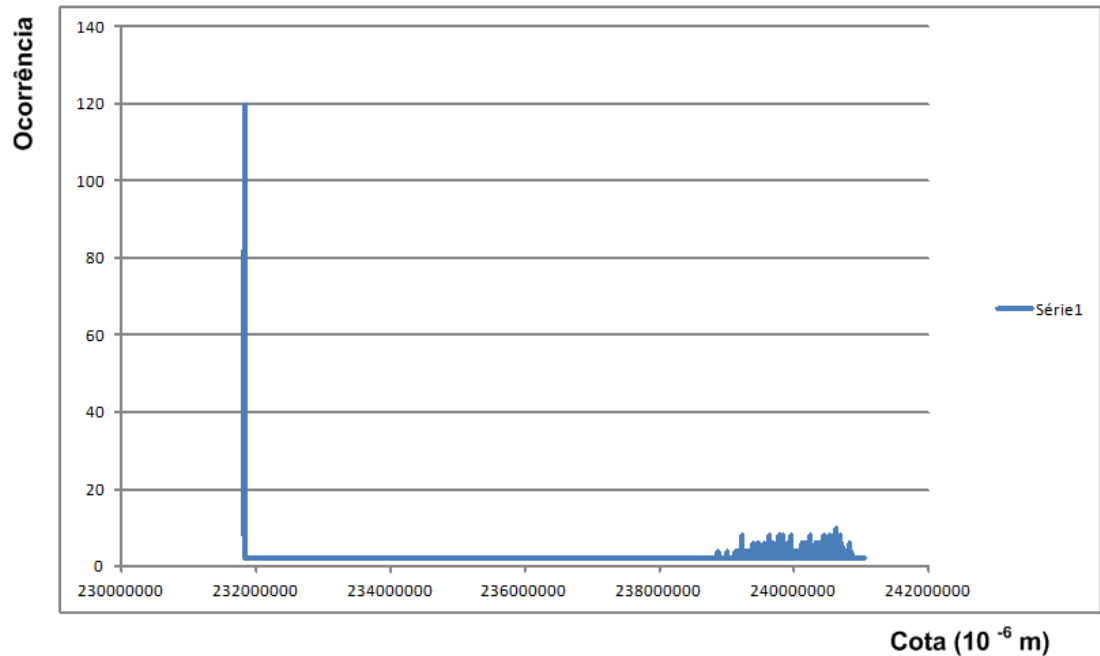
Figura 42- Pontos estudados nesta aplicação no piezômetro



Fonte: Autor

Utilizando o mesmo conjunto de dados, um histograma ilustra a ocorrência de cada valor individual medido. A figura 43 mostra o histograma construído com o conjunto de medições provenientes do piezômetro, onde eixo y representa a quantidade de vezes da ocorrência de um determinado valor, enquanto eixo dos x indica o valor das cotas.

Figura 43- Histograma com os dados da série temporal



Fonte: Autor

Cabe notar que a série temporal é utilizada em um método denominado Básico (no caso, o MAIM), que gera mensagens contendo: a identificação do instrumento de medição, a medição realizada (número sequencial da medição, data, e valor) e a indicação da interpretação inicial de uma anomalia. Essa mensagem é recebida pelo método complementar.

O método básico forneceu ao método complementar os seguintes dados:

- Identificação do Instrumento de Medição;
- Candidato à anomalia;
- Identificação do Instrumento de Medição Piezômetro PZ1762;
- Sequência na série temporal original: 946;
- Data/hora da ocorrência: 08/09/2014 12:43:00;
- Valor medição: 231852345;

- Inclinação crescente crítica;
- Valor próximo ao mínimo do projeto;
- Texto: Inclinação gráfica abrupta com valor próximo ao mínimo do projeto entre 232000000 e 23100000.

Ao utilizar o método complementar, as mensagens originais são inseridas no *software* utilizado como proposto na arquitetura mostrada no item 1.1.2 na figura 2, assim uma nova inferência é realizada, resultando em nova mensagem com mais detalhes sobre a anomalia detectada pelo método principal.

Cabe ressaltar que neste exemplo de aplicação o ponto de medição é monitorado por um piezômetro, cuja Base de Conhecimento completa é composta por informações oriundas: do manual técnico do instrumento, de entrevistas técnicas com especialista, de um banco de dados da base Nacional de Meteorologia, do mapa conceitual desenvolvido, além de outras fontes. No processamento, parte da Base de Conhecimento é utilizado na forma de regras.

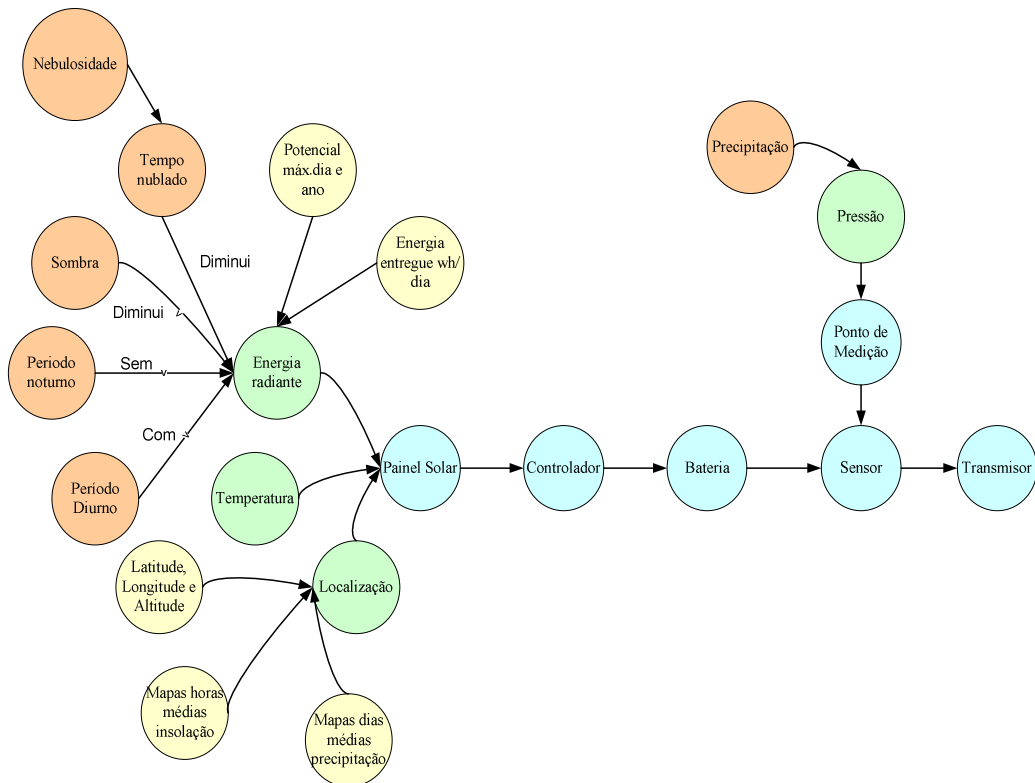
O mapa conceitual utilizado é composto pelas influências existentes nos arredores do conjunto do piezômetro utilizado através do estabelecimento das relações existentes entre os conhecimentos pertinentes como mostrado na figura 42.

Para o piezômetro instalado, há muitas possibilidades capazes de causar uma falha. Supondo que uma falha de energia ocorra, o método complementar pode listar uma série de possibilidades para melhorar a assertividade e a precisão da mensagem, o que resulta em uma análise que pode possibilitar mensagens, alertas e ações de manutenção.

No mapa apresentado na figura 44, são indicadas em círculos azuis o painel solar, controlador, bateria, sensor, transmissor e o ponto de medição que compreendem todo o instrumento. Os círculos verdes representam as variáveis que influenciam os dispositivos, tais como a pressão que desloca o valor medido pelo sensor, bem como a localização, a temperatura e a energia radiante do sol, que afetam diretamente o desempenho do painel solar. Em círculos amarelos e laranja, estão as variáveis que podem influenciar outros parâmetros, de acordo com a direção das setas mostradas no mapa. A partir do mapa

conceitual, a relação de causa e efeito pode ser estabelecida. Por exemplo: uma sombra causa a diminuição de energia radiante no painel solar, o que pode implicar na redução da energia gerada, que pode levar ou não a um problema sobre o funcionamento do aparelho.

Figura 44 – Parte de um mapa conceitual tendo como elemento central o sensor



Fonte: Autor

Além disso, são consideradas as informações sobre a precipitação de chuva na região da barragem e sobre o índice de nebulosidade, como em (WOYTE; BELMANS; NIJS, 2007) e como as encontradas no banco de dados do Instituto Nacional de Meteorologia (INMET, 2016). Essas informações são organizadas em tabelas para a precipitação de chuva da estação de medição, data, hora e a taxa de precipitação de chuva são os dados principais, enquanto que, para o índice de nebulosidade atmosférica, os mesmos parâmetros são utilizados com a adição de parâmetros sobre a umidade do ar. Como exemplos são apresentados trechos das tabelas de nebulosidade e de precipitação,

nas tabelas 7 e 8. Para a data 08/09/2014 e hora 12:43:00, a correlação com a nebulosidade corresponde ao valor 0 indicando um dia totalmente claro; já no caso da precipitação, corresponde também ao valor zero, o que é, pela lógica, céu claro e sem chuva.

Tabela 7- Exemplo de Trecho da tabela de nebulosidade

<b>Estação</b>	<b>Data</b>	<b>Hora</b>	<b>Temperatura do Bulbo Seco</b>	<b>Temperatura do Bulbo Úmido</b>	<b>Nebulosidade</b>
83033	01/01/2014	0	26.9	24.9	8.00
83033	01/01/2014	1200	26.9	24.4	0.00
83033	01/01/2014	1800	32.3	26.1	3.75
83033	02/01/2014	0	27.5	25.3	0.00
83033	02/01/2014	1200	27.5	24.3	6.25
83033	02/01/2014	1800	34.1	26.3	5.00

Fonte: Autor

Tabela 8 – Exemplo de Trecho da tabela de Precipitação

<b>Estação</b>	<b>Data</b>	<b>Hora</b>	<b>Precipitação</b>
83033	01/01/2014	0	
83033	01/01/2014	1200	1.4
83033	02/01/2014	0	
83033	02/01/2014	1200	11.4
83033	03/01/2014	0	
83033	03/01/2014	1200	5.8
83033	04/01/2014	0	

Fonte: Autor

Assim, correlacionando todos os dados disponíveis na referida data 08/09/2014 - 00:43, o sistema complementar usa as regras estabelecidas, como mostrado no quadro 8, que indica a relação sequencial das regras utilizadas. A aplicação ocorreu da seguinte forma: com a informação originária do método principal foi realizada a correlação com informações remotas de nebulosidade e precipitação relativa ao período da obtenção de dados, além de regras já existentes na base de conhecimento foi inferida a questão de dia claro.



Quadro 8 – Relação sequencial das regras utilizadas

<b>Regra aplicada</b>	<b>Origem da regra</b>
data/hora = <b>informada na tabela nebulosidade</b> → nebulosidade = 0	Correlação entre eventos periódicos e períodos equivalentes usados no método principal.
data/hora = <b>informada na tabela precipitação</b> → precipitação = 0	Correlação entre eventos periódicos e períodos equivalentes usados no método principal.
Se data/hora = <b>informada</b> e nebulosidade = 0 e precipitação = 0 → céu claro = <b>sim</b>	Inferência obtida a partir das duas regras anteriores
Se horário informado > <b>hora início dia</b> e horário informado < <b>hora fim dia</b> → horário correspondente = <b>dia</b>	Inferência com dados informados pelo método anterior
Se valor informado muito baixo = <b>sim</b> e inclinação decrescente crítica informada = <b>sim</b> e céu claro = <b>sim</b> e horário correspondente = <b>dia</b> → Bateria carga baixa = <b>sim</b> e sombra no painel = <b>talvez</b> e verificar instalação = <b>sim</b>	Inferência com dados classificados a partir de informação do método anterior e inferências anteriores
Se sombra no painel = <b>talvez</b> → <b>verificar possibilidade de sombra</b>	Ação a partir de dados inferidos
Se Bateria carga baixa = <b>sim</b> → <b>verificar bateria</b>	Ação a partir de dados inferidos

Fonte: Autor

A partir do processamento do método complementar, para a ação de verificar a bateria, muitas recomendações podem ser incorporadas na mensagem, por exemplo: verificar rachaduras, deformações e vazamento de água; apertar conexão; verificar oxidação; verificar o desempenho; realizar trabalhos de limpeza. Observa-se que dados de serviço, dentre outras opções, podem ser incorporados na base de conhecimento, enquanto o sistema computacional evolui ao longo do tempo. Como resultado, a mensagem final será mais assertiva e precisa em comparação com a mensagem original para o usuário final, pois pode incorporar justificativa e ações recomendadas.

#### 4.4 CONSIDERAÇÕES FINAIS SOBRE O MÉTODO COMPLEMENTAR

O método Complementar deve ser utilizado após o método MAIM ou após outro método que apresente dados candidatos à anomalia. Desse modo, os resultados obtidos no

MAIM seriam enviados para Método Complementar, que efetuará suas operações complementando os resultados do método MAIM e enviando quando necessário, aos usuários tomadores de decisões.

Como mostrado na seção de Estudo de Casos, o Sistema Complementar proposto para mensagens é capaz de receber dados provenientes de várias fontes, tais como a experiência humana, os dados históricos armazenados em web ou em uma rede local, realizando uma inferência e gerando novas mensagens com mais detalhes para o usuário final. Com base em experiências em um aplicativo de monitoramento real, o sistema revela capacidade de melhorar a mensagem final através da base de conhecimento.

O Sistema Complementar processa os dados imediatamente após o recebimento do método Básico, trazendo melhorias em termos de qualidade para a mensagem final sem consumir grandes recursos computacionais.

Novas regras, dados e aprendizagem também podem ser atualizadas a qualquer momento pela equipe de engenharia, a fim de evoluir o sistema computacional ao longo do tempo, o que pode trazer muitos benefícios para todas as pessoas envolvidas na operação desse sistema. Finalmente, não foi encontrado sistema computacional semelhante na literatura, portanto, nenhuma comparação em termos de desempenho e precisão pôde ser estabelecida. No entanto, em termos de qualidade, o Sistema Complementar é capaz de atingir seu objetivo de suplementar as mensagens provenientes do método Básico, como proposto.

## 5 APLICAÇÕES DO MÉTODO MAIM

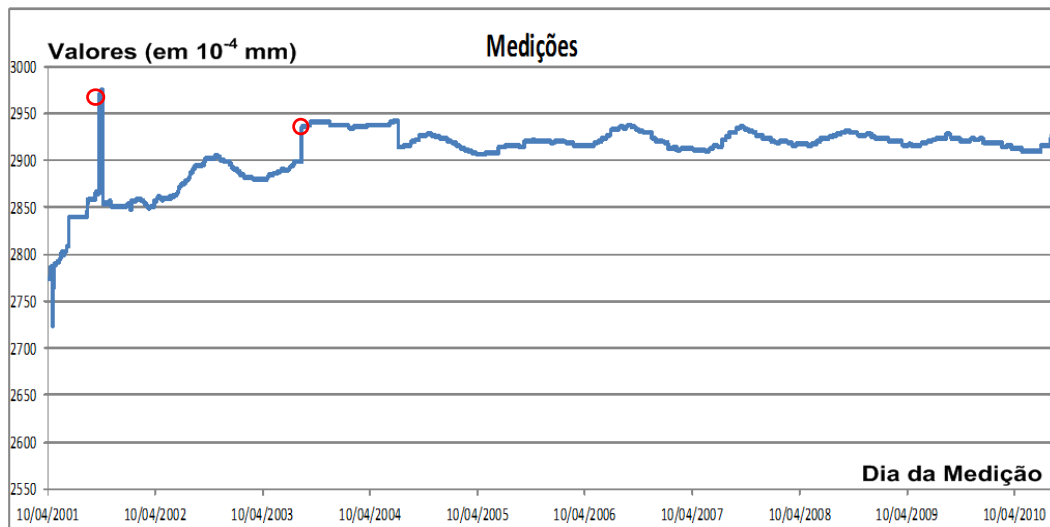
As aplicações apresentadas neste item consideram que ao longo do tempo um Ambiente é mutável em função das alterações de sua realidade, e ele é classificado em três categorias: Novo, em Evolução e Estável, conforme apresentado no item 4.1.1.2. Considera-se também que uma Novidade corresponde a uma informação que ocorreu uma única vez em um ambiente. Dessa forma sua primeira ocorrência é uma novidade, item 4.1.1.1. A aplicação desenvolvida tem como embasamento as validações realizadas no item 4.1.2, envolvendo Agrupamentos de valores no item 4.1.3, envolvendo Estabilidade e Ambiente e, finalmente, no item 4.1.7, envolvendo as tabelas de decisões simples e adaptativa.

### 5.1 INFORMAÇÕES GERAIS DA APLICAÇÃO QUE ENVOLVE O SENSOR TRIORTOGONAL

Nesta aplicação, são apresentadas as análises referentes a dois pontos de uma série histórica que podem ser candidatos a conter anomalia, correspondendo a medições realizadas por um sensor triortogonal, cujo código na instalação é (MT-V2-3) eixo um localizado no vertedouro da Usina Hidroelétrica de Lajeado, com período de medições de 15 dias, tomados entre 10/04/2001 a 26/08/2010, num total de 3426 medições, cujo traçado gráfico esta apresentado na figura 45.

No gráfico da figura 45, é possível identificar visualmente os dois pontos importantes escolhidos para a análise, destacados em círculos vermelhos. O primeiro ponto corresponde a um ponto máximo formando um pico, e o segundo é um ponto alto iniciando um patamar bem formado visualmente. Cabe ressaltar, neste momento, que os registros da série temporal também são numerados de forma crescente, e a unidade de deslocamento é dada em  $10^{-4}$  (mm). Ressalta-se que, além dos valores testados, outros pontos são detectados como candidatos a anomalia durante a aplicação do modelo computacional.

Figura 45 – Serie Temporal do Medidor Triortogonal MT-V2-3 eixo Um



Fonte: Autor

Na análise não foram considerados os limites do instrumento. No processamento do método MAIM com os dados históricos do instrumento MT-V2-3 eixo um, as informações foram processadas e analisadas individualmente em uma simulação que interpreta os dados históricos como se fossem em tempo real.

Cabe ressaltar que nos registros da série temporal as medições são numeradas de forma crescente. Com relação ao Pico, os pontos de medição estudados envolvem as medições de número 181, 182 e 183, 209 e 210 e, no caso que envolve o ponto de início do patamar escolhido, as medições correspondem aos pontos de 863, 864 e 865. A utilização dos pontos 181, 182, 863 e 864 serve como referência no texto para o entendimento das sequencias candidatas à anomalia, no processamento normal, elas não são exibidas ao usuário.

A seguir são apresentadas informações que envolvem o processamento e a análise utilizada no sensor Triortogonal.

### 5.1.1 Informações sobre o processamento e análise

Na figura 46, há exemplos de informações que, quando utilizadas em sequência com a tabela de decisões adaptativa ativam ações que disparam funções específicas apresentadas na figura 47.

A figura 46 apresenta quatro registros relativos a quatro entradas numeradas de 133 a 136 (que não correspondem aos dois pontos analisados), acrescentados por informações complementares obtidas durante o processamento. Uma entrada, normalmente, possui uma informação constituída pelos campos: sequência, código do sensor no caso “T”, instante do registro, e valor. Durante o processamento, são acrescentados novos campos que indicam: diferença de tempo entre os registros, diferença de valores, inclinação, indicação da identificação do ambiente, o estado do ambiente, e indicações relativas à indicação de Novidade por grupo e só do grupo1.

Figura 46 - Informações utilizadas na Tabela de decisões

SEQUENCIA	SENSOR	DIA	VALOR	DIAS_APOS_ANTERIOR	DIFERENCA_ATUAL_ANT
133	T	20/08/2001 00:00	2840	1	0,00
134	T	21/08/2001 00:00	2840	1	0,00
135	T	22/08/2001 00:00	2858	1	18,00
136	T	23/08/2001 00:00	2858	1	0,00

INCLINACAO_COM_ANTERIOR	INDICACAO	AMBIENTE	ESTADO_AMBIENTE	INDICACAO_GRUPO1
0,00		1	ESTAVEL	
0,00		1	ESTAVEL	
1,52	N5	2	NOVO	N1
0,00	N5	2	EVOLUCAO	

Fonte: Autor

Na figura 47, aparece a tabela de decisões simplificada da aplicação resultante das condições existentes com regras estabelecidas principalmente pelo reconhecimento da indicação de uma Novidade e do estado do ambiente, o período, os limites operacionais de valor e angular, que geram ações que por sua vez ativam funções. Cada célula resultante do encontro com uma regra e uma condição pode ter ou não algum conteúdo; se não tiver

nenhum conteúdo, ela é desconsiderada na tomada de decisão. Os possíveis conteúdos são para:

- Novidade que reflete a pergunta se é ou não novidade: sim, não;
- Estados do Ambiente: novo, em evolução, estável;
- Período, a diferença com dois ou mais períodos (igual 2, maior 2) e igual a um período;
- No caso dos limites Operacionais de valor e angular: menor, maior, dentro;
- Suporte (índice relativo à regra de associação).

No caso de ser realizada a busca, considerando-se as prioridades pré-definidas no conjunto de regras existentes, a regra identificada corresponde àquela em que todas as condições são satisfeitas. Por exemplo, se a regra R6 é a escolhida, duas condições seriam ativadas: uma para mudar a identificação do ambiente (executando a função  $f_{ma}$ ) e outra para considerar o valor da medição como candidato à anomalia (executando a função  $f_{ca}$ ).

Cabe ressaltar que, na representação da figura 47, os componentes opcionais da Ação Adaptativa Anterior e Geradores não estão indicados devido ao fato de serem opcionais como apresentados no item 2.3.2 e retratados por (TCHEMRA, 2009). Na figura 47, estão indicadas funções relacionadas às ações da tabela de decisões, que corresponde ao dispositivo subjacente.

Figura 47 - Tabela de decisões da aplicação com as funções relacionadas às ações

Regras	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	
Prioridades	0	0	0	0	0	1	1	1	1	1	1	0
Condições:												
Novidade					Não	Sim	Sim	Não	Não	Sim		
Ambiente					Estável	Estável	Evolução	Evolução	Novo	Evolução		
Períodos entre registros	Igual 2	Maior 2			Igual 1	Igual 1	Igual 1	Igual 1	Igual 1	Igual 1		
Limite Operacional Valor anterior				Maior	Dentro	Dentro	Dentro	Dentro	Dentro	Dentro		
Limite Operacional angular adotad			Maior		Menor	Menor	Menor	Menor	Menor	Menor	Maior	
Limite suporte G2					< SUP.	< SUP.	< SUP.	< SUP.	< SUP.	< SUP.		
Limite de Projeto Inferior												Próximo
Ações												Funções
Sinalizar Mensagem Anomalia de periodo	xxx											f_smap()
Sinalizar Mudança de Ambiente									xxx			f_smam()
Sinalizar verificação		xxx										f_sv()
Sinalizar Alerta			xxx	xxx								f_sa()
Mudar cod.Ambiente						xxx						f_ma()
Candidato anomalia						xxx	xxx			xxx		f_ca()
Sinalizar valor minimo de projeto											xxx	f_vmp()
Não sinalizar					xxx			xxx				

Fonte: Autor

A seguir são realizadas as análises dos pontos escolhidos.

### 5.1.2 Verificação se o Primeiro ponto, o pico, corresponde a uma Anomalia.

O pico formado é provavelmente candidato à anomalia, pois corresponde à sequencia 182 da série temporal, indicando e indica o dia 08/10/2001, com valor de 2975 a identificação do ambiente 3 tipo evolução. Sabe-se de antemão que é uma região do gráfico que corresponde ao assentamento do terreno na Usina. Na tabela 9, estão apresentados os registros da série temporal, sendo que a sequencia 181 corresponde aos dados anteriores ao inicio do pico e a sequênciã 183 contém os dados seguintes a 182 (que também é um pequeno patamar).

Tabela 9 - Trecho da série temporal com os registros das sequencias 181, 182 e 183.

<b>Registros da Série Histórica (tbSERIE_HISTORICA)</b>			
<b>Sequencia</b>	181	182	183
<b>Sensor</b>	MT-V2-3 eixo Um	MT-V2-3 eixo Um	MT-V2-3 eixo Um
<b>Dia</b>	07/10/2001	08/10/2001	09/10/2001
<b>Valor</b>	2970	2975	2975
<b>Dias após anterior</b>	1	1	1
<b>Diferença atual anterior</b>	0,00	5,00	0,00
<b>Inclinação com anterior (rd)</b>	0,00	1,37	0,00
<b>Inclinação com anterior (graus)</b>	0,00	78,69	0,00
<b>Indicação</b>		N5	N5
<b>Identificação do Ambiente</b>	3	3	3
<b>Estado do Ambiente</b>	Evolução	Evolução	Evolução
<b>Indicação Grupo 1</b>		N1	

Fonte: Autor

Cabe ressaltar que a mudança da identificação do ambiente (é sequencial) do número 3 para 4 só ocorrerá na sequencia 210, isso devido ao aparecimento do valor 2857, conforme mostrado na tabela 10.

Tabela 10 - Registros da série Histórica com a indicação da mudança da identificação do ambiente 3 para o 4

<b>Registros da Série Histórica (tbSERIE_HISTORICA)</b>		
<b>Sequencia</b>	<b>209</b>	<b>210</b>
<b>Sensor</b>	MT-V2-3 eixo Um	MT-V2-3 eixo Um
<b>Dia</b>	04/11/2001	05/11/2001
<b>Valor</b>	2855	2857
<b>Dias após anterior</b>	1	1
<b>Diferença atual anterior</b>	0,00	2,00
<b>Inclinação com anterior (rd)</b>	0,00	1,11
<b>Inclinação com anterior (graus)</b>	0,00	64,44
<b>Indicação</b>		N5
<b>Identificação do Ambiente</b>	3	4
<b>Estado do Ambiente</b>	Estável	Novo
<b>Indicação Grupo 1</b>		N1

Fonte: Autor



Ao mesmo tempo em que a Serie Histórica é completada com valores, também são registrados conteúdos em uma tabela do banco de dados contendo os últimos cinco movimentos, a última identificação do ambiente e os limites operacionais do ambiente até a última sequencia processada. A tabela 11 apresenta esses conteúdos.

Tabela 11 – Cinco últimos registros medidos após as sequencia 181, 182 e 183.

<b>Registros de resumos de referência da Série Histórica (tbResumo_Cinco_Ultimos)</b>			
<b>Após Sequencia</b>	<b>181</b>	<b>182</b>	<b>183</b>
<b>Sensor</b>	MT-V2-3 eixo Um	MT-V2-3 eixo Um	MT-V2-3 eixo Um
<b>Descrição</b>	Triortogonal	Triortogonal	Triortogonal
<b>Atual</b>	2970	2975	2975
<b>Segundo</b>	2970	2970	2975
<b>Terceiro</b>	2970	2970	2970
<b>Quarto</b>	2970	2970	2970
<b>Quinto</b>	2970	2970	2970
<b>Identificação do Ambiente</b>	3	3	3
<b>Limite Operacional Menor</b>	2724	2724	2724
<b>Limite Operacional Maior</b>	2970	2975	2975
<b>Medias QL</b>			
<b>Data da Ultima Novidade</b>			
<b>Data Atual</b>			

Fonte: Autor

Nota-se que, durante o processamento das sequencias 181, 182 e 183, os grupos passaram a conter informações, conforme apresentado na tabela 12, onde o cabeçalho Q corresponde à quantidade de ocorrências de cada agrupamento.

Tabela 12 Grupos registrados durante cada sequencia

<b>Sequencia 181</b>		<b>Sequencia 182</b>		<b>Sequencia 183</b>	
<b>Grupo</b>	<b>Q</b>	<b>Grupo</b>	<b>Q</b>	<b>Grupo</b>	<b>Q</b>
2970	14	2975	1	2975	1
2970_2970	12	2975_2970	1	2975_2975	1
2970_2970_2970	10	2975_2970_2970	1	2975_2975_2970	1
2970_2970_2970_2970	8	2975_2970_2970_2970	1	2975_2975_2970_2970	1
2970_2970_2970_2970_2970	6	2975_2970_2970_2970_2970	1	2975_2975_2970_2970_2970	1
70		0		2970	

Fonte: Autor

Assim, os dados obtidos serão reunidos utilizando a tabela de decisões adaptativa, implantada segundo a representação simplificada no modelo da figura 42.

No quadro 9, estão os dados utilizados e a inferência realizada durante o processamento de cada uma das sequencias 181, 182 e 183. Observa-se que o ponto 182 corresponde a uma candidata à anomalia, pois o MAIM recomenda sinalizar a partir da regra R4 um alerta.

Quadro 9 – Resultados após o processamento das sequencias de 181 a 183

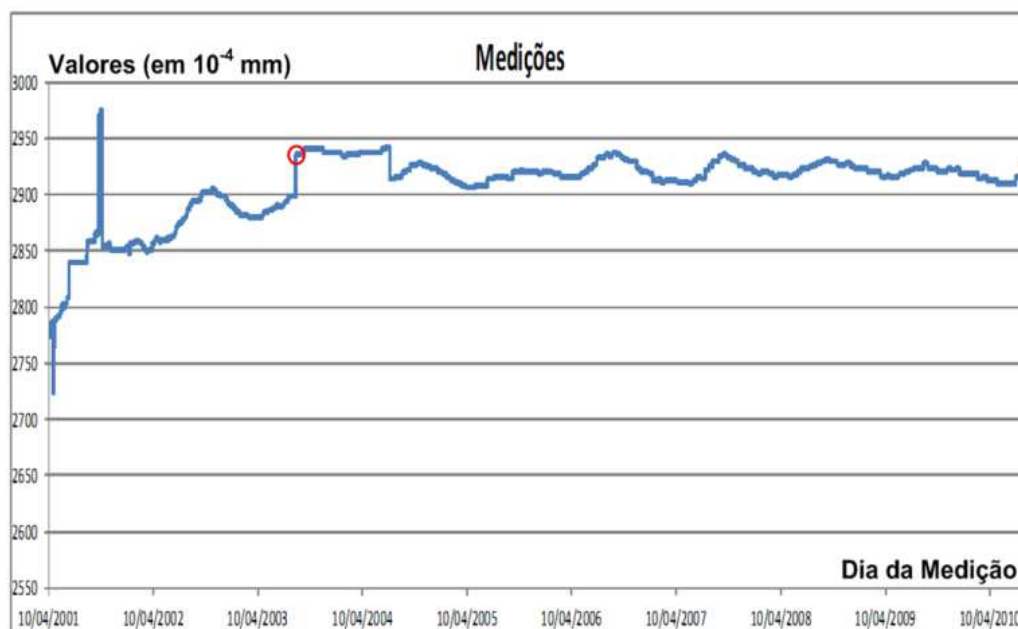
Sequencia 181	Sequencia 182	Sequencia 183
<b>Dados reunidos para a inferência</b> <b>Valor = 2970</b> <b>Período entre os registros = 1</b> <b>Identificação Ambiente = 3</b> <b>Novidade = não</b> <b>Estado do ambiente = Evolução</b> <b>Limite operacional anterior</b> <b>Menor = 2724 e Maior = 2970 (da sequencia 180)</b> <b>Limite Operacional Angular adotado = 85,94° ou 1,5 rad.</b> <b>Inclinação = 0 rd</b> <b>Inclinação graus = 0°</b> <b>Suporte = não foi necessário, só se incluir previsão imediata.</b>	<b>Dados reunidos para a inferência</b> <b>Valor = 2975</b> <b>Período entre os registros = 1</b> <b>Identificação Ambiente = 3</b> <b>Novidade = sim (surgimento de um novo grupo)</b> <b>Estado do ambiente = Evolução</b> <b>Limite operacional anterior</b> <b>Menor = 2724 e Maior = 2970 (da sequencia 181)</b> <b>Limite Operacional Angular adotado = 85,94° ou 1,5 rad.</b> <b>Inclinação = 1,37 rd</b> <b>Inclinação graus = 78,69°</b> <b>Suporte = não foi necessário, só se incluir previsão imediata.</b>	<b>Dados reunidos para a inferência</b> <b>Valor = 2975</b> <b>Período entre os registros = 1</b> <b>Identificação Ambiente = 3</b> <b>Novidade = não</b> <b>Estado do ambiente = Evolução</b> <b>Limite operacional anterior</b> <b>Menor = 2724 e Maior = 2975 (da sequencia 182)</b> <b>Limite Operacional Angular adotado = 85,94° ou 1,5 rad.</b> <b>Inclinação = 0 rd</b> <b>Inclinação graus = 0°</b> <b>Suporte = não foi necessário, só se incluir previsão imediata.</b>
<b>A Inferência Resultante</b> Aplicação da <b>regra R8</b> sem nenhuma sinalização	<b>A Inferência Resultante</b> Aplicação da <b>regra R4</b> sinalizar alerta	<b>A Inferência Resultante</b> Aplicação da <b>regra R8</b> sem nenhuma sinalização

Fonte: Autor

### 5.1.3 Verificação se o Segundo ponto, corresponde a uma anomalia

O segundo ponto corresponde a 21/08/2003, com valor 2936 registrado com sequência 864, Identificação do ambiente é 20 e o tipo de ambiente é novo. Seu gráfico está apresentado na figura 48 destacando melhor os detalhes analisados. Os dados gravados na tabela do banco de dados e utilizados na Tabela de Decisões Adaptativa estão apresentados na figura 49 e na tabela 13.

Figura 48 - Série temporal de um eixo do sensor triortogonal instalado no vertedouro



Fonte: Autor

Os dados apresentados na figura 49 correspondem à imagem obtida da interface nativa do Banco de Dados utilizado. Nela estão apresentados campos registrados nas sequências estudadas da tabela de serie temporal tbSERIE\_HISTORICA.

Figura 49 - Tabela de decisões da aplicação

SEQUENCIA	SENSOR	DIA	VALOR	DIAS_APOS_ANTERIOR	DIFERENCA_ATUAL_ANT	INCLINACAO_COM_ANTERIOR
863	T	20/08/2003	2899	1	0,00	0,00000
864	T	21/08/2003	2936	1	37,00	1,54378
865	T	22/08/2003	2936	1	0,00	0,00000

INCLINACAO_COM_ANT_GRAU	INDICACAO	AMBIENTE	ESTADO_AMBIENTE	INDICACAO_GRUPO1
0,00000		19	ESTAVEL	
88,45184	N5		20 NOVO	N1
0,00000	N5		20 EVOLUCAO	

Fonte: Autor

Tabela 13 – Registros da série temporal das sequencias 863, 864 e 865.

Registros da Série Histórica (tbSERIE_HISTORICA)			
Sequencia	863	864	865
Sensor	MT-V2-3 eixo Um	MT-V2-3 eixo Um	MT-V2-3 eixo Um
Dia	20/08/2003	21/08/2003	22/08/2003
Valor	2899	2936	2936
Dias após anterior	1	1	1
Diferença atual anterior	0,00	37,00	0,00
Inclinação com anterior (rd)	0,00	1,54	0,00
Inclinação com anterior (graus)	0,00	88,45	0,00
Indicação		N5	N5
Identificação do Ambiente	19	20	20
Estado do Ambiente	Estável	Novo	Evolução
Indicação Grupo 1		N1	

Fonte: Autor

Na tabela 14 estão apresentados os registros da sequencia 863, anteriores aos da sequencia 864.

Tabela 14 - Registros dos 5 últimos valores após as sequencias 863, 864 e 865

<b>Registros de resumos da Série Histórica (tbResumo_Cinco_Ultimos)</b>			
<b>Após Sequencia</b>	<b>863</b>	<b>864</b>	<b>865</b>
<b>Sensor</b>	MT-V2-3 eixo Um	MT-V2-3 eixo Um	MT-V2-3 eixo Um
<b>Descrição</b>	Triortogonal	Triortogonal	Triortogonal
<b>Atual</b>	2899	2936	2936
<b>Segundo</b>	2899	2899	2936
<b>Terceiro</b>	2899	2899	2899
<b>Quarto</b>	2899	2899	2899
<b>Quinto</b>	2899	2899	2899
<b>Identificação do Ambiente</b>	19	20	20
<b>Limite Operacional Menor</b>	2724	2724	2724
<b>Limite Operacional Maior</b>	2975	2975	2975
<b>Medias QL</b>			
<b>Data da Última Novidade</b>			
<b>Data Atual</b>			

Fonte: Autor

Observa-se que, durante o processamento das sequencias 863, 864 e 865 os grupos passaram a conter informações, conforme apresentadas na tabela 15.

Tabela 15 - Grupos registrados durante cada sequencia

<b>Sequencia 863</b>		<b>Sequencia 864</b>		<b>Sequencia 865</b>	
<b>Grupo</b>	<b>Q</b>	<b>Grupo</b>	<b>Q</b>	<b>Grupo</b>	<b>Q</b>
2899	37	2936	1	2936	2
2899_2899	35	2936_2899	1	2936_2936	1
2899_2899_2899	33	2936_2899_2899	1	2936_2936_2899	1
2899_2899_2899_2899	31	2936_2899_2899_2899	1	2936_2936_2899_2899	1
2899_2899_2899_2899_2899	29	2936_2899_2899_2899_2899	1	2936_2936_2899_2899_2899	1

Fonte: Autor

No quadro 10, são mostrados os dados utilizados e a inferência realizada durante o processamento de cada uma das sequencias 863, 864 e 865. Na sequencia 864 foi aplicada

a regra R6 indicando mudança para próxima sequência do ambiente e estado Novo além de indicação de candidato a anomalia.

Quadro 10 - Resultados após o processamento de cada sequência de 863 a 865

Sequencia 863	Sequencia 864	Sequencia 865
<b>Dados reunidos para a inferência</b> <b>Valor</b> = 2899 <b>Período entre os registros</b> = 1 <b>Identificação Ambiente</b> = 19 <b>Novidade</b> = não <b>Estado do ambiente</b> = Estável <b>Limite operacional anterior Menor</b> = 2724 e <b>Maior</b> = 2975 (da sequência 862) <b>Limite Operacional Angular adotado</b> = 85,94° ou 1,5 rad. <b>Inclinação</b> = 0 rd <b>Inclinação graus</b> = 0° <b>Suporte</b> = não foi necessário, só se incluir previsão imediata.	<b>Dados reunidos para a inferência</b> <b>Valor</b> = 2936 <b>Período entre os registros</b> = 1 <b>Identificação Ambiente</b> = 19 <b>Novidade</b> = sim (surgimento de um novo grupo) <b>Estado do ambiente</b> = Estável <b>Limite operacional anterior Menor</b> = 2724 e <b>Maior</b> = 2975 (da sequência 863) <b>Limite Operacional Angular adotado</b> = 85,94° ou 1,5 rad. <b>Inclinação</b> = 1,54 rd <b>Inclinação graus</b> = 88,45° <b>Suporte</b> = não foi necessário, só se incluir previsão imediata.	<b>Dados reunidos para a inferência</b> <b>Valor</b> = 2936 <b>Período entre os registros</b> = 1 <b>Identificação Ambiente</b> = 20 <b>Novidade</b> = não <b>Estado do ambiente</b> = Novo <b>Limite operacional anterior Menor</b> = 2724 e <b>Maior</b> = 2975 (da sequência 864) <b>Limite Operacional Angular adotado</b> = 85,94° ou 1,5 rad. <b>Inclinação</b> = 0 rd <b>Inclinação graus</b> = 0° <b>Suporte</b> = não foi necessário, só se incluir previsão imediata.
<b>A Inferência Resultante</b> Aplicação da <b>regra R8</b> sem nenhuma sinalização	<b>A Inferência Resultante</b> Aplicação da <b>regra R6</b> com a mudança da tabela para estado do ambiente Novo e indicação de candidato à anomalia, além de mudança de regras relativas à anomalia (caso implementadas).	<b>A Inferência Resultante</b> Aplicação da <b>regra R9</b> sem nenhuma sinalização, mas mudança do estado do ambiente de Novo para em Evolução.

Fonte: Autor

#### 5.1.4 Considerações finais sobre os pontos analisados

A análise da aplicação no entorno dos dois pontos propostos apresentou a evolução da informação com relação a agrupamento de dados, a existência de novidades e ao surgimento de ambientes de referência no processamento da série histórica até a utilização do processamento da Tabela de Decisões Adaptativa, apresentada na figura 40. Tudo ocorreu conforme o previsto para o método desenvolvido.

O processamento foi realizado de forma contínua com a evolução das informações armazenadas no sistema computacional. Nos pontos estudados foram capturadas as verificações. Desse modo, as regras identificadas nos dois casos são diferentes, sendo que,

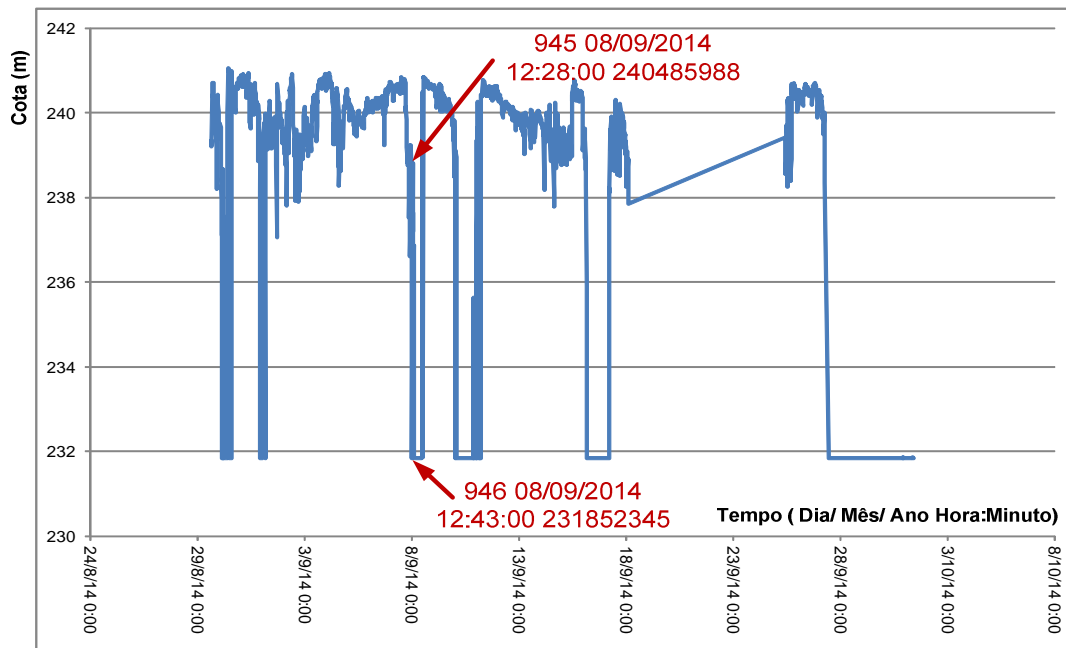
no primeiro caso (o pico), somente indica a sinalização de alerta e, no segundo caso (início do patamar), indica, além da sinalização de alarme, a mudança para uma nova identificação do ambiente (próxima sequência) e estado do ambiente Novo.

Dessa maneira, a aplicação foi satisfatória nas análises e/ou inferências com uma nova perspectiva empregando as novidades no reconhecimento de ambientes e na utilização de regras capazes de reconhecerem a situação da série temporal com relação a pontos candidatos à anomalia e às anomalias.

## 5.2 INFORMAÇÕES GERAIS DA APLICAÇÃO QUE ENVOLVE O SENSOR PIEZÔMETRO

Nesta aplicação, são utilizadas medições oriundas de instrumento de medição civil do tipo piezômetro, alimentado por energia armazenada a partir de painel solar fotovoltaico, elas envolvem o período de agosto a outubro de 2014, constituindo uma série temporal com 2289 valores. O período de aquisição das medições é 15 minutos de 29/08/2014 16h 07min a 01/10/2014 10h 04min. A Figura 50 mostra a série temporal que representa as cotas obtidas em função da coluna d'água (m) obtida a partir dos valores lidos pelo piezômetro instalado, onde o eixo Y representa os valores medidos em metro de coluna de água e referenciados a uma cota da planta da usina utilizando assim a cota em metros, em que o eixo x representa o tempo. Nota-se que, na parte inferior do traçado do gráfico, existem patamares que iniciam com leitura de fundo relacionada à diminuição de energia para o instrumento.

Figura 50 – Série Temporal do sensor Piezométrico PZ1762



Fonte: Autor

Considerando a figura 50, constata-se que existe uma sequência das medições da série (armazenada nos registros), a data/hora e o valor inteiro (utilizado assim para facilitar o processamento), o valor real da cota corresponde, para efeito de uso no computador ao valor inteiro dividido por 1000000, assim, um valor inteiro da medição é 231852345, corresponde a um valor real de 231,852345 que, por sua vez corresponde a uma leitura com um valor negativo (sendo considerado um valor de fundo) antes de ser transformada em uma medição. Com a intenção de facilitar o entendimento do processamento, os pontos utilizados para processamento são: 944, 945 e 946, sendo que as informações a serem processadas no *software* são mostradas na tabela 16.

Na base de conhecimento, há informação sobre o limite inferior e superior do projeto instalado do Piezômetro PZ1762, numa tabela de banco de dados que possui informações de parâmetros para o instrumento de medições esses valores são, respectivamente, cota mínima 231000000 e cota máxima 242000000.



Tabela 16 - Dados lidos pelo piezômetro para processamento

<b>Registros da Série Histórica (tbSERIE_HISTORICA)</b>			
<b>Sequência</b>	<b>944</b>	<b>945</b>	<b>946</b>
<b>Sensor</b>	PZ1762	PZ1762	PZ1762
<b>Dia</b>	08/09/2014 12:13:00	08/09/2014 12:28:00	08/09/2014 12:43:00
<b>Valor</b>	240085295	240485988	231852345
<b>Períodos Após Anterior</b>	1	1	1
<b>Diferença atual anterior</b>	-344416	400693	-8633643
<b>Inclinação com anterior</b>	-1,57	1,57	-1,57
<b>Inclinação com anterior grau</b>	-89,99	89,99	-89,99
<b>Indicação</b>	N5	N5	N5
<b>Identificação Ambiente</b>	3	3	3
<b>Estado Ambiente</b>	EVOLUÇÃO	EVOLUÇÃO	EVOLUÇÃO
<b>Indicação grupo 1 (novidade)</b>		N1	

Fonte: Autor

A aplicação considerou para o Piezômetro as regras como as apresentadas na Tabela de Decisões da figura 51 semelhantes às utilizadas no instrumento de medição Triortogonal, com a diferença de incluir as regras R10 e R11 no processamento.

Cabe ressaltar que o processamento das regras opera verificando se existe uma função anterior para ser aplicada a uma regra, executa a regra e constata se há uma função a ser realizada posterior à regra utilizada. Além disso, segue o modelo apresentado na figura 4, lembrando também que, em uma tabela de Decisões Adaptativa, a utilização de variáveis e geradores é opcional.

Ressalta-se que além dos valores testados, outros pontos são detectados como candidatos a anomalia durante a aplicação do modelo computacionalmente.

Figura 51 - T.D simplificada aplicada ao sensor Piezométrico com funções relacionadas às ações

Regras	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11
Prioridades	0	0	0	0	1	1	1	1	1	1	0
Condições:											
Novidade					Não	Sim	Sim	Não	Não	Sim	
Ambiente					Estável	Estável	Evolução	Evolução	Novo	Evolução	
Períodos entre registros	Igual 2	Maior 2			igual 1	igual 1	igual 1	igual 1	igual 1	igual 1	
Limite Operacional Valor anterior				Maior	Dentro	Dentro	Dentro	Dentro	Dentro	Dentro	
Limite Operacional angular adotad			Maior		Menor	Menor	Menor	Menor	Menor	Maior	
Limite suporte G2					< SUP.	< SUP.	< SUP.	< SUP.	< SUP.	< SUP.	
Limite de Projeto Inferior											Próximo
Ações											Funções
Sinalizar Mensagem Anomalia de período	xxx										f_smap()
Sinalisar Mudança de Ambiente									xxx		f_smam()
Sinalizar verificação		xxx									f_sv()
Sinalizar Alerta			xxx	xxx							f_sa()
Mudar cod.Ambiente						xxx					f_ma()
Candidato anomalia						xxx	xxx			xxx	f_ca()
Sinalizar valor minimo de projeto											xxx
Não sinalizar					xxx			xxx			

Fonte: Autor.

### 5.2.1 Informações sobre o processamento e análise

A partir do processamento dos dados com o *software* proposto pelo método MAIM e baseado nos conceitos de adaptividade, os resultados são apresentados no quadro 11, indicando os resultados para as três sequencias analisadas: 944, 945, e 946, indicando que a regra R3 foi aplicada nas três sequencias, a regra R10 na 945, e a R13 na 946.

Quadro 11- Resultado do Processamento por sequencia

Sequencia 944	Sequencia 945	Sequencia 946
<p><b>Dados para a inferência</b>  <b>Valor</b> = 24008529  <b>Período entre os registros</b> = 1  <b>Identificação Ambiente</b> = 3  <b>Novidade</b> = não</p> <p><b>Estado do ambiente</b> = Evolução  <b>Limite operacional anterior</b>  <b>Menor</b> = 231838088 e <b>Maior</b> = 241057012 (da sequencia 943)  <b>Limite Operacional Angular adotado</b> = 85,94° ou 1,5 rad.  <b>Inclinação</b> = -1,57 rd  <b>Inclinação graus</b> = -89,99°  <b>Suporte</b> = não foi necessário, só se incluir previsão imediata.</p>	<p><b>Dados para a inferência</b>  <b>Valor</b> = 240485988  <b>Período entre os registros</b> = 1  <b>Identificação Ambiente</b> = 3  <b>Novidade</b> = sim (surgimento de um novo grupo)</p> <p><b>Estado do ambiente</b> = Evolução  <b>Limite operacional anterior</b>  <b>Menor</b> = 231838088 e <b>Maior</b> = 241057012 (da sequencia 944)  <b>Limite Operacional Angular adotado</b> = 85,94° ou 1,5 rad.  <b>Inclinação</b> = 1,57rd  <b>Inclinação graus</b> = 89,99°  <b>Suporte</b> = não foi necessário, só se incluir previsão imediata.</p>	<p><b>Dados para a inferência</b>  <b>Valor</b> = 231852345  <b>Período entre os registros</b> = 1  <b>Identificação Ambiente</b> = 3  <b>Novidade</b> = não</p> <p><b>Estado do ambiente</b> = Evolução  <b>Limite operacional anterior</b>  <b>Menor</b> = 231838088 e <b>Maior</b> = 241057012 (da sequencia 945)  <b>Limite Operacional Angular adotado</b> = 85,94° ou 1,5 rad.  <b>Inclinação</b> = -1,57 rd  <b>Inclinação graus</b> = -89,99°  <b>Suporte</b> = não foi necessário, só se incluir previsão imediata.</p>
<p><b>A Inferência Resultante</b>  Aplicação da <b>regra R3</b> limite operacional angular adotado maior</p>	<p><b>A Inferência Resultante</b>  Aplicação da <b>regra R3</b> limite operacional angular adotado maior.   Aplicação da <b>regra R10</b> com novidade e limite operacional angular adotado maior.</p>	<p><b>A Inferência Resultante</b>  Aplicação da <b>regra R3</b> limite operacional angular adotado maior   Aplicação da <b>regra R11</b> valor próximo ao limite inferior de projeto</p>

Fonte: Autor.

### 5.2.2 Verificação se o ponto corresponde a uma Anomalia

Como resultado do processamento, a mensagem a ser emitida na medição 945 da série temporal é composta pelas seguintes informações:

- Candidato à anomalia;
- Identificação do Instrumento de Medição Piezômetro PZ1762;
- Sequência na série temporal original: 945;
- Data/hora da ocorrência: 08/09/2014 12:28:00;

- Valor medição: 240485988;
- Inclinação crescente crítica;
- Valor novo no ambiente (novidade) / série temporal;
- Texto: Inclinação gráfica abrupta com novidade.

No caso da mensagem emitida na medição da sequência 946 da série temporal, estão as seguintes informações:

- Candidato à anomalia;
- Identificação do Instrumento de Medição Piezômetro PZ1762;
- Sequência na série temporal original: 946;
- Data/hora da ocorrência: 08/09/2014 12:43:00;
- Valor medição: 231852345;
- Inclinação crescente crítica;
- Valor próximo ao mínimo do projeto;
- Texto: Inclinação gráfica abrupta com valor próximo ao mínimo do projeto entre 232000000 e 23100000.

Desse modo, está constatada a identificação de regras e sua aplicação no envio de mensagens indicando que um ponto é candidato a uma anomalia.

### **5.2.3 Consideração final sobre os pontos analisados**

A análise dos pontos mostrou que não ocorreu a mudança para um novo ambiente, mas um ponto indicou uma novidade e mudança acentuada na inclinação. O que elegeu o ponto 945 como candidato, a uma anomalia. Por causa disso, é necessária uma investigação por pessoal preparado ou a utilização de um *software* específico como o utilizado no Método Complementar.

### 5.3 CONSIDERAÇÃO FINAL

O sistema computacional foi desenvolvido baseado na necessidade de que dados podem ser monitorados e interpretados por um método não supervisionado trabalhando com um fluxo contínuo de medições. Neste capítulo, foi realizada a simulação e foram apresentados os resultados referentes a pontos analisados com relação a duas séries históricas diferentes envolvendo respectivamente, os instrumentos de medição do tipo Triortogonal e Piezômetro.

As medições, em ambos os casos, estão reunidas em uma série histórica, o que possibilitou a simulação de uma medição por vez no sistema MAIM.

A análise foi realizada em pontos selecionados para demonstrar a funcionalidade do sistema MAIM. Para cada medição escolhida, também foram utilizadas as medições anterior e posterior para demonstrar a evolução do ponto escolhido.

Assim em cada medição, foram reunidos os conteúdos necessários para a utilização da TDA, obtendo os resultados para a indicação de medições candidatas à anomalia.

Os resultados obtidos corresponderam às expectativas, pois possibilitaram a emissão de mensagens adequadas aos pontos analisados, considerando de forma adequada a identificação de novidades, ambientes e regras aplicáveis, além de realizarem alterações nos conteúdos das variáveis das regras e a aplicação de funções adaptativas ou não seguindo a teoria que envolve as TDA. Dessa forma os testes demonstraram a possibilidade da aplicação do método MAIM de forma autônoma e contínua.

## **6 COMPARAÇÃO DOS RESULTADOS OBTIDOS PELO MAIM E HOT SAX MODIFICADO**

A razão para serem realizadas essas comparações entre os resultados do HOX SAX modificado e do MAIM é que ambos foram praticamente desenvolvidos sob o mesmo núcleo de requisitos principais (por exemplo: ser não supervisionado), estimulados pelo mesmo projeto e patrocinadores. Os requisitos comuns e os diferenciados estão apresentados ao longo desta tese, sendo que os do HOT SAX modificados estão no apêndice B.

No item B 1 2, são mostrados os testes realizados pelo método HOT SAX modificado e apresentados em Nicoli et al (2013), indicando a análise do método em quatro séries históricas realizadas em quatro instrumentos civis diferentes, a saber: Extensômetro de Haste (EH, EHF30), Medidor Triortogonal (MT, MTTL2) em um de seus eixos, Piezômetro (PZ, PZAM2DA), e Medidor de Vazão (MV, MVAM1).

Como já explicado no item B 1 2, o método HOT SAX modificado verifica a existência de anomalia (ou não conformidade) a partir das últimas entradas, enquanto que o HOT SAX original verifica a partir do início da série histórica. Em ambas as versões do HOT SAX, os segmentos analisados iniciam uma numeração sequencial das medições em uma série histórica partir de 1. Cabe ressaltar que somente com a inversão da série, é possível utilizar o HOT SAX modificado em tempo real.

Os resultados apresentados estão resumidos no item B 1 2 para cada uma das séries históricas, estão apresentados no quadro 12, destacando que a identificação de anomalia é indicada na última coluna e é constituída por um conjunto de pontos.

Cabe notar que, no quadro 12, as séries estão indicadas com as abreviaturas dos instrumentos de medição, e está indicada a quantidade de medições disponíveis em cada série histórica bem como as sequencias identificadas nos testes com o HOT SAX modificado, além da indicação, se foi indicada ou não anomalia nos pontos testados.

Quadro 12 – Resultados obtidos nos testes com HOT SAX modificado.

Série	Quantidade de pontos (semanas)	Sequência de referência	Sequência de maior distância	Indicação de anomalia na sequência
EH	271	0	42	sim
MT	352	7	91	não
PZ	450	0	8	sim
MV	416	12	102	não

Fonte: Autor

Para comparar os resultados obtidos no MAIM, serão analisados os resultados dos quatro últimos valores correspondentes à janela de quatro pontos utilizada no teste do HOT SAX modificado em Nicoli et al (2013), para cada uma das quatro séries históricas utilizadas.

Para tanto, são utilizadas as regras apresentadas na Tabela de Decisão Subjacente da figura 52, com suas condições e ações, lembrando que antes da utilização das regras, para as mesmas condições existentes são consideradas ações que podem ser executadas antes e/ou depois das ações subjacentes (que fazem parte da Tabela de Decisões Adaptativas).

Figura 52 - Tabela de Dados Subjacente

Regras	R1	R2	R3	R4	R5	R6	R7	R8	R9
Prioridades	0	0	1	1	1	1	1	1	0
Condições:									
Novidade			Não	Sim	Sim	Não	Não	Sim	
Ambiente			Estável	Estável	Evolução	Evolução	Novo	Evolução	
Limite Operacional Valor anterior		Maior	Dentro	Dentro	Dentro	Dentro	Dentro	Dentro	
Limite Operacional angular adotado	Maior		Menor	Menor	Menor	Menor	Menor	Maior	
Limite suporte G2			<SUP.	<SUP.	<SUP.	<SUP.	<SUP.	<SUP.	
Limite de Projeto Inferior									Próximo
Ações									
Sinalizar Mensagem Anomalia de periodo									
Sinalizar Mudança de Ambiente							xxx		
Sinalizar verificação									
Sinalizar Alerta	xxx	xxx							
Mudar cod.Ambiente				xxx					
Candidato anomalia				xxx	xxx			xxx	
Sinalizar valor minimo de projeto									xxx
Não sinalizar			xxx			xxx			

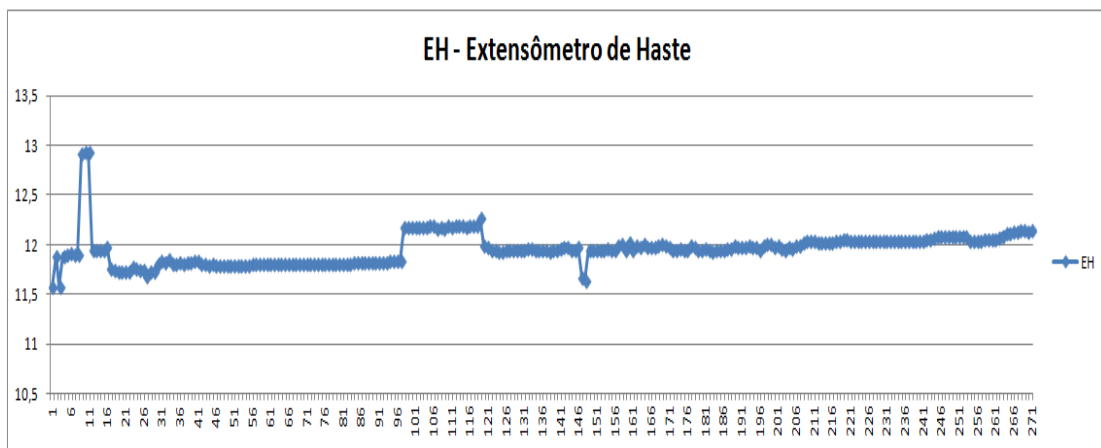
Fonte: Autor

Ressalta-se que, além dos valores testados, outros pontos são detectados como candidatos à anomalia ou à anomalia, tanto para o HOT SAX como para o MAIM, durante a aplicação dos modelos computacionalmente. Com relação aos gráficos apresentados, os valores no eixo das ordenadas correspondem a sequência de entrada das medições nas respectivas aplicações, considerando períodos iguais de tempo, como foram considerados em Nicoli et al (2013).

## 6.1 TESTE REALIZADO COM O EXTENSÔMETRO

Na série histórica do Extensômetro de Haste existem 271 medições (em mm) numeradas a partir de 1. Os pontos investigados correspondem às sequencias de números 268, 269, 270, e 271 (que coincide com a janela HOT SAX). Os valores originais da medição foram multiplicados por 100. Na figura 53 está o gráfico traçado com esses dados.

Figura 53 – Série histórica do Extensômetro de Haste, com medições em mm



Fonte: Nicoli et al (2013)

Na figura 54 estão informações dos registros do banco de dados do MAIM sobre as sequencias analisadas.



Figura 54- Registros das sequencias analisadas do Extensômetro de Haste

Resumo									
Sequencia	Atual	Segundo	Terceiro	Quarto	Quinto	Ambiente	LO.MENOR	LO.MAIOR	
268	1215	1214	1214	1213	1212	8	1158	1295	
269	1215	1215	1214	1214	1213	8	1158	1295	
270	1214	1215	1215	1214	1214	8	1158	1295	
271	1215	1214	1215	1215	1214	8	1158	1295	

Registro da Série Histórica									
	Dia	Valor	Diferença atual anterior	Inclinação com anterior	Inclinação com anterior graus	Indicação	Ambiente	Estado do ambiente	Indicação do grupo N1
268	07/07/2010 00:00	1215	1,00	0,00	0,04	N5	8	EVOLUCAO	N1
269	23/07/2010 00:00	1215	0,00	0,00	0,00	N5	8	EVOLUCAO	
270	11/08/2010 00:00	1214	-1,00	0,00	-0,03	N5	8	EVOLUCAO	
271	26/08/2010 00:00	1215	1,00	0,00	0,04	N5	8	EVOLUCAO	

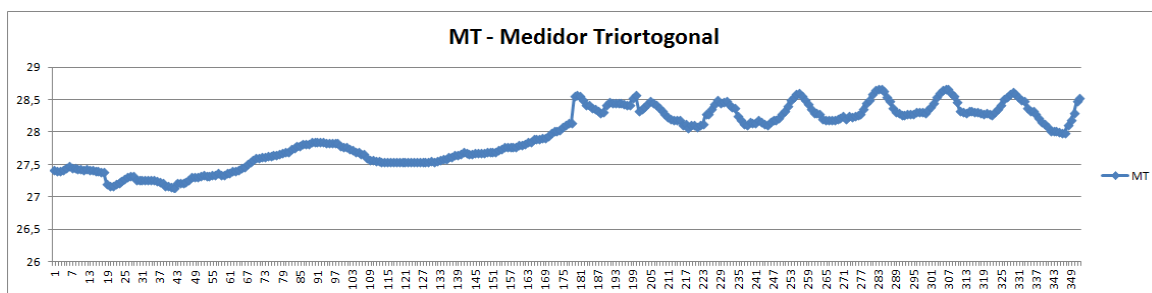
Fonte: Autor

Ao aplicar a TDA são encontrados os seguintes pares de sequencias e regras: (268, R5), (269, R6), (270, R6), e (271, R6), lembrando que a regra R5 indica “candidato à anomalia” e a regra R6 indica “não sinalizar”. Assim, somente ocorreu sinalização no ponto 268.

## 6.2 TESTE REALIZADO COM O MEDIDOR TRIORTOGONAL EM UM DE SEUS EIXOS

No caso da série histórica do Medidor Triortogonal correspondente a um de seus eixos de Haste, existem 349 medições em  $10^{-2}$  mm numeradas a partir de 1. Os pontos investigados correspondem às sequencias de números 349, 350, 351, e 352 (que coincide com a janela HOT SAX). Os valores originais da medição foram multiplicados por 100. Na figura 55 está o gráfico traçado com esses dados.

Figura 55 - Série histórica do Medidor Triortogonal em um de seus eixos, com medições em  $10^{-2}$  mm



Fonte: Nicoli et al (2013)

Na figura 56 estão informações dos registros do banco de dados do MAIM sobre as sequencias analisadas.

Figura 56 - Registros das sequencias analisadas de um eixo do Medidor Triortogonal

Resumo									
Sequencia	Atual	Segundo	Terceiro	Quarto	Quinto	Ambiente	LO.MENOR	LO.MAIOR	
349	2819	2812	2800	2800	2801	7	2715	2868	
350	2830	2819	2812	2800	2800	7	2715	2868	
351	2849	2830	2819	2812	2800	7	2715	2868	
352	2853	2849	2830	2819	2812	7	2715	2868	

Registro da Série Histórica									
	Dia	Valor	Diferença atual anterior	Inclinação com anterior	Inclinação com anterior graus	Indicação	Ambiente	Estado do ambiente	Indicação do grupo N1
349	07/07/2010 00:00	2819	7,00	0,01	0,30	N5	7	EVOLUCAO	
350	23/07/2010 00:00	2830	11,00	0,01	0,41	N5	7	EVOLUCAO	
351	12/08/2010 00:00	2849	19,00	0,01	0,57	N5	7	EVOLUCAO	N1
352	26/08/2010 00:00	2853	4,00	0,00	0,17	N5	7	EVOLUCAO	

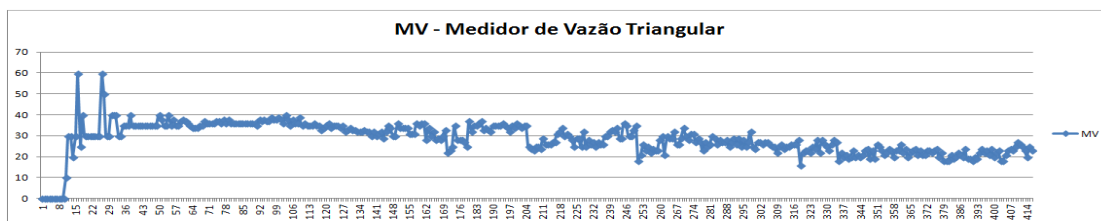
Fonte: Autor

Ao aplicar a TDA são encontrados os seguintes pares de sequencias e regras: (349, R6), (350, R6), (351, R5), e (352, R6), na qual a regra R5 indica “candidato à anomalia” e a regra R6 indica “não sinalizar”.

### 6.3 TESTE REALIZADO COM O MEDIDOR DE VAZÃO TRIANGULAR

As medições correspondentes ao Medidor de Vazão Triangular constituem uma série histórica com 416 medições (em mm) numeradas a partir de 1. Os pontos investigados correspondem às sequencias de números 413, 414, 415, e 416 (que coincide com a janela HOT SAX). Os valores originais da medição foram multiplicados por 1. Na figura 57 está o gráfico traçado com esses dados.

Figura 57 - Série histórica do Medidor de Vazão Triangular, com medições em mm.



Fonte: Nicoli et al (2013)

Na figura 58 estão informações dos registros do banco de dados do MAIM sobre as sequencias analisadas.

Figura 58 - Registros das sequencias analisadas do Medidor de vazão triangular

Resumo								
Sequencia	Atual	Segundo	Terceiro	Quarto	Quinto	Ambiente	LO.MENOR	LO.MAIOR
413	23	25	26	27	25	10	0	60
414	20	23	25	26	27	10	0	60
415	25	20	23	25	26	10	0	60
416	23	25	20	23	25	10	0	60

Registro da Série Histórica									
	Dia	Valor	Diferença atual anterior	Inclinação com anterior	Inclinação com anterior graus	Indicação	Ambiente	Estado do ambiente	Indicação do grupo N1
413	07/07/2010 00:00	23	-2,00	0,00	-0,09	N5	10	ESTAVEL	
414	26/07/2010 00:00	20	-3,00	0,00	-0,09	N5	10	ESTAVEL	
415	17/08/2010 00:00	25	5,00	0,00	0,14	N5	10	ESTAVEL	
416	26/08/2010 00:00	23	-2,00	0,00	-0,13	N5	10	ESTAVEL	

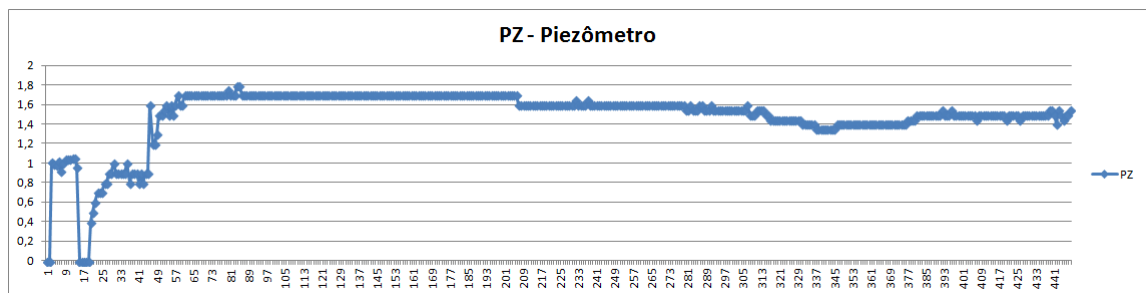
Fonte: Autor

Ao aplicar a TDA são encontrados os seguintes pares de sequencias e regras: (413, R3), (414, R3), (415, R3), e (416, R3), na qual a regra R3 indica “não sinalizar”, o que significa não enviar nenhuma mensagem ou alarme.

#### 6.4 TESTE REALIZADO COM O PIEZÔMETRO

Na série histórica do Piezômetro contêm 450 medições (em m) numeradas a partir de 1. Os pontos investigados correspondem as sequencias numéricas de 447, 448, 449, e 450 (que coincide com a janela HOT SAX). Os valores originais da medição foram multiplicados por 1000 ( $10^{-3}$  m). Na figura 59 está o gráfico traçado com esses dados

Figura 59 - Série histórica do Piezômetro, em ( $10^{-3}$  m)



Fonte: Nicoli et al (2013)

Na figura 60 estão informações dos registros do banco de dados do MAIM sobre as sequencias analisadas.

Figura 60 - Registros das sequencias analisadas do Piezomêtro

Resumo								
Sequencia	Atual	Segundo	Terceiro	Quarto	Quinto	Ambiente	LO.MENOR	LO.MAIOR
447	2300	2200	2100	1550	1500	7	0	2300
448	2500	2300	2200	2100	1550	7	0	2500
449	3700	2500	2300	2200	2100	7	0	3700
450	3700	3700	2500	2300	2200	7	0	3700

Registro da Série Histórica									
	Dia	Valor	Diferença atual anterior	Inclinação com anterior	Inclinação com anterior graus	Indicação	Ambiente	Estado do ambiente	Indicação do grupo N1
447	16/08/2010 00:00	2300	100,00	0,03	1,81	N5	7	EVOLUCAO	N1
448	31/08/2010 00:00	2500	200,00	0,14	7,91	N5	7	EVOLUCAO	N1
449	01/09/2010 00:00	3700	1200,00	1,49	85,43	N5	7	EVOLUCAO	N1
450	15/09/2010 00:00	3700	0,00	0,00	0,00	N5	7	EVOLUCAO	

Fonte: Autor

Ao aplicar a TDA são encontrados os seguintes pares de sequencias e regras: (447, R5), (448, R5), (449, R5), e (450, R6), lembrando que a regra R5 indica “candidato à anomalia” e a regra R6 indica “não sinalizar”.

## 6.5 CONSIDERAÇÕES FINAIS

A comparação das análises automáticas (inferência) realizada pelos dois aplicativos está resumida no quadro 13, no qual se observa a ocorrência de uma coincidência.

Quadro 13 – Resumo das indicações de anomalias

Sensor	Anomalias			
	Hot Sax	MAIM	Sequencias candidatas	Regras indicadas
EH	sim	sim	268	R5
MT	não	sim	351	R5
MV	sim	não	Sem candidatas	R3
PZ	não	sim		R5

Fonte: Autor

Os resultados obtidos no HOT SAX modificado são devido principalmente a uma busca do tipo “força bruta” ou do tipo heurística (existem 4 sugeridas), em função diretamente dos passos e critérios do método codificado. Já no aplicativo MAIM, esse conhecimento está distribuído no código de programação, na base de conhecimento (formada incrementalmente a cada leitura), e na Tabela de Decisões Adaptativa. Do ponto de vista de programação o HOT SAX modificado é mais complexo (em parte devido a outros métodos embutidos nele).

No caso do aplicativo MAIM, há uma reunião de conteúdos de variáveis que são utilizadas pela TDA nas regras existentes. Assim só são aplicadas regras que possuem conteúdos pertinentes ao processamento do momento, logo, o resultado depende das variáveis utilizadas nas condições. Desse modo, é possível aperfeiçoar a TDA automaticamente ou inserindo explicitamente uma nova regra.

Diante dos testes realizados, o MAIM indica uma maior flexibilidade com relação ao HOT SAX modificado, pois tem a possibilidade de identificar candidatos à anomalia com maior assertividade devido a responder a uma maior quantidade de comparações e poder se adaptar diante das circunstâncias do processamento.

## 7 CONCLUSÃO

Neste capítulo, estão sintetizados os principais aspectos apresentados nesta tese incluindo conclusões, contribuições e trabalhos futuros.

### 7.1 ASPECTOS GERAIS DO MÉTODO

O método está adequado aos estágios apresentados por Porter e Heppelmenn (2014), isto é, está preparado para monitoração, controle, otimização e autonomia, pois os dois aplicativos desenvolvidos, que também são considerados métodos nesta tese, o MAIM e o Complementar, seguem essa linha, apesar de neste momento atenderem principalmente à monitoração. Ambos os aplicativos estão preparados para emitir mensagens e alertas, sendo que o Complementar é capaz de melhorar as mensagens oriundas de outros aplicativos com justificativas e aumentando a assertividade se a medição envolvida for somente uma candidata à anomalia ou uma anomalia.

Além das considerações de Porter; Heppelmenn (2014) foram estabelecidos requisitos que contemplam os métodos de serem não supervisionados, de não exigir treinamento em sua operação inicial, trabalhando com um fluxo contínuo de medições, além de considerar uma estrutura de dados e informações que contemplam o uso de regras com uma busca adequada e adaptatividade e/ou uma inferência que ajudem no aprendizado continuamente, enriquecendo e estando adequado à base de conhecimento.

Dentre os principais requisitos atendidos, estão:

- Aplicativo não ser supervisionado;
- Considerar séries históricas;
- Registrar, interpretar e analisar cada medição recebida;
- Identificar novidades, ambientes, e suas mudanças e anomalias;
- Aprender com a série histórica;
- Emitir mensagens adequadas aos interessados;
- Melhoria contínua na forma de aprendizado e adaptação.

O método pode ser aplicado a outros instrumentos de medição, além dos aplicados à área de usinas hidroelétricas e sistemas elétricos.

## 7.2 CONTRIBUIÇÕES

A principal contribuição propõe um método adaptativo que trabalha com uma base de conhecimento que considera o uso de regras que podem ser parametrizadas ou não. Em sua adaptatividade, é possível excluir e criar novas regras com valores novos ou baseados na combinação de novas condições.

Outras contribuições estão relacionadas à identificação de novidades e o uso delas para reconhecer mudanças no ambiente envolvendo o ponto de medição, sendo consideradas como condições que podem contribuir para decidir se um valor medido seja considerado ou não candidato à anomalia.

Uma contribuição que surgiu durante o desenvolvimento do método MAIM foi o Método Complementar, que pode ser aplicado quando existem informações pertinentes para a obtenção de uma assertividade maior nas respostas oriundas de outros métodos, além de incluir justificativas e complementos.

Nos dois métodos desenvolvidos, é proposta uma forma adequada para representar e trabalhar com regras utilizando uma estrutura própria para banco de dados relacional.

Uma contribuição indireta é o fato de possibilitar uma forma de organização de conhecimentos relativos a um ponto de medição, ao utilizar o mapeamento de influências e relações, que podem auxiliar no estabelecimento de regras, as quais foram utilizadas na forma de regras para as tabelas de decisão adaptativas ou não, regras de associação (ajudando em previsões) e regras de produção, tornando o método como um todo inteligente.



Além disso, uma contribuição decorrente dos estudos realizados transformou o método HOT SAX original supervisionado em um método não supervisionado. O HOT SAX Adaptado foi fornecido para uso em hidroelétrica de um dos patrocinadores, tendo sido publicado em Nicoli et al. (2013), juntamente com dois outros métodos não apresentados neste trabalho.

O método para monitoramento pode ser sintetizado no seguinte:

- Obter inicialmente e eventualmente informações do instrumento de medição e registrá-las no aplicativo MAIM, que é adaptativo;
- Identificar e implantar as regras inicialmente identificadas para o instrumento de medição e eventualmente para uso na Tabela de Decisões Adaptativa que, em operação automática, poderá se adaptar às novas situações que ocorrerão ao longo do tempo. Parte das regras considera as novidades em medições e, indiretamente, as mudanças de ambiente no entorno do ponto (e instrumento) de medição;
- Estabelecer inicialmente o recebimento da série histórica no aplicativo MAIM, que passará a ter recebimento automático;
- O aplicativo MAIM operará sem supervisão enviando alertas e mensagens aos interessados;
- As mensagens poderão ser opcionalmente enviadas ao método complementar que utiliza inferência;
- Devem ser registradas, inicialmente e eventualmente, informações e regras relativas ao ambiente do ponto de medição e de outras fontes de informação no aplicativo Complementar relacionadas ao instrumento de medição;
- Caso o método Complementar esteja operando para o método MAIM, este poderá enviar mensagens com justificativas e complementações aos interessados.

### 7.3 TRABALHOS FUTUROS PARA A EVOLUÇÃO DESTA PESQUISA

Durante o desenvolvimento dos métodos, foi percebido que a aplicação deles dependia muito do ambiente organizacional, sendo que cada organização possuía características favoráveis diferentes. Dessa forma, surgiu a necessidade de estabelecer o estágio de maturidade de uma organização para a utilização dos métodos, possibilitando uma referência para as questões de melhoria contínua e ao enquadramento apresentado por Porter; Heppelmenn (2014), com relação aos estágios de monitoramento, controle, otimização e autonomia, mesmo que o controle seja desconsiderado em certas situações utilizadas em sistemas elétricos.

Entre outros trabalhos propostos, podem-se:

- Realizar testes comparativos com outros métodos;
- Incorporar o módulo adaptativo no Método Complementar;
- Estabelecer uma metodologia envolvendo a maturidade e os métodos desenvolvidos neste trabalho;
- Realizar testes para categorizar as aplicações e otimizar os procedimentos computadorizados;
- Incluir procedimentos nos métodos que contemplem o controle;
- Investigar questões envolvendo a otimização e a autonomia;
- Investigar, com relação ao ambiente, o uso de percentagem em vez de quantidade de medições para passar de um estado em evolução para um estado estável.

## REFERÊNCIAS BIBLIOGRÁFICAS

AGRAWAL, R.; IMIELŃSKI, T.; SWAMI, A. Mining Association Rules between sets of items in large databases. In: SIGMOD, 5. 1993. **Conference...** Washington, 1993, p. 207-216.

AGRAWAL, R.; SRIKANT, R. Fast algorithms for mining association rules. In: INTERNATIONAL CONFERENCE VERY LARGE DATA BASES, VLDB, 20. 1994. **Proceedings ...** p. 487-499, 1994.

AGRAWAL, R.; SRIKANT, R. Mining generalized Association Rules. In: INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES, 21., 1995, **Proceedings ...** p. 407-419.

AGRAWAL, R.; SRIKANT, R. Mining quantitative Association Rules in large relational tables. In: **ACM SIGMOD Record**. ACM, 1996, p. 1-12.

ALCIATORE, D. G.; HISTAND, M. B. **Introdução à mecatrônica e aos sistemas de medições**. Porto Alegre: Amgh Editora, 2014.

ALENCAR, A. B. **Mineração e visualização de coleções de séries temporais**. 2007. 127p. Dissertação (Mestrado) - Instituto de Ciências Matemáticas e de Computação USP, São Carlos, 2007.

BARAÚNA, H. **Classificação de textos com árvore de decisão adaptativa**. Disponível em: <<http://www.bitabit.eng.br/2010/11/16/classificacao-de-textos-com-arvore-de-decisao-adaptativa/>>. Acesso em: 16 Out. 2010.

BARROS, M. **Processos estocásticos**. s.l.: Papel Virtual Editora, 2009.

BASHA, R.; AMEEN, J. Unusual sub-sequence identifications in time series with periodicity. **International Journal of Innovative Computing, Information and Control**. v. 3, n 2, p. 471- 480, April 2007.

BITTEM COURT, G. **Inteligência artificial: ferramentas e teorias**. Universidade Federal de Santa Catarina, Florianópolis: UFSC, 2006.

BOLTON, W. **Mecatrônica: uma abordagem multidisciplinar**. Porto Alegre: Bookman Editora, 2010.

BRITAIN, G. et al. **ITIL service design**. London: Tso, 2011.

BU, YINGYI et al. Finding Top-K Discords in Time Series Database. In: PROCEEDINGS OF THE SIAM INTERNATIONAL CONFERENCE ON DATA MINING, SDM07, **Proceedings...** 2007, p. 449-454.

BUNKE, H.; KRAETZL, M. Classification and Detection of Abnormal Events in Time Series of Graphs. **Data Mining in Time Series Databases**, v. 57, p. 127, 2004.

BUSINESS DICTIONARY. **Proof of Concept**. Disponível em <<http://www.businessdictionary.com/definition/proof-of-concept.html>>. Acesso em: 10 de Janeiro 2017.

CAMARGOS, F. L. **Lógica Nebulosa**: uma abordagem filosófica e aplicada. Universidade Federal de Santa Catarina, Departamento de Informática e Estatística, 2002.

CARHUANINA, R. E. C.; NETO, JOÃO JOSÉ. Personalização, "customização", adaptabilidade e adaptatividade. In: MEMÓRIAS DO WORKSHOP DE TECNOLOGIA ADAPTATIVA - WTA 2016, 10. **Anais...** São Paulo. ISBN: 978-85-86686-86-3, p. 52-59. 28 e 29 de Janeiro, 2016.

CARVALHO, M. V. G. S. A. **Um modelo para o dimensionamento de calçadas considerando o nível de satisfação do pedestre**. 2006. 176p. Tese (Doutorado). Escola de Engenharia de São Carlos da Universidade de São Paulo, São Carlos, 2006.

CATAE, F.; ROCHA, L. A. Introdução a árvore de decisão adaptativa. IN: WORKSHOP DE TECNOLOGIA ADAPTATIVA - WTA, 5. **Anais...** Escola Politécnica da Universidade de São Paulo, São Paulo, 2011.

DORF, R. C.; BISHOP, R. H. **Sistemas de controle modernos**. 12. ed. Rio de Janeiro: LTC, 2013.

FANG, C.; ZHANG, X. BP. Neural network prediction model of rock surface displacement caused by anchor-hold change. In: INTELLIGENT SYSTEMS AND APPLICATIONS (ISA), 2011 3. **International Workshop on. IEEE**, 2011, p. 1-4.

FERREIRA, A. B. H. **Dicionário Aurélio Online**. Disponível em <<http://dicionariodoaurelio.com/>>. Acesso em: 27 Maio, 2015.

FACELI, K. et al. **Inteligência Artificial**: Uma abordagem de aprendizado de máquina. Rio de Janeiro: LTC, 2011.

FIALHO, A. B. **Instrumentação Industrial**: conceitos, aplicações e análises. São Paulo: Editora Érica, 2010.

FONG, S.; ZHOU, N. Towards an adaptive forecasting of earthquake time series from decomposable and salient characteristics. In: INTERNATIONAL CONFERENCES ON PERVASIVE PATTERNS AND APPLICATIONS, 3. **Conference ...** Rome, Italy, 2011, p. 53-60.

FU, A. W.-C. et al. Finding time series discords based on haar transform. In: INTERNATIONAL CONFERENCE ON ADVANCED DATA MINING AND APPLICATIONS. **Proceedings...** Springer Berlin Heidelberg, 2006, p. 31-41.

FREESCALE. MPXV7002 Integrated Silicon Pressure Sensor On-Chip Signal Conditioned, Temperature Compensated and Calibrated. **Catalog.** Freescale Semiconductor, Inc, 2015.

GADDAM, S. R.; PHOHA, V. V.; BALAGANI, K. S. K-Means+ID3: A novel method for supervised anomaly detection by cascading K-Means clustering and ID3 decision tree learning methods. **IEEE Transactions on Knowledge and Data Engineering**, v. 19, n. 3, p. 345-354, 2007.

GROOVER, M. P. **Automação industrial e sistemas de manufatura.** São Paulo: Pearson Education do Brasil, 2011.

HACHICH, W. Probabilidade, estatística, decisões e segurança em engenharia geotécnica. Universidade de São Paulo, 1999. **Apresentação de aula**, disciplina Probabilidade, Estatística, Decisões e Segurança em Engenharia Geotécnica (PEF5837). Disponível em <<http://www.lmc.ep.usp.br/PEOPLE/whachich/pef5837/default.htm>>. Acesso em 30 Maio, 2014.

INMET. Instituto Nacional de Meteorologia. **Tabelas da Base de Dados de Nebulosidade e de Precipitação.** Disponível em <<http://www.inmet.gov.br/portal/>>. Acesso em: em 7 Maio. 2016.

KAWANO, B. R. et al. Modelo de um sistema de conservação de alimentos baseado na IoT-A e seleção de elementos para Tabela de Decisão Adaptativa (TDA). In: WORKSHOP DE TECNOLOGIA ADAPTATIVA - WTA 9. **Anais...** Escola Politécnica da Universidade de São Paulo, São Paulo. 2015, p. 64.

KEOGH, E. et al. Locally Adaptive Dimensionality Reduction for Indexing Large Time Series Databases. **ACM SIGMOD Record**, v. 30, n. 2, p. 151-162, 2001.

KEOGH, E.; LIN, J.; FU, A. Hot Sax: efficiently finding the most unusual time series subsequence. In: IEEE INTERNATIONAL CONFERENCE ON DATA MINING (ICDM'05), 5. **Proceedings...** IEEE, 2005.

KNORN, F.; LEITH, D. J. Adaptive Kalman Filtering for Anomaly Detection. in Software Appliance. In: INFOCOM Workshops 2008, **Proceedings...** IEEE, 2008, p. 1-6.

KOUSHANFAR, F.; POTKONJAK, M.; SANGIOVANNI-VINCENTELLI, A. On-line Fault Detection of Sensor Measurements. In: SENSORS, 2003. **Proceedings...** 2003, p. 974-979.

LAFRAIA, J. R. B. **Manual de confiabilidade, manutenibilidade e disponibilidade**. Rio de Janeiro: Qualitymark Editora; Petrobras, 2014.

LEMOS, A. P. **Proposta de um algoritmo genérico de detecção de novidades em séries temporais utilizando modelos de previsão**. 2007. 102 p. Dissertação (Mestrado) – Escola de Engenharia Elétrica da Universidade Federal de Minas Gerais, Belo Horizonte, 2007.

LIN, J. et al. A symbolic representation of time series, with implications for streaming algorithms. In: ACM SIGMOD WORKSHOP ON RESEARCH ISSUES IN DATA MINING AND KNOWLEDGE DISCOVERY, 8. **Proceedings...** ACM, 2003, p. 2-11.

LIN, J. et al. Finding motifs in time series. In: WORKSHOP ON TEMPORAL DATA MINING, 2. **Proceedings...** 2002, p. 53-68.

LUGER, G. F. **Artificial Intelligence: structures and strategies for complex problem solving**. New York: Addison Wesley, 2008.

LUGER, G. F.; STUBBLEFIELD W. A. **Artificial Intelligence: structures and strategies for complex problem solving**. Palo Alto: The Benjamin Cummings Publishing Co., 1993.

LUGER, G. F. **Inteligência artificial**. São Paulo: Pearson Education, 2013.

MAHONEY, M. V.; CHAN, P. K. **Learning Rules for Time Series Anomaly Detection**. Technical report, Florida Institute of Technology, Melbourne, Florida, October, 2004. Disponível em <<http://cs.fit.edu/~mmahoney/nasa/siam.pdf>>. Acesso em: 12 Dez. 2015.

MALETZKE, ANDRÉ GUSTAVO et al. Uma avaliação sobre a identificação de motifs em séries temporais. In: CONGRESSO DA ACADEMIA TRINACIONAL DE CIÊNCIAS, **Anais...** 2008, p. 1-10.

MALETZKE, A. G. et al. Mineração de séries temporais por meio da extração de características e da identificação de Motifs. In: ENCONTRO NACIONAL DE INTELIGÊNCIA ARTIFICIAL (ENIA), RIO GRANDE DO SUL, BENTO GONÇALVES, 7. **Anais...** 2009, p. 1-10.

MARÈ, R. M. et al. Operação sustentável de sistemas de ar condicionado usando tecnologia adaptativa. In: WORKSHOP DE TECNOLOGIAS ADAPTATIVAS - WTA, 9. **Anais...** Escola Politécnica da Universidade de São Paulo, São Paulo, 2015.

MARTINELLI, D. P. et al. **Teoria geral dos sistemas**. São Paulo: Saraiva, 2012.

MIGUEL, D. S.; OLIVEIRA, D. A. Uso de métricas de roteamento através de Tabela de Decisão Adaptativa para redes de sensores sem fio. In: WORKSHOP DE TECNOLOGIA ADAPTATIVA - WTA 10. **Anais...**, São Paulo. ISBN: 978-85-86686-90-0, p. 7. 26 e 27 de Janeiro, 2017.

MOREIRA, L. P.; BARRETTO, M. R. P. Aplicação de Tabelas de Decisão Adaptativas em sistemas de controle de crescimento de plantas. In: WORKSHOP DE TECNOLOGIA ADAPTATIVA - WTA 10. **Anais...** São Paulo. ISBN: 978-85-86686-90-0, p. 100. 26 e 27 de Janeiro, 2017.

MORETTIN, P. A.; TOLOI, C. **Análise de séries temporais**. São Paulo: Blucher, 2006.

NETO, J. J. Adaptive rule-driven devices - general formulation and case study. In: CIAA'2001 SIXTH INTERNATIONAL CONFERENCE ON IMPLEMENTATION AND APPLICATION OF AUTOMATA. **Pre-proceedings....** Pretoria- Rep. of South Africa, 2001. p. 158-176.

NETO, J. J. **Um levantamento da evolução da Adaptatividade e da Tecnologia Adaptativa**. Revista IEEE América Latina, v. 5, n. 7, p. 1548-0992, 2007.

NETO, J. J. Adaptatividade: generalização conceitual. In: WORKSHOP DE TECNOLOGIA ADAPTATIVA - WTA, 3. **Anais...** Escola Politécnica da Universidade de São Paulo, São Paulo. 2009.

NETO, J. J. Adaptatividade: um glossário sobre adaptatividade. In: WORKSHOP DE TECNOLOGIA ADAPTATIVA - WTA, 3. **Anais...** Escola Politécnica da Universidade de São Paulo, São Paulo. 2009.

NICOLI, S. et al. Detection of nonconformities in monitoring system's measurements. In: INNOVATIVE SMART GRID TECHNOLOGIES LATIN AMERICA (ISGT LA), 2013 **Conference...** IEEE PES, 2013. p. 1-6. DOI:10.1109/ISGT-LA.2013.6554414.

NONAKA, I.; TAKEUCHI, H. **The Knowledge-creating company: how Japanese companies create the dynamics of innovation**. Local: Oxford University press, 1995.

NOVAK, J. D. **Learning, creating, and using knowledge: concept maps as facilitative tools in schools and corporations**. Lawrence Erlbaum Associates, 1998 / Routledge, 2010.

OLIVEIRA JUNIOR, H. A. **Lógica difusa: aspectos práticos e aplicações**. Rio de Janeiro: Interciência, 1999.

POLANYI, M. **The tacit dimension**. Londres: Routledge & Kegan Paul, 1966.

PMI. **A guide to the project management body of knowledge**. 5th ed. [s.l.: s.n.], 2012.

PORTER, M. E.; HEPPELMENN, J. E. How smart, connected products are transforming competition. **Harvard Business Review**, v. 92, n. 11, p.64-88, 2014.

POST, E. L. Formal reductions of the general combinatorial decision problem. **American Journal of Mathematics**, v. 65, n. 2, p. 197-215, 1943. Disponível em: [https://www.jstor.org/stable/2275590?seq=1#page\\_scan\\_tab\\_contents](https://www.jstor.org/stable/2275590?seq=1#page_scan_tab_contents). Acesso em: 12/11/2016.

PROÁGUA. **Manual de segurança e inspeção de barragens**. Brasília: Ministério da Integração Nacional – Secretária de Infraestrutura Hídrica, 2002. 148 p.

PROBST, G.; RAUB, S. R.; ROMHARDT, K. **Gestão do Conhecimento, os elementos construtivos do sucesso**. Porto Alegre: Artmed Editora, 2002.

RAMOS, M. V. M.; NETO, J. J.; VEGA, I. S. **Linguagens formais: teoria, modelagem e implementação**. Porto Alegre: Bookman Editora, 2009.

RIBEIRO, C. V.; GOLDSCHMIDT, R. R.; CHOREN, R. **Métodos para previsão de séries temporais e suas tendências de desenvolvimento**. Rio de Janeiro: Instituto Militar de Engenharia, setembro, 2009.

RICH, E.; KNIGHT, K. **Inteligência artificial**. São Paulo: Makron Books, 1993.

ROBLES-KELLY, A.; HANCOCK, E. R. Edit distance from graph spectra. In: Computer Vision, 2003. In: IEEE INTERNATIONAL CONFERENCE ON. IEEE, 9. 2003, **Proceedings...** p. 234-241.

RUSSELL, S. J.; NORVIG, P. **Artificial intelligence: a modern approach**. Upper Saddle River, New Jersey: Pearson Education, 2010.

SCHWABACHER, M.; OZA, N.; MATTHEWS, B. Unsupervised anomaly detection for liquid-fueled rocket propulsion health monitoring. **Journal of Aerospace Computing, Information, and Communication**, v. 6, n. 7, p. 464-482, 2009.

SHA, M.; SHENTANG, D. Treatment method on vary interval data serial based on chaotic neural network and improved chaotic wavelet. In: E-PRODUCT E-SERVICE AND E-ENTERTAINMENT (ICEEE), 2010 INTERNATIONAL CONFERENCE ON. IEEE, 2010, p. 1-4.

SILVA, P. O. M. P. et al. **Previsão de séries temporais utilizando lógica nebulosa**. In: 4º CONTECSI, 4. **Anais...** - Universidade de São Paulo, 2007.



SILVEIRA, J. F. A. Teria sido possível evitar a ruptura da barragem de Teton através de um bom plano de instrumentação? **Revista Brasileira de Barragens**, v. 1, n.1, 01 Abril de 2014.

SIMÕES, M. G.; SHAW, I. S. **Controle e modelagem fuzzy**. São Paulo: Blucher; FAPESP, 2007.

SPINOSA, E. J.; CARVALHO, A. P. L. F.; GAMA, J. Detecção de novidade com aplicação a fluxos contínuos de dados. In: WORKSHOP DE TESES E DISSERTAÇÕES DO ICMC, 13. **Anais...** São Paulo, ICMC, 2008.

STANGE, R. L.; NETO, J. J. Mecanização da aprendizagem com dispositivos adaptativos: conceitos e aplicação. In: WORKSHOP DE TECNOLOGIA ADAPTATIVA- WTA 7, **Anais...** Escola Politécnica da Universidade de São Paulo, 2013.

TCHEMRA, A. H. **Tabela de decisão adaptativa na tomada de decisão multicritério**. 2009. 172 p. Tese (Doutorado). Escola Politécnica da Universidade de São Paulo. São Paulo, 2009.

TCHEMRA, A. H. Adaptatividade na tomada de decisão multicritério. In: WORKSHOP DE TECNOLOGIA ADAPTATIVA - WTA 4, **Anais...** Escola Politécnica da Universidade de São Paulo, 2010.

TCHEMRA, A. H.; CAMARGO, R. Descoberta de padrões em bases de dados utilizando Técnicas Adaptativas. In: WORKSHOP DE TECNOLOGIA ADAPTATIVA - WTA 3, **Anais...** Escola Politécnica da Universidade de São Paulo, São Paulo. 2009.

TIESHENG, W.; ZHENXING, R. Fuzzy neural network algorithm and its application. In: INTERNATIONAL CONFERENCE ON ELECTRIC TECHNOLOGY AND CIVIL ENGINEERING (ICETCE), **Proceedings...** 2011.

TOSHNIWAL, D.; YADAV, S. Adaptive outlier detection in streaming time series. In: INTERNATIONAL CONFERENCE ON ASIA AGRICULTURE AND ANIMAL IPCBEE, **Proceedings...** IACSIT Press, Singapore, 2011, v.13.

WÄRME. **Linha automação**: série WTP-4010 WTP-4010. Novembro, 2005 / 2016.

WEKA. **Weka 3**: data mining software in java. Disponível em: <http://www.cs.waikato.ac.nz/~ml/weka/>. Acesso em: 21/02/2015.

WOYTE, A.; BELMANS, R.; NIJS, J. Fluctuations in instantaneous clearness index: analysis and statistics. **Solar Energy**, v. 81, n. 2, p. 195–206, 2007.

WU, C. et al. Study on Applications of Neural Network to Deformation Control of Dam. In: INTERNATIONAL CONFERENCE ON COMPUTER APPLICATION AND SYSTEM MODELING (ICCASM). **Proceedings...** 2010.

WU, X. et al. **Top 10 algorithms in data mining**. Knowledge and Information Systems, v. 14, n. 1, p. 1-37, 2007.

XIAODONG, L. et al. Application of artificial neural network in the typhoon flood prediction system – a case study in Shanghai, China. In: ARTIFICIAL INTELLIGENCE AND COMPUTATIONAL INTELLIGENCE (AICI), 2010 INTERNATIONAL CONFERENCE ON. **Proceedings...** IEEE, 2010, p. 191-195.

YE, M.; SONG, L. Weak signal detection in chaos using adaptative Neuro-Fuzzy inference system. In: ELECTRICAL AND CONTROL ENGINEERING (ICECE), 2011 INTERNATIONAL CONFERENCE ON. IEEE, 2011, p. 928-931.

## APÉNDICE

## APÊNDICE A – COMPLEMENTAÇÃO DO ESTADO DA ARTE

Neste apêndice estão contidas pesquisas complementares ao estado da arte, envolvendo conceitos que não contribuíram diretamente no desenvolvimento do método, mas são esclarecedoras com relação à utilização de outros métodos inteligentes e representações de conhecimentos que também utilizam regras.

### A 1 BASE DE CONHECIMENTO

Nesta tese é utilizado o termo Base de Conhecimento de forma abrangente, possibilitando as diversas considerações existentes nas áreas de inteligência artificial / computacional, comunicação, administração, dentre outras.

Assim, considera-se que, em geral, a base de conhecimento é um repositório público que pode ser centralizado ou não, relacionado a um ou vários assuntos (domínios). Esses assuntos podem estar embutidos em um código de programa, depositados em arquivos, arquivos relacionados, banco de dados, dentre outros. Eles podem ser estruturados (bem organizados, facilitando a busca de suas partes), semiestruturados (somente uma parte é organizada) ou sem nenhuma estrutura (como em um texto livre).

No caso de uma base de conhecimento organizacional, segundo Probst; Raub; Romhardt (2002), é a somatória dos ativos de conhecimentos individuais e coletivos, incluindo dados e informações que os constroem e que a organização pode utilizar para realizar suas tarefas. O PMBOK (*Project Management Body of Knowledge*) está dentro dessa linha, já que é uma base de conhecimento em gerenciamento de projetos organizado pelo PMI (2012).

A mesma preocupação é compartilhada pela área de governança de serviços em Tecnologia da Informação por meio da definição dada pela ITIL (*Information Technology Infrastructure Library*), que considera base de conhecimento “Um banco de dados lógico

contendo dados e informações usadas pelo sistema de gerenciamento de conhecimento de serviço” (BRITAIN et al, 2011). O ITIL foi desenvolvido originalmente em 1980, no Reino Unido, pela *Central Computer and Telecommunications Agency*.

No caso de Inteligência Artificial e computacional o termo ficou popularmente relacionado a regras e fatos, utilizados em aplicações do tipo Sistemas Especialista discutido no item 2.4 com profundidade, mas existem várias conotações. Assim, segundo Bitemcourt (2006), a base de conhecimento contém informação específica sobre um domínio e será tão complexa quanto à capacidade cognitiva a ser simulada, o que não implica a existência explícita dessa base de conhecimento, ela pode ser compilada através de procedimentos ou autômatos.

## A 2 MÉTODOS ESTATÍSTICOS

Diversos métodos para o tratamento de séries temporais são puramente estatísticos, outros, além de utilizarem a estatística consideram técnicas não estatísticas como, por exemplo, árvore de decisões.

Dessa forma, nos artigos estudados que envolvem estatística, são encontrados os seguintes tópicos (combinados entre si): média, variância, desvio padrão, normalização, probabilidades, filtros de Kalman, teste-t, ki-quadrado, quartil, e *outliers*. Estes elementos foram encontrados e estudados a partir dos seguintes artigos:

- (KEOGH; LIN; FU, 2005) que trata do uso do método HOT SAX para encontrar sequências incomuns em subsequências de series temporais, apresentando algoritmos e aplicações;
- (BU et al., 2011) e (FU et al., 2006) que além do uso da transformada de Haar, utilizam técnicas estatísticas como: média, desvio padrão e normalização utilizadas no HOT SAX;
- (TOSHNIWAL; YADAV, 2011) que além das técnicas estatísticas utilizadas no HOT SAX utiliza também quartil e *outliers*;

- (KNORN; LEITH, 2008) que utiliza o filtro de Kalman;
- (KOUSHANFAR; POTKONJAK; SANGIOVANNI-VINCENTELLI, 2003) que utiliza estatística na detecção de falhas em medidas obtidas em sensores;
- (BASHA; AMEEN, 2007) que em seu método utiliza o teste-t e ki-quadrado.

## A 3 MÉTODOS INTELIGENTES

Neste item estão apresentadas as investigações realizadas por esta pesquisa envolvendo métodos inteligentes. Assim estão destacadas as seguintes abordagens: Inteligência Artificial, Inteligência Computacional, Mineração de Dados, e Técnicas utilizadas em métodos inteligentes.

### A 3 1 Inteligência Artificial

Desde sua criação a inteligência artificial está envolvida em polemicas e possui diversas linhas de pensamento que conduzem a uma diversidade de definições, de tal forma que é necessário apresentar algumas delas, como a seguir:

- (RICH; KNIGHT, 1993) definem como “o estudo de como fazer os computadores realizarem coisas que, no momento, as pessoas fazem melhor”;
- (LUGER; STUBBLEFIELD, 1993) definem como “o ramo da ciência da computação que se preocupa com a automação do comportamento inteligente”;
- (ROUSSELL; NORVIG, 2010) dividem os enfoques da inteligência artificial entre aqueles que retratam o desempenho humano através do pensamento e comportamento e aqueles que consideram a racionalidade e a atuação através da lógica utilizada na computação.

Assim na inteligência artificial existem diversas técnicas capazes de inferir sobre uma representação do conhecimento através de buscas e até de realizarem o aprendizado, acumulando novos conhecimentos. Dentre as técnicas estão: Sistemas Especialistas, redes neurais, Sistemas Baseados em Casos, Lógica Fuzzy, Algoritmos Genéticos, etc. Em Russell; Norvig (2010), é expressa essa ideia através das expressões: processamento em linguagem natural, representação do conhecimento, raciocínio automatizado e aprendizado de máquina. Um exemplo de técnica é apresentado em (MAHONEY; CHAN, 2004), cuja finalidade é o aprendizado de regras para a detecção de anomalias em séries temporais.

### A 3 2 Inteligência Computacional

A Inteligência Computacional é um termo mais recente que busca, através de técnicas inspiradas na Natureza, o desenvolvimento de sistemas inteligentes que imitem aspectos do comportamento humano, tais como: aprendizado, percepção, raciocínio, evolução e adaptação. No quadro 14, estão apresentados alguns exemplos. Nota-se que aí ocorre a apropriação de técnicas que originalmente estavam sobre a esfera da Inteligência Artificial.

Quadro 14- Exemplos de técnicas computacionais x Inspiração na natureza.

<b>Técnica Computacional</b>	<b>Inspiração na Natureza</b>
Redes Neurais	Neurônios biológicos
Computação Evolucionária	Evolução biológica
Lógica Fuzzy	Processamento linguístico
Sistemas Especialistas	Processo de Inferência

Fonte: Adaptado de Roussel e Norvig, (2010), Luger (2013).e Bitemcourt (2006)

### A 3 3 Mineração de Dados

A Mineração de dados e textos (*Data Mining e Text Mining*) envolve o processo de "descoberta do conhecimento", em geral, em grandes bases de dados. Normalmente, são

aplicados determinados algoritmos para a extração de padrões, anomalias e regras a partir dos dados. Também são utilizadas técnicas de Inteligência Artificial para a procura de similaridades ou discordância nos dados. Um exemplo pode ser encontrado no artigo de (GADDAM; PHOHA; BALAGANI, 2007) que utiliza a técnica do *K-means* e a Árvore de decisões (com o ID3), caracterizando um método em cascata com enfoque na mineração de dados de séries temporais.

## A 4 TÉCNICAS UTILIZADAS EM MÉTODOS INTELIGENTES

Cabe ressaltar, neste item, o termo “técnica”, que em algumas situações é utilizado para verdadeiros métodos, a razão disso esta relacionada às abordagens inteligentes que contêm métodos que, por sua vez, contêm técnicas. Os métodos inteligentes possuem uma série de procedimentos embutidos em ferramentas computacionais específicas, na forma de algoritmos e/ou heurísticas. Neles são aplicadas técnicas desenvolvidas a partir de uma ou mais abordagens conceituais; algumas dessas, por sua vez, são usadas em métodos de abordagens diferentes.

### A 4 1 Árvores de Decisões

A árvore de decisão está presente em algumas das abordagens estudadas como, por exemplo, na mineração de dados.

Basicamente a árvore de decisões possui uma estrutura bem conhecida composta por nodos e folhas relacionadas aos atributos dos dados, em que cada nó interno corresponde a uma decisão sobre um atributo.

A árvore de decisão é gerada a partir de um conjunto de dados de treino e da definição de uma propriedade (atributo) objetivo, gerando estruturas abstratas (modelos) que resumem a característica essencial para se chegar aos valores do atributo objetivo.



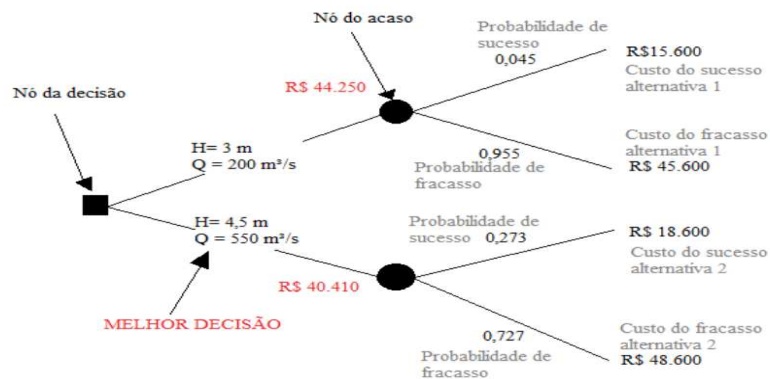
Assim, os modelos gerados, quando expostos a novos casos podem originar previsões para o atributo objetivo do novo exemplo.

Além das considerações já apresentadas para a construção de uma árvore de decisão, existem outras, como: discretização, entropia, ganho informativo, navalha de Occan, ruídos, tratamento de valores nulos, matriz de erros e acerto.

Nas arvore de decisões pode ocorrer a existência de dados inadequados, que podem ser minimizados ou eliminados com consideração adequada, tanto para os dados de treinamento quanto para com os dados gerados. Para tanto são utilizados filtros que separam os dados errados, existentes na base de treinamento. No caso dos dados gerados serem inadequados, utiliza-se a Poda.

A figura 61 apresenta um exemplo desenvolvido para a tomada de decisão na construção de uma Ensecadeira (entre duas alternativas de altura) a ser utilizada para proteção de um canteiro de obras.

Figura 61- Árvore de decisão entre duas alternativas de altura de Ensecadeira



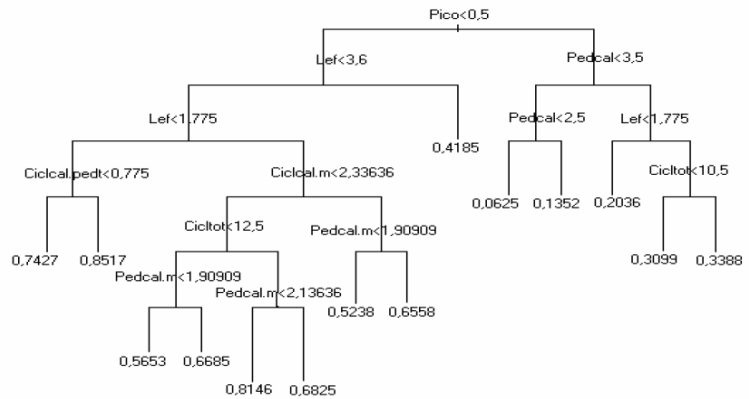
Fonte: Autor, adaptado de Hachich (1999).

Na figura 62, está outro exemplo de árvore de decisão obtida em Carvalho (2006), a partir de vários dados e testes para o dimensionamento de calçadas para pedestres e

ciclistas e considerando o nível de satisfação de pedestres. A árvore de decisão é resultado do uso do *software* S-Plus 6.1

Figura 62 – Árvore de decisões

Lef = largura efetiva da calçada em metros;  
 Ciclotot = Número total de ciclistas;  
 Pedcal = Número de pedestres na calçada;  
 Ciclical/pedt = No de ciclistas na calçada / No de pedestres na via  
 Ciccal/m = No de ciclistas na calçada / metro (largura de calçada)  
 Pedcal/m = No de pedestres na calçada / metro (largura de calçada)



Fonte: Carvalho (2006)

Cabe ressaltar que árvore de decisão também tem outras denominações reservadas para casos específicos, como:

- Árvore de classificação: utilizada quando o objetivo é encontrar o grau de pertinência de um objeto em uma classe específica;
- Árvore de regressão: utilizada quando o objetivo é prever um valor futuro de uma variável com base em seu comportamento passado.

#### A 4 2 Árvore de decisão adaptativa - ADA

Este item tem por base o trabalho desenvolvido por (CATAE; ROCHA, 2011), no qual é indicado que uma ADA possui os mesmos conceitos da Árvore de Decisão tradicional mais a capacidade de assimilar novas informações por meio da inclusão de nodos e ramos em sua estrutura. Na prática, são criados novos atributos em função daqueles já existentes. A incorporação dessa mudança é feita de forma incremental sem a

necessidade de ser refeito o autômato original. Em outras palavras, uma ADA é uma extensão da árvore de decisão, que atua como núcleo, acrescido de uma camada adaptativa.

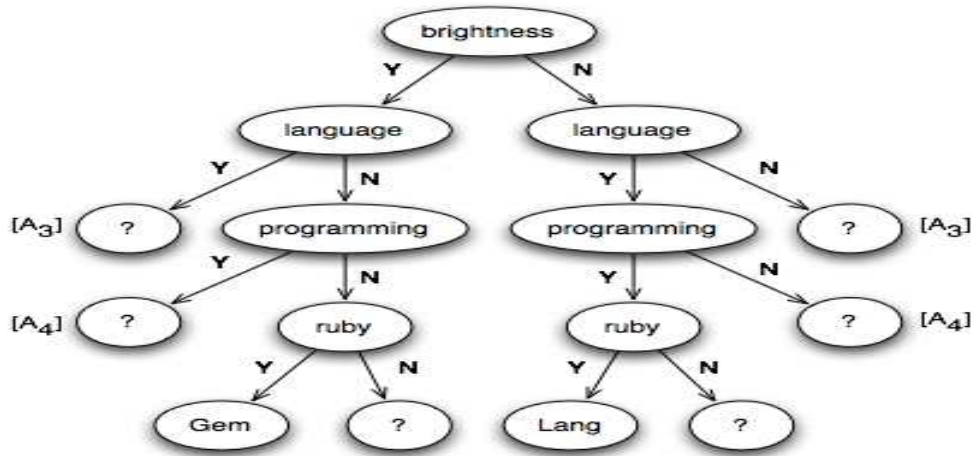
Uma árvore de decisão adaptativa pode ser desenvolvida através do método incremental e completo. O método completo avalia todos os elementos antes de iniciar a criação da árvore de decisão. Após a fase de treinamento, a estrutura continua reagindo aos dados fornecidos e pode se automodificar. Esse processo de adquirir novas informações a partir do meio externo é chamado de aprendizagem incremental.

A árvore de decisão incorpora novas informações em sua estrutura através de automodificações, também modelando adequadamente quando da existência de paridade, ou exclusivo (“xor”) e multiplexação.

Em nosso caso, existe a dependência da ordem dos dados que é a chamada influência ou tendência (bias), na qual resultam em diferentes estruturas de acordo com a ordem de apresentação dos dados. A título de exemplificação está apresentado, na figura 61, um trecho de uma árvore de decisão adaptativa utilizada em (BARAÚNA, 2010) em um classificador de dados conforme seus atributos, a partir do conjunto de treinamento, utilizando o algoritmo AdapTree.

A árvore de decisão da figura 63 é obtida pela aprendizagem do seguinte conjunto de treinamento:  $\{(ruby\ programming\ language, Lang), (ruby\ brightness, Gem)\}$ , onde *Lang* e *Gem* são as categorias dos documentos em questão. Na representação da figura 63, o símbolo de interrogação indica os locais de possíveis adaptações.

Figura 63 - Exemplo de representação de uma árvore de decisão adaptativa



Fonte: Baraúna (2010)

### A 4 3 Lógica Fuzzy

O texto apresentado neste item está embasado nas referências de (SIMÕES e SHAW, 2007) e de (OLIVEIRA, 1999).

A lógica Fuzzy foi desenvolvida por Lofti A. Zadeh na Universidade da Califórnia, em Berkeley, na década de 60 do século XX, e é utilizada em aplicações que envolvem conhecimento impreciso e naquelas que envolvem impossibilidades de custo, dificuldades técnicas, tempo de desenvolvimento curto, dentre outras.

Em (CAMARGOS, 2002), foram reunidas características, vantagens e desvantagens da Lógica Fuzzy apresentada no quadro 15.

Quadro 15 - Característica da Lógica Fuzzy: vantagens e desvantagens

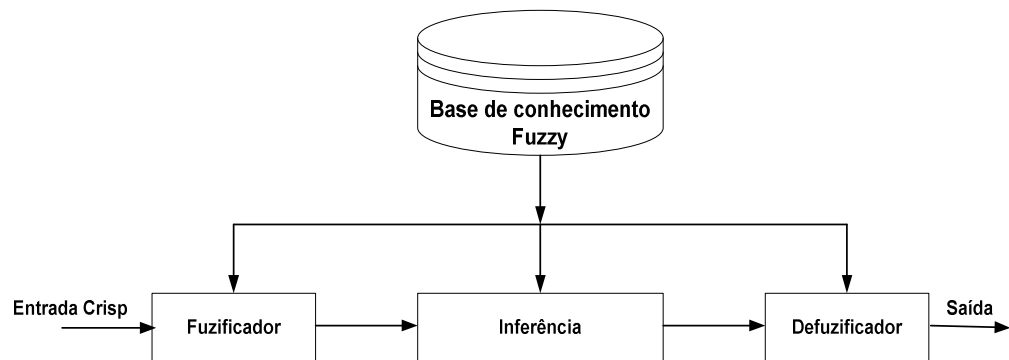
Características	Vantagens	Desvantagens
A Lógica Nebulosa está baseada em palavras e não em números, ou seja, os valores verdades são expressos linguisticamente. Por exemplo: quente, muito frio, verdade, longe, perto, rápido, vagaroso, médio.	O uso de variáveis linguísticas nos deixa mais perto do pensamento humano.	Necessitam de mais simulação e testes.
Possui vários modificadores de predicado, tais como: muito, mais ou menos, pouco, bastante, médio.	Requer poucas regras, valores e decisões.	Não aprendem facilmente.
Possui também um amplo conjunto de quantificadores como: poucos, vários, em torno de, usualmente.	Simplifica a solução de problemas e a aquisição da base do conhecimento.	Dificuldades de estabelecer regras corretamente.
Faz uso das probabilidades linguísticas (como, PE, provável e improvável) que são interpretados como números nebulosos e manipulados pela aritmética.	Mais variáveis observáveis podem ser valoradas.	Não há uma definição matemática precisa.
Manuseia todos os valores entre 0 e 1, tomando estes, como um limite apenas.	Mais fáceis de entender, manter e testar.	
	São robustos. Operam com falta de regras ou com regras defeituosas.	
	Acumulam evidências contra e a favor.	
	Proporciona um rápido protótipo dos sistemas.	

Fonte: Camargos (2002)

A imprecisão é expressa por graus e/ou funções de pertinência, que indicam o quanto uma sentença é verdadeira. Em um conjunto Fuzzy o grau de pertinência considera um intervalo  $[0,1]$ , no qual o valor zero é totalmente falso e o valor um é totalmente verdade, e entre os dois existem infinitas possibilidades, de valores, como por exemplo, 0,7.

Na arquitetura de inferência Fuzzy apresentada na figura 64, pode-se observar a existência de uma base de conhecimento que contém regras Fuzzy do tipo “Se antecedente então consequente” ou “Se condições então conclusão”, e uma base de dados que contém as funções de pertinência dos conjuntos Fuzzy utilizados nas regras. A base de conhecimento fornece as informações necessárias para as operações de fuzzificação, inferência e defuzzificação.

Figura 64 – Arquitetura de Inferência Fuzzy



Fonte: Autor

As etapas que envolvem as operações de inferência são:

- Obter a partir das entradas (nas variáveis e suas funções de pertinência no lado antecedente), os valores de pertinência de cada variável linguística, (fuzzificação);
- Combinar os valores obtidos com parte consequente para disparar cada regra;
- Gerar os consequentes habilitados (Fuzzy ou Crisp) para cada regra;
- Agregar os consequentes habilitados para produzir uma saída Crisp, (defuzzificação).

Cabe ressaltar que a inferência pode ser realizada de muitas formas, como por exemplo, utilizando a técnica MÁX-MÍN ou Mandani.

Com relação às regras, elas são obtidas de diversos modos, e dois deles se destacam. O primeiro é baseado em conhecimento intuitivo e em resultados experimentais, dependendo da habilidade do desenvolvedor e do conhecimento dele sobre a aplicação. O segundo modo é mais formal e visa cobrir todas as possibilidades, combinando as variáveis linguísticas de entrada com as de saída, isto é, relacionando as variáveis antecedentes com as consequentes.

## A 4 4 Redes Neurais

As redes neurais podem ser utilizadas no estudo de monitoramento, inclusive no *smart grid*, com base em: auto-organização, adaptatividade, autoaprendizagem, memória associativa, alto grau de tolerância a falhas, processamento paralelo, um alto grau de não linearidade na capacidade de mapeamento, e com as características da dinâmica linear.

Basicamente, uma rede neural é composta por camadas, sendo uma de entrada, outra de saída e uma ou mais camadas ocultas. Cada camada possui nodos que são interligados com os nodos de outras camadas, esses nodos não possuem ligação com os nodos da mesma camada.

Os estudos que envolvem redes neurais e deslocamentos em barragens em sua maioria necessitam de treinamentos, assim são descartados devido aos requisitos desta Tese. Dentre os artigos estudados estão:

- “Detecção de sinais fracos caóticos usando um sistema de inferência adaptativo neuro-fuzzy” (YE; SONG, 2011);
- “Método para o tratamento de uma série de dados em intervalo variável com o uso de redes neurais caóticas e wavelet caóticas” (SHA, 2010);
- “Modelo de Previsão de Redes Neurais BP em deslocamentos na superfície da rocha provocado por troca de ancoragem” (CHUNHUI; XIAOYUE, 2011);
- “Algoritmo de rede neural fuzzy e sua aplicação” de (TIESHENG; ZHENXING, 2011);
- “Aplicação da Rede Neural Artificial no sistema de previsão de inundações tufão: Um estudo de caso em Xangai, China” (XIAODONG et al., 2010);
- “Estudo das aplicações de redes neurais no controle de deformações em reservatórios” (WU et al., 2010);
- “Modelo de deslocamento de reservatórios em redes neurais baseado no algoritmo colônia de formigas” (JIANG; WANG, 2009).

## A.5 CONSIDERAÇÕES SOBRE TÉCNICAS DIVERSAS UTILIZADAS NO TRATAMENTO DE SÉRIES TEMPORAIS

As técnicas apresentadas neste item são consideradas em diversos métodos aplicados a séries temporais e podem estar utilizadas em abordagens: Estatísticas, Inteligência Artificial, Inteligência Computacional, Mineração de Dados, dentre outros.

Dentre as técnicas discutidas neste item, estão as janelas deslizantes e a função de distância.

Muitas dessas técnicas são avançadas e para:

(RIBEIRO; GOLDSCHMIDT; CHOREN, 2009) “a previsão de valores futuro com base em valores passados, é necessário que se disponha de uma memória histórica de dados ocorridos anteriormente. Todavia o conjunto de dados, por si só, não permite a previsão dos valores futuros. Para isso, é necessário a utilização de algoritmos, técnicas ou métodos de previsão de series temporais, que podem envolver cálculos simples ou procedimentos avançados”.

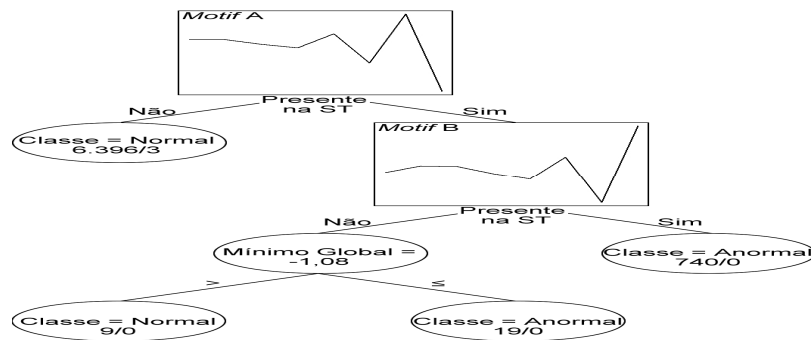
Atualmente existem muitos trabalhos acadêmicos que buscam em séries temporais a extração de padrões e de características existentes positivas, além daqueles que visam à obtenção de anomalias, sequências incomuns, discórdias, divergências e novidades.

Grande parte dos sistemas existentes, conforme Maletzke et al. (2009), tratam os dados de maneira particular, adhoc, e uma dessas abordagens é utilizar uma janela deslizante que personaliza cada deslocamento da janela de forma a predizer um valor individual (na realidade, é predizer um atributo  $x$  a partir de outro atributo  $y$ ), os exemplos são tratados independentemente e identicamente distribuídos. Existe um viés neste tratamento, no qual nem tudo é independentemente e identicamente distribuído e nem todo deslocamento pode ser personalizado, pois não é independente em seus atributos, podendo acarretar em perda de informação. Por isso, em Maletzke et al. (2009), é proposto um método e, conseqüentemente, um algoritmo que utiliza dados estatísticos globais (média, variância, mínimos e máximos) como características e Motifs (que são subsequências que se repetem em uma série temporal representando fenômenos locais de interesse). Os Motifs também são denominados de *Frequent Pattern*.



No método de (MALETZKE et al., 2009) com referencias a (MALETZKE et al., 2008), existe uma janela (Motif) correspondendo 10% do tamanho da sequência a ser discretizada. Uma evolução para o método é utilizar Motifs de tamanho distinto e a utilização de outras características para obter modelos mais precisos. Na figura 65, é mostrado um exemplo de árvore de decisão gerada com o método apresentado em (MALETZKE et al., 2009).

Figura 65 - Árvore de decisão gerada com o método apresentado por Maletzke



Fonte: Maletzke et al. (2009)

Outras formas de análise de séries temporais visam obter eventos anormais e sequencias não padrões conforme apresentados em:

- (ROBLES-KELLY; HANCOCK, 2003), que trata da distância de edição de um grafo espectral, no qual a *Edit Distance* (distância de edição) calcula o número de inserções, remoções ou substituições para transformar um conteúdo em outro (na realidade uma *string* em outra), capaz de medir a diferença e semelhança entre duas sequencias. O autor considera essa ideia básica com relação a grafos, após a serialização de um grafo. Também é utilizada a teoria espectral que estuda os autovalores (valor próprio) e autovetores (vetor próprio) em operadores lineares;
- (BUNKE; KRAETZL, 2004), que trata da classificação e detecção de eventos anormais em series temporais de grafos, utilizando a distância de edição em grafo;

- (KEOGH; LIN; FU, 2005) que trata do uso do método HOT SAX para encontrar sequências incomuns em subsequências de series temporais, apresentando algoritmos e aplicações. Também utiliza janela deslizante.

Nos métodos considerados anteriormente são utilizados o conceito de medida de dissimilaridade (também denominados como função de distância), em Alencar (2007) é comentado que essa função tem como parâmetros de entrada duas séries temporais ou duas subsequências, e retorna um valor não negativo que quantifica a dissimilaridade (ou distância) de uma série com outra. Quanto menor o valor da distância, maior é a similaridade entre as duas séries. Para que essa função seja simétrica, não importa a ordem dos parâmetros de entrada na função, dessa forma, é possível identificar que uma subsequência é similar à outra. Uma das funções de distância utilizada é a Euclidiana, sendo válida para séries de mesmo tamanho.

Outro conceito utilizado é o Match (casamento), conforme Alencar (2007): dados um número real positivo (chamado de tolerância) e uma série temporal contendo uma subsequência que começa em uma posição e outra subsequência que começa em outra posição, se a função distância entre elas for calculada e o resultado for menor que a tolerância, então esse resultado indica o chamado casamento das subsequências.

Reforçando as ideias apresentadas em (KEOGH; LIN; FU, 2005) para encontrar séries incomuns, denominadas por ele de Discórdias, definidas como a subsequência da serie temporal que mais difere do restante das subsequências dada uma métrica de distância.

O algoritmo apresentado em (KEOGH; LIN; FU, 2005) e comentado em (LE MOS, 2007), como segue, utiliza uma janela deslizante com passo de deslocamento 1 para gerar janelas de pontos que serão convertidas em uma representação simbólica utilizando o método SAX (*Symbolic Aggregate ApproXimation*), com o objetivo de reduzir a dimensionalidade dos dados. Cada uma das janelas convertidas é considerada uma subsequência. Em seguida, é utilizada a distância Euclidiana como métrica de distância entre subsequências, ignorando aquelas que possuem sobreposição de dados. As subsequências utilizadas são uma subsequência particular e a subsequência mais similar,

definindo como discórdia aquela que tiver o maior valor de distância. Segundo Lemos (2007), na prática, o método HOT SAX, conforme apresentado em Keogh; Lin; Fu (2005), não dá indicações de quantas Discórdias podem ser encontradas pelo algoritmo, não possui uma metodologia de determinação do limiar de distância, de tal forma que uma subsequência possa ultrapassar o limiar e ser classificada como Discórdia.

## A 6 CONSIDERAÇÕES SOBRE MEDIÇÕES

O método e os aplicativos desenvolvidos neste trabalho realizam as suas análises após o recebimento de uma medição em um conjunto de dados que compõe uma série histórica (ou temporal). Este item esclarece o que acontece antes das medições serem recebidas pelo aplicativo MAIM desenvolvido por esta tese.

O caminho dos valores envolvidos é iniciado com a leitura realizada por um instrumento de medição de forma analógica e segue até sua transformação em medição na forma digital. Duas possibilidades estão consideradas: o tratamento dado ao processo manualmente e dado automaticamente.

No tratamento dado manualmente, os dados são coletados e eventualmente envolvem cálculos por um operador antes de serem digitados no sistema computacional que, praticamente transforma o valor inserido em cadastros que são armazenados digitalmente, não havendo perda de informação, e todos os dados coletados sendo utilizados.

No caso do tratamento automático, os dados são coletados por instrumento de medição e transferidos para um sensor que conduz à transformação em valores digitais, que pode envolver um conversor analógico-digital (A/D). Os valores transformados são enviados (através de comunicação) a um sistema computacional.

A seguir estão apresentadas informações sobre os sinais analógicos e digitais.

## A 6 1 Medidas e sinais analógicos

Os instrumentos de medição podem obter os seus dados diretamente ou indiretamente. No caso da grandeza ser obtida diretamente é utilizado o instrumento de medida sem nenhuma transformação da grandeza em outra, por exemplo, a utilização de uma régua em centímetros. No caso da aquisição de dados de forma indireta, uma grandeza é transformada em outra. Desse modo, com base em Fialho (2010), um exemplo para a obtenção indireta são as leituras realizadas na natureza que podem ser convertidas em sinais elétricos proporcionais e lidos por instrumentos apropriados como: voltímetros, amperímetros, frequencímetros, dentre outros.

Sabe-se que erros acontecem e propagam ao longo de todo o fluxo de dados em um processo, sejam eles realizados de forma manual e/ou digital. Segundo Fialho (2010), as medidas obtidas refletem a precisão de como foram medidas e a incerteza (estimativa de seu erro); assim é comum encontrá-las representadas por seu valor numérico, incluindo a indeterminação e a unidade de medida (algumas são adimensionais), por exemplo: a temperatura  $(500 \pm 3) ^\circ\text{C}$ .

Para Fialho (2010) quanto maior a precisão requerida para uma medida, mais demorada e cara ele é em um processo, dessa forma recomenda sempre utilizar o método que forneça a informação com a precisão necessária ao processo.

Ainda segundo Fialho (2010), um dos métodos utilizados para o cálculo do erro é o do Kleine e McClintock, que como a grande maioria dos métodos semelhantes, tem por base conceitos da estatística. Nele, um erro resultante é o resultado de uma função contendo todas as variáveis independentes do processo de medição, onde para cada uma delas é considerado o seu próprio erro.

(FIALHO,2010) afirma ainda que “o maior erro presente em um processo de medição contribui mais significativamente para o resultado final”, essa consideração pode ajudar na escolha de componentes e na determinação do processo.

Em instrumentos analógicos, além dos erros que podem envolver as variáveis existentes, também podem aparecer erros característicos como os existentes em instrumento com mostradores baseados em ponteiros, que são o erro de paralaxe (devido a posição do observador) e erro de interpolação (devido a posição do ponteiro com relação à escala de medida).

## **A 6 2 Sinais analógicos**



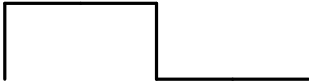
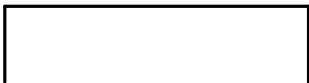
No caso do uso de instrumentos digitais, grande parte dos erros apresentados, devido à características analógicas, não existe (como no caso dos erros de paralaxe e interpolação), mas dependendo da forma como são realizadas as conversões analógicas para digitais, outros tipos de erros existem como os originados em alguns componentes do processo. Esses erros podem ser propagados para a forma digital e serem percebidos, por exemplo: em figuras por meio de desenhos serrilhados, em áudio, por meio do som de apito, além de, em geral envolverem a perda de informações e distorções em valores.

Atualmente na conversão analógico-digital existem quatro passos: amostragem, retenção, quantificação, e codificação.

Antigamente a amostragem e a retenção utilizavam amplificadores e, a quantificação e a codificação os conversores A/D. Hoje estão embutidas diretamente em microprocessadores, microcontroladores, processadores e computadores.

No quadro 16, está apresentado um exemplo de conversão de dados analógicos para digitais, mostrando uma entrada de dados em Volts que é transformada em uma palavra (byte) de 4 bits, e o correspondente sinal gerado.

Quadro 16 – Exemplo de transformação de entrada analógica em saída digital

Entrada em Volts	Palavra (4 Bits)	Sinal
0,0	0000	
.....		
0,7	0111	
.....		
1,2	1100	
.....		
1,5	1111	

Fonte: Adaptado de Fialho (2010)

## A 6 2 1 Amostragem

A Amostragem consiste em dividir parte do sinal em amostras, o mais próximo uma das outras e representar o valor na forma digital. Quanto maior o número de amostras melhor, pois será fácil reconstruir o sinal original.

Para determinar a taxa de amostragem (sua frequência), segundo Bolton (2010) e Alciotore ; Histan (2014), é utilizado o critério (ou frequência) de Nyquist ou teorema da Amostragem (ou amostragem de Shannon). Esse teorema indica que para reconstituir o sinal, é necessária pelo menos uma taxa de amostragem duas vezes maior que a frequência do sinal analógico. Alguns autores como Fialho (2010), enfocam como mínima a velocidade de amostragem, pois o número de amostras depende da velocidade do circuito. Comercialmente e na prática é recomendado multiplicar por dez diminuindo bem a perda de informação.

No caso de a taxa de amostragem ser menor que o recomendado, isto é, não corresponder ao teorema da amostragem, pode ocorrer o chamado sinal de *Aliasing*, que faz com que sinais diferentes se tornem indistinguíveis ou correspondam a uma distorção, tendo como consequência uma reconstrução do sinal contínuo diferente do original.

Devido a essa questão nos processos de conversão A/D é utilizado um componente denominado filtro Passa-baixos (ou *Anti-Aliasing*).

### **A 6 2 2 Retenção**

Retenção ocorre devido ao tempo de conversão (que é função dos componentes e da velocidade da amostragem), que começa na quantificação e vai até a saída do código digital, referência para a armazenagem de uma palavra (Byte) ou mais.

### **A 6 2 3 Quantificação**

Para Fialho (2010), a quantificação dá uma grandeza para cada uma das amostragens, como os dados são binários. Na realidade, para Alciatore; Histan (2014), um sinal digital contínuo é transformado em um conjunto de estados de saída discretos. Cabe ressaltar que a resolução ou sensibilidade de um conversor A/D, segundo (FIALHO,2010), é “a mínima variação do sinal analógico que provoca uma variação do código de saída até ao imediatamente superior ou inferior”.

### **A 6 2 4 Codificação**

Com base em Fialho (2010) e Alciatore e Histan (2014), a codificação no sistema binário é o número de diferentes bits existentes no conversor para quantificação (na prática é o número de bits de saída), é atribuído um número para cada estado de saída.

#### **A 6 2 4 Eliminação de alguns tipos de erros na comunicação**

No processamento digital de sinais segundo Bolton (2010), é possível utilizar duas faixas de tensões para distinguir entre dois estados binários para cada bit, por exemplo, para 5 V, normalmente existe a diferença de 3V no mínimo, assim os sinais podem ser (0 e 5V) ou (1 e 4V) distinguindo os bits 0 e 1. Esse permite que a precisão dos dados seja menos afetada por: ruído, deriva, tolerância dos componentes e outros fatores que provocam flutuações nas tensões.

No caso da transmissão de dados, segundo Bolton (2010), existe a detecção de erro pelo método da paridade, que evita os erros devido aos ruídos elétricos que mudam o sinal. Consiste no acréscimo na palavra de um bit extra, se o bit menos significativo é 0, é acrescentado o bit 1 na palavra após o bit menos significativo e, se for 1 é 0. Desse modo, se algum bit for alterado, um erro é detectado e solicitada a retransmissão do sinal. Outro tipo de verificação é a soma (*Sum Check*).



## APÊNDICE B – DETALHES DAS ANÁLISES E DESENVOLVIMENTOS

A pesquisa tem como resultados diversos experimentos realizados, sendo que alguns apresentam modificações que os tornam funcionais, segundo alguns requisitos da Tese. Para os experimentos, existe um padrão de dados criado de forma a ser utilizado na repetição de experimentos e na realização de novos estudos, por meio do uso de uma sequência numérica básica contendo uma sequência de pontos característicos.

### B 1 MÉTODO HOT SAX ORIGINAL

Para Keogh; Lin; Fu (2005), o método HOT SAX basicamente diminui o tamanho de uma série temporal em segmentos com o mesmo número de pontos, obtendo suas médias, e agrupando-as em intervalos estabelecidos e nomeados com símbolos (caracteres ou letras). A junção desses símbolos gera uma palavra, que poderá ser subdividida em novos segmentos de igual tamanho, nos quais é investigada a existência de padrões e não padrões. A existência dessas possibilidades levou a utilização do HOT SAX em diversas aplicações, dentre elas, a obtenção de um segmento que destoa dos outros segmentos existentes na série temporal, podendo ter um significado particular para cada aplicação, como inconsistência nos dados, não conformidade, distúrbio, dentre outras possibilidades.

O HOT SAX é composto por uma série de etapas e um algoritmo básico para buscar algo incomum ou padrões. As duas primeiras etapas são compostas por outros dois métodos anteriores ao HOT SAX e com aplicações consagradas. O primeiro método é denominado segundo Keogh et al. (2001), de PAA (*Piecewise Aggregate Approximation*), que consiste em normalizar uma série temporal, subtraindo de cada ponto da série seu valor médio e dividindo pelo seu desvio padrão, obtendo uma nova série temporal dita normalizada. A nova série é subdividida em N segmentos igualmente espaçados; para cada segmento, é calculado um valor médio, obtendo outra série composta somente com os valores médios de cada segmento. Na prática, é dito que o PAA reduziu a dimensionalidade de uma série de dados.

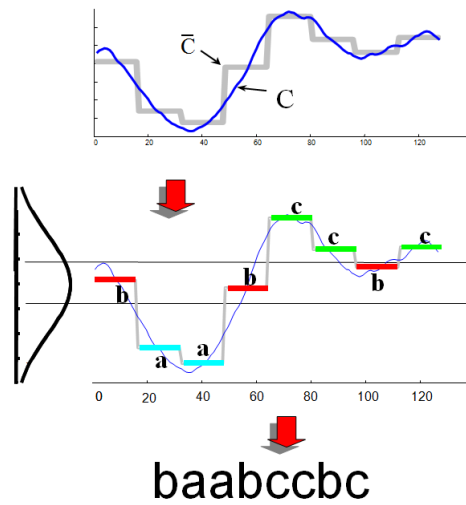
O segundo método é denominado segundo Lin et al. (2003), de SAX (*Symbolic Aggregate ApproXimation*), que inclui a série gerada pelo PAA e uma curva Gaussiana dividida em um determinado número de intervalos, que originalmente são considerados de 3 a 10.

A aplicação do SAX permite a redução da dimensionalidade e a representação simbólica dos dados de uma série temporal e também apresenta uma medida de distância eficaz para cálculo da similaridade entre a série discretizada e a série de dados originais (*Lower Bounding*). Outra das características do SAX é identificar as sequências que ocorrem mais frequentemente em uma série temporal.

Na prática o SAX permite que uma série de tamanho  $n$  seja reduzida para uma palavra de tamanho  $w$ , onde  $w$  pode ser muito menor que  $n$ . O tamanho do alfabeto é um número inteiro representado por  $a$ , onde  $a > 2$ . Cabe ressaltar que parte do método do SAX corresponde à utilização do PAA. Para a aplicação do SAX, normalmente, são seguidos diversos passos que levam a transformações da série temporal, conforme a técnica PAA e, posteriormente, para uma representação simbólica em uma palavra com mostrado em um exemplo ilustrativo na figura 66.

Na figura 66 o gráfico superior corresponde, inicialmente, a uma transformação da série temporal em azul na curva gerada pelo método PAA, que trabalha a média entre pontos para chegar a uma curva cheia de patamares. Posteriormente, uma curva Gaussiana (utilizada para escalonar os dados, para se enquadrarem em uma faixa pequena de valores) é utilizada para categorizar as regiões da curva obtida em um alfabeto determinado. A curva Gaussiana é dividida em áreas iguais e os patamares existentes nesta área correspondem a determinados caracteres do alfabeto utilizado. Em geral esse alfabeto corresponde de 3 a 10 elementos (nada impede de se utilizar um alfabeto maior que 10), no caso da figura 64 esses elementos são: a, b, c. A concatenação dos símbolos do alfabeto utilizado na curva final irá corresponder a uma palavra que, no caso da figura 66, é baabccbc.

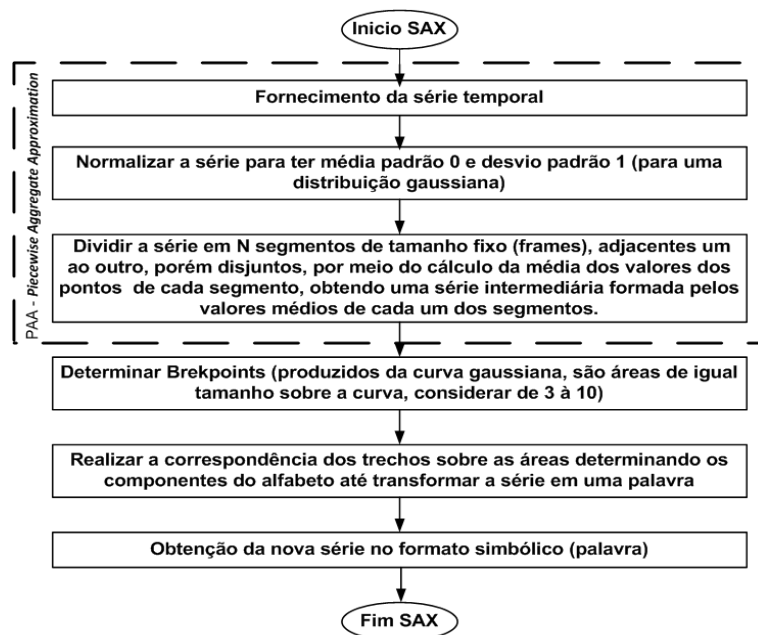
Figura 66 - Transformações de uma série temporal em uma palavra



Fonte: Autor

Para chegar à transformação final de uma série temporal em palavra SAX, são seguidos os diversos passos apresentados no fluxograma da figura 67.

Figura 67 - Fluxograma com os passos utilizados no SAX

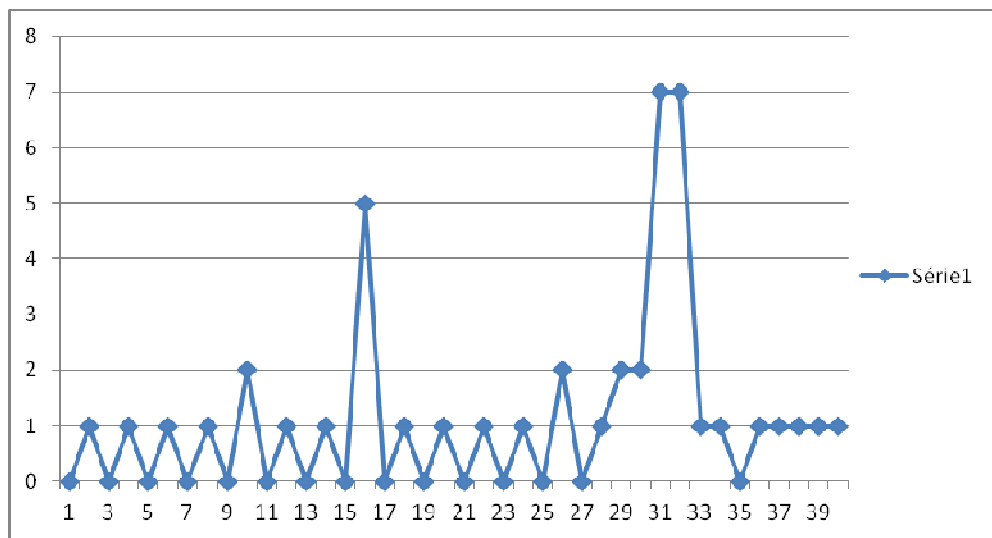


Fonte: Autor

Todos os intervalos irão corresponder ao mesmo valor de área da curva. O objetivo é classificar os valores das médias nos intervalos correspondentes. Para facilitar, a cada intervalo é atribuído um símbolo diferente; o conjunto desses símbolos é denominado alfabeto. Os valores da série temporal obtida pelo PAA são alocados no eixo das abscissas e o tempo alocado no eixo das ordenadas.

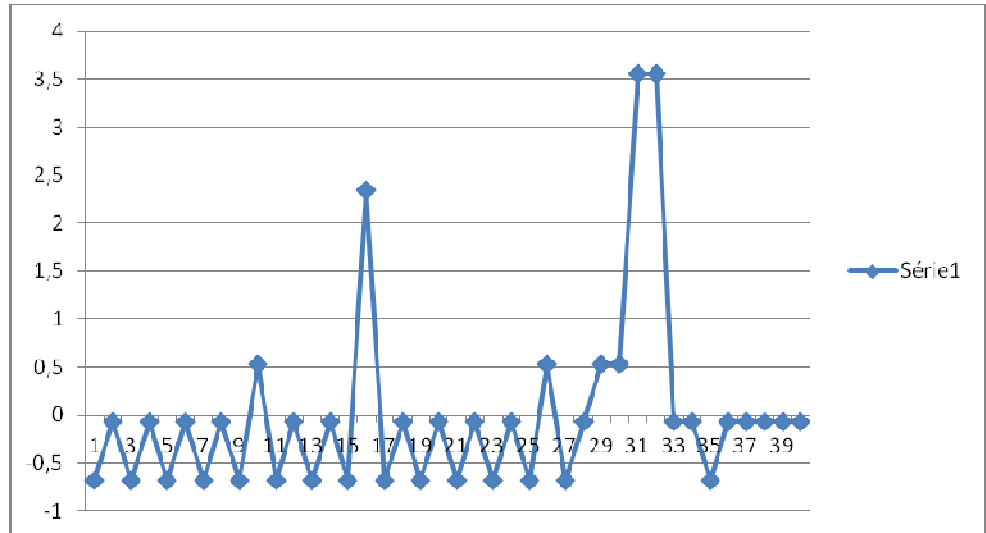
Em seguida o SAX sobrepõe, no eixo das ordenadas, os valores dos intervalos da curva Gaussiana. O SAX percorre os valores médios existentes no tempo e vai identificando em qual intervalo ele está, atribuindo o símbolo correspondente ao intervalo; caso todas as letras forem concatenadas, será formada a palavra SAX. Um exemplo dessa transformação pode ser vista nas figuras 68, 69, e 70. Nelas foi utilizada uma série padrão experimental considerada para testes com valores adimensionais. A figura 66 apresenta um gráfico traçado com a série original, na figura 70, está apresentado um gráfico traçado com a série normalizada, e a figura 68 apresenta um gráfico traçado com as médias e os respectivos símbolos de cada grupo no tempo, que é composto por dois pontos. No caso do exemplo é obtida a palavra SAX “bbbbbbdbbbbbbcdbbbb”.

Figura 68 - Representação gráfica da série original (valores adimensionais)



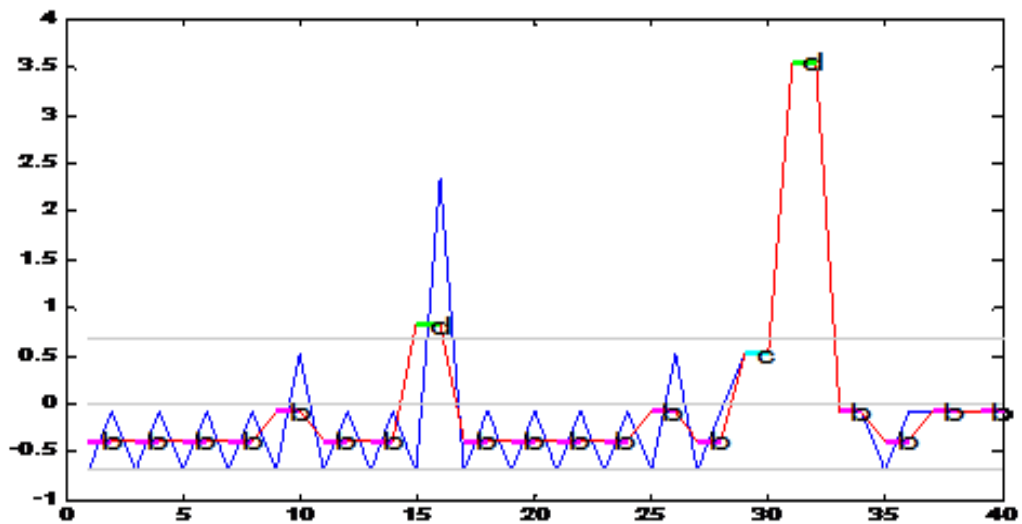
Fonte: Nicoli et al (2013)

Figura 69 - Representação gráfica com a série normalizada (valores adimensionais)



Fonte: Nicoli et al (2013)r

Figura 70 - Representação gráfica da aplicação com o SAX, (valores adimensionais)



Fonte: Nicoli et al (2013)

O método HOT SAX propriamente dito é aplicado após o SAX estar concluído. Para isso estabelece uma janela móvel com um determinado comprimento fixo em

símbolos, cuja finalidade é percorrer a palavra SAX, por exemplo, um símbolo por vez, identificando segmentos de símbolos. Na figura 71, estão apresentados os segmentos obtidos com a janela móvel, indexando e indicando a quantidade de ocorrências iguais, além de indicar a estrutura de uma árvore de busca.

Figura 71 - Estrutura obtida do exemplo

Índice	Segmento	Número Ocorrências	Árvore de busca
1	bbbb	8	
2	bbbb	8	
3	bbbb	8	
4	bbbb	8	
5	bbbd	1	
6	bbdb	1	
7	bdbb	1	
8	dbbb	2	
9	bbbb	8	
10	bbbb	8	
11	bbbb	8	
12	bbbc	1	
13	bbcd	1	
14	bcd b	1	
15	cdbb	1	
16	dbbb	2	
17	bbbb	8	

Fonte: Nicoli et al (2013)

De posse desse conjunto de segmentos, conforme pode ser observado na tabela 30, realiza-se a comparação entre esses segmentos por meio do cálculo da distância Euclidiana, que fornece um valor que quantifica o quanto dois segmentos são diferentes. Essa comparação pode ser feita em sua pior hipótese, com o uso da chamada força bruta,

da seguinte forma: para cada segmento, calcula-se sua distância relativa aos demais. Com o intuito de diminuir a quantidade de comparações, o método HOT SAX utiliza as chamadas heurísticas.

Com a intenção de facilitar o trabalho computacional e a demonstração do exemplo, são utilizadas as tabelas 17 e 18, que simplificarão o cálculo da distância entre seguimentos com os resultados apresentados na figura 72.

Para a obtenção da distância entre os segmentos é utilizada a tabela 17 que apresenta os valores que dividem a distribuição Gaussiana em regiões equiprováveis em números de áreas arbitrárias de 3 a 10. Na tabela 17 a coluna denotada por  $\beta/a$  indica na vertical os valores dos pontos limites das áreas denominadas em (KEOGH; LIN; FU, 2005), como “*Breakpoints*” e na horizontal, a quantidade de áreas em que é dividida a curva Gaussiana.

Os *Breakpoints* da tabela 17 são utilizados para determinar o conteúdo das matrizes apresentadas na tabela 18, que pode ser gerado no momento de atribuição dos símbolos do alfabeto. Esse conteúdo varia em função da quantidade de símbolos. Na tabela 18, são mostradas as matrizes para alfabetos com três e quatro símbolos. No exemplo apresentado neste texto é utilizada a matriz para quatro símbolos.

Tabela 17 - Distribuição gaussiana em regiões equiprováveis

$\beta/a$	3	4	5	6	7	8	9	10
$\beta_1$	-0,43	-0,67	-0,84	-0,97	-1,07	-1,15	-1,22	-1,28
$\beta_2$	+0,43	0	-0,25	-0,43	-0,57	-0,67	-0,76	-0,84
$\beta_3$		0,67	+0,25	0	-0,18	-0,32	-0,43	-0,52
$\beta_4$			+0,84	+0,43	+0,18	0	-0,14	-0,25
$\beta_5$				+0,97	+0,57	+0,32	+0,14	0
$\beta_6$					+1,07	+0,67	+0,43	+0,25
$\beta_7$						+1,15	+0,76	+0,52
$\beta_8$							+1,22	+0,84
$\beta_9$								+1,28

Fonte: Nicoli et al (2013)

Tabela 18 - Distâncias entre símbolos do alfabeto

Três Símbolos				Quatro Símbolos				
	<b>a</b>	<b>b</b>	<b>c</b>		<b>a</b>	<b>b</b>	<b>c</b>	<b>d</b>
<b>a</b>	0	0	0,86	<b>a</b>	0	0	0,67	1,34
<b>b</b>	0	0	0	<b>b</b>	0	0	0	0,67
<b>c</b>	0,86	0	0	<b>c</b>	0,67	0	0	0
				<b>d</b>	1,34	0,67	0	0

Fonte: Nicoli et al (2013)

A matriz para quatro símbolos apresentada na tabela 17 é obtida realizando cálculos, para cada uma de suas células, conforme a equação abaixo:

$$C_{l,c} = 0, \text{ Se } |l-c| \leq 1, \text{ ou Caso contrário: } \beta_{\max(l,c)} - 1 - \beta_{\min(l,c)}$$

Onde C é o conteúdo da célula e l e c são os índices referentes à linha e coluna da matriz.

O conteúdo da célula da matriz corresponde à distância entre dois símbolos, por exemplo: a distância entre “a” e “d”, ad, na matriz para quatro símbolos é 1,34. Assim, com o uso da matriz, por exemplo, a distância entre os segmentos “bbdb” e o segmento “bbbd” é calculado pela somatória dos valores (bb + bb + db + bd = 0 + 0 + 0,67 + 0,67) que resulta 1,34.

Os cálculos realizados até agora estão incluídos na expressão (1), apresentada em (LIN et al., 2002), em que a distância dos segmentos corresponde à raiz quadrada da somatória ao quadrado da distância entre seus símbolos ao quadrado multiplicado pela raiz quadrada do comprimento da série dividida pelo tamanho do segmento. Na prática, essa expressão é uma aproximação da distância Euclidiana entre os segmentos, que podem ser estabelecidos para montar a tabela 19 com o auxílio da tabela 17 e 18 ou simplesmente serem calculados computacionalmente.



$$MINDIST(\hat{Q}, \hat{C}) \equiv \sqrt{\frac{n}{w}} \sqrt{\sum_{i=1}^w (dist(\hat{q}_i, \hat{c}_i))^2} \quad (1)$$

Mas para efeito de simplificação das operações de cálculo, só é considerada a distância dada pela somatória que se encontra no interior da raiz quadrada, que simplifica os cálculos.

Assim pode ser obtida a tabela 19, que corresponde aos resultados obtidos pelo cálculo simplificado.

Nos dados da tabela 19, é feita a busca para encontrar o valor da maior distância (diferença) correspondendo a não conformidade. A eficiência do método está na forma como essa busca é realizada. Para isso, o método estabelece um algoritmo que compara as distâncias existentes entre os segmentos e indica a primeira ocorrência de algo incomum. Para evitar o uso da força bruta, são criadas otimizações denominadas de heurísticas que diminuem a quantidade de cálculos entre segmentos, evitando, por exemplo, cálculos repetitivos, tais como a distancia entre “aaaa” e “bbbb” e seu simétrico “bbbb” e “aaaa”. É possível utilizar, segundo o método original (KEOGH; LIN; FU, 2005), uma das três heurísticas denominadas de: *Random*, *Magic* e *Perverse*.

Tabela 19 - Resultados do cálculo simplificado da distância entre os segmentos

	bbbb	bbbb	bbbb	bbbb	bbbd	bbdb	bdbb	dbbb	bbbb	bbbb	bbbb	bbbc	bbcd	bcbd	cdbb	dbbb	bbbb
bbbb	0	0	0	0	0,67	0,67	0,67	0,67	0	0	0	0	0,67	0,67	0,67	0,67	0
bbbb	0	0	0	0	0,67	0,67	0,67	0,67	0	0	0	0	0,67	0,67	0,67	0,67	0
bbbb	0	0	0	0	0,67	0,67	0,67	0,67	0	0	0	0	0,67	0,67	0,67	0,67	0
bbbb	0	0	0	0	0,67	0,67	0,67	0,67	0	0	0	0	0,67	0,67	0,67	0,67	0
bbbd	0,67	0,67	0,67	0,67	0	1,34	1,34	1,34	0,67	0,67	0,67	0,67	0	1,34	1,34	1,34	0,67
bbdb	0,67	0,67	0,67	0,67	1,34	0	1,34	1,34	0,67	0,67	0,67	0,67	1,34	0	1,34	1,34	0,67
bdbb	0,67	0,67	0,67	0,67	1,34	1,34	0	1,34	0,67	0,67	0,67	0,67	1,34	1,34	0	1,34	0,67
dbbb	0,67	0,67	0,67	0,67	1,34	1,34	1,34	0	0,67	0,67	0,67	0,67	1,34	1,34	1,34	0	0,67
bbbb	0	0	0	0	0,67	0,67	0,67	0,67	0	0	0	0	0,67	0,67	0,67	0,67	0
bbbb	0	0	0	0	0,67	0,67	0,67	0,67	0	0	0	0	0,67	0,67	0,67	0,67	0
bbbc	0	0	0	0	0,67	0,67	0,67	0,67	0	0	0	0	0,67	0,67	0,67	0,67	0
bbcd	0,67	0,67	0,67	0,67	0	1,34	1,34	1,34	0,67	0,67	0,67	0,67	0	1,34	1,34	1,34	0,67
bcbd	0,67	0,67	0,67	0,67	1,34	0	1,34	1,34	0,67	0,67	0,67	0,67	1,34	0	1,34	1,34	0,67
cdbb	0,67	0,67	0,67	0,67	1,34	1,34	0	1,34	0,67	0,67	0,67	0,67	1,34	1,34	0	1,34	0,67
dbbb	0,67	0,67	0,67	0,67	1,34	1,34	1,34	0	0,67	0,67	0,67	0,67	1,34	1,34	1,34	0	0,67
bbbb	0	0	0	0	0,67	0,67	0,67	0,67	0	0	0	0	0,67	0,67	0,67	0,67	0

Fonte: Nicoli et al (2013)r

A heurística *Random* trabalha com duas séries ordenadas aleatoriamente. É como se os elementos da série original dessem origem a duas séries (de tamanho igual a da original) com seus elementos alocados ao acaso. No caso da heurística *Magic*, também existem duas séries: na primeira, denominada exterior, os segmentos são classificados em ordem descendente da distância do segmento com o vizinho mais próximo. Na segunda série, denominada interior, os segmentos são classificados em ordem ascendente da distância ao atual candidato. A primeira invocação da série exterior será executada até sua conclusão, a partir daí todas as subsequentes vão sendo abandonadas na primeira iteração. Já na heurística *Perverse*, as séries possuem as piores ordenações possíveis, também existindo a série exterior e interior. A ordenação da série exterior é ascendente e a da série interior é descendente.

No final da aplicação do HOT SAX, é obtida uma única resposta correspondendo à primeira ocorrência da maior distância encontrada, isto é, só é considerada a informação encontrada na primeira vez, desprezando as informações de mesma distância encontradas posteriormente.

## B 2 MÉTODO HOT SAX MODIFICADO

Neste item está apresentado, conforme Nicoli et al. (2013), o Método HOT SAX Modificado desenvolvido para atender a identificação em tempo real de um ponto candidato a anomalia em instrumentos civis de uma usina hidroelétrica.

Originalmente em (KEOGH; LIN; FU, 2005), o HOT SAX foi concebido para analisar uma única série temporal de valores. Em geral essa série está armazenada em arquivos sequenciais. Assim, as primeiras adaptações foram realizadas para comportarem várias séries temporais, cada uma correspondendo a um sensor diferente, que podem representar tipos diferentes de sensores.

Devido ao fato da coleta de medições ser realizada com uma periodicidade pré-estabelecida, foi necessária a manutenção de um conjunto de medições por sensor constantemente atualizado.

O HOT SAX, originalmente, sinaliza numa série temporal, como segmento incomum a primeira ocorrência da maior distância. Nesta aplicação o sistema computacional executa a coleta periódica de medições e realiza a análise de séries temporais que incorporam a todo instante uma nova leitura, dessa forma as medidas mais recentes são as mais importantes para a interpretação da atual série, tendo o método sido adaptado para considerar o processamento da medição mais recente para a mais antiga, correspondendo a uma inversão no uso da série temporal.

Com relação ao processamento do HOT SAX original, ele considera uma relação entre o tamanho da série temporal e o tamanho dos segmentos que dividirão essa série e que tem como resultado um número inteiro. Como novos dados serão constantemente acrescentados à série temporal original, essa relação deve ser mantida constantemente, independentemente do tamanho da série. Assim, foi realizada a inclusão de uma função que está adequada ao tamanho da série (desprezando alguns valores mais antigos) para manter a relação inteira.

Outra consideração é a diminuição da sensibilidade do algoritmo quanto à indicação de algo incomum, de forma a sinalizar somente mudanças relevantes de comportamento.

Também foi considerado manter no algoritmo adaptado um número de até vinte símbolos para o alfabeto, permitindo dessa forma um maior número de classes de distâncias entre os segmentos.

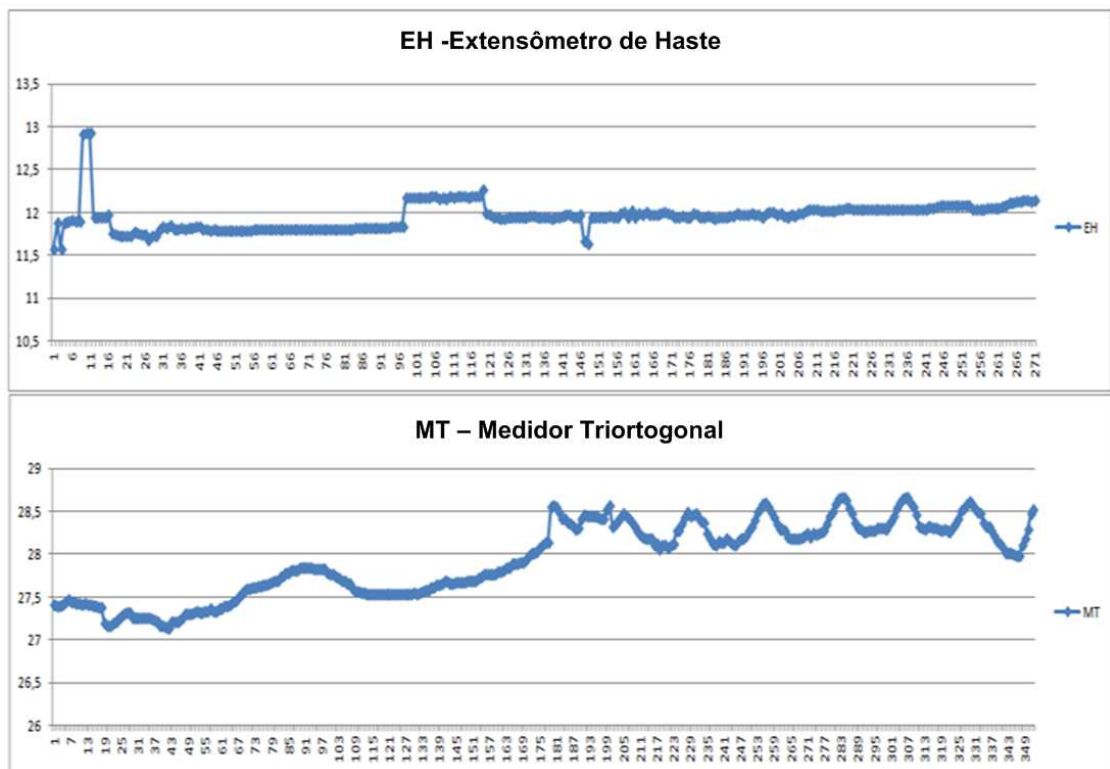
Com base nas considerações apresentadas, foram realizadas as alterações no algoritmo de forma a possibilitar sua aplicação em sistema de coleta periódica de dados, tais como os sistemas SCADA.

## B 2 1 Testes realizados

Os testes utilizam séries temporais reais dos seguintes tipos de sensores: Extensômetro de Haste (compreendendo as leituras de três hastes), Medidor Triortogonal (que correspondem às leituras de três eixos), Medidor de Vazão Triangular, e Piezômetro.

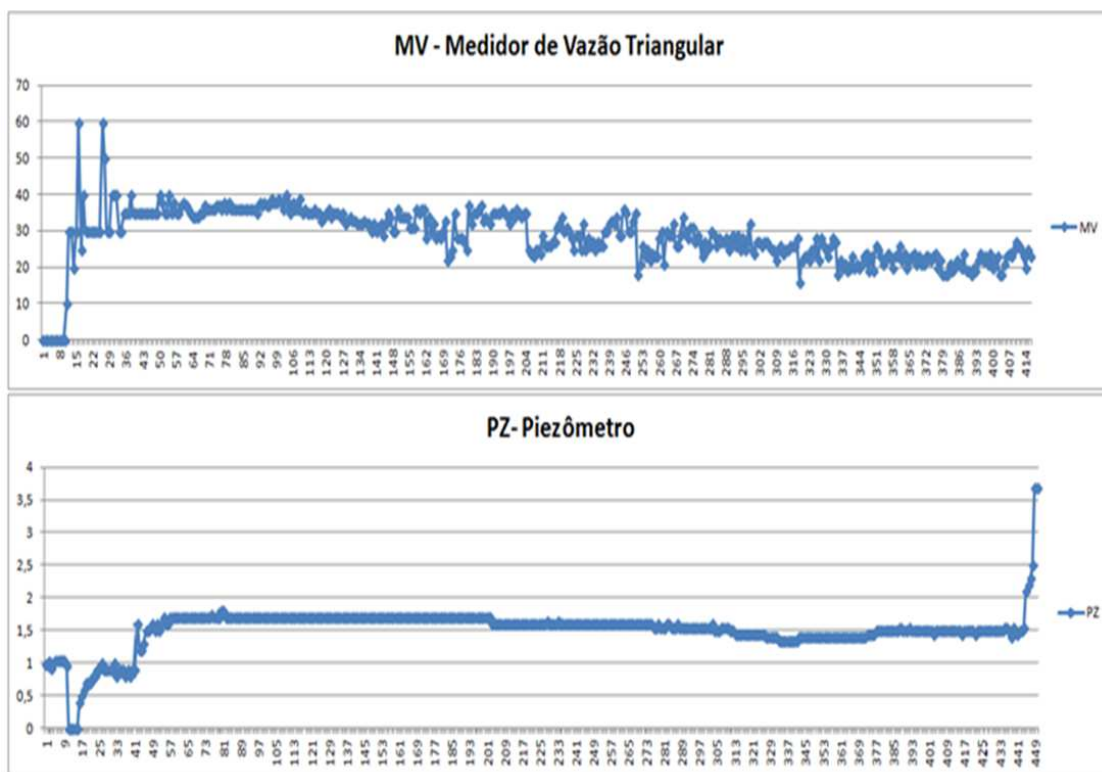
A partir das medidas reais, são traçadas curvas para cada sensor indicadas nas figuras 72 e 73, na qual o eixo das abscissas corresponde ao número de semanas e o eixo das ordenadas aos valores das leituras. Cabe ressaltar que as leituras estão relacionadas aos instrumentos Extensômetro de Haste EH, ao Medidor Triortogonal MT, ao Medidor de Vazão Triangular MV, e ao Piezômetro PZ, correspondendo respectivamente as seguintes quantidades de semanas: 271, 352, 416 e 450.

Figura 72- Medições referentes ao Extensômetro de Haste (em mm) e ao Medidor Triortogonal ( $10^{-2}$  mm).



Fonte: Nicoli et al (2013).

Figura 73 - Medições do Medidor de Vazões Triangular (em mm) e Piezômetro (em  $10^{-3}$  m)



Fonte: Nicoli et al (2013)

Nesses testes são considerados os seguintes parâmetros:

- O tamanho dos segmentos é de dois pontos;
- A janela móvel é composta por quatro pontos;
- Alfabeto com quatro símbolos.

O método adaptado do HOT SAX considera todas as medições existentes para todos os sensores e suas dimensões a partir do ponto mais recente para o mais antigo, a fim de obter pontos incomuns correspondentes à identificação do segmento com a maior distância com relação a outros segmentos. Dessa forma, nos instrumentos MT e MV são obtidos pontos incomuns distantes das leituras mais recentes, indicando que esses pontos não constituem motivos para o envio de mensagens, pois já foram sinalizados nos processamentos anteriores. Já no caso dos instrumentos EH e PZ a indicação de ponto

incomum recai sobre o primeiro segmento (de número zero), que contém as leituras mais recentes, apesar de existirem valores com desvios significativos no passado, como pode ser percebido nos gráficos correspondentes a esses sensores nas figuras 72 e 73.

A fim de elucidar melhor os resultados apresentados pelo método, na tabela 20 são apresentados quais segmentos possuem a maior distância entre si para cada sensor. Assim, a coluna “Segmento de referência” é o segmento a partir do qual foi calculada a distância com relação a outros segmentos e a coluna “Segmento da maior distância” é o segmento que corresponde à maior distância. No caso dos instrumentos EH, o segmento de referência corresponde ao 0 (que coincide com o segmento inicial) com a maior distância encontrada calculada para o segmento 42 e, para o instrumento PZ, está o segmento de referência 0 com o segmento 8. Por outro lado, o instrumento MT apresenta a relação entre os segmentos 7 e 91, e o instrumento MV, a relação entre os segmentos 12 e 102.

Desse modo, devido ao fato dos instrumentos EH e PZ terem como segmento de referência a indicação do segmento 0, a última medição realizada em cada instrumento deve ser considerada como candidata ao envio de mensagem indicando uma condição de não conformidade.

Tabela 20 - segmentos correspondentes a maior distância por sensor.

<b>SENSOR</b>	<b>Segmento de referência</b>	<b>Segmento da maior distância</b>
EH	0	42
MT	7	91
PZ	0	8
MV	12	102

Fonte: Nicoli et al (2013)

## B 2 2 Considerações sobre o método HOT SAX Modificado

As utilizações do HOT SAX Modificado para análise das medições coletadas em barragens hidroelétricas conduzem, devido às características da aplicação, às seguintes adaptações:

- Passa a ser um método não supervisionado;
- Redefinição da série temporal com a inclusão de novas medições periódicas;
- Mecanismo computacional capaz de filtrar os dados coletados de forma a manter a integridade da série temporal;
- Utilização de banco de dados, de forma a comportar várias séries temporais, vários sensores e tipos de sensores;
- Ajuste em tempo de execução do tamanho da série temporal, devido à mudança de tamanho com a incorporação de novas medições;
- Inversão no tratamento da série temporal, considerando as medições a partir da mais recente para a mais antiga;
- Possibilidade de generalização para a utilização em outros tipos de aplicações do sistema de aquisição de dados periódicos semelhantes.

As adaptações realizadas permitem a sua utilização em sistemas de monitoramento e supervisão comumente utilizados em *Smart Grids*, que envolvem uma quantidade significativa de sensores e de tipos de sensores e não implicaram em diminuição do desempenho na velocidade de processamento em relação ao HOT SAX original.

## B 3 MÉTODO DE BASHA E AMEEN

Este método é criação de (BASHA et al., 2007) e lida com séries temporais com periodicidade. Em sua estratégia de busca heurística, o método tira partido dos padrões repetitivos, da janela deslizante e de testes estatísticos. Também lida com os conceitos de periodicidade, normalização, distância euclidiana, o conceito do vizinho mais próximo.

Uma das justificativas desse método é a de que as periodicidades possibilitam procedimentos mais rápidos na busca heurística tirando partido dos padrões repetitivos, janelas deslizantes, cálculo da distância ao vizinho mais próximo e testes estatísticos.

Na figura 74, está representado o fluxograma básico do método. Nele estão representadas as principais etapas existentes no método. Note que o fluxograma está dividido em três blocos de processamento. No processo apresentado na figura 74, após o fornecimento da série temporal com periodicidades ocorre a aplicação do Bloco 1 com busca pela identificação de padrões básicos existentes na série com a identificação do tamanho dos padrões com a finalidade de identificar o tamanho de uma janela de deslizamento. No Bloco 2 a janela deslizante é aplicada para auxiliar no cálculo das distâncias entre uma subsequência e as suas vizinhas mais próximas. Após esse cálculo, no Bloco 3, são determinadas as subsequências candidatas a discórdias e a partir daí, em função do objetivo do pesquisador ou sistema computacional, são determinadas as discórdias através procedimentos estatísticos.

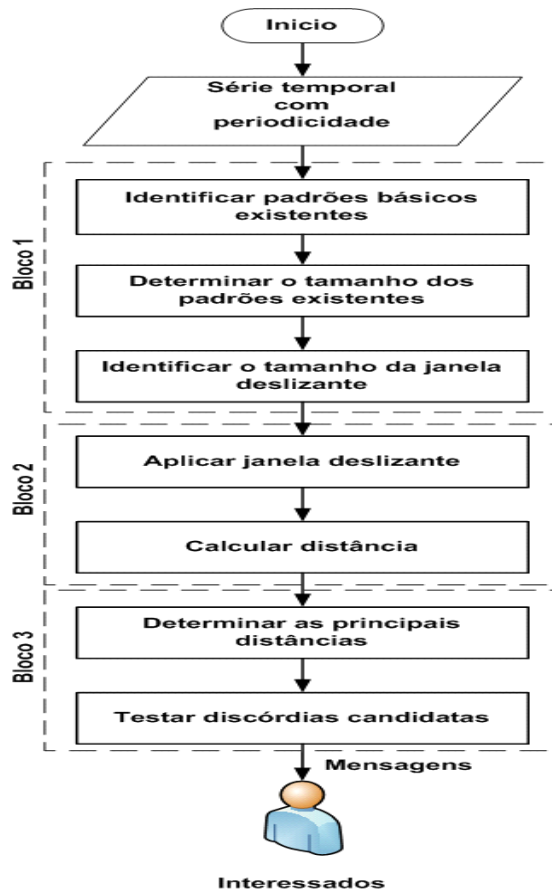
A seguir é aplicado o método considerando-se a seguinte série temporal (com dados obtidos em intervalos de tempos iguais indo de zero a 40): A nova série é: {0, 1, 0, 1, 0, 1, 0, 1, 0, 2, 0, 1, 0, 1, 0, 5, 0, 1, 0, 1, 0, 1, 0, 1, 0, 2, 0, 1, 2, 2, 7, 7, 1, 1, 0, 1, 1, 1, 1, 1}. A série está apresentada no gráfico da figura 73.

Para efeito de estudo, considera-se a existência de uma periodicidade de leitura de dados, contendo 8 registros cada, distribuídos regularmente no eixo do tempo, o que na prática conduz para a existência de cinco grupos no exemplo.

Na figura 76 estão traçados os dados após uma normalização para cada grupo, considerando-se média igual a 0 e variância igual a 1. No gráfico, estão traçados cinco séries, sendo que a série 1 corresponde às 8 primeiras leituras da série original e, a série 5, às 8 últimas leituras.

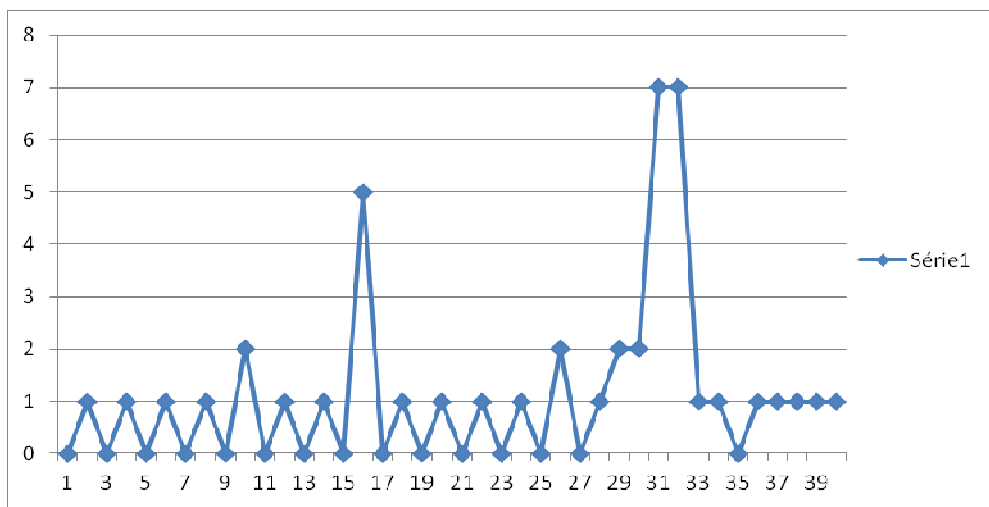


Figura 74 – Fluxograma básico do método de Basha e Ameen.



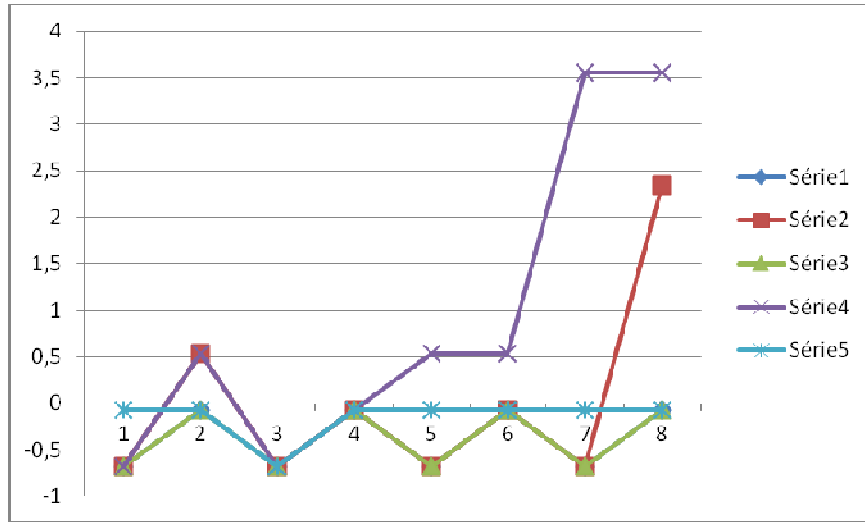
Fonte: Autor

Figura 75 - Dados do exemplo, considerando uma série experimental (valores adimensionais).



Fonte: Autor

Figura 76 - Dados do exemplo normalizados para cada grupo (valores adimensionais)



Fonte: Autor

De posse das cinco subsequências, foi calculada a distância Euclidiana entre elas, gerando a tabela figura 77. Observa-se que a numeração das colunas e linhas corresponde às respectivas sequências. Na tabela, existe uma região que está sem valores, pois é simétrica à região com valores.

Figura 77 - Resultados do cálculo da distância Euclidiana

Cálculo da Distância Euclidiana					
Sequência	1	2	3	4	5
1	0	2,49654046	0	5,776101865	1,04875676
2		0	2,49654046	4,611350616	2,70787832
3			0	5,776101865	1,04875676
4				0	5,27862662
5					0

Fonte: Autor

Ao analisar a tabela existente na figura 77, percebe-se que as relações da série 4 com as outras séries apresentam valores de distância maiores, indicando que ela deve ser considerada em uma possível análise estatística.

No caso da existência de muitas subsequências, várias delas poderão ser consideradas e assim a confirmação delas como discrepâncias serão obtidas através de testes estatísticos como o t-teste e o Qui-quadrado.

### **B 3 1 Considerações sobre o método Basha e Ameen**

No caso do método de (BASHA; AMEEN, 2007), ele é um método que necessita da intervenção posterior do operador para decidir alguns procedimentos, como por exemplo, a escolha de um método estatístico em função da aplicação. Dessa forma, em seu estado atual, ele é supervisionado, o que não é muito vantajoso em procedimentos automatizados e simulados automaticamente. Para estar adequado a um trabalho deste tipo, sua análise deveria ocorrer a toda atualização de leituras, além de incluir decisões automáticas, o que pode restringir suas aplicações.

## **B 4 ANÁLISE SINTETIZADA DE ALGUNS MÉTODOS INVESTIGADOS**

Este item apresenta, de forma sintetizada, análises com relação a outros métodos, inclusive os existentes no item anterior. Para cada método, são apresentados: seu objetivo, suas características principais, a base conceitual, a indicação de ser ou não adaptativo, quais algoritmos e/ou heurísticas utilizadas, e comentários. Em alguns deles foram realizados testes para a obtenção de pontos localizados em uma situação extrema.

#### **B 4 1 Método HOT SAX**

Apresentado em (KEOGH; LIN; FU, 2005). Está analisado em item acima inclusive com adaptação para rodar em tempo real. Tem em seu objetivo obter discórdias e anomalias, sendo que entre suas características principais tem uma abordagem estatística e possui boa velocidade em seu processamento computacional. Está conceitualmente baseado no PAA (*Piecewise Aggregate Approximation*) e SAX (*Symbolic Aggregate ApproXimation*), obtendo uma discretização (com conseqüente diminuição do tamanho da série temporal, compactação) e uso da distância euclidiana. Usa também janela deslizante (com passo variável, mas um é o ideal), cria tabela indexada e árvore de busca com folhas contendo os índices da tabela. Não é adaptativo e só encontra uma discórdia por vez. Em seu processamento utiliza o algoritmos da força bruta, além de otimizar a poda (descarte de itens); essa poda tem uma consideração maior as heurísticas utilizadas: randômica, mágica, perversa.

Nos testes realizados com módulo SAX os resultados são satisfatórios, mas é necessária complementação com a evolução do SAX que é o próprio HOT SAX, com a identificação da anomalia na série temporal original, além de definir o alfabeto e o tamanho da palavra para as palavras discretizadas.

#### **B 4 2 Método WAT**

Esse método é apresentado por em (BU et al. 2011). Seu objetivo é obter discórdias e anomalias e, entre suas características principais está o uso da estatística (como herança do uso do HOT SAX) e a utilização de *Wavelet*. Conceitualmente, está baseado no uso do SAX e do algoritmo base do HOT SAX, incluindo uma heurística com a transformada *Wavelet* Haar e utilizando também uma árvore de prefixos (*trie*), e a correlação de Pearson (LP). É considerado adaptativo pela forma como encontra as discórdias ditas top-k, além de ter uma abordagem dinâmica adaptativa para obter o tamanho da palavra adequada para conjuntos de dados diferentes. Adaptável a diferentes aplicações sem as tarefas de

configuração de parâmetros intuitivos dos usuários. Seu algoritmo é denominado de *Find\_Unusual* e tem como base o HOT SAX, incluindo a transformada Haar.

Os testes não apresentam respostas adequadas. O algoritmo identificado na internet é um pouco diferente do artigo. Em ambos: o artigo e o algoritmo da internet, existem informações incompletas e dúbias. O WAT automatiza a seleção de tamanho de palavra, mas ainda é preciso definir o tamanho do alfabeto. Usa ordem fixa / aleatória no laço externo para pesquisar todas as discórdias top-K. Também não está claro como o conhecimento adquirido em encontrar a discórdia no k-ésimo pode ajudar a encontrar a discórdia (k +1)-ésimo.

#### **B 4 3 Método de Fu pré WAT**

Apresentado em (FU et al., 2006), com o objetivo de obter discórdias e anomalias, também, entre suas características principais, estão o uso da estatística (segundo o HOT SAX) e a utilização de *Wavelet*. Baseado no SAX, utiliza o mesmo algoritmo de base do HOT SAX, e algoritmo semelhante ao WAT apresentado em artigo posterior. Opera como heurística a Transformada de Haar (com suas médias e diferenças) tanto nas repetições (*looping*) externa como interna. Não é adaptativo e só encontra uma discórdia por vez. O algoritmo obtido está em C para a transformada de Haar é idêntico ao apresentado no artigo e utiliza o núcleo de base do HOT SAX.

Existe uma aparente contradição com o exemplo dado para a resolução manual da transformada de Haar e o existente no algoritmo. Também é proposto um algoritmo que pode determinar dinamicamente o tamanho da palavra para a compressão. Na realidade, usa a propriedade do *Wavelet* de Haar para ajustar dinamicamente o comprimento da palavra de acordo com as características dos dados.

#### **B 4 4 Método de Toshniwal e Yadav**

Método apresentado em (TOSHNIWAL; YADAV, 2011) com o título “Detecção adaptativa de *Outliers* em *streaming* de séries temporais”. Seu objetivo é obter *Outliers*. Entre suas características principais, está o uso de estatística (herança do HOT SAX, além de: quartis, e *Outliers*). Tem como base o uso de *Buffer* de dados, o que requer apenas um parâmetro que é o comprimento da subsequência de discórdia. Utiliza o HOT SAX estendido para encontrar discrepâncias locais nos fluxos locais. A distribuição das discordâncias considera os quartis. Realiza a criação de regras com bases nos quartis, sendo adaptativo. Em seu algoritmo aplica o HOT SAX para obter discórdias locais, em geral, a sequência local é do tamanho do Buffer utilizado. Por fim, realiza uma busca para a obtenção dos quartis nas discórdias locais encontradas.

Nesse método o algoritmo HOT SAX foi estendido, também foi introduzido o conceito de "*Outlierness*" utilizado para capturar o grau de comportamento anormal mostrado pelos valores aberrantes. Outro conceito criado é o "*Type*" de discrepância que define o desvio das discrepâncias acima ou abaixo do comportamento normal. Foram realizados testes e obtidos os *Outliers* que excluem as respostas esperadas. Em outras palavras, o critério para obter o *Outliers* não permite que os resultados esperados correspondam a anomalias (ou discórdias).

#### **B 4 5 Método de Knorn e Leith**

O método apresentado em (KNORN; LEITH, 2008), cujo objetivo é obter anomalia utilizando sua característica estatística no uso do filtro de Kalman. Tem por base, além do filtro de Kalman, o uso dos modelos: *Form-Free*, orientado a evento, e Geral (mistura dos dois anteriores), além de um modelo estatístico (em vez da série temporal pura). Os modelos auxiliares tem a finalidade preditiva. Este método é adaptativo. O artigo não apresenta o algoritmo que também não é encontrado em outras fontes, mas foram estabelecidas as sequências utilizadas no método.

Utilizado para produtos de missão crítica, mais especificamente onde a disponibilidade e confiabilidade são características importantes como em *software* de aparelhos, *firewalls*, servidores web, etc. Com o Filtro de Kalman implantado e em execução, pode-se simplesmente declarar uma anomalia quando uma nova medição está fora do intervalo de confiança em torno do seu valor.

#### **B 4 6 Método K-Means + ID3**

Em (GADDAM; PHOHA; BALAGANI, 2007), está apresentado um método que tem por objetivo identificar anomalias e realizar aprendizado. Em suas características principais, estão a mineração de dados (utilizando o K-means) e a Inteligência artificial (com árvore de decisões). Este método está baseado em K-means, e Árvore de decisões (com o ID3), caracterizando-se como um método em cascata. Utiliza os conceitos da regra do vizinho mais próximo e da regra do consenso mais próximo. Não é um método adaptativo. Utiliza o algoritmo do ID3, sendo que o texto apresenta um pseudocódigo para detecção de anomalia com agrupamento k-means. Identifica a detecção de anomalia com árvore de decisão ID3; em transparência do artigo, está o pseudocódigo para a seleção da anomalia candidata. Também realiza aprendizado.

Esse método é patenteado, assim faltam algumas informações básicas para a sua reprodução e teste.

#### **B 4 7 Método de Mahoney e Chan**

O método de (MAHONEY; CHAN, 2004) tem por objetivo obter anomalias. O método é matemático com a geração de regras. Não é adaptativo. Apresenta em seu artigo um pseudocódigo para regras de aproximação e de um modelo sequencial. É um Método multidimensional, envolvendo valor, inclinação e curvatura. Representa um caminho (tridimensional).

Nesse método, cada ponto de teste deve ser comparado com todos os pontos de treinamento. Cada caixa é um conjunto de regras delimitado por  $y$  (valor),  $dy$  (inclinação) e  $dy^2$  (curvatura). É um método multidimensional de detecção de anomalias em séries temporais, em que para cada ponto da série de testes é necessário a coincidência do conteúdo (valor), da inclinação e da curvatura de um ponto de vista em treinamento (com uma restrição sequencial opcional). Possui um método de geração de um modelo na forma de regras concisas, compreensível para humanos e que pode ser editado. O método generaliza para múltiplas series de treinamento, permitindo que os pontos de teste possam estar dentro de uma faixa limitada pelos dados de treinamento.

#### **B 4 8 Método de Koushanfa e Sangiovanni-Vincentelli**

O método, segundo Koushanfa; Potkonjak; Sangiovanni-Vincentelli (2003), tem por objetivo realizar testes on-line em redes de sensores e obter Falhas, utilizando principalmente estatística (não paramétrica). Para tanto, considera a realização da otimização a partir do método de minimização não linear da função Powell, também é capaz de realizar autoteste e considera a taxonomia de modelos de falha. A técnica estatística não paramétrica é baseada em testes on-line dos nós em uma rede de sensores. Não é um método adaptativo.

Na detecção de falhas on-line em redes de sensores, é introduzida uma taxonomia para classificação dessas falhas e a primeira técnica é baseada em modelo de teste on-line. São aplicados métodos estatísticos não paramétricos para identificar os sensores que têm a maior probabilidade de serem defeituosos. Testes no artigo utilizam a introdução de ruídos aleatórios. O objetivo principal da avaliação é analisar o impacto da percentagem de sensores defeituosos, a relação sinal-ruído, bem como a duração do teste sobre as detenções falsas de falha de erro positivo e negativo. Não foram realizados testes nesta tese, mas é um método que completaria as considerações de teste em sensores on-line.



#### **B 4 9 Método de Basha e Ammeen**

O método de (BASHA; AMEEN, 2007) está em tópico anterior. Seu objetivo são Subseqüências não usuais, discórdias em séries com periodicidade. Para tanto utiliza a estatística (teste-t, ki-quadrado). Não constitui um método adaptativo. Baseado na identificação de padrões, utiliza janela deslizante (coincidindo com o tamanho do padrão identificado), distância Euclidiana, divergência de Kullback-Leibler, função de autocorrelação, teste-t, Ki-quadrado. É um método em cascata, após identificar, seqüências consideradas de não usuais, verifica quais delas correspondem a discórdias.

#### **B 4 10 Detecção de Novidades com aplicação a fluxos contínuos de dados**

Segundo Spinosa; Carvalho; Gama (2008), o objetivo do seu trabalho é identificar de maneira não supervisionada o aparecimento de novos conceitos em dados que podem ser apresentados sob a forma de um fluxo contínuo, ocorrendo de forma continua à aprendizagem de máquina. Para tal, considera a detecção de novidade - DN - como a capacidade de identificar novos conceitos à medida que, dados adicionais são recebidos. (SPINOSA; CARVALHO; GAMA, 2008) relaciona que, dentre as aplicações de DN, estão: detecção de falhas em máquinas, novos tópicos em textos, invasões em redes de computadores, regiões de interesse em imagens médicas, mudanças em um ambiente no uso de robôs, estouros (*bursts*) em séries temporais, novos perfis de expressão gênica, dentre outras.

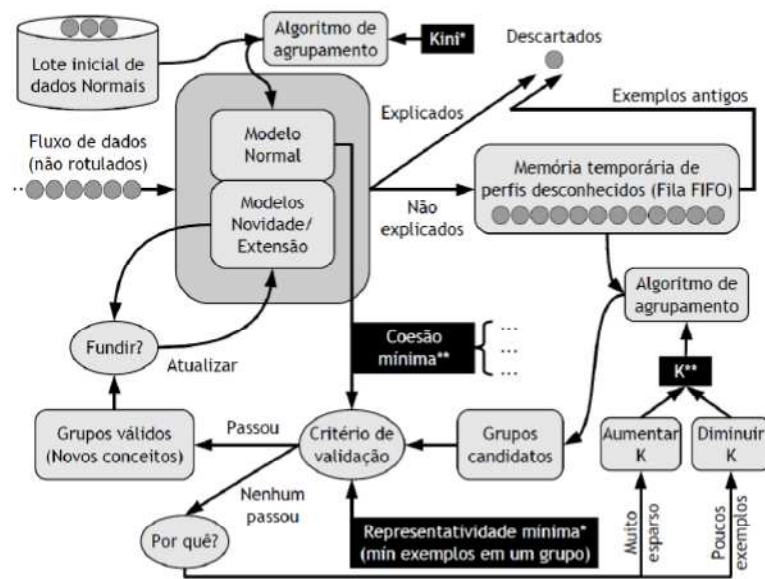
Para Spinosa; Carvalho; Gama (2008) a detecção de anomalia também é um termo utilizado com o mesmo sentido DN, normalmente fazendo referência a uma aplicação em que a novidade representa uma condição indesejável no domínio que está sendo analisado. Em Spinosa; Carvalho; Gama (2008), os *Outliers* são entendidos como exemplos isolados, não representativos, distribuídos de maneira esparsa, que, portanto, não compartilham características com os demais. Ainda com relação a Novidades e *Outliers*, em Spinosa; Carvalho; Gama (2008) é ressaltado que os *Outliers* podem ser vistos como elementos

contaminantes, que interferem negativamente na representação de uma população. Novidades, por outro lado, induzem naturalmente à ideia de novos conceitos.

Em Spinosa; Carvalho; Gama (2008) é comentado com relação à DN que ela é tratada como um problema de Classificação com 1 Classe (C1C), em Inglês *one class classification* ou *single class classification*. A C1C, caracteriza-se pelo aprendizado de um único conceito-alvo a partir apenas de exemplos desse conceito. O maior desafio da C1C está em obter o grau adequado de generalização na descrição do conceito alvo. Isso ocorre porque a ausência de contra exemplos dificulta o posicionamento das fronteiras de decisão. Um Classificador do tipo C1C é, portanto, capaz de identificar exemplos que não pertencem ao conceito alvo que foi aprendido. Por esse motivo, o problema de DN é frequentemente tratado ou referido na literatura como um problema de C1C.

O sistema computacional desenvolvido em (SPINOSA; CARVALHO; GAMA, 2008) é denominado OLINDDA apresentado na figura 78 Nele estão incluídos: o fluxo contínuo de dados, agrupamento K-Means, conceito de modelo normal e de extensão do modelo normal, conceito de modelo novidade (diferente do normal), área de trabalho (memória temporária), validação utilizando critérios de coesão (densidade, distância, dentre outros), representatividade, fusão de dois ou mais conceitos, e descarte.

Figura 78 - Esquema do Sistema computacional OLINDDA



Fonte: Spinosa; Carvalho; Gama (2008)

## APÊNDICE C - CONFIGURAÇÃO DE *HARDWARE* E *SOFTWARE* UTILIZADOS

Neste apêndice estão informações sobre as configurações utilizadas no ambiente de desenvolvimento para as provas de conceitos e protótipos dos métodos desenvolvidos.

### C 1 HARDWARE

Constituído por um computador Notebook com:

- Windows 7 Home Premium;
- Processador Intel ® Core™ i5-2410M CPU @ 2.30 GHz;
- Modelo Lenovo Win7 PC;
- Memória instalada (RAM) de 4,00 GB;
- Com um computador sobrecarregado em aplicativos e em memória de disco rígido de 500 GB.

### C 2 SOFTWARE

Composto por:

- Sistema operacional de 64 Bits;
- Com Windows 7 Home Premium;
- Linguagem VBA (*Visual Basic for Applications* ) existente no gerenciador de Banco de Dados;
- Gerenciador de Banco de Dados Microsoft Access 2007;
- Banco de Dados Access 2007;
- SQL (*Structured Query Language*) padrão;
- e modelo.

O modelo de dados desenvolvido permite a portabilidade para outros gerenciadores de banco de dados existentes no mercado.

Os algoritmos desenvolvidos possibilitam a sua utilização em várias linguagens e proporcionam considerações para linguagens voltadas a objeto.

Os testes realizados consideram a simulação das entradas de dados desde o início da série histórica até um instante determinado, permitindo analisar o conjunto de resultados em vários momentos, de forma a validar os modelos como um todo.