Board games group recomendation using metadata

Eduardo Salinas [1] **Alfonso Badilla** [1] **Nicolás Gutiérrez** [1]

Abstract

In this document we propose a group recommendation system for board games using data from the Board Game Geek website, which includes attributes such as the number of players, average playtime, complexity, and user ratings. The implementation employs Python and the scikit-learn library, offering an efficient and scalable solution for group-based game recommendations.

1. Motivation and state of the art

Board games have been a popular form of entertainment for centuries, and their popularity has only increased in recent years. Even as digital games have become more prevalent, board games have remained a favorite pastime for many people, and the market for board games has grown significantly giving rise to plataforms such as Board Game Geek[1] (BGG). This website is a popular platform for board game enthusiasts, where users can rate and review games, as well as access information about them.

1.1. State of the art and related work

This plataform has a vast amount of data that has been used in the past to create recommendation systems. For example, github user richengo (2021) proposed a recommendation system for board games based on user ratings, using a collaborative filtering approach. However, this approach does not take into account the characteristics of many of the games, that are supposed to be played in groups. As such, we propose a group recommendation system to make recommendations tailored to groups of players.

Another example of group recommendation systems is the work of (Peska et al., 2023), who proposed a group recommendation system for tourism and movies[2]. This system uses regular clustering algorithms to group users based on their ratings, such as user k-nearest, item k-nearest and support vector decomposition (SVD), and then recommends movies and touristic locations based on the preferences of the group.

1.2. Proposal

We propose a refined version of that last system, using metadata-centric clustering to group board games based on their characteristics, and then recommending games to groups of players based on their summed preferences used the functions detailed on the previous work. This system is designed to be more accuarate than the previous one, as it takes into account the characteristics of the games, and not just the ratings of the users.

2. Data analysis

The dataset that will be used for this project[3], obtained from the website kaggle, has 10 different files, each one tith different information about board games. The main file is `user_ratings.csv`, which contains 19 million rows. Due to the limited processing power of Google Collab and our own local machines, we will use a reduced set of items and users for this work. The other files are:

- `games.csv`: This file contains 22 attributes about the rated games and includes 22,000 different games. Among the attributes, we have aspects such as the game description, which could be evaluated using language models, the year of publication (which distributes as shown in figure 1), the average rating received, among others.
- `ratings_distribution.csv`: Contains the total ratings for each board game. In general, these ratings distribute as shown in figure 2.
- `temes.csv`: Contains the themes of each game as a binary flag. The 50 most common themes can be seen in the graph in figure 3.
- `mechanics.csv`: Contains the game mechanics as

[1]Department of Computation, Pontificia Universidad Católica de Chile, Santiago, Chile. Correspondence to: Eduardo Salinas <esalinasbarros@uc.cl>, Alfonso Badilla <alfonso.badilla@uc.cl>, Nicolás Gutiérrez <njgutierrez@uc.cl>.

Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021.

[1](BoardGameGeek, 2024)

[2]This work is hosted on a github repo and is intended to be used as a basic tutorial.

[3](Wadkins, 2023)

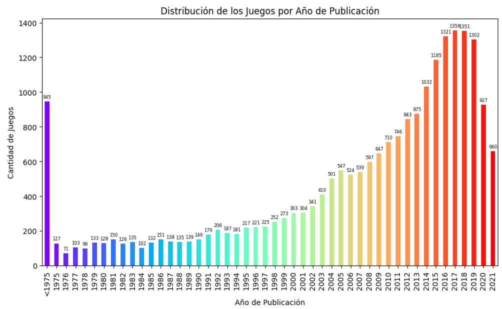

Figure 1. Juegos por año de publicación

a binary flag. The 50 most common mechanics can be seen in the graph in figure 4.

- `subcategories.csv`: Contains the subcategories of each game in binary flag format.
- `artists_reduced.csv`: Contains information about which artist created a certain game in binary flag format. Only artists with more than 3 games created will be considered.
- `designer_reduced.csv`: Contains information about which designer designed a certain game in binary flag format. Only designers with more than 3 games created are considered.
- `publishers_reduced.csv`: Contains information about the companies that sell these games in binary flag format. Only companies that sell more than 3 games are considered.

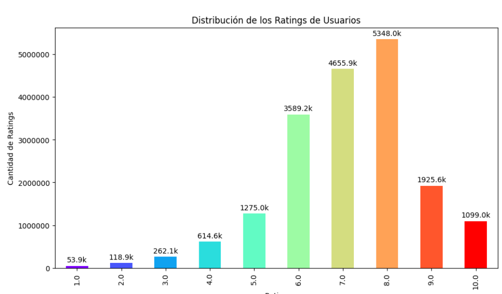

Figure 2. Distribución de ratings de juegos

3. Methodology

The methodology for this project is divided into three main stages: data preprocessing, group making, and recommendation.

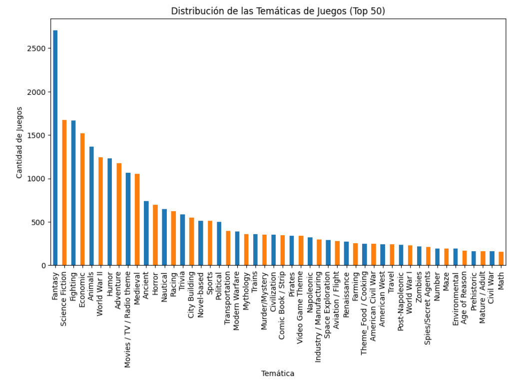

Figure 3. Temáticas más comunes en el dataset

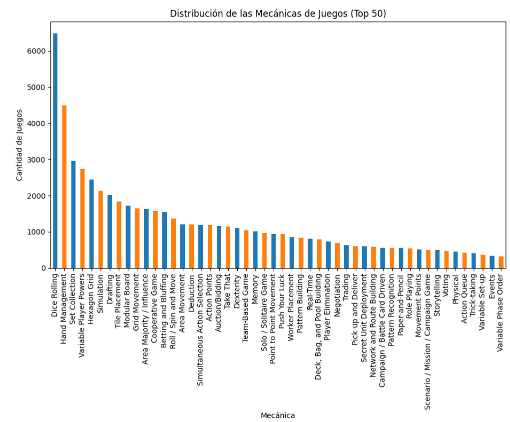

Figure 4. Mecánicas más comunes en el dataset

3.1. Data preprocessing

The first stage consists of reducing the dataset by not picking items without interactions and then dividing the...

3.2. Group making

The second stage consists of grouping the users based on their characteristics by using a correlation matrix. The function used is an adaptation of the one used in the ”” work, which is based on the cosine similarity between the users. The function is as follows:

(1)

3.3. Recommendation

The third stage consists of recommending games to groups of players based on their preferences. Since the recommendations are supposed to be made to groups, the preferences of the group are calculated using a variety of methods, that can be described as follows:

1. Simple average: The preferences of the group are calculated as the simple average of the preferences of the individual players.
2. Least misery: The preferences of the group are calculated as the minimum of the preferences of the individual players.
3. ...

4. Results

To evaluate the results of our model, we will compare it to most popular as a sanity check, to iknn applied to the same dataset and to SVD applied to the same dataset.

The metrics that will be used to evaluate the performance of each model are the following:

- Precision at k: This metric measures the proportion of recommended items that are relevant to the user, where relevance is defined as the items that the user has interacted with.
- Recall at k: This metric measures the proportion of relevant items that are recommended to the user.
- F1 score: This metric is the harmonic mean of precision and recall, and is used to measure the overall performance of the model.

5. Sensitivity Analysis

6. Conclusions

7. Figures

You may want to include figures in the paper to illustrate your approach and results. Such artwork should be centered, legible, and separated from the text. Lines should be dark and at least 0.5 points thick for purposes of reproduction, and text should not appear on a gray background.

Label all distinct components of each figure. If the figure takes the form of a graph, then give a name for each axis and include a legend that briefly describes each curve. Do not include a title inside the figure; instead, the caption should serve this function.

7.1. Citations and References

Please use APA reference format regardless of your formatter or word processor. If you rely on the LaTeX bibliographic facility, use `natbib.sty` and `icml2021.bst` included in the style-file package to obtain this format.

Citations within the text should include the authors' last names and year. If the authors' names are included in the sentence, place only the year in parentheses using yrcite.

Alphabetize references by the surnames of the first authors, with single author entries preceding multiple author entries. Order references for the same authors by year of publication, with the earliest first. Make sure that each reference includes all relevant information (e.g., page numbers).

Please put some effort into making references complete, presentable, and consistent. If using bibtex, please protect capital letters of names and abbreviations in titles, for example, use {B}ayesian or {L}ipschitz in your .bib file.

References

BoardGameGeek. Boardgamegeek: The world's largest board game database, 2024. URL https://boardgamegeek.com. Accesed december 7th 2024.

Ngo, R. Board-games-recommender, 2021. URL https://github.com/richengo/Board-Games-Recommender. Accesed december 7th 2024.

Peska, L., Antonetti, L., et al. Group recommenders: Offline evaluation, 2023. URL https://github.com/barnap/group-recommenders-offline-evaluation. Accesed december 7th 2024.

Wadkins, J. Board games database from boardgamegeek, 2023. URL https://www.kaggle.com/datasets/threnjen/board-games-database-from-boardgamegeek. Accesed december 7th 2024.