UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

A Cognitive Diagnosis Approach for Recommending Items Based on Polytomous Responses and Latent Attributes

Fernanda Tostes Marana

Dissertação de Mestrado do Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional (PPG-CCMC)

ICMC USP
SÃO CARLOS

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: ______________________

Fernanda Tostes Marana

A Cognitive Diagnosis Approach for Recommending Items Based on Polytomous Responses and Latent Attributes

Dissertation submitted to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP – in accordance with the requirements of the Computer and Mathematical Sciences Graduate Program, for the degree of Master in Science. *EXAMINATION BOARD PRESENTATION COPY*

Concentration Area: Computer Science and Computational Mathematics

Advisor: Profa. Dra. Mariana Curi
Co-advisor: Prof. Dr. André Carlos Ponce de Leon Ferreira de Carvalho

USP – São Carlos
April 2023

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

Tc

Tostes Marana, Fernanda
A Cognitive Diagnosis Approach for Recommending Items Based on Polytomous Responses and Latent Attributes / Fernanda Tostes Marana; orientadora Mariana Cúri; coorientador André Carlos Ponce de Leon Ferreira de Carvalho. -- São Carlos, 2023.
55 p.

Dissertação (Mestrado - Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional) -- Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2023.

1. Sistemas de Recomendação . 2. Modelos de Diagnóstico Cognitivo. 3. Atributos Latentes. 4. Respostas Politômicas. I. Cúri, Mariana , orient. II. Carlos Ponce de Leon Ferreira de Carvalho, André, coorient. III. Título.

Bibliotecários responsáveis pela estrutura de catalogação da publicação de acordo com a AACR2:
Gláucia Maria Saia Cristianini - CRB - 8/4938
Juliana de Souza Moraes - CRB - 8/6176

Fernanda Tostes Marana

Uma Abordagem de Diagnóstico Cognitivo para Recomendar Itens com Base em Respostas Politômicas e Atributos Latentes

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestra em Ciências – Ciências de Computação e Matemática Computacional. *EXEMPLAR DE DEFESA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientadora: Profa. Dra. Mariana Curi
Coorientador: Prof. Dr. André Carlos Ponce de Leon Ferreira de Carvalho

USP – São Carlos
Abril de 2023

This study is dedicated to my beloved parents Denise and Nilceu, who have been a source of inspiration and strength through all my life. They provided their moral, spiritual, emotional support and taught me the value of hard work.

I also dedicate this project to my sister Carolina who never left my side and has always been my best friend.

Lastly, I dedicated this study to my family, friends and my advisors, who contributed immensely with their words to encourage me in finishing this project.

ACKNOWLEDGEMENTS

I would first like to express my sincere gratitude to my advisor Professor Mariana Curi from University of São Paulo. Her patience, enthusiasm and spectacular guidance helped me in overcoming numerous obstacles I have been facing through the project. She consistently allowed this project to be my own work, but steered me in the right the direction whenever she thought I needed it.

I would also like to thank Professor André Carlos P. L. F. de Carvalho for his encouragement, his guidance and insightful comments that helped me to target the main points on my study. His knowledge on topics of Recommendation Systems was invaluable in the formulating of the research topic and methodology in particular.

I would also like to thank my colleagues Lohan Rodrigues and Renato Fernandes for the help at creating and testing my proposed recommendation system.

I am also indebted to CAPES program for their financial support, as I could never do this research on my own. The opportunity of this Masters program has helped grow a lot and will definetely have a great impact on my personal and professional life.

I would also like to acknowledge the Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo for providing the knowledge I needed to acquire this achievement. The experience I have in Computer Science today would not be possible without your support.

Finally, I must express my very profound gratitude to my parents Denise and Nilceu, and my sister Carolina for provinding me with continuous support and encouragment through my process of research and my personal experience. To my family, thank you for your love and belief in me.

RESUMO

MARANA, F. T. **Uma Abordagem de Diagnóstico Cognitivo para Recomendar Itens com Base em Respostas Politômicas e Atributos Latentes**. 2023. 55 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2023.

Sistemas de Recomendação tornaram-se predominantes nos últimos anos, atraindo a atenção de pesquisadores para investigar diferentes métodos de filtragem de informações relevantes para os usuários. Estas informações nem sempre são explícitas e diferentes propostas surgiram para obter os valores latentes de indivíduos por meio de seu comportamento. Nas áreas educacionais, os atributos latentes de estudantes podem ser obtidos por modelos psicométricos como o Modelo de Diagnóstico Cognitivo. Esses modelos tentam criar um perfil de usuário para explorar as conexões entre alunos e disciplinas, assim como um sistema de recomendação faz com seus usuários e os produtos a serem recomendados. O objetivo deste trabalho é desenvolver uma nova abordagem em sistemas de recomendação que incorpore Modelos de Diagnóstico Cognitivo aplicados a dados de mídias definidas por conteúdos discretos (como gêneros em filmes e séries) para gerar respostas politômicas na forma de previsões de notas que um usuário daria a um item. A abordagem proposta foi aplicada a dois conjuntos de dados (MovieLens 20M Dataset and Anime Recommendation Database) e, devido à esparsidade de dados, obtidos em alguns casos resultados melhores do que um método clássico de recomendação de filtragem baseada em conteúdo. Em seguida, o sistema de recomendação com a abordagem proposta por este projeto foi integrada junto com um modelo de recomendação clássico e o sistema de recomendação híbrido criado obteve alguns resultados melhores quando comparados com os adquiridos pelos sistemas individuais. Por fim, este trabalho também explorou o desempenho dos modelos no ranqueamento de itens a serem recomendados aos usuários. Alguns pontos interessantes foram observados e o modelo proposto teve o melhor desempenho mesmo comparado ao modelo híbrido.

Palavras-chave: Sistemas de Recomendação, Modelos de Diagnóstico Cognitivo, Modelo Híbrido, Respostas Politômicas, Atributos Latentes.

ABSTRACT

MARANA, F. T. **A Cognitive Diagnosis Approach for Recommending Items Based on Polytomous Responses and Latent Attributes**. 2023. 55 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2023.

Recommendation Systems have become prevalent in recent years, attracting the attention of researchers to investigate different methods to filter relevant information for users. This information is not always explicit and different proposals have emerged to obtain the latent values of individuals through their behavior. In educational areas, latent attributes of test-takers can be acquired by psychometric models such as the Cognitive Diagnostic Model. These models attempt to create a user's profile in order to explore the connections between students and subjects, just like a recommendation system does with its users and the products to be recommended. The objective of this work is to develop a new recommendation approach that incorporates Cognitive Diagnostic Models applied to data from media defined by discrete content (such as genres in movies and series) in order to generate its polytomous response in the form of the rating prediction that a user would give to each item. The proposed approach was applied to two datasets (MovieLens 20M Dataset and Anime Recommendation Database) and, due to the sparsity of the data, obtained in some cases better results than a classic content-based filtering recommendation method. Then, our new recommendation approach was fused to the classic recommendation model and this hybrid recommendation system obtained some results that were better when compared with the ones acquired by the individual systems. Finally, this work also explored the performance of the models in ranking items to be recommended for the users. Some interesting points were observed and the proposed model had the best performance even compared to the hybrid model.

Keywords: Recommendation System, Cognitive Diagnostic Model, Hybrid Model, Polytomous Response, Latent Attributes.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

CHAPTER 1

INTRODUCTION

Due to the popularization of the internet and its drastic annual growth around the world, huge amounts of data and information have been created daily. This exorbitant amount of unbalanced information develops the need to organize and separate the relevant and reliable content to generate meaningful search results.

In the field of electronic commerce (or e-commerce), the problem of large datasets stands out in the numerous offers of similar products, turning the customers' choice into a great challenge. Since it's important for companies, big or small, to understand their client's needs and desires in order to create marketing personas and provide services and products of their interest (HISANABE, 2009), many types of filtering information techniques have been gradually gaining popularity in various applications. Some of these techniques are known as Recommendation Systems.

Recommendation Systems are used as information filters that attempt to reduce the difficulty of choosing items among an immense set of options by proposing products that attract and satisfy the user's needs and interests. A recommendation system builds a user's profile based on their past records and extracts similar characteristics of different users and items in order to suggest newer products. Most of the approaches of these systems use the products (which can be books, movies, and services, among others) and the users' ratings of past products in an effort to create a list in ascending order filled with products that have a great chance to attract the user's attention (RICCI; ROKACH; SHAPIRA, 2010).

Although the traditional recommendation systems apply the user-item historical records in order to evaluate the user's preferences, many recent articles propose additional information in their implementation to increase performance and alleviate problems such as the cold-start effect and the matrix sparsity. While the cold-start effect occurs when the system has very few amounts of information from some users, the sparsity problem comes from the common fact that most users don't rate the majority of the products offered, leaving the matrix of ratings with a lot of empty values. All of these missing values in the dataset provide a challenge to recommend products, as there is not sufficient information about the user to help infer their preferences.

There are two types of additional information that can be collected: explicit and implicit feedback data(LIU *et al.*, 2010). Explicit feedback data are quantifiable. By using the ratings of the user or their comments about the products, this type of data takes into consideration the input from the user about how they liked or disliked a product. The downside of the explicit data is that it's very limited and doesn't take into consideration the context of the situation. As for the implicit feedback data, this type doesn't directly reflect the interest of the user but it can help to infer the user's personal preferences. Some examples of it are the count of clicks on a link page, browsing history, how many times you have watched a movie or listened to a song.

Some projects acquire the implicit data for the purpose of obtaining the user latent preferences, an estimation of the user latent interest from ratings. This supplementary information can alleviate data sparsity and enhance recommendation accuracy because it enriches the aspects of the items in an effort to understand what attributes are stronger in each user's preference.

Many approaches have explored the area of cognitive modelling in order to bridge recommendation systems and psychological characteristics, helping companies that rely heavily on users' aspects, such as personality, behaviour, and attitude, to suggest better products for their clients. Gonzalez-Carrasco et al. (GONZALEZ-CARRASCO *et al.*, 2012), for instance, proposed a multi-investment Recommendation System, PB-ADVISOR, to obtain the semantic features of the investments by using semantic and fuzzy logic in order to support private bankers with decision-making on complex investment processes. Chen et al.(CHEN; DUH; LIU, 2004) developed a personalized courseware recommendation system with the application of fuzzy item response theory to suggest courseware with an appropriate difficulty level to each student. Egglestone et al. (EGGLESTONE *et al.*, 2010) proposed that visitors of thematic parks were categorized on different orthogonal personality dimensions according to two psychometric measures known as Big Five and Sensation Seeking Scale as a means to predict the ride experience for each visitor. Hu et al. (HU *et al.*, 2011) proposed that users' latent preferences were obtained by computing psychometric models such as Rasch model to enhance the representation of user preferences and, therefore, the recommendation's accuracy.

In the psychological and educational fields, latent factors such as user latent preferences have been applied with similar objectives but in different executions. Parallel to recommendation systems aiming to identify multiple fine-grained characteristics of their users, statistic models in education and mental health aim to understand human behaviour and performance during specialized tests. The Cognitive Diagnostic Model (CDM) (SVETINA, 2011), a discrete latent variable model, analyzes students' and patients' patterns and estimates through them a level of proficiency in different objects or the presence of distinct psychological conditions (TORRE, 2009a). This relationship between the items present in the educational or psychological assessments and the competencies is registered by a binary matrix known as Q matrix with dimensions

$J \times K$, where $q_{jk} = 1$ indicates the required proficiency or condition from item j in attribute k, and $q_{jk} = 0$ indicates the opposite. Bradshaw et al. (BRADSHAW *et al.*, 2013) use CDM to evaluate the attribute mastery of middle-grade students through a test grounded in research on cognition. In a psychological study, de la Torre et. al. (TORRE; ARK; ROSSI, 2018) apply CDM to provide a diagnostic of psychological disorders by capturing the interactions among the disorders of each patient. In fact, even similar problems from recommendation systems, such as missing data, appear in these latent variable models. These problems are removed or dealt with by different techniques like the genetic algorithms proposed by Ordóñez Galán et al.(Ordóñez Galán *et al.*, 2017).

Therefore, both Recommendation and Cognitive Diagnostic models aid the decision-making process in their respective contexts. Professors can define ways to improve student learning by focusing on areas that are lacking (BRADSHAW *et al.*, 2013). Recommendation Systems can have more distinguished and personalized recommendations for their users. Both approaches aim to find information about the individuals in order to achieve accurate results and, in the end, indicate a possibility to bridge these different fields into a single model.

The main objective of this work is to explore the possibility of incorporating Cognitive Diagnostic models into Recommendation Systems. In other words, the idea is to apply the CDM in recommender databases that have ratings and additional information about their items in order to identify the level of interest a user has in a product. We hypothesize that, by incorporating cognitive attributes into the recommendation process, the proposed system can make more personalized and relevant recommendations that are better aligned with the user's preferences and decision-making processes. An additional objective is to integrate this proposed approach with another recommendation system in a hybrid model and evaluate its performance. Ultimately, our interest lies in exploring how the proposed recommendation approach and hybrid recommendation model perform compared to a classic content-based filtering in relation to the prediction of ratings and the ranking of items.

This study is divided into five chapters. In Chapter II, there will be an exploration of the background study for this project, including Recommendation Systems and Cognitive Diagnostic Models studies. Chapter III will discuss in detail the implementation of the proposed model as well as the database used, addressing, additionally, some issues encountered during the development process. Afterwards, Chapter IV will present the results obtained, bringing possible justifications. Finally, Chapter V will highlight the important points, the limitations found and possible improvements that can be done.

CHAPTER 2

BACKGROUND

This chapter presents the State of the Art regarding several research issues important to this Thesis: Recommender Systems (Section 2.1) and Cognitive Diagnostic Models (Section 2.2).

2.1 Recommendation Systems

With the shift towards the online economy, introduced by the rapid development of social networks and e-commerce platforms, a new phenomenon has appeared to be faced by sellers and buyers online: information overload (BOBADILLA *et al.*, 2013). Since every platform has more information than its users can consume, the excess of irrelevant choices decreases the user experience rates and makes it more difficult to obtain the desired products, information, and services. Therefore, strategies like Recommendation Systems (Knotzer, 2018) aim to understand how items are related to one another by analyzing their patterns and recommending the ones that should interest each user in particular based on their preferences or past behaviour.

By evaluating the types of data available in the database, different strategies can be planned to build the user's profile and, subsequently, recommend items of their likeness. One widely-known challenge faced in recommendation systems is the fact that items with few user interactions (also known as long-tail items) have very few chances of being recommended due to the popularity bias these systems tend to have (SREEPADA; PATRA, 2020). These long-tail items can include niche products, speciality items, or items that have a smaller market share but still have a loyal customer base and even though these items may not be as visible or promoted as heavily as the top-selling items they still can be profitable for businesses (OESTREICHER-SINGER; SUNDARARAJAN, 2012)(ELBERSE; OBERHOLZER-GEE, 2007).

The cold-start effect is an additional problem that can provide bad recommendations since it happens because newer users to the database have yet very few ratings and therefore a weak user profile.

Ratings are one of the most common ways of user feedback, though, in some cases, they can be too basic to help the recommendation system obtain good results. That's why it is important to differentiate the two types of feedback that can be collected as well as the advantages and disadvantages of each one (LIU *et al.*, 2010). User feedback can be divided in:

Explicit feedback: Requires the direct participation of the users as a means to obtain user preferences. It can be collected by functionalities such as a like/dislike button on a website, a rating system from one to five, or even a comment section where users can leave their personal opinions. Although this data is very transparent, it is not often easy to collect and can be very limited with sparsity problems;

Implicit feedback: Requires analyzing different aspects of the user's behaviour to obtain user preferences. Some examples of this data can be how many times a link was clicked or how long a video was watched by a user. A disadvantage of this type of data is that it only provides positive feedback, as there is no way to differentiate whether a user did not like an item or did not notice that item.

In an attempt to acquire better performances and accurate results, the literature presents a wide variety of proposed recommendation systems, from basic ideas such as always recommending the most popular items to complex designs like applying deep learning concepts. The most popular examples of recommendation systems remain as the three classic models: Collaborative Filtering, Content-Based Filtering and Hybrid Filtering.

2.1.0.1 Collaborative Filtering

Collaborative Filtering (CF) is the process of filtering information by exploiting relationships and recognizing patterns of interactions between users and items for the purpose of making recommendations of products and services that the user has not yet checked but will probably like.

It can be done in two ways: user-based and item-based. User-based collaborative filtering analyzes the behaviour of similar users to recommend items to a target user(CAI *et al.*, 2014). For example, if two users have rated or interacted positively with similar items in the past, then it is likely that they have similar preferences, and therefore, the system can recommend items that one user has liked to the other user. Item-based collaborative filtering, on the other hand, analyzes the similarities between items to recommend them to a user (HUANG; DAI, 2015). For example, if a user has liked a particular item in the past, then the system can recommend similar items to the user.

In this type of model, user's preferences are expressed as ratings provided by a collection of these users.

The data structure used in CF is known as the rating matrix R. The rows of this matrix store the users' ratings and the columns the items' ratings. As presented in Figure 1, this matrix is defined as R_{UI}, representing a set of users U, where $u_j \in U, j \in \{1, ..., N\}$, and a set of items I, where $i_k \in I, k \in \{1, ..., M\}$. If a user u_j has rated an item i_k, the corresponding value will be allocated in position R_{jk}.

	i_1	i_2	...	i_M
u_1	$r_{1,1}$	$r_{1,2}$	...	$r_{1,M}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
u_N	$r_{N,1}$	$r_{N,1}$	...	$r_{N,M}$

Figure 1 – The rating matrix R.

With the information provided by this rating matrix, it is possible to obtain the user similarities represented in a similarity matrix with $N \times N$ dimensions. Each cell of a similarity matrix shows the similarity between two users, calculated by metrics such as the Cosine Similarity, defined by Equation 2.1, where r_1 and r_2 are the ratings of two users. It's also noticeable that the similarity between items can also be calculated in a similarity matrix with $M \times M$ dimensions.

$$sim_{cossine}(r_1, r_2) = \frac{r_1 \cdot r_2}{\|r_1\|_2 \, \|r_2\|_2} \tag{2.1}$$

where $r_1 = (r_{11}, ..., r_{1M})$ and $r_2 = (r_{21}, ..., r_{2M})$.

One of the advantages of collaborative filtering is that it can recommend items that are not similar to the ones a user has already interacted with. For example, if a user has only interacted with items of a certain genre or category, collaborative filtering can recommend items that are not in that genre or category but that are similar to the ones the user has interacted with in terms of the preferences of similar users.

In addition, algorithms from collaborative filtering can be divided into two classes: memory-based and model-based (TRAN *et al.*, 2019). While memory-based types rely heavily on heuristics to measure similarities between users or items, as Nearest Neighbor algorithms (ELAHI; RICCI; RUBENS, 2016), model-based attempt to guess how much a user will like an item that they did not encounter before by inducing a model from a rating matrix, as Matrix Factorization algorithms (BOBADILLA *et al.*, 2011). While memory-based models are easy to implement and allow accommodating new data, in large datasets this type of collaborative filtering has scalability problems and its performance decreases. On the other hand, model-based collaborative filtering has better predictions and has solutions to the scalability problem, but systems of this type require a lot of time and memory to develop and present the problem of the loss of information when using dimensionality reduction (SU; KHOSHGOFTAAR, 2009).

2.1.0.2 Content Based Filtering

Content-Based Filtering (CBF) recommendation systems are also based on the similarity of recommended items, but by the content they have. In other words, the model attempts to learn the user's preferences in order to recommend items with aspects similar to those already evaluated, since, if a user likes a certain item, it is very possible that they will also like one with similar content.

Several works have used content-based filtering in recommendation systems. For example, in music recommendation systems, content-based filtering is used to recommend songs that are similar in genre (SOLEYMANI *et al.*, 2015), artist (BAUER; KHOLODYLO; STRAUSS, 2017), or even acoustic similarity of musical compositions (NIYAZOV; MIKHAILOVA; EGOROVA, 2021). In movie, it is used to recommend movies that are similar in genre or stylistic features (such lighting, color, and motion) (DELDJOO *et al.*, 2016) to the ones that a user has watched before. In e-commerce, it is used to recommend products that are similar in features or attributes to the ones that a user has purchased before (TAWFIQ; RAHMA; WAHAB, 2021).

This filtering method utilizes multiple properties of a domain that are used to describe an item in order to perform similarity calculations and recommend similar items, taking into consideration the initial ratings of the user as initial feedback (BOBADILLA *et al.*, 2013). The properties can have binary, nominal, or numerical attributes and can be obtained from the information about the users' preferences or the content of the items.

There are several ways to represent items and their properties. A representation method that is often used is the vector space model (SALTON; WONG; YANG, 1975), in which each row represents a different item and each column is a property of this item. With this structure, it is possible to apply heuristics between rows in order to build a similarity matrix $M \times M$. This matrix contrasts with the one in the Collaborative Filtering by obtaining its similarity values between items using the content they have related to themselves (for example, genres).

The heuristics used can be the Euclidean distance or Cosine similarity for numerical attributes or the Jaccard distance for binary ones. The Jaccard distance is present in Equation 2.2, where $i_1 = (i_{11},...,i_{1M})$ and $i_2 = (i_{21},...,i_{2M})$ are different vector space models representing the attributes present or not in two different items. So, for example, a movie i_{11} has the genres Adventure, Comedy and Mystery and a movie i_{21} has the genres, Mystery and Terror. If we are working with only these 4 genres then the attribute vector for the first movie will be $i_{11} = (1,1,1,0)$ and for the second movie $i_{21} = (0,0,1,1)$. The Jaccard distance will calculate how many genres will have in common to obtain the similarity between the two.

$$sim_{jaccard}(i_1,i_2) = \frac{|i_1 \cap i_2|}{|i_1 \cup i_2|} = \frac{|i_1 \cap i_2|}{|i_1| + |i_2| + |i_1 \cap i_2|} \tag{2.2}$$

These attributes can have different contents of an item. Usually, tags or descriptions are the simplest components that can be extracted to use in CBF systems and although tags are a more direct way to cluster items into different groups, Natural Language Processing methods can be used in descriptions to create feature vectors for each item (RENUKA; KIRAN; ROHIT, 2021) (BEUTEL *et al.*, 2018).

One of the main challenges this model presents is the overspecialization problem, which occurs when the system only recommends items that are too similar to the ones the user has already interacted with (B.THORAT; GOUDAR; BARVE, 2015). This can lead to a lack of diversity in the recommendations, and the user may miss out on discovering new items that are outside their comfort zone. Another challenge is the feature engineering problem, which occurs when the system relies on a limited set of features to describe the items, and some relevant features are missing or hard to extract (MOHAMED; KHAFAGY; IBRAHIM, 2019).

2.1.0.3 Hybrid Filtering

Although collaborative filtering and content filtering provide simple but interesting methods of suggesting items based on similarities between users or items, both still have some difficulties when dealing with the sparsity matrix from a recommendation system database. It is very rare for these datasets to have users rating the majority of items and, as a consequence, there is a lot of missing data present. Thus, it is common for the methods discussed so far to have problems when there aren't enough similar users or items that also have ratings. Some of these methods involve integrating different recommendation systems in a model known as Hybrid Filtering for the purpose of getting the best characteristics from each method (ÇANO, 2017).

This combination can be done in different phases of the recommendation processing. Either by collecting scores and predictions from different filtering strategies to be used to create a single result (a method known as Weighted) (SURIATI; DWIASTUTI; TULUS, 2017)(DO; LE; YOON, 2020) or by merging data from different strategies into a single recommendation algorithm (as it happens in the Feature Combination method)(ZANKER; JESSENITSCHNIG, 2009)(VALL *et al.*, 2019), experiments discuss the improved performance obtained compared to more simplistic approach. Additionally, these hybrid filtering methods don't apply exclusively to CF and CBF methods and don't always merge one technique with another (BURKE, 2002). Switching (GHAZANFAR; PRUGEL-BENNETT, 2010), for example, is a recommendation system that selects which filtering method will be used depending on the situation. Mixed also provides recommendations from several different recommenders without merging them. Cascade method (REBELO *et al.*, 2022) is done by one recommendation refining the recommendations done by a previous filtering method, and Metalevel (IMMANENI *et al.*, 2017) happens when the model learned by one strategy is used as input to another.

Although the hybrid models have several advantages by leveraging the strengths of different algorithms to overcome their limitations, they also present some challenges. They can be more complex to implement and maintain, and they require more computational resources. Additionally, hybrid models can suffer from the overfitting problem, where the system is too specialized to the training data and performs poorly on new data.

2.2 Cognitive Diagnostic Models

Psychometric paradigms have long been used to analyze how humans perceive different situations. For instance, Fragoso and Cúri (FRAGOSO; CúRI, 2013) proposed a study using Multidimensional Item Response Theory models to evaluate depression's symptom evolution by identifying cognitive and somatic-affective latent traits. Although the predominant psychometric paradigm is the Item Response Theory (IRT), which focuses on measuring a single unidimensional factor to obtain the latent trait, the interest in Cognitive Diagnostic Model (CDM) has been recently renewed due to its different proposal of approach.

Suitable for modeling categorical response variables, CDM aims to obtain multi-dimensional latent attributes in order to build a profile for each individual that highlights which personal characteristics are present and which are not (RUPP; TEMPLIN, 2007). This allows the cognitive model to deliver individualized and elaborated feedback and help set priorities during decision-making tasks.

Commonly applied in the educational field as a means to evaluate test-takers, the literature focuses on analyzing the student's mastery over some skills, commonly called competencies, to answer correctly a set of quizzes that depend on these competencies. As an example, in the study carried out by Tatsuoka (TATSUOKA, 1984), the domain of fraction was broken down into eight skills and each skill had selected items necessary for covering its mastery. Then, by using CDM to analyze which students had possession of such competencies, it was possible to determine which ones would need better orientation to improve in future tests.

CDMs can also be divided into different types of models. Among the most common ones used in the literature, DINA (Deterministic Inputs, Noisy "And" gate), DINO (Deterministic Inputs, Noisy "Or" gate) and G-DINA (Generalized Deterministic Inputs, Noisy "And" gate) are present. Considering N individuals that answer J items in a questionnaire, DINA models assume that for an item to be answered positively by an individual, that individual must possess all the competencies on which the item depends (TORRE, 2009b). As a compensatory counterpart of the former, DINO models assume that it is enough to have one of the competencies of the item to obtain a positive answer (TEMPLIN; HENSON, 2006). Due to the underlying stochastic nature of a questionnaire answering process, statistical noise is introduced DINA and DINO models (TORRE, 2009a). Thereby, both models have two parameters related to each item observed:

Slipping (s_j): Represents the probability that an individual, who has the required characteristics to correctly answer an item j, gets the answer wrong;

Guessing (g_j): Represents the probability of an individual, who does not have the required characteristics for responding item j, correctly gets the right answer anyway.

A problem with these two models is their assumptions on the probability of correctly answering an item being the same in the attribute vectors of the same group, which is not always the case. In other words, these models assume that students either know or do not know a particular skill or attribute required to answer an item. However, in reality, students may possess partial knowledge or guess an item's answer even if they do not have the required attributes. For example, a student may answer a math problem correctly by using a formula that they partially understand or guess the correct answer without knowing the necessary math concepts. Therefore, the DINA and DINO model's assumption of all-or-nothing knowledge may not always hold in practice. On that account, de La Torre (TORRE, 2011) presented the generalized DINA (G-DINA) model, which proposes a relaxation in the DINA model's assumption of equal probability of success for all attribute vectors, relying on a design matrix M_j to convert parameter estimates into its matricial components.

When considering a questionnaire with J items that evaluate K attributes answered by N individuals, Y is a matrix of dimension $J \times K$ in which each component Y_{ij}, with $j = 1, ..., J$, is an observable binary variable defined as:

$$Y_{ij} = \begin{cases} 1, & \text{if individual } i \text{ answers item } j \text{ correctly;} \\ 0, & \text{otherwise.} \end{cases} \tag{2.3}$$

Q is a $J \times K$ binary matrix of known values, normally defined by experts in the subject(s) of the questionnaire, and $\alpha_i = (\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{iK})$ is the vector of attributes of the individual. The components q_{jk}, $j = 1, ..., J$, and $k = 1, ..., K$, and $\alpha_{ik}, k = 1, ..., K$, are defined by Equation 2.4 and Equation 2.5, respectively.

$$q_{jk} = \begin{cases} 1, & \text{if item } j \text{ requires attribute } k\text{;} \\ 0, & \text{otherwise.} \end{cases} \tag{2.4}$$

$$\alpha_{ik} = \begin{cases} 1, & \text{if individual } i \text{ has latent attribute } k\text{;} \\ 0, & \text{otherwise.} \end{cases} \tag{2.5}$$

There are $C = 2^K$ possible values for the vector α_{ik}, which indicates that the attribute vector α_{ik} can be represented as a categorical variable. α_c will be used to denote the attribute profile of the latent class $c = 1, 2, ..., C$.

In this model, irrelevant attributes for each item can be omitted and hence the attribute vector α_i of individual i with respect to item j can be simplified into the reduced attribute vector $\alpha_i^* = (\alpha_{i1}^*, \alpha_{i2}^*, ...\alpha_{iK}*)$, where $K_j^* = \sum_{k=1}^{K} q_{ik}$ is the number of required attributes for item j. This new vector is obtained by taking the complete array of attributes α_i and keeping only the components required to answer the item j represented by $q_{jk} = 1$. If there are $2^{K_j^*}$ possible values for α_{ij}^*, this model will use $\alpha_{(l)j}^*$, where $l = 1, ..., 2^{K_j^*}$, to denote the latent group related to item j.

Furthermore, Equation 2.6 demonstrates how, in the G-DINA model, the probability $P_j(\alpha_i) = P(Y_{ij} = 1|\alpha_i)$, for each individual i and item j, is dependent only on the reduced vector of attributes α^*_{ij}.

$$P_j(\alpha_i) = P(Y_{ij} = 1|\alpha_i) = P(Y_{ij} = 1|\alpha^*_i) = P(\alpha^*_{ij}) \tag{2.6}$$

Let $\eta_{jh} = (\eta_{h1}, \ldots, \eta_{hK^*_j})$, $h = 1, \ldots, G_j$ where $G_j = 2^{K^*_j}$, represent the h-th possible value of α^*_{ij}. The item response function (IRF) of the G-DINA model can be decomposed into Equation 2.7, considering the interaction between different attributes.

$$g(P(\eta_{jh})) = \delta_{jc0} + \sum_{k=1}^{G_j} \delta_{jk}\eta_{hk} + \sum_{k=1}^{G_j-1} \sum_{k'=k+1}^{G_j} \delta_{jkk'}\eta_{hk}\eta_{hk'} + \ldots + \delta_{j12\ldots G_j} \prod_{k=1}^{G_j} \eta_{hk} \tag{2.7}$$

The components of the reduced attribute vector η_{hk} contribute as the predictor variables, g is a link function (for example, a identity function, a log or a logit) and $\delta_{j0}, \delta_{j1}, \ldots, \delta_{j12\ldots K^*_j}$ are the structural parameters of item j (FERNANDES; BAZáN; CúRI, 2023). δ_{j0} represents the baseline probability of answering a question correctly with any of the required attributes, δ_{jk} is the changed probability after mastering a single attribute, $\delta_{jkk'}$ represents the interaction effect due to the mastery of both α_{jk} and $\alpha_{jk'}$, and $\delta_{j12\ldots K^*_j}$ indicated the probability of a correct response after the mastery of all the required attributes for the item (de la Torre, 2011).

Without any restrictions, each item j has $2^{K^*_j}$ parameters to be estimated and $P(\eta_{jh})$ can assume any value between 0 and 1 . This lack of constraints allows some counter intuitive situations, such as individuals who possess all required attributes for item j having a lower success probability in this item than individuals that not possess any of the required attributes (FERNANDES; BAZáN; CúRI, 2023). To avoid some model's unexpected behaviour, we must impose constraints such as the monotonic constraint in which the domain of an additional attribute implies an increase (or no change) in the probability of success on that item (de la Torre, 2011). Figure 2 provides a visual explanation of this constrain: from the bottom where there is no mastery over any attribute to the top where all the required attributes are mastered, the gain of mastery over one additional attribute can improve this probability of success one step higher.

The three models (DINO, DINA and G-DINA) discussed so far present flexibility in obtaining the probabilities of dichotomous responses. There is a generalization for DINA and DINO for polytomous responses, but their limitations lie in the requirement for each item category to be associated with specific attributes explicitly, which doesn't always occur.

Chen et. al. (CHEN; TORRE, 2018) propose a solution fit for this restriction with the saturated General Polytomous Diagnosis Model (GPDM).

The difference in the setup of the GPDM model compared to the G-DINA is that the item responses are ordinal polytomous with C_j different levels. Therefore, $Y_{ij} = 1$ changes

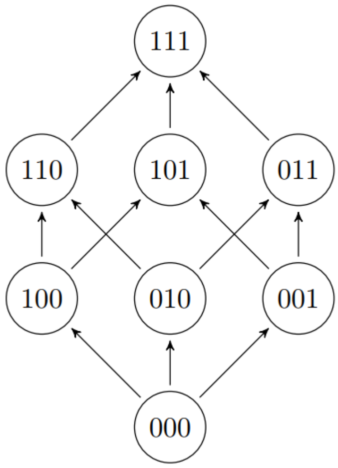


Figure 2 – Monotonic constrain visual example (FERNANDES; BAZáN; CúRI, 2023).

to $Y_{ij} = 0, 1, ..., C_j - 1$ while the Q matrix and the latent attribute vectors continue with the same definition and interpretation. The probability function also changes to Equation 2.8 where $c = 0, \ldots, C_j - 1$ represents the level of answer given to item j and $\sum_{c=0}^{C_j-1} P_c(\eta_j h) = 1$.

$$P_c(\eta_{jh}) = P_r(Y_{ij} = c|\eta_{jh}) \tag{2.8}$$

Additionally, $P_c^*(\eta_{jh}) = P_r(Y_{ij} \geq c|\eta_{jh})$ represents the cumulative probability of a response equal to or greater than level C in item j.

Similar to G-DINA, the item response function of the GPDM model is expressed by Equation 2.9.

$$g(P_c(\eta_{jh})) = \delta_{jc0} + \sum_{k=1}^{G_j} \delta_{jck}\eta_{hk} + \sum_{k=1}^{G_j-1} \sum_{k'=k+1}^{G_j} \delta_{jckk'}\eta_{hk}\eta_{hk'} + \ldots + \delta_{jc12\ldots G_j} \prod_{k=1}^{G_j} \eta_{hk} \tag{2.9}$$

where $\delta_{jc12\ldots G_j}$ are the structural parameters of the response level c of item j and g is a link function. The interpretation of the parameters $\delta_{jc12\ldots G_j}$ are the same as the G-DINA model, except of being regarding to the c-th response level of item j.

Likewise, the monotonic construction presented for G-DINA is adapted for the context of polytomous responses. Within the context of polytomous responses, we have that monotonic construction is defined so that if $\eta_{jh} < \eta_{jh'}$ then $P_c^*(\eta_{jh}) \leq P_c^*(\eta_{jh'})$ for every c and every j.

Like in G-DINA, the estimated parameters are $P_c^*(\eta_{ij})$, and the parameters η_{jc} are obtained through a linear transformation after the estimation. The estimation of the parameters of GPDM (and G-DINA) is obtained by maximizing the marginalized log-Likelihood of the model with an Expectation-Maximization (EM) algorithm.

CHAPTER

3

METHODS

This chapter describes the new recommendation approach proposed in this work. Additionally, it describes the databases and the metrics used to evaluate the proposed approach.

3.1 Proposed Approaches

As our goal is to explore the connection between the cognitive diagnosis model and recommendation systems that use latent factors, the approach proposed in this work uses the GPDM for suggesting items based on polytomous responses. Figure 3 shows a diagram that depicts the overall process of the proposed recommendation system.

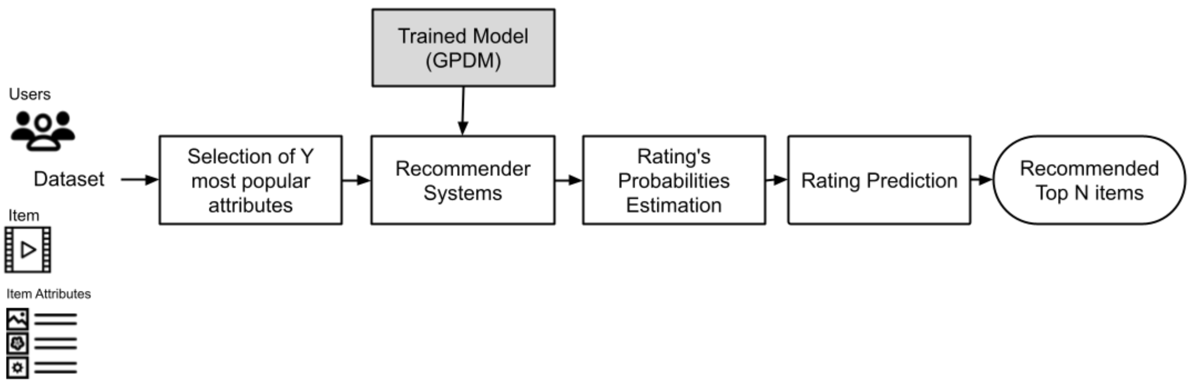


Figure 3 – Diagram of the proposed approach for the recommendation that uses GPDM for cognitive diagnosis.

As previously discussed, the GPDM applies reduced attribute vectors to obtain the estimation of all 2^Y probabilities of success. Therefore, the more attributes, the longer the algorithm takes to process. So the first step of this model is to select the Y most popular attributes present in the input items. Consequently, items that didn't have any of these attributes are excluded and users that only have ratings from the excluded items are also not counted in the process.

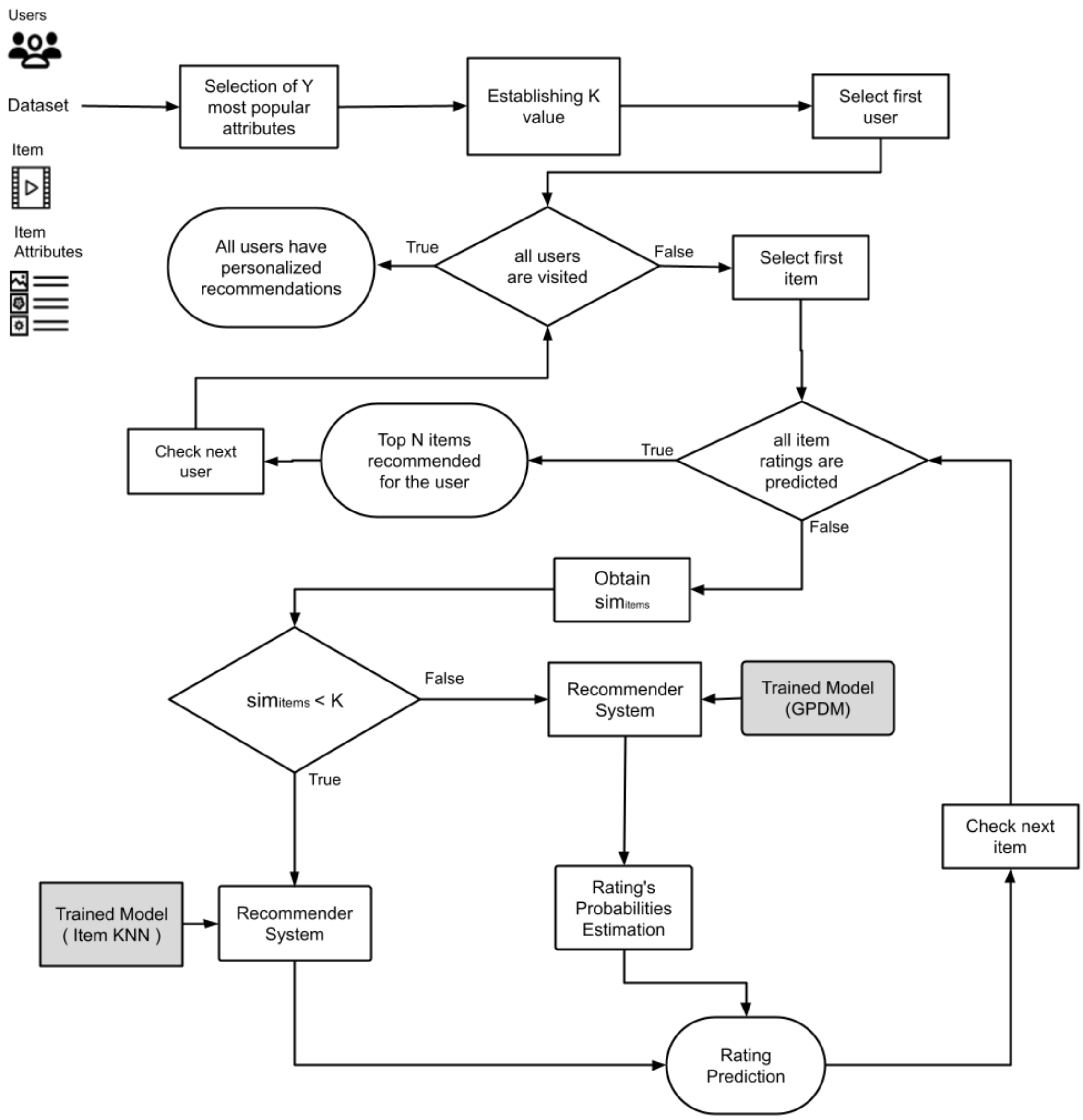


Figure 4 – Hybrid model with content filtering and GPDM.

With the data selected, the algorithm goes through every user and every item the users had rated using the GPDM-trained model to make a rating prediction. The predictions are done by two techniques: selecting the rating with the greater probability (Max Probability method) or computing the weighted average between all items and probabilities as the suggested rating (Weight Average method). These predictions are compared to the real ratings of the displayed items and evaluated with a metric (for instance, MAE or RMSE).

With all the ratings predicted for a user, the items are sorted in ascending order and the first N items on the list are displayed as the suggested items for each user. This ranking system is analyzed by a metric (for instance, Mean Average Precision).

As we are also interested in evaluating if the GPDM can be integrated with other recommendation systems, we have also proposed a hybrid model that fuses the GPDM and the K-Nearest Neighbors classifier as a content filtering model. Figure 4 shows a flowchart that depicts this proposed hybrid model.

In the hybrid model, after selecting the Y most popular attributes, a K value is established for the content filtering model to work, where K represents the number of similar items that would be used.

The algorithm goes through every user and every movie the user had rated. For each movie, a list of similar movies is obtained using the Jaccard metric to calculate the similarities. This list is then reduced to only the movies the user had rated. sim_{items} represented the final size of the list. Therefore, if a user had rated less than the required K value for similar movies, the GPDM model is used instead to make the predictions. Ifsim_{items} is equal to or greater than K, the item's ratings are predicted using Equation 3.1.

$$P_{ui} = \frac{\sum_{k=1}^{K} r_{uk} * sim_{ik}}{sim_{ik}} \tag{3.1}$$

where P_{ui} is the prediction from user u to item i, r_{uk} is the real rating from user u to item k and sim_{ik} is the similarity percentage between items i and k.

3.2 Databases

Two databases were used for this study in order to assess the proposed approaches, MovieLens 20M and Anime Recommendation Database, and it is noticeable some similarities related to the type of data they carry. For instance, both contain an archive for the user-item matrix that has the ratings for each user to each movie/anime (if it exists) and a second archive for the items and genres indicating by binary values if each item includes or not each genre.

It is important to notice that these databases were slightly modified to fit the proposed study. The first step was to select the most rated items and the most popular genres among the selected movies or animes. This method contributed to reducing a little bit of the sparsity problem presented in recommendation systems by guaranteeing the presence of ratings in the majority of users.

3.2.1 MovieLens 20M Dataset

The MovieLens dataset MovieLens, publicly available on GroupLens, has 20 million ratings estimated from 138,000 users with 27,000 movies. In each movie, there is a set of categories to which the movie belongs and a set of tags.

All assessments include the following features:

movieId: A unique identifier of the rated movie;

movieTitle: The movie title evaluated with the year of release in parentheses;

movieGenres: A sequence of genres to which the rated movie belongs;

userId: A unique identifier of the user who made the assessment;

userRating: The rating of the review on a 0.5 to 5-star scale;

timestamp: The timestamp of the evaluations, represented in seconds since midnight Coordinated Universal Time (UTC) January 1, 1970.

3.2.2 Anime Recommendation Dataset

The Anime Recommendation dataset AnimeRecommenderDatabase, publicly available on the Kaggle platform, contains information on user preference data from 73,516 users on 12,294 animes with 84 genres, found on myanimelist.net. Each user is able to add anime to their completed list and give it a rating.

The following features are presented along with the ratings:

animeId: Unique identifier of the rated anime from myanimelist.net;

animeTitle: The anime title in Japanese or English;

animeGenres: A sequence of genres to which the rated movie belongs stored in a comma-separated list;

userId: A unique identifier of the user who made the assessment;

userRating: The rating of the review on a 1 to 10-star scale. If the user watched it but didn't assign a rating a value of -1 is assigned.

3.3 Evaluation Metrics

The Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) metrics were applied in the analysis of the predictions of ratings. With the predictions, it was then possible to analyze the recommendation of the available items by using the Mean Average Predictions (MAP@X).

3.3.1 Root Mean Squared Error

Root Mean Squared Error (RMSE), expressed by Equation 3.2, represents the quadratic mean of the differences between predicted values and observed values.

$$RMSE = \sqrt{\frac{\sum_{n=1}^{N}(\overline{r}_n - r_n)^2}{N}} \tag{3.2}$$

where $\overline{r}_i$ is the predicted value, r_i is the real rating and N is total amount of ratings to predict. Since this metric takes the square root of the mean of the squared differences between the predicted ratings and actual ratings, RMSE is more sensitive to large errors than the next metric, which it's useful to evaluate a model by penalizing it more heavily. So a smaller RMSE indicates better prediction accuracy.

3.3.2 Mean Absolute Error

The Mean Absolute Error (MAE) measures the average absolute difference between the predicted ratings and the actual ratings. It aims to obtain the average of all the errors in a set in order to evaluate the difference between the measured value and the true value. Its result is calculated by Equation 3.3.

$$MAE = \frac{\sum_{n=1}^{N}|\overline{r}_n - r_n|}{N} \tag{3.3}$$

where $\overline{r}_n$ is the predicted value, r_n is the real rating and N is total amount of ratings to predict.

A lower MAE value indicates better accuracy, as it means that the model's predictions are closer to the actual ratings.

3.3.3 MAP@X

As one of the most popular metrics used to evaluate recommendation systems, the Mean Average Precision (MAP@X) aims to evaluate how many of the recommended items are relevant and showing at the top. It analyzes the quality of item ranking in recommender systems and other ranking tasks by measuring the average precision at different levels of recall, where recall is the fraction of relevant items that have been retrieved over the total number of relevant items.

Firstly, it calculates the $Precision@X$ which is the number of relevant items in top X results divided by the X values we are working with. The second step is to gather the average precision by different X values. This calculation is obtained by Equation 3.4 in which rel(x)=1 if x^{th} item is relevant.

$$AP@X = \frac{1}{r_x}\sum_{x=1}^{X} Precision@x * rel(x) \tag{3.4}$$

Finally, the mean average precision gathers the mean of these AP values across all queries or users, where X is the maximum rank considered. For example, if X=5, then only the top 5 items in the list are considered for calculating AP. MAP@X ranges from 0 to 1, where a higher value indicates better performance in ranking the relevant items higher. Therefore, the goal is to maximize the number of relevant items in the top X positions of the list, where X is the number of items recommended to the user. MAP@X provides a way to compare the performance of different ranking algorithms and optimize their parameters to achieve better performance.

CHAPTER

4

RESULTS

The GPDM-based recommendation approach, as well as the hybrid approaches (Max Probability and Weight Average), proposed in this work were assessed on two datasets, Database 1 (DB1) and Database 2 (BD2), obtained from MovieLens 20M Dataset and Anime Recommendation Database, respectively, as follows:

Database 1 (DB1): A subset of MovieLens 20M, containing:

- **Training set:** 6978 users, 1680 movies, and 4 genres per movie;
- **Test set:** 2000 users for the same 1680 movies, and 4 genres;

Database 2 (DB2): A subset of Anime Recommendation Database, containing:

- **Training set:** 6540 users, 1457 animes, and 4 genres;
- **Test set:** 2000 users for the same 1457 animes, and 4 genres.

The choice to have subsets from the original datasets was selected in an attempt to balance the amount of information between the two databases so comparisons in the performance of the models would be easier to interpret. In addition, the original MovieLens 20M Dataset had far more users and movies than the dataset with animes, but it counted only 21 genres in total while, in Anime Recommendation Database, the genres amounted to 84. This amount of genres was reduced to 4 in the subsets since, as explained in section 2.2, the GPDM has to calculate all 2^{genres} possible values for the vector α_{ik}, which leads to exponential growth with the increase of genres used. By excluding the majority of genres, movies that didn't have at least one of the remaining genres were also taken out of the datasets and users that didn't have at least 10 ratings from the movies that remained were also not considered in this work.

With all the data filtered and prepared, the models were trained and tested with the results presented and discussed below.

For the sake of comparison with traditional models, we have also applied on the same datasets the CFG (Content Filtering of Genres) model. Table 1 presents the results obtained for each dataset and for each tested model in the rating prediction phase.

For CFG model and the hybrid models, two different k values were used, $k = 100$ and $k = 840$. For each user and each item (movie or anime, in our datasets) we want to predict the rating, the k value represents the k most similar items to a specific item that was also rated by the user.

Metric & Dataset	**CFG**		**GPDM**		**Hybrid Max.Prob**		**Hybrid Wei.Avg**	
	k=100	k=840	**Max.Prob**	**Wei.Avg**	k=100	k=840	k=100	k=840
RMSE-DB1	1.1386	1.1478	1.0002	0.9163	0.9836	1.0150	0.9346	0.9282
MAE-DB1	0.9047	0.9227	0.6862	0.7112	0.7166	0.7003	0.7321	0.7253
RMSE-DB2	0.8619	1.0875	0.7954	0.7261	0.7061	0.7954	0.7020	0.7250
MAE-DB2	0.6175	0.8592	0.5235	0.5726	0.5346	0.5261	0.5371	0.5697

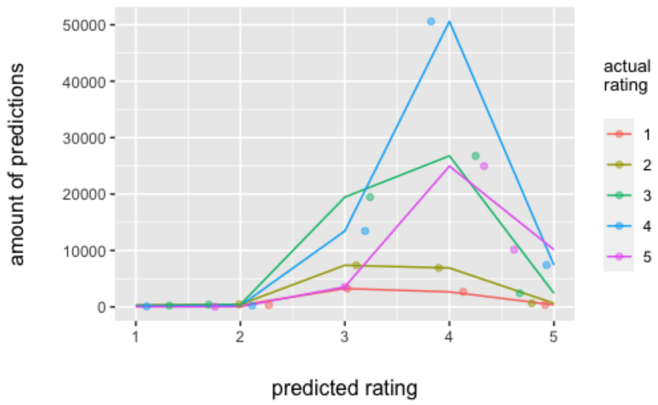
Table 1 – RMSE and MAE error rates obtained by CFG, GPDM, and both Hybrid models on DB1 and DB2 datasets, in the rating prediction phase.

It is important to point out that when there are not enough similar items, the CFG model applies the difference between the mean average rating of the whole dataset and the bias from the user's average ratings and the movie's average rating. Because of that, one can observe in Table 1 that the higher the k value, the lower the performance of the CFG model. One can also observe in Table 1 that the GPDM-based model, proposed in this work, outperforms the CFG model in both datasets.

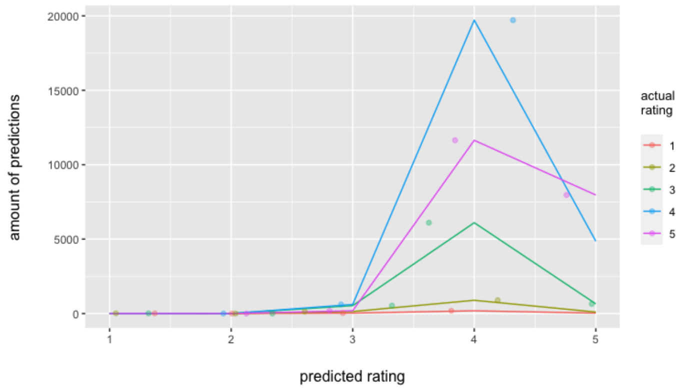
Although the Weight Average hybrid approach obtained better RMSE values than the Max Probability hybrid approach, it obtained worse MAE values. This implies that if we are interested in choosing a model that performs better with a metric that heavily penalizes outliers from the predictions, then the Weight Average approach is a better choice than the Max Probability. The reason for this is that the Weight Average can predict floating values, while the Maximum Probability only predicts integer ratings from the range we are working on.

However, having better results in the MAE metric indicates that the magnitude of the errors from the Max Probability is smaller than the one from the Weight Average. These results can also be seen in Figures 5 and 6 that indicate how many predictions each model makes.

Figure 6 (a) and Figure 6 (b) show a lot more points and thus a greater variety of predictions from the model. However, none of the points are in 1 and 5, the limits from the interval we are working on. Since the predictions are obtained by the weighted average from all probabilities, the only way to have predictions in the limits are if the other probabilities are zero, which is very uncommon. Meanwhile, Figure 5 (a) and Figure 5 (b) have less points but the predictions are always between the established interval). Therefore, we can conclude that the Weight Average has more accurate results in ratings between the interval limits, but not in



(a) DB1



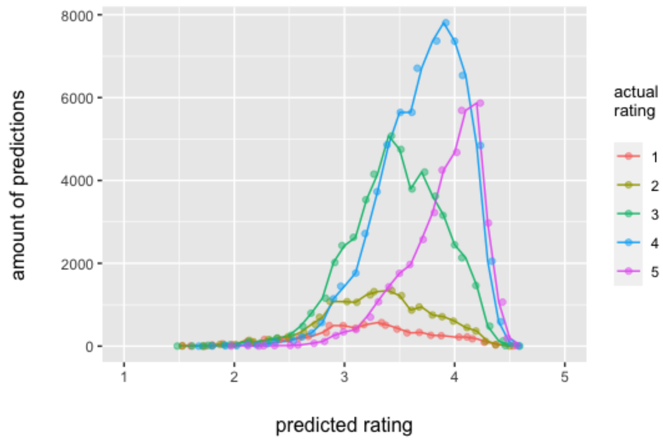
(b) DB2

Figure 5 – Histogram of predictions from GPDM.MaxProb model on DB1 and DB2.

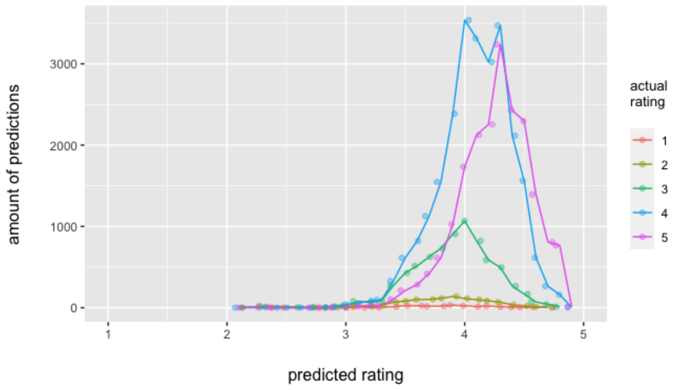
the limits themselves, while the Max Probability doesn't perform so well but has less errors in general by including the limits in the possible predictions.

Figure 7 exhibits the confusion matrices from the Max Probability GPDM in the two datasets. Even though Figure 7 (a) presents more accurate results from the model in the DB1 compared to the results DB2 shown in Figure 7(b), it's possible to evaluate that, in general, the model predicts very close ratings to the real ones. For example, if a movie's score is 4, the model has more predictions between 3, 4, and 5 compared to predictions of 1 or 2. In spite of Figure 7(b) showing that this doesn't always happen (for example, the model predicting a lot of 3 and 5 scores for anime's ratings that are actual 2), in the context of this work that aims to recommend items that the user might like, it's preferable that the model brings more positive predictions than negative ones. There's also the fact that this database had its rating interval modified (from 1 to 10 to 1 to 5) to fit the comparison with the other database (DB1) and, as a consequence, there was a loss of information. Since rating is subjective, it's possible that a user who has rated an anime 5 or 6 on a scale 1-10 might not have rated it 3 on a scale of 1-5.

Figures 8 and 9 introduce the behavior coming from systems that integrates the CFG



(a) DB1



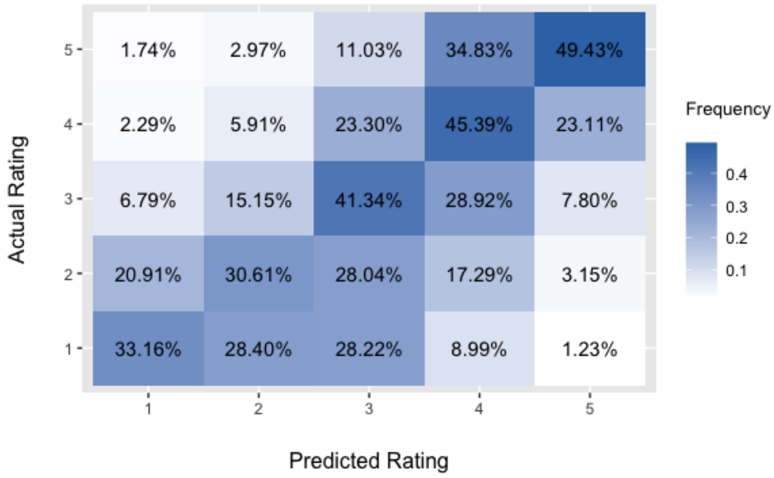
(b) DB2

Figure 6 – Histogram of predictions from GPDM.WeiAvg model on DB1 and DB2.

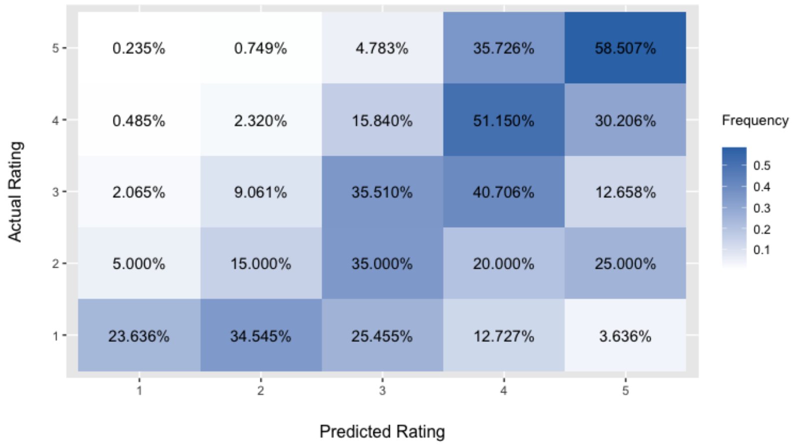
model and the GPDM in two hybrid models: one with the Max Probability GPDM and the other with the Weight Average GPDM. A few differences between the Max Probability GPDM and the Max Probability Hybrid model can be very noticeable. For instance, the hybrid model can predict floating ratings and thus has a better performance compared to the CFG model. The Weight Average GPDM and the Weight Average Hybrid model have more similar curves with some increases in the variety of ratings, but the hybrid model also has problems with predicting the values on the limits.

It's also worth mentioning that, as shown in Table 1, the hybrid model improves on the problem with big k values on the metric MAE. Although the RMSE metric still increases with the increase of the required amount of similar items, the average error of predictions decreases in both datasets. Additionally, both methods improve on the CFG model, but the GPDM alone still has better results overall.

The final step of this work was to analyze the performance of all models in recommending relevant and interesting items for the users. Figures 10 and 11 show the results obtained in both



(a) DB1



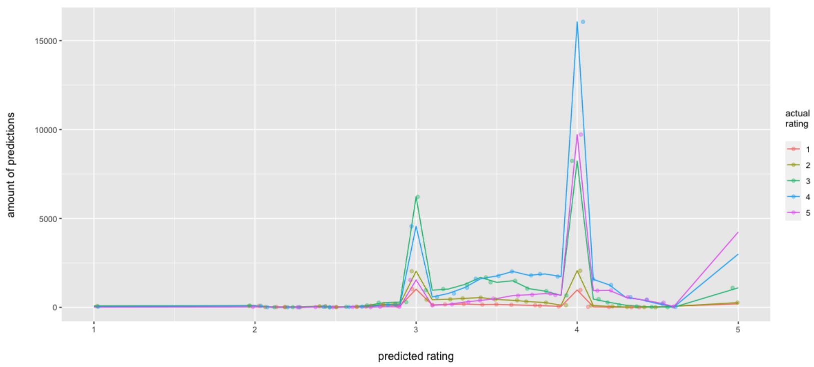
(b) DB2

Figure 7 – Confusion-matrix of ratings and predictions from GPDM.MaxProb model on DB1 and DB2.

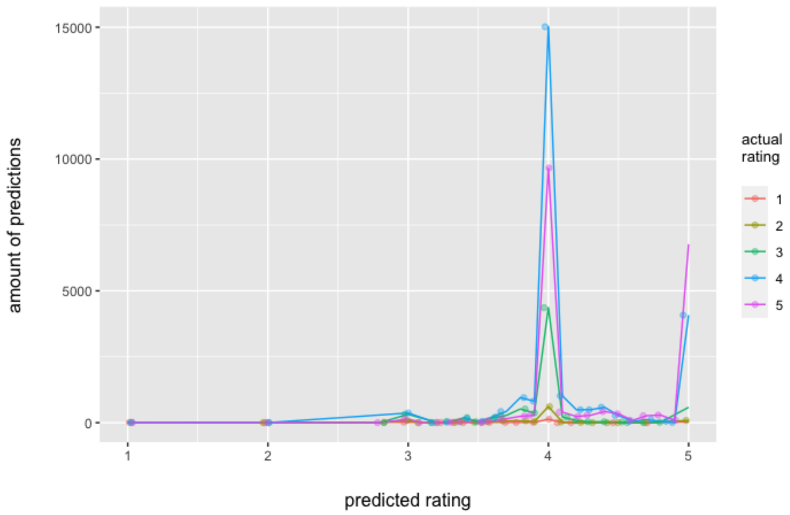
datasets by varying the amount of required k similar movies in the models that required them (CFG and the hybrid models).

In Figure 10, there is a greater difference between the models when $k = 100$ and when $k = 840$. In this case, the hybrid model is still used in the majority of the CFG and thus their performance is closer to the classic model. By incrementing k, CFG improves in recommending items and the hybrid models also start using more the GPDM. As a consequence of this increase in the k value, the mean average precision of the recommendations improves, with the GPDMs still having greater results. Figure 11 doesn't show much difference between the variance of k, but it's still possible to see the GPDMs with higher results between all the tested ones.

Although, in the hybrid model, we intended to integrate the GPDMs with the CFG to take advantage of their strong characteristics and, as consequence, get better performances, the results presented in this work show otherwise. The GPDM is outperforming the hybrid model in



(a) DB1

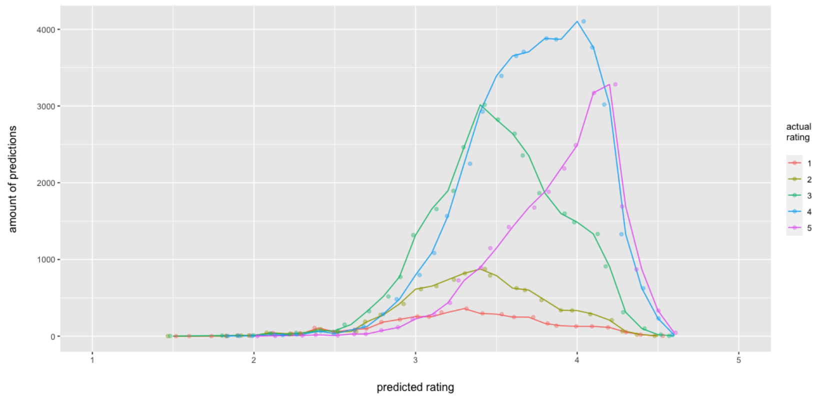


(b) DB2

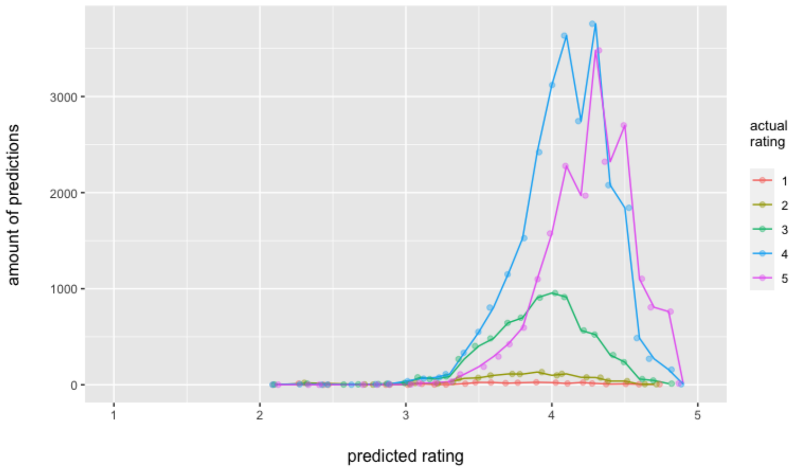
Figure 8 – Histogram of predictions from Hybrid Max Probability (sim_i=100) model on DB1 and DB2.

almost all aspects and one explanation for this is the classic recommendation system selected to be integrated with the proposed model.

The CFG was chosen because it also uses genres to obtain the user's preferences (even though it is not by calculating their latent preferences) and it would be interesting to compare these different models that are based on discrete attributes side by side. However, the number of required genres for one model to effectively work is inversely proportional to the required amount for the other. In other words, the CFG calculates the similarity between items by using the genres, so it needs more genres to give more accurate values in the similarity between items. The GPDM, in contrast, needs fewer attributes due to the exponential increase it has when calculating all possible variations of users' latent attributes. So by limiting our subsets to 4 genres for the GPDM to work we are also interfering with the performance of the CBF. Therefore, when integrating the two models in a hybrid system, the results are improved from CBF's ones but not from the GPDM's. One way to try to solve this problem in future works would be to select a collaborative filtering model to be integrated into the hybrid model since it is independent of



(a) DB1

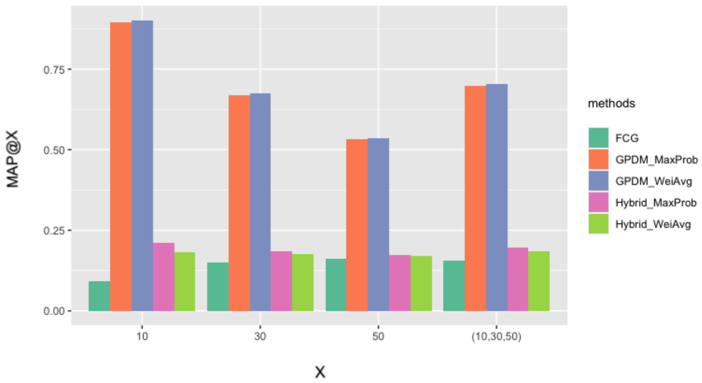


(b) DB2

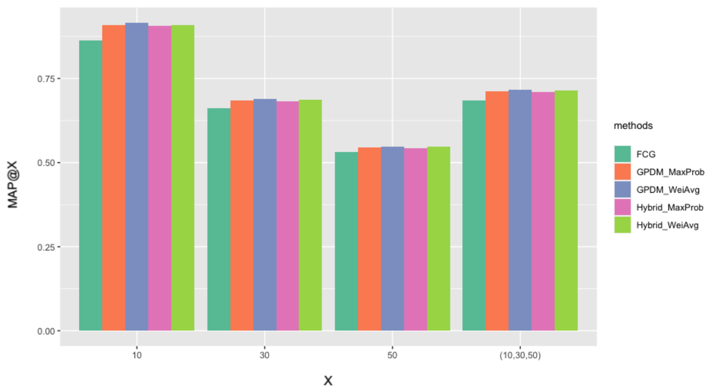
Figure 9 – Histogram of predictions from Hybrid Weight Average (sim_i=100) model on DB1 and DB2.

the number of genres. Nonetheless, the improvement of the hybrid models presented in Table 1 compared to the classic CFG model as well as the increase in the number of possible ratings to be predicted by the Max Probability model in a hybrid system shows there are still some advantages in exploring these two models together.

In addition, by the GPDM having such great performances in both rating prediction and ranking of items, it's indicative that this CDM is possible to cross other domains, not being restricted to only educational and psychological tests. It also offers the point that others cognitive diagnostic models can also be applied to recommendation systems to indicate the user's interests in the form of polytomous responses.

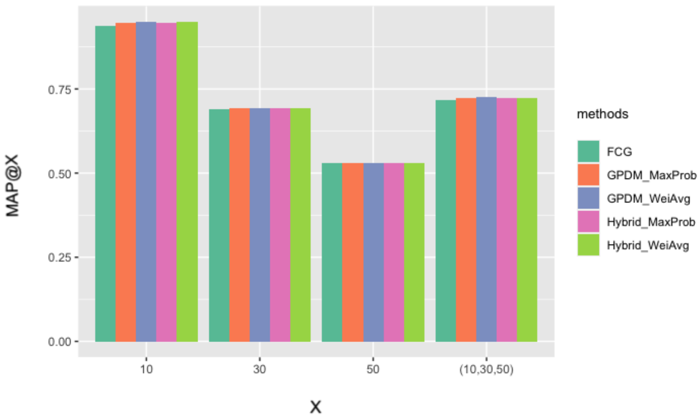


(a) k=100

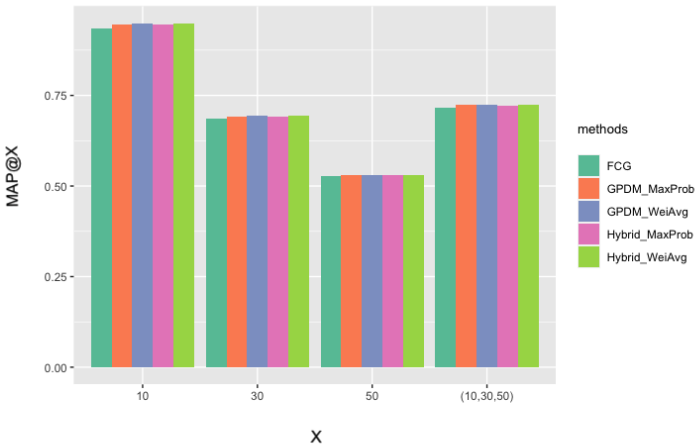


(b) k=840

Figure 10 – Barplot from $MAP@x$ results for k=100 and k=840 on DB1.



(a) k=100



(b) k=840

Figure 11 – Barplot from $MAP@x$ results k=100 and k=840 on DB2.

CHAPTER 5

CONCLUSIONS

The aim of this project was to explore the Cognitive Diagnostic Models in the application of Recommendation Systems and, consequently, propose a new model that suggests media defined by discrete attributes and predicts polytomous responses in the form of ratings for each user by analyzing their latent preference. In order to obtain a better understanding of the proposed model's impact, there's also the objective of comparing and merging it with other recommendation systems that were studied so far.

An experimental analysis has been performed in which three models have been included: this work's proposed approach (the GPDM), a K Nearest Neighbors method as a content-based filtering model and the hybrid model which integrates the previous two. The results have shown that the proposed approach has a better performance compared to the content-based filtering, a classic literature model, not only on the prediction of ratings but also on the ranking of items to recommend. This indicates that the model is an improvement to be considered in future recommendation studies as it's obtaining implicit information about users by analyzing their behaviour in the rating of items.

As for the hybrid model, in the rating prediction tests, it shows improvements in the RMSE metric when compared to the other models, but in the ranking of items, it outperforms the classic model but not the GPDM.

One limitation of this study is the comparison with only one classic recommendation system from the literature. In future works, it would be interesting to compare it with more models using more evaluation metrics available. However, the results shown so far indicate a promising method to be explored as a powerful tool that can obtain additional information from the user's latent preferences in a non-intrusive way: just by evaluating their rating behavior and the content present in the items available.

BIBLIOGRAPHY

BAUER, C.; KHOLODYLO, M.; STRAUSS, C. Music recommender systems challenges and opportunities for non-superstar artists. In: **Bled eConference**. [S.l.: s.n.], 2017. Citation on page 26.

BEUTEL, A.; COVINGTON, P.; JAIN, S.; XU, C.; LI, J.; GATTO, V.; CHI, E. H. Latent cross: Making use of context in recurrent recommender systems. In: . New York, NY, USA: Association for Computing Machinery, 2018. (WSDM '18), p. 46–54. ISBN 9781450355810. Available: <https://doi.org/10.1145/3159652.3159727>. Citation on page 27.

BOBADILLA, J.; HERNANDO, A.; ORTEGA, F.; BERNAL, J. A framework for collaborative filtering recommender systems. **Expert Systems with Applications**, v. 38, n. 12, p. 14609–14623, 2011. ISSN 0957-4174. Available: <https://www.sciencedirect.com/science/article/pii/S0957417411008049>. Citation on page 25.

BOBADILLA, J.; ORTEGA, F.; HERNANDO, A.; GUTIéRREZ, A. Recommender systems survey. **Knowledge-Based Systems**, v. 46, p. 109–132, 2013. ISSN 0950-7051. Available: <https://www.sciencedirect.com/science/article/pii/S0950705113001044>. Citations on pages 23 and 26.

BRADSHAW, L.; IZSáK, A.; TEMPLIN, J.; JACOBSON, E. Diagnosing teachers' understandings of rational numbers: Building a multidimensional test within the diagnostic classification framework. **Educational Measurement: Issues and Practice**, v. 33, 11 2013. Citation on page 21.

B.THORAT, P.; GOUDAR, R.; BARVE, S. Survey on collaborative filtering, content-based filtering and hybrid recommendation system. **International Journal of Computer Applications**, v. 110, p. 31–36, 01 2015. Citation on page 27.

BURKE, R. Hybrid recommender systems: Survey and experiments. **User Modeling and User-Adapted Interaction**, v. 12, 11 2002. Citation on page 27.

CAI, Y.; LEUNG, H.-f.; LI, Q.; MIN, H.; TANG, J.; LI, J. Typicality-based collaborative filtering recommendation. **IEEE Transactions on Knowledge and Data Engineering**, v. 26, n. 3, p. 766–779, 2014. Citation on page 24.

CHEN, C.-M.; DUH, L.-J.; LIU, C.-Y. A personalized courseware recommendation system based on fuzzy item response theory. In: **IEEE International Conference on e-Technology, e-Commerce and e-Service, 2004. EEE '04. 2004**. [S.l.: s.n.], 2004. p. 305–308. Citation on page 20.

CHEN, J.; TORRE, J. de la. Introducing the general polytomous diagnosis modeling framework. **Frontiers in Psychology**, v. 9, 2018. ISSN 1664-1078. Available: <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01474>. Citation on page 30.

de la Torre, J. The generalized dina model framework. **Psychometrika**, v. 76, n. 2, p. 179–199, April 2011. Citation on page 30.

DELDJOO, Y.; ELAHI, M.; CREMONESI, P.; GARZOTTO, F.; PIAZZOLLA, P.; QUADRANA, M. Content-based video recommendation system based on stylistic visual features. **Journal on Data Semantics**, v. 5, p. 1–15, 06 2016. Citation on page 26.

DO, H.-Q.; LE, T.-H.; YOON, B. Dynamic weighted hybrid recommender systems. In: **2020 22nd International Conference on Advanced Communication Technology (ICACT)**. [S.l.: s.n.], 2020. p. 644–650. Citation on page 27.

EGGLESTONE, S.; WHITBROOK, A.; GREENSMITH, J.; WALKER, B.; BENFORD, S.; MARSHALL, J.; KIRK, D.; SCHNäDELBACH, H.; IRUNE, A.; ROWLAND, D. Recommending rides: Psychometric profiling in the theme park. **Computers in Entertainment**, v. 8, p. 21, 12 2010. Citation on page 20.

ELAHI, M.; RICCI, F.; RUBENS, N. A survey of active learning in collaborative filtering recommender systems. **Computer Science Review**, v. 20, p. 29–50, 2016. ISSN 1574-0137. Available: <https://www.sciencedirect.com/science/article/pii/S1574013715300150>. Citation on page 25.

ELBERSE, A.; OBERHOLZER-GEE, F. Superstars and underdogs: An examination of the long tail phenomenon in video sales. **MSI Reports**, v. 45, p. 7–15, 01 2007. Citation on page 23.

FERNANDES, R.; BAZáN, J.; CúRI, M. A bayesian approach for the g-dina model. Manuscript submitted for publication. 2023. Citations on pages 13, 30, and 31.

FRAGOSO, T.; CúRI, M. Improving psychometric assessment of the beck depression inventory using multidimensional item response theory. **Biometrical journal. Biometrische Zeitschrift**, v. 55, 07 2013. Citation on page 28.

GHAZANFAR, M.; PRUGEL-BENNETT, A. An improved switching hybrid recommender system using naive bayes classifier and collaborative filtering. In: **The 2010 IAENG International Conference on Data Mining and Applications (17/03/10 - 19/03/10)**. [s.n.], 2010. Event Dates: 17-19 March, 2010. Available: <https://eprints.soton.ac.uk/268483/>. Citation on page 27.

GONZALEZ-CARRASCO, I.; COLOMO-PALACIOS, R.; LOPEZ-CUADRADO, J. L.; GARCı´A-CRESPO Ángel; RUIZ-MEZCUA, B. Pb-advisor: A private banking multi-investment portfolio advisor. In: . [s.n.], 2012. v. 206, p. 63–82. ISSN 0020-0255. Available: <https://www.sciencedirect.com/science/article/pii/S0020025512002526>. Citation on page 20.

HISANABE, Y. Persona marketing for fujitsu kids site. **Fujitsu scientific technical journal**, v. 45, 04 2009. Citation on page 19.

HU, B.; LI, Z.; CHAO, W.-H.; HU, X.; WANG, J. User preference representation based on psychometric models. 2011. Citation on page 20.

HUANG, B.-H.; DAI, B.-R. A weighted distance similarity model to improve the accuracy of collaborative recommender system. In: **2015 16th IEEE International Conference on Mobile Data Management**. [S.l.: s.n.], 2015. v. 2, p. 104–109. Citation on page 24.

IMMANENI, N.; PADMANABAN, I.; RAMASUBRAMANIAN, B.; SRIDHAR, R. A meta-level hybridization approach to personalized movie recommendation. In: **2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)**. [S.l.: s.n.], 2017. p. 2193–2200. Citation on page 27.

Knotzer, N. **Product Recommendations in E-Commerce Retailing Applications**. Berlin, Germany: Peter Lang Verlag, 2018. Available: <https://www.peterlang.com/document/1068515>. Citation on page 23.

LIU, N. N.; XIANG, E. W.; ZHAO, M.; YANG, Q. Unifying explicit and implicit feedback for collaborative filtering. In: **Proceedings of the 19th ACM International Conference on Information and Knowledge Management**. New York, NY, USA: Association for Computing Machinery, 2010. (CIKM '10), p. 1445–1448. ISBN 9781450300995. Available: <https://doi.org/10.1145/1871437.1871643>. Citations on pages 20 and 24.

MOHAMED, M. H.; KHAFAGY, M. H.; IBRAHIM, M. H. Recommender systems challenges and solutions survey. In: **2019 International Conference on Innovative Trends in Computer Engineering (ITCE)**. [S.l.: s.n.], 2019. p. 149–155. Citation on page 27.

NIYAZOV, A.; MIKHAILOVA, E.; EGOROVA, O. Content-based music recommendation system. In: **2021 29th Conference of Open Innovations Association (FRUCT)**. [S.l.: s.n.], 2021. p. 274–279. Citation on page 26.

OESTREICHER-SINGER, G.; SUNDARARAJAN, A. Recommendation networks and the long tail of electronic commerce. **MIS Quarterly**, Management Information Systems Research Center, University of Minnesota, v. 36, n. 1, p. 65–83, 2012. ISSN 02767783. Available: <http://www.jstor.org/stable/41410406>. Citation on page 23.

Ordóñez Galán, C.; Sánchez Lasheras, F.; de Cos Juez, F. J.; Bernardo Sánchez, A. Missing data imputation of questionnaires by means of genetic algorithms with different fitness functions. **Journal of Computational and Applied Mathematics**, v. 311, p. 704–717, 2017. ISSN 0377-0427. Available: <https://www.sciencedirect.com/science/article/pii/S0377042716303727>. Citation on page 21.

REBELO, M. Â.; COELHO, D.; PEREIRA, I.; FERNANDES, F. A new cascade-hybrid recommender system approach for the retail market. In: ABRAHAM, A.; MADUREIRA, A. M.; KAKLAUSKAS, A.; GANDHI, N.; BAJAJ, A.; MUDA, A. K.; KRIKSCIUNIENE, D.; FERREIRA, J. C. (Ed.). **Innovations in Bio-Inspired Computing and Applications**. Cham: Springer International Publishing, 2022. p. 371–380. ISBN 978-3-030-96299-9. Citation on page 27.

RENUKA, S.; KIRAN, G. S. S. R.; ROHIT, P. An unsupervised content-based article recommendation system using natural language processing. In: JACOB, I. J.; SHANMUGAM, S. K.; PIRAMUTHU, S.; FALKOWSKI-GILSKI, P. (Ed.). **Data Intelligence and Cognitive Informatics**. Singapore: Springer Singapore, 2021. p. 165–180. Citation on page 27.

RICCI, F.; ROKACH, L.; SHAPIRA, B. Recommender systems handbook. In: ____. [S.l.: s.n.], 2010. v. 1-35, p. 1–35. Citation on page 19.

RUPP, A.; TEMPLIN, J. Unique characteristics of cognitive diagnosis models. 01 2007. Citation on page 28.

SALTON, G.; WONG, A.; YANG, C. S. A vector space model for automatic indexing. **Commun. ACM**, Association for Computing Machinery, New York, NY, USA, v. 18, n. 11, p. 613–620, nov 1975. ISSN 0001-0782. Available: <https://doi.org/10.1145/361219.361220>. Citation on page 26.

SOLEYMANI, M.; ALJANAKI, A.; WIERING, F.; VELTKAMP, R. Content-based music recommendation using underlying music preference structure. In: . [S.l.: s.n.], 2015. p. 1–6. Citation on page 26.

SREEPADA, R. S.; PATRA, B. K. Mitigating long tail effect in recommendations using few shot learning technique. **Expert Systems with Applications**, v. 140, p. 112887, 2020. ISSN 0957-4174. Available: <https://www.sciencedirect.com/science/article/pii/S0957417419305974>. Citation on page 23.

SU, X.; KHOSHGOFTAAR, T. M. A survey of collaborative filtering techniques. **Adv. in Artif. Intell.**, Hindawi Limited, London, GBR, v. 2009, jan 2009. ISSN 1687-7470. Available: <https://doi.org/10.1155/2009/421425>. Citation on page 25.

SURIATI, S.; DWIASTUTI, M.; TULUS, T. Weighted hybrid technique for recommender system. **Journal of Physics: Conference Series**, IOP Publishing, v. 930, n. 1, p. 012050, dec 2017. Available: <https://dx.doi.org/10.1088/1742-6596/930/1/012050>. Citation on page 27.

SVETINA, D. Diagnostic measurement: Theory, methods, and applications by rupp, andré a., templin, jonathan, henson, robert a. **Journal of Educational Measurement**, v. 48, 06 2011. Citation on page 20.

TATSUOKA, K. K. Analysis of errors in fraction addition and subtraction problems. final report. In: . [S.l.: s.n.], 1984. Citation on page 28.

TAWFIQ, F.; RAHMA, A. M.; WAHAB, H. Recommendation systems for e-commerce systems an overview. **Journal of Physics: Conference Series**, v. 1897, p. 012024, 05 2021. Citation on page 26.

TEMPLIN, J.; HENSON, R. Measurement of psychological disorders using cognitive diagnosis models. psychological methods, 11(3), 287-305. **Psychological methods**, v. 11, p. 287–305, 10 2006. Citation on page 28.

TORRE, J. The generalized dina model framework. **Psychometrika**, v. 76, p. 179–199, 04 2011. Citation on page 29.

TORRE, J. de la. Dina model and parameter estimation: A didactic. **Journal of Educational and Behavioral Statistics**, v. 34, n. 1, p. 115–130, 2009. Available: <https://doi.org/10.3102/1076998607309474>. Citations on pages 20 and 28.

____. Dina model and parameter estimation: A didactic. **Journal of Educational and Behavioral Statistics**, v. 34, n. 1, p. 115–130, 2009. Citation on page 28.

TORRE, J. de la; ARK, L. A. van der; ROSSI, G. Analysis of clinical data from a cognitive diagnosis modeling framework. **Measurement and Evaluation in Counseling and Development**, Routledge, v. 51, n. 4, p. 281–296, 2018. Available: <https://doi.org/10.1080/07481756.2017.1327286>. Citation on page 21.

TRAN, C.; KIM, J.-Y.; SHIN, W.-Y.; KIM, S.-W. Clustering-based collaborative filtering using an incentivized/penalized user model. **IEEE Access**, v. 7, p. 62115–62125, 2019. Citation on page 25.

VALL, A.; DORFER, M.; EGHBAL-ZADEH, H.; SCHEDL, M.; BURJORJEE, K.; WIDMER, G. Feature-combination hybrid recommender systems for automated music playlist continuation. **User Modeling and User-Adapted Interaction**, Kluwer Academic Publishers, USA, v. 29, n. 2, p. 527–572, apr 2019. ISSN 0924-1868. Available: <https://doi.org/10.1007/s11257-018-9215-8>. Citation on page 27.

ZANKER, M.; JESSENITSCHNIG, M. Collaborative feature-combination recommender exploiting explicit and implicit user feedback. In: **2009 IEEE Conference on Commerce and Enterprise Computing**. [S.l.: s.n.], 2009. p. 49–56. Citation on page 27.

ÇANO, E. Hybrid recommender systems: A systematic literature review. **Intelligent Data Analysis**, v. 21, p. 1487–1524, 11 2017. Citation on page 27.

ICMC USP
SÃO CARLOS