Unveiling modern board games: an ML-based approach to BoardGameGeek data analysis

Dien Nguyen, Joshua Kritz, Raluca D. Gaina, Diego Perez-Liebana
Queen Mary University of London - Game AI Group
d.l.nguyen;j.s.kritz;r.d.gaina;diego.perez@qmul.ac.uk

***Abstract*—There has been growing interest in modern board games, which have been increasing in complexity with respect to their classic counterparts (e.g. Chess, Go), by utilizing new mechanics and novel ways to interact with them, resulting in richer player interaction. Boardgamegeek.com (BGG) is the biggest forum for board games and it now has registered 191 different mechanics. Users can rate games on the forum and BGG will rank them accordingly. This work aims to investigate how mechanics relate to player ratings using a Decision Regression Tree (RT) to predict the expected rating based on a game's mechanics. To achieve this we collect mechanics and player ratings data of all ranked games on BGG and train our Regression Tree. After training the RT and further extending it with Random Forest (RF), we use Mean Decrease in Impurity (MDI) and Permutation Feature Importance (PFI) to evaluate how much each mechanic influences the player ratings.**

We show that, using only game mechanics, Regression Tree and Random Forest can account for 28% and 32% of the variance in games' ratings, respectively. We highlight the interpretability of RT and how it can be used to gain insights into the relationship between game mechanics and player ratings.

***Index Terms*—board games, player ratings, machine learning, decision tree, regression tree, random forest, feature importance**

I. Introduction

Modern board games are a part of the games market that has seen a steep increase in interest and size since 2010 [1]. For researchers, board games have been a topic of interest even earlier. They have been using them applied to fields like health and education. Even before that, classical board games like Chess, Go, Checkers, etc. have seen a lot of research interest, particularly in Artificial Intelligence (AI), where the focus is on creating game-playing agents [2].

However, serious research on the structure and design of modern board games is more recent than those applications. The first-ever catalog of game mechanics was published in 2019 [3] while the Game Ontology Project, a big catalog for digital games, [4] started in the early 2000s.

One problem of board game research is the difficulty of obtaining data. BoardGameGeek.com (BGG) is a forum with over 100,000 board game entries and thousands of users [5] and the only major source of board game data. Although it is readily available and has a substantial amount of data, it is still lacking in many aspects, most of the data provided are crowd-sourced by users and not experts. While this does make it invaluable as a source of player opinion, it lacks the polish necessary for some research. Many researchers need to adapt the data to fit their intended work [1].

BGG data has been used to inform research, as either a selection criteria for games based on user ratings [6], or to better differentiate between mechanisms [7]. Other authors analyze BGG data directly, with mixed results, to evaluate the relation between game attributes and player motivation, or to highlight a positive correlation between game attributes and success (and related board game trends) [8].

Samarashinghe et al. [1] pushed this idea by making a thorough data analysis, limited to only the top 10,000 games on BGG. They investigated the correlation between mechanics and player-voted attributes, such as complexity, rating, and duration. Using correlation and co-occurrence analysis, they conclude that neither the number of mechanics in a game, nor the average mutual information of mechanics contributes to complexity and rating. However, they draw this conclusion using simple statistical methods. On the other hand, our work considers every ranked game (over 20,000) and applies machine learning methods to analyze the importance of each individual mechanic in predicting a game's rating.

For this work, we obtain the data on board games available on BGG using their public API and filter it to games with at least 30 user ratings (the minimum requirement for a game to be ranked on BGG). Regression Tree [9] and Random Forest [10] are fitted on the filtered data to predict the player rating of a game based on its mechanics. Using mean decrease in impurity (MDI) and permutation feature importance (PFI) [10], we analyze the relationship between each mechanic and game ratings through feature importance, i.e. how much does the presence (or absence) of a mechanic affect how predictive the model is.

We aim to improve on existing research by using a Machine Learning (ML) approach to obtain new insights into how mechanics can be used to understand player ratings. This contrasts with previous works where only traditional statistical techniques were used.

Additionally, we highlight the advantages of using a simple Regression Tree model to understand how mechanics interact with each other to affect player ratings. The simplicity and interpretability of the tree structure will provide insight for designers if combinations of mechanics behave differently than their separate implementations.

Our objective is to analyze the relationship between BGG ratings and the game mechanics. We can then identify what mechanics have a high impact on ratings and which have little impact. Game designers will be able to make a more informed decision when creating their game's mechanics, and publishers will also benefit from understanding which mechanics yield better results.

With this, we provide important insight for future research by investigating how mechanics relate to ratings and introducing a new framework for analyzing features of board games. Furthermore, this can provoke designers to investigate particular mechanics and why they are successful.

II. Data

The data provided by the BGG API contains all information shown on an item's BGG page[1]. The Python code used for data collection is available on GitHub[2] The fields we collected are detailed in the metadata.yaml file in this same repository. BGG differentiates games with a unique identifier *gameID* for every game. Our raw data had 117,859 unique IDs.

We are mainly interested in the data fields: ratings, mechanics, and re-implementations.

Rating average is the average rating of all BGG users who rated the game. This is our target for prediction. We filter out games with less than 30 ratings (i.e. we take only the ranked games).

The **mechanics** in each game are features that we use to predict the rating of a game. The raw data has 191 different mechanics, which means it features all possible mechanics in the BGG list. The most frequent mechanic (Dice Rolling) is present in $26,774$ games while the least frequent (Auction: Compensation) is only in 4 games. Notably, half of the mechanics appear in only 264 games, less than 0.01% of the full dataset.

BGG has also a record of game **re-implementations**, which is a game that has been re-released either as a new edition with revised rules or with a different theme. As they are the same game, to avoid giving more weight to a particular set of mechanics, we merge re-implementations with identical mechanics into a single entry with a rating equal to the weighted average of all the merged entries.

After removing non-ranked games and filtering the re-implementations we have a total of $23,113$ games. The highest rating average in the resulting dataset is 9.70 and the lowest is 1.16. Its histogram (omitted due to space) adheres approximately to a normal distribution with a mean of around 6.45 and a standard deviation of 0.93.

III. Methodology

We use Decision Regression Tree [9] and Random Forest [10] to predict the rating average of a game based on its mechanics. Each mechanic is a binary feature, valued 0 and 1 for its absence and presence, respectively.

[1]Catan, for example: https://boardgamegeek.com/boardgame/13/catan

[2]https://github.com/HuntedSouls/BGGMechsAndRatings

Decision Regression Tree (RT) fits our data and purposes well with its nonparametric approach, inherent binary structure, and interpretability. The nonparametric approach of RT makes little assumption about the relationship between the features, the response variable and the underlying data distribution, which is essential for our investigation into the relationship between mechanics and players' ratings.

Our tree is built to maximize the variance reduction of the training samples, i.e. the difference between the training samples variance and the weighted average of the variance at each leaf node:

$$I_S(\mathcal{T}) = Var(S) - \sum_i \frac{|L_i|}{|S|} Var(L_i) \tag{1}$$

Here, $\mathcal{T}$ is the regression tree model, S is the set of training samples, and L_i is the set of samples at the i^{th} leaf node. $Var(X)$ is the ratings' variance of samples in X. Our goal is to maximize $I_S(\mathcal{T})$.

Therefore, each leaf node is optimized to contain board games with overlapped mechanics and similar ratings to minimize its variance. Consequently, RT automatically selects mechanics that are most relevant to the ratings in a hierarchical structure.

Since searching for the optimal structure for RT is NP-complete [11], we use a greedy best-first search to decide what to split. At each step, we select which node and which feature to split the node to maximize the variance reduction on the leaf nodes.

Random Forest (RF) is an extension of RT, where predictions are based on ensembles of regression trees. Each tree in RF is intentionally built to overfit a subset of data sampled with replacement from the training data (bootstrapping). At each split of a tree in RF, only a random subset of features are considered as candidates (*candidate set*).

Fitting multiple trees within the model makes RF less interpretable than RT, but in turn more robust with better predictive power.

To analyze feature importance we use **Mean Decrease in Impurity** (MDI) and **Permutation Feature Importance** (PFI). MDI [9] is simply the normalized total variance reduction of a feature whenever it is used for splitting a node. In other words, the MDI of a feature is how much of variance reduction by the model can be attributed to that feature. For example, a mechanic with MDI 0.09 means 9% of the variance reduction of the model is due to a game having or not having that mechanic.

The permutation feature importance [10] of a feature measures the decrease of the model's score on test data when values of the feature are permuted in the data.

Due to the weakness of greedy search and MDI's bias for local variance reduction, it may not be the most reliable metric to measure feature importance. On the other hand, PFI makes no assumption on the structure of the model, which results in a more robust and general metric.

Model	R^2 score
Regression Tree (RT)	0.277 ± 0.01
Random Forest (RF)	0.315 ± 0.008

TABLE I
REGRESSION TREE AND RANDOM FOREST AVERAGE R^2 SCORE AND STANDARD DEVIATION ON TEST DATA IN 100 RUNS

IV. EXPERIMENT AND ANALYSIS

Scoring We use R^2 as the metric to evaluate how predictive our models are on the test data, as shown in Equation 2:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - f(x_i))^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{2}$$

N is the number of data points, x_i is the mechanic feature vector and y_i is the rating of the i^{th} game, respectively, $\bar{y}$ is the average of ratings, and f is our model. The fraction is the ratio between the mean-squared error of the base model (i.e. all inputs are predicted with the data average) and our trained model. Thus, we interpret R^2 as the portion of the rating variance that our models can explain using only mechanics. That is, given a set of games, R^2 is how much their rating variance can be accounted for by the differences in mechanics.

Hyperparameters The Regression Tree model is constrained to have a minimum of 20 samples per leaf node. The Random Forest model uses 50 trees, candidate set of length $\lfloor\sqrt{D}\rfloor = 13$ where D is the total number of mechanics (191), and bootstrapped samples of size $N_{Tr}/4$ where N_{Tr} is the size of the training data. These hyperparameters are found to perform the best based on preliminary trial and error.

We shuffle and split our data into a training set and a test set; the training set has $N_{Tr} = 18,000$ games and the test set has $N_T = 5,113$ games. We fit our models on training data and evaluate their R^2 score on test data for 100 runs; for each run, the data is shuffled differently. Table I shows each model's average R^2 score and its standard deviation.

The average R^2 score of RF and RT are 0.315 and 0.277, respectively. RF performs better while also appearing to be more stable against data shuffling across runs, evident by RF's lower variance. Nonetheless, RT by itself is not much worse, capable of explaining 28% of the variance in board games' ratings from their mechanics alone.

The R^2 scores of both models suggest that there are relationships between the mechanics of a game and its ratings.

Feature Importance Analysis With the same 100 runs, we collected the MDI of each feature in each run. PFI of both RT and RF is evaluated with 30 permutations per feature. PFI is done only once due to computational cost. Table II illustrate the top-5 scoring in each analysis.

The top-5 mechanics are virtually unchanged across all metrics of feature importance. *Roll / Spin and Move* is consistently the most important mechanic by a large margin in predicting the rating of a game. Splitting the entire dataset by the presence/absence of this mechanic shows that games with it have an average rating of 5.45 ± 0.026, while games without it have an average rating of 6.48 ± 0.007 (Figure 1).

	Mechanic	Tree	Random forest	%
MDI	Roll / Spin and Move	0.220 ± 0.005	$0.107 \pm^0 .006$	5.4
	Solo / Solitaire Game	0.143 ± 0.004	0.066 ± 0.005	5.4
	Variable Player Powers	0.081 ± 0.004	0.044 ± 0.003	10.9
	Simulation	0.077 ± 0.005	0.042 ± 0.003	8.5
	Hexagon Grid	0.045 ± 0.004	0.044 ± 0.003	9.9
PFI	Roll / Spin and Move	0.104 ± 0.007	0.064 ± 0.004	5.4
	Solo / Solitaire Game	0.053 ± 0.006	0.047 ± 0.004	5.4
	Hexagon Grid	0.048 ± 0.004	0.041 ± 0.004	9.9
	Variable Player Powers	0.039 ± 0.003	0.032 ± 0.003	10.9
	Simulation	0.036 ± 0.004	0.031 ± 0.003	8.5

TABLE II
FIVE MOST RELEVANT MECHANICS ACCORDING TO MDI AND PFI ACROSS 100 RUNS. THE PERCENTAGES REPRESENT THEIR PRESENCE IN THE DATASET. RED HIGHLIGHT FOR NEGATIVE EFFECT ON RATINGS AND GREEN FOR POSITIVE.

This echoes the consensus among many board game players and designers that *Roll / Spin and Move* is a bad mechanic [12].

On the other hand, every other mechanic in the top-5 has a positive impact on the ratings when they do appear. Tree nodes that are split by these mechanics tend to have higher averages on the child nodes where these mechanics are present.

Another interesting fact is that the second most represented mechanic, *Hand Management*, does not feature in the top-5 mechanics. Even though it is largely present in games it does not have a big impact on the rating of the game. This mechanic is present in very high-rated games like Brass: Birmingham and Gloomhaven, both rated above 8.5. Whilst also featured in games like Boogie Beast (5.1 rating) and Druids (4.8 rating).

Comparing the MDI and PFI numeric values between RT and RF also supports the narrative that RF is more robust than RT. Although the ranking of features is similar, the importance of the first few features is not as extreme in RF as they are in RT, implying RF is less dependent on the top mechanics to make accurate predictions on the rating.

Regression Tree as a data exploration and analysis tool. With such a comprehensive dataset, one might want to query a subset of mechanics with the best rating. An exhaustive analysis of every subset of mechanics is infeasible as there are 2^{191} possible subsets.

The tree recursively partitions the dataset in a way that the set of board games at any node is optimized to be as similar as possible in both mechanics and ratings. Each node in a tree represents a subset of mechanics that meaningfully contribute to ratings, along with its estimated average for that subset.

In order to find the subset of mechanics with the best rating, the reader can observe Figure 1 and descend from the root node of the tree to the node with the highest average. In this example, this node is the green-colored one (7.958 rating) representing games with no *Roll and Move*, but with a *Solo Game* option, *Variable Player Powers*, and *Variable Set-Up*. This subset of mechanics is most likely not the best, but it is a decent approximation. Due to the tree structure, it is possible to look at each node on the path, see how tweaking a mechanic affects ratings, and analyze its effect on the destination node.

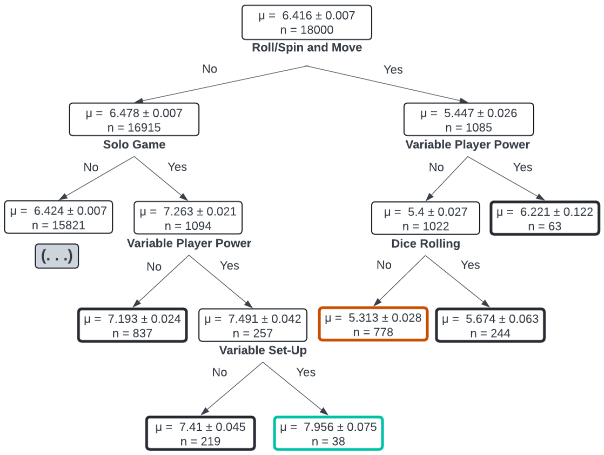

Fig. 1. Regression tree (top nodes). In each node, μ is the average rating with its standard error, n is the number of games. Bolded nodes are leaf nodes. The green and orange outline indicate the nodes with the overall highest and lowest average rating, respectively.

V. Conclusion

This research focuses on understanding and analyzing the relationship between mechanics and player ratings using the BoardGameGeek (BGG) data.

First, we analyzed the entirety of the BGG database, deviating from previous research [1] where only limited data was analyzed. We define our approach focusing on player ratings as a function of individual mechanics. This allows for more granularity in our analysis and deeper insights into the relationship between mechanics in player ratings.

Second, we show that, by using only mechanics as the input features, a simple Regression Tree can achieve validation R^2 score of 0.28 in predicting the ratings of board games, while Random Forest improves the score to 0.31. This implies a relationship between mechanics and player ratings, contrasting with previous research. Using the Mean Decrease in Impurity (MDI) and Permutation Feature Importance (PFI) of the trained RT and RF, we demonstrate that some mechanics have more impact on player ratings than others.

Third, we introduce Decision Regression Tree as a tool and framework for data exploration and analysis of board game data. We explain how one can interpret the tree structure in the context of mechanics and how the algorithm automatically learns an important subset of mechanics.

VI. Future Work

Results taken from quantitative analysis of BGG data should be considered concerning a niche of players that rate games in BGG. Although this can be a useful proxy to measure players' reception, it does not provide an accurate measurement of the whole population. In particular, it is not a good indicator for the more casual player who does not use BGG often. Research conducted directly with this audience could cover this gap, perhaps by using questionnaires at places this audience goes to, like board game cafes and local stores.

Furthermore, BGG's current state of mechanics tags is not consistent, as identified by its extreme redundancies and under-classification. Predicting ratings based on mechanics is a valuable result in itself. The fact that at least 27% of the variance in ratings can be explained by a simple Regression Tree based only on mechanics bears further investigation by both researchers and designers. We aim to expand on this with further analysis to identify if the mechanics are dependent on each other concerning the ratings, that is, each mechanic contributes individually or they depend on which other mechanics are present in the game.

Using tree-based models for analyzing board game data is a promising avenue of research to address the explainability issue of other ML methods. Future research should focus on using extensions to the basic Regression Tree that address many of its weaknesses and apply them to the same dataset.

Improving tagging by creating a proper set of tags and retroactively re-evaluating all games for this new set will prove invaluable for future research. This can be done either through expert curation throughout the database, or by using Information Retrieval techniques on rulebooks or game synopsis [13].

Acknowledgments

For the purpose of open access, the author(s) has applied a Creative Commons Attribution (CC BY) license to any Accepted Manuscript version arising. This work was supported by the EPSRC Centre for Doctoral Training in Intelligent Games & Games Intelligence (IGGI) EP/S022325/1.

References

[1] D. Samarasinghe, M. Barlow, E. Lakshika, T. Lynar, N. Moustafa, T. Townsend, and B. Turnbull, "A Data Driven Review of Board Game Design and Interactions of Their Mechanics," *IEEE Access*, vol. 9, pp. 114 051–114 069, 2021, conference Name: IEEE Access.

[2] M. Genesereth, N. Love, and B. Pell, "General Game Playing: Overview of the AAAI Competition," *AI Magazine*, vol. 26, no. 2, pp. 62–62, Jun. 2005, number: 2.

[3] G. Engelstein and I. Shalev, *Building Blocks of Tabletop Game Design: An Encyclopedia of Mechanisms*. CRC Press, Mar. 2022, google-Books-ID: cnNYEAAAQBAJ.

[4] J. Zagal, "gameontology," 2005. [Online]. Available: https://www.gameontology.com/index.php/Main_Page

[5] "BoardGameGeek." [Online]. Available: https://boardgamegeek.com/

[6] M. Sousa, "Finding Ludemes in Modern Board Games: Analyzing the Top Number One Games of Board Game Geek," *Board Game Studies Journal*, vol. 17, no. 1, pp. 205–229, Oct. 2023.

[7] M. Sousa, N. Zagalo, and A. P. Oliveira, "Mechanics or Mechanisms: defining differences in analog games to support game design," in *2021 IEEE Conference on Games (CoG)*, Aug. 2021, pp. 1–8, iSSN: 2325-4289.

[8] Z. Du, "Data Analysis of Board Games' Fashion Trends and Relationships between Their Successes and Attributes," *Highlights in Science, Engineering and Technology*, vol. 24, pp. 131–142, Dec. 2022.

[9] L. Breiman, J. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. New York: Chapman and Hall/CRC, Oct. 1984.

[10] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[11] L. Hyafil and R. L. Rivest, "Constructing optimal binary decision trees is NP-complete," *Information Processing Letters*, vol. 5, no. 1, pp. 15–17, May 1976.

[12] M. O. K. Games, "Why Roll and Move is a Bad Idea," Aug. 2020. [Online]. Available: https://www.manokentgames.com/single-post/why-roll-and-move-is-a-bad-idea

[13] M. Mitra and B. Chaudhuri, "Information Retrieval from Documents: A Survey," *Information Retrieval*, vol. 2, no. 2, pp. 141–163, May 2000.