Peer Consulting Report

Board game rating predictor

31 December 2023

Module 3 project of Ramona Herz & Lara Nonis

Consultation report by: Patricio Becerra

Content

1 Project Summary

BoardGameGeek [1] is an online forum and games database for board game hobbyists. It holds information on over 125’000 tabletop games, including game data, reviews, images and videos. BoardGameGeek (BGG) allows users to rate games on a scale of 1-10 and publishes a ranked list of board games based on these ratings. This ranking is a measure of the games’ popularity and is considered an important reference value among hobbyists (and thus a proxy for market success of a game).

The goal of the “Board game rating predictor” project by R. Herz and L. Nonic is to develop an algorithm that can predict the BGG rating of a new game or a game under development using a series of characteristics or features of the target game. The project was conceived as a tool for game developers to get an idea of the potential popularity or market value of their game before its launch, allowing them to fine-tune certain aspects of the game to improve its chances of success.

The project is split into two phases. Each phase will make use of different machine learning techniques:

- Phase I has the objective of finding the features of a game (as recorded in BGG) that have the most effect on the game’s rating. These features could be e.g. number of players, mechanism of play, average play time. This phase will use unsupervised learning to find the features that most influence the target variable (rating).
- Phase II will use the information from Phase I to develop the predictor algorithm and provide a single predicted rating value for any game. This phase will use regression techniques (i.e., supervised learning) to find the rating prediction from multiple game features.

Each game entry can be identified by a game ID, and example game features on the BGG database include game name, minimal and maximal number of players, rated “weight” (or difficulty), year published, mechanisms, categories, and several others. Naturally, the target features are the game ratings, for which BGG provides two values, the “average rating” (the arithmetic mean of the user ratings), and the “Bayesian rating”, which uses several parameters (not disclosed by BGG), but presumably is calculated using additional known information about the game.

The data for the main project is obtained via web scraping of the BoardGamesGeek platform through its API [2]. However, in addition, the project’s authors performed two exploratory data analyses (EDAs), the first of which was performed on a subset of the BGG dataset from February 2023 that had been uploaded to Kaggle [3]. They compared the variability of the data to that of their own, up-to-date web-scraped data, which provided them with an idea of how often their model would need to be re-trained to prevent data-drift. The total dataset contains 2000 game entries (rows) and 39 features, but after data reduction and cleaning, 28 native features were used, plus an additional 5 created by the authors, amounting to a total of 33 features.

The next step in the data flow after web-scraping and cleaning, consists of the two EDA’s (in which statistical tendencies on the Kaggle and the updated datasets were explored). At this point the data would be ready for Phase I, for which the output is the identification of the features that most impact a game’s rating. Subsequently, a second cleaning step reduces the dataset to make use of only these relevant features. This final cleaned dataset is used in Phase II, which consists of a multiple regression model that predicts a single rating for a game based on the multiple features that were downselected at the end of Phase I. This second Phase follows the usual supervised learning flow of Training (on a subset "training dataset") → Validation (on a subset "validation dataset") → Optimisation → Testing (on "test data").

The output of the complete algorithm is the predicted Bayesian rating for any game for which the relevant features or characteristics can be known.

2 Key strengths

The proposed project and its preliminary results (up to Module 3) have many strengths to highlight. The main overarching strength is the existence of a clear goal with a potentially useful and even marketable outcome: A "Board game rating predictor" would be a unique tool that (1) uses readily available interesting data that is continuously updated (2) has the potential to make an impact in its field of use (one could easily see the usefulness of such a tool for game designers) and (3) makes use of most of the tools learned during the CAS up to Module 3, such that the authors can acquire substantial experience in these areas.

Key strengths in the approach to the project are as follows:

- The Phase split is logical, easy to understand, and advantageous to the objective of the project. It makes sense to perform unsupervised learning to pull out the most relevant features before applying a regression model. Although a direct supervised learning approach with all features available is probably feasible, it would take a significantly larger amount of computational time for likely little gain in terms of model accuracy and precision.
- Performing the two EDAs (first on the Feb 2023 Kaggle dataset and then on the current web-scraped set) is a clever way of finding out, from the very beginning of the project, how the end tool would be affected by data-drift, and therefore how it would need to be used in a realistic scenario (i.e., frequency of re-training).
- The authors used several ML methods in both the unsupervised (Phase I) and supervised (Phase II) learning phases of their project. During Phase I, the authors tested PCA, UMAP and GMM clustering, and finally settled on a best approach combining UMAP and dimensionality reduction techniques. During Phase II, linear regression, decision trees, random forests, and XGBoost were all tested, compared and used to feed knowledge into each other, with Random Forests and XGBoost showing the best results, and XGBoost being the algorithm chosen by the authors for optimisation. This approach is comprehensive, ensures a good predictor, and also allows the authors to familiarise themselves with many ML methods during the development of their project.
- The use of the Bayesian rating as an output value rather than the average rating is a great choice of target feature by the authors, as it would likely be a much more useful value for game developers to predict the success of a future game. As per ChatGPT: "the Bayesian rating given by BGG is a more nuanced approach that considers both the average rating and the reliability of that average based on the number of ratings. The Bayesian rating is particularly useful in situations where some games might have a small number of highly polarized reviews, which could skew the results if not properly accounted for."

Finally, it is worth noting that the code and Jupyter notebook written by the authors is well documented and relatively easy to understand and follow. The authors developed their code in an orderly, and advanced manner, saving their “final model” as an object using the joblib Python library.

3 Suggestions for Improvements

A few suggestions for improvements of the approach were brought up during the Module 5 session, and are noted here. In a general sense, the main area for improvement was deemed to be the data visualisation.

The first suggestion was to add graphical views and outputs of the intermediate model results, in order to make the model and code itself more user friendly and customizable. This suggestion was followed by the authors before their Module 3 presentation and is visible in their final notebook.

Secondly, an evaluation of the level of importance of each original feature was not present in the analysis, especially in a graphical display. I consider this quite important for the user to attain a complete understanding of the output of Phase I, i.e., why certain features are considered over others, and how much more important certain features are than others. This suggestion was followed by the authors to a limited extent before their Module 3 presentation, i.e. only after fine-tuning the hyperparamters for the XGBoost regressor, and limited only to a visualisation of categorical game characteristics, such as “role-playing” or “cooperative game” (see their Jupyter notebook). It would be useful to also add intermediate explanatory visualisations that explain why and how the full dataset features are selected by each type of clustering method in Phase I. In addition to better data visualisation, this could potentially add a certain degree of user customisation, e.g. a game developer could choose to rely on a particular feature over another based on their individual goals or vision for the game under development.

Finally, the model could benefit from further exploration of the dataset, e.g. large subsets of games could be divided into macrocategories or major “types” of games (e.g. Board Games, Abstract Games, Customizable Games, Children’s games, etc.). A rating predictor that considers only the data pertaining to a specific macrocategory may provide a more accurate and useful prediction in a realistic scenario. In their Jupyter notebook, the authors briefly mention the possibility of implementing this improvement as a selection widget, which would indeed be an optimal way to add such a functionality.

4 Conclusions

The “Board Game Rating Predictor” project is based on a compelling concept and goal. Up to this point, it is being developed with a professional and well-ordered approach. Aside from the risks noted by the authors in their Module 1 report [4], only minor weaknesses are readily observable at this point, though further improvements may be necessary as the project develops.

References

[1] BoardGameGeek. (2023). BGG. https://boardgamegeek.com

[2] BoardGameGeek API, https://boardgamegeek.com/wiki/page/BGG_XML_API2

[3] Kaggle, Top 2000 Board Games Ratings, Nikita Fedorov, https://www.kaggle.com/datasets/nfedorov/top-2000-board-games-ratings

[4] R. Herz & L. Nonis (2023), Board games' rating predictor, Conceptual Design Report. CAS Module 1 Project Report.