BGRS: A Board Games Recommender System

Camila Denecken
Pontificia Universidad Católica de Chile
Santiago, Chile
camila.denecken@uc.cl

Ignacio Diaz
Pontificia Universidad Católica de Chile
Santiago, Chile
iediaz@uc.cl

Sofía Escobedo
Pontificia Universidad Católica de Chile
Santiago, Chile
siescobedo@uc.cl

Rafaela Kara
Pontificia Universidad Católica de Chile
Santiago, Chile
rkara@uc.cl

Figure 1: Internationale Spieltage SPIEL, 2021.

ABSTRACT

This study focused on developing a Board Games Recommender System (BGRS), with the aim of achieving an advanced content-based system. A dataset from the BoardGameGeek's Board Game Database [1], encompassing information on 22,000 games, 411,000 users, and 19 million ratings was utilized. For method evaluation, significant metrics such as MAP@K and NDCG@K were employed, with K values set at 10, 20, and 30. Results were compared with baseline methods Most Popular, Random, ALS, BPR and SVD to gage the performance of content-based recommendation. Though results were suboptimal, analysing the recommendations given by BGRS for certain user profiles gave valuable insight about the problem at hand, being a good starting point for future work in building a recommender system for board games.

ACM Reference Format:
Camila Denecken, Ignacio Diaz, Sofía Escobedo, and Rafaela Kara. 2023. BGRS: A Board Games Recommender System. In *Proceedings of A Board Games Recommender System (BGRS)*. Santiago, Chile, 4 pages.

1 INTRODUCTION

Within the growing community of board game enthusiasts, research has identified a lack of specialized recommendation systems. This study aims to overcome the limitations of common models such as most popular, random, SVD, ALS, and BPR by introducing an advanced content-based recommendation approach.

The essence of this method lies in the careful analysis of user profiles and game attributes, with the aim of providing customized suggestions to improve the board game selection process.

This innovative system is designed to revolutionize the way board game enthusiasts discover and select their games, offering a distinctly personalized and enriching selection experience.

2 STATE OF ART

Content-based filtering focuses on predicting item relevance by analyzing the content of data items in relation to a user's profile. This approach, which is situated at the intersection of Information Retrieval and Artificial Intelligence, addresses the diverse nature of items and the intricacies of user information needs. Drawing from Information Retrieval concepts, user's information needs in content-based recommender systems are encapsulated within their profiles, and the challenge arises in describing items like web pages or documents with unstructured text. Document modelling rechniques rooted in Information Retrieval become essential in such scenarios. From an Artificial Intelligence perspective, the recommendation task is treated as a learning problem that harnesses historical user knowledge. User profiles, often in the form of keywords or rules reflecting longterm interests, are learned rather than explicitly provided by users. Machine Learning techniques play a crucial role in categorizing new information items based on past

BGRS, December 14, 2023, Santiago, Chile

user preferences, generating predictive models to determine the likelihood of an item being of interest to a user.

Alternative item representation techniques ranging from traditional text representation to more advanced methods integrating ontologies and encyclopedic knowledge have been explored. These techniques are vital for effectively capturing the diverse attributes of items, especially those described through unstructured text. There are recommendation algorithms tailored to these diverse item representations, each with computational strategies that underlie content-based recommender systems. [3]

3 DATASET

The dataset utilized for this project, sourced from Kaggle, is the Board Game Database from BoardGameGeek [1]. It includes data on 22.000 games, 411.000 users, and 19 million ratings, with various attributes like theme, mechanics, and subcategories of a specific game. The key datasets used were ratings, user ratings, games, and mechanics. These datasets were cleansed of null values and filtered to target specific games and users, improving algorithm efficiency. Due to the extensive size, ratings were limited to 300.000, resulting in a dataset of 20.061 games and their respective mechanincs, and a dataset of 300.000 user ratings from 590 users.

In the data segmentation for training, validation and testing, users were split into 80% for training and 20% for testing, leading to 472 and 118 users respectively. This segmentation guided the creation of datasets by filtering the dataset of ratings, ensuring no user overlap in training and testing. The validation set maintained one ´(user, item, rating)´ for each user in the training set, with both sets containing ratings from 472 users. The final dataset sizes were 193.426 for training, 48.357 for validation, and 57.330 for testing

4 METHODS

We propose a Board Games Recommender System (BGRS), which is a content-based recommendation system that uses some of the games' features to identify which games to recommend to a user, based on games they have already interacted with.

We began by using a BERT encoder [2] for the embedding of the main text features of the games table. The initial features we considered were `Description`, `Category`, `Mechanics`, `Themes` and `Subcategories`. We consider these features to be able to characterise a game in such a way that BGRS is going to be able to identify similar games and pair them to make a recommendation.

Using the embeddings for said features we calculated the cosine similarity between all of them to create a matrix that would later allow us to compare games to each other.

We explored various combinations of the mentioned features looking for the best results, which resulted in a pipeline shown in Figure 2. The pipeline starts by using the `Categories` feature to generate 60 recommendations for a game, from those it selects 40 games given their similarity in the `Mechanics` feature. Finally, from those 40 recommendations it selects 30, but now uses the `Description` feature for similarity.

Using this pipeline, BGRS generates recommendations for a user by applying it to each game the user has interacted with, then sorting the results by the number of times each game is repeated,

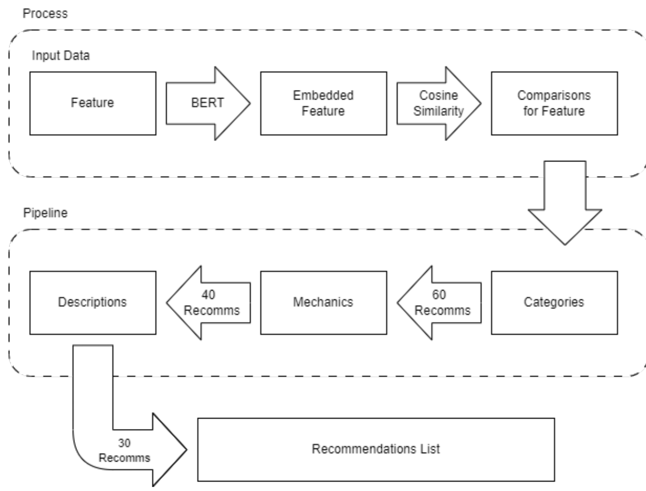

Figure 2: Diagram illustrating the functioning of the pipeline.

i.e. a game is higher on the recommendations list when it was recommended for more games the user has interacted with. Then it gives the top 30 games on the resulting list.

5 RESULTS

For the assesment of the different methods, two evaluation metrics were utilized: MAP and NDCG. The reaserch analyzed performance metrics including NDCG@10, NDCG@20, NDCG@30, MAP@10, MAP@20 and MAP@30. For ALS, BPR, and SVD, a range of latent factors — 50, 100, 200, 500, and 1000 — were explored to determine the configurations that produced the best result. Specifically, SVD and ALS achieved optimal results with 1000 latent factors and BPR with 200 latent factors. The results can be observed in Figure 3 and Figure 4.

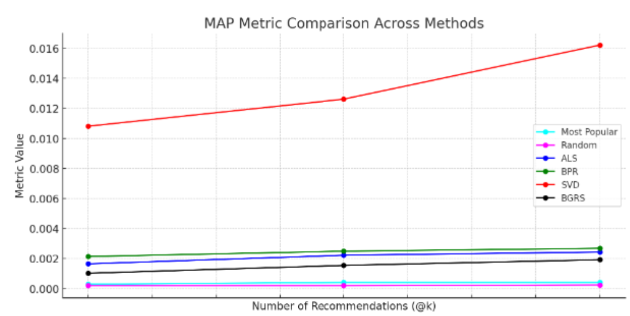

Figure 3: Comparison of MAP metric across different recommendation methods.

For the MAP@k metric, SVD has the best results by far, with BPR and ALS following. These three methods outperform our content-based method, although it does manage to have better evaluation than Most Popular and Random.

For the NDCG@k metric, once again SVD, ALS and BPR outperform our content-based method, this time ALS and BPR joining

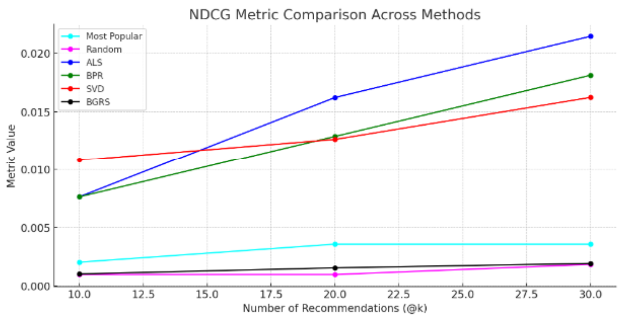

Figure 4: Comparison of NDCG metric across different recommendation methods.

SVD in having far better results. Most Popular is slightly better than our approach, and only Random is evaluated worse.

6 DISCUSSION

The first aspect we consider relevant to discuss is the decision of which features to consider for the recommendations. The main reason was the results we got when experimenting with different combinations and orders for the 5 features initially considered. In most cases the metrics worsen when using `Subcategories` and `Themes`, which we mostly ascribe to the sparsity of those features. This lack of data may have translated to worse recommendations since it frequently meant BGRS had no information to work with.

Secondly, to address the low values obtained in the metrics shown in section 5, we did a qualitative analysis of some examples from our train and validation set. Here, we found two cases, which repeated a lot in said datasets, that may explain the poor performance of BGRS.

The first case is when the user choices are consistent in train and validation, but the games' metadata does not reflect it. Table 2 shows an example, where a user liked war related games, which is noticeable by their titles and further investigation, but in terms of metadata, the one from the training set has Dice Rolling as its mechanism and the one from validation has no information for said feature, therefore, BGRS recommends a Dice Rolling game, but that does not match with the ground truth answer form the validation set. Also, neither the train or validation set games had information for their category, but both had the word 'war' in their names and the recommended game is categorised as a War game. With this example, we can see that the system is able to recognise some of the main features of the games interacted with by the user, but since not all the information is given it ends up failing anyway.

The second case is when a user has eclectic tastes, which means, the games used to generate the recommendations are very different from the one used for validation and in most cases there is not enough information in the metadata that fully characterises the games the user interacted with. Table 1 shows the example where the games for training and validation have completely different categories and mechanisms, and the recommended game shares a category with the training one, but without information for its mechanism. Nevertheless, by their name we can see that the recommended game belongs to the same series as the one in the training set, which shows that the recommendations are consistent with the information used to generate them.

Feature	Training Sample	Validation Sample	Recommendation Sample
Name	Vanguard of Wars	Alien Wars	Paths of Glory
Mechanism	Dice Rolling	-	Dice Rolling
Category	-	-	Wars

Table 1: Example from user 6372, where the Training and Validation samples are not similar, so the model fails since its recommendation is similar to the Training sample, not the Validation

Feature	Training Sample	Validation Sample	Recommendation Sample
Name	1825 unit 2	Pisa	1825 unit 1
Mechanism	Set Collection	Simultaneous Action Selection	-
Category	Strategy	-	Strategy

Table 2: Example from user 13035, where the Training and Validation samples are similar but the recommendation is not able to reflect said similarity due to lack of information

Thus, we can see that the recommendations made by BGRS are consistent in terms of content with the given games, still this seems to be insufficient when recommending for a user, who may have more complex taste than what can be modelled by each game they have interacted with separately. Hence, we conclude that, to reach better results it is necessary to include information about the user that allows BGRS to consider a wider perspective on the user's preferences in a way that it can give better recommendations according to them.

7 CONCLUSIONS

Even though the recommendations obtained with the content-based method developed were in line with what was expected from the user profile based on their previous games and their categories, mechanics and descriptions, the results were suboptimal and only surpassed the random baseline method consistently, with the methods SVD, ALS and BPR outperforming. By analysing further, we attributed this to two main cases: when the user choices are consistent in train and validation, but the game's metadata does not reflect it; and when the user has eclectic tastes, thus the games used to generate the recommendation are different from the ones in validation and could not be reached with only the information gathered by the game metadata through content-based recommendation.

We can conclude that the implemented method may not be the best to work with this data, because some crucial information is missing from columns used to gather information but also because users can have a varied taste on board games. For future work, the studied method could be used as a tool to join a different recommendation method such as Collaborative Filtering.

REFERENCES

[1] 2021. Kaggle. https://www.kaggle.com/datasets/threnjen/board-games-database-from-boardgamegeek?select=bgg_data_documentation.txt Accessed: october 2023.

[2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and*

Short Papers). Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186. https://doi.org/10.18653/v1/N19-1423

[3] Marco de Gemmis Pasquale Lops and Giovanni Semeraro. 2011. Content-based Recommender Systems: State of the Art and Trends. In *Recommender Systems Handbook*. 73–105. https://doi.org/10.1007/978-0-387-85820-3_3