Data mining 7 Wonders, the board game

Joaquim Assunção, Gabriel Pereira, Júlia Acosta, Rafael Vales, Lana Rossato
Department of Applied Computing
Universidade Federal de Santa Maria (UFSM)
Santa Maria, Brazil
{joaquim, ggpereira, jgacosta, rvales, lbrossato}@inf.ufsm.br

***Abstract*—Board games can be used as objects of study for artificial intelligence (AI), as their rules are well defined and good players can challenge ordinary AI. However, unlike Chess or Go, most board games are not deterministic, as they rely on dice or cards. The stochastic factor is a challenge as a supervised learner should learn rules that lead to victory. These rules should be, ultimately, knowledge extracted from data. In this article, we describe a knowledge discovery and data mining process (KDD) for the board game 7 Wonders. From data selection to analysis and evaluation, we mined the data to get the implications of playing specific types of cards in specific situations.**

***Keywords*-Data mining, boardgames, 7Wonders**

I. Introduction

The state-of-the-art in artificial intelligence for games are using reinforcement learning supported by an initial supervised learning step [1]. Such combination of techniques means that even deep learning models need a first step of supervised learning, so the model can learn the essential rules for a complex game. Furthermore, deep learning models often heavily relies on GPUs and are time consuming, which makes the use of an initial supervised learning useful for practical purposes.

In a game of checkers, chess, or go, we have a relatively simple structure and tons of logs to be learned from an algorithm. In more complex electronic games, such as StarCraftII, we have vast amouts of data and we can use bots to play thousands of matches in a few days [1]. However, for many other board and card games, we do not have enough data for fed complex models. Furthermore, we also lack a fast forward machine vs machine for deep learning training. In fact, we do not even have prior information for a winning strategy or, in many cases, even a game log (hand by hand). For many games, we only have match statistics, which describes a match as a whole (not action by action).

Considering such scenario, in this work we concern with a step prior to the supervised learning. We focus on data mining game data to get the basic knowledge for a winning strategy, which can be further fed as a class for a supervised learner.

II. Related Work

There is a myriad of works that uses AI to play games. However, our work distinguishes by a couple of reasons. First, it is for data mining for a specific board game, 7 Wonders, with a set of restrictions described in Section I. Second, as far as we know, it is the first work to get association rules for 7 Wonders. Third, it is based on the match data of the best players from a large community of players.

Although our contribution might become evident for the unique aspects of the work, we highlight a few related works that have a similar approach in some aspect.

Regarding card games, Poker, for instance, a few works use Case-based reasoning (CBR) to improve the bot performance against humans, which can be compared to ordinary human players [2], [3]. CBR, however, requires a large number of cases and a considerable modeling effort and initial knowledge, which might impact on the performance if the model has to be changed [4]. When the initial knowledge is a crucial factor, and there are no game logs, a data mining process can be useful to acquire the initial knowledge for another technique.

Siqueira *et al.,* [5] and Odierna *et al.,* [6] used data mining to analyze and identify patterns in the players' behavior, as well as to classify these behaviors and players. Siqueira *et al.,* worked on mining the game World of Warcraft. They used methods clustering and regression models to identify patterns of players in order to indicate if the player will stop playing in the near future.

Oliveira *et. al.,* [7] worked to form good team compositions to increase the victory rate in the game League of Legends. The data were collected from professional matches, observing the choice of characters and the outcome of the game. They generated a tree with several combinations of possible good compositions. By using linear regression, they identified the team with a higher chance of winning, helping the player to put together a composition that increases a team overall chances of winning.

Robilliard *et. al.,* [8] wrote about the Monte-Carlo tree search algorithm for creating an AI for the 7 Wonders game. They implemented using a Monte-Carlo tree search with susceptible levels, in which the nodes correspond to the possibilities of plays. A second AI was implemented deterministic, using fixed rules. In the end, they compared the two AIs, showing that the first had better results.

We use data mining process to achieve winning patterns from players' behavior (like in [5], [6]), using the best players, considering different amounts and combinations of players per table (Similar to [7]) of 7 Wonders. However, unlike Robilliard *et. al.,* [8] we collected prior knowledge, for the creation of a bot, through data mining association rules.

III. 7 Wonders Overview

7 Wonders is a board game where 3 to 7 players receive a board representing one of the seven Wonders of the ancient world. The game is split into three ages. In each era, cards are distributed. These cards are divided into seven types: civil, scientific, commercial and military structures, raw materials, manufactured goods, and guilds. Figure 1 displays how is the board of a player during a match from Board Game Arena (BGA) [1].

Figure 1. Mausoleum of Halicarnassus' board along with the cards played. A screenshot, from BGA, that shows only the part of the cards that matters; productions, shields, commerce, points, etc.

Civil, scientific and guild cards generate victory points, commercial cards provide coins or advantages in purchases in the commerce, and military cards grant shields for conflicts. Cards related to raw materials and manufactured goods generate the necessary resources for the construction of other structures.

Like most card games, 7 Wonders is a stochastic strategic game in which players deals with other players action. After each age, the players have to ponder or change their strategies based on the situation in the first age. In a stochastic formalism, considering the computational cost, the different types and values for every combination of cards could generate millions of states. At the end of the third age, the victory points of each player are counted. Detailed rules can be found in the game manual [2].

IV. Mining 7 Wonders Match Logs

Since our game logs correspond to the final results, we do not have the information hand by hand. Thus, we need to find useful information using only the data corresponding to the final statistics of each match. Figure 2 shows the steps to extract information regarding the cards that make the difference for a player victory. We can use these pieces of information as an input for a supervised learning approach; *i.e.,* use this data as a class for a learner algorithm to learn the best card to play.

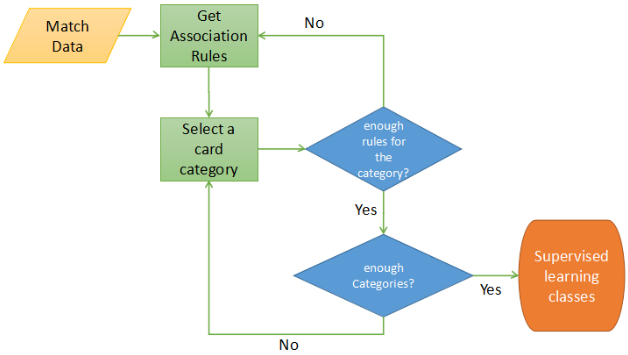

Figure 2. Workflow to generate decisions for the supervised learning algorithm.

To define a minimum number of rules and categories, we created a series of questions regarding the game. These questions were based on a few analysis made by experienced players and the board game community, for instance, the *boardgamegeek* forums [3].

The next subsections illustrate our data mining process, from data selection to evaluation, which was based on the KDD process. Figure 3 illustrates the process.

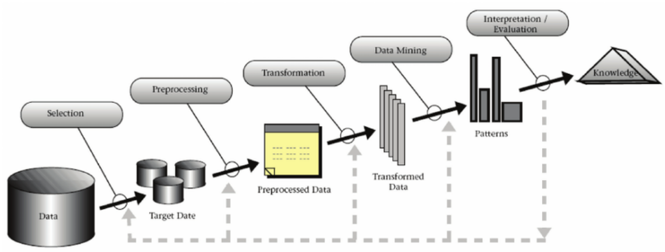

Figure 3. Knowledge discovery process, illustrated by Fayyad et al., [9]

A. Data selection

Our data were collected from Board Game Arena (BGA) [1], a well-known website that has many online implementations of board games, among them, 7 Wonders, which have a large number of games played daily. At the end of each match, the site provides statistics of the game regarding the actions of each player. There are 25 different attributes in the statistics, such as victories points from each type of cards, the amount of each type of cards played, the number of stages built of wonder, etc.

In order to extract and format the data, we implemented a script that generates a CSV file. We pick the best player in the BGA ranking as a source for collecting the data. Thus, we have the results belonging to the best players in the entire

[1] www.boardgamearena.com

[2] www.7wonders.net/wp-content/uploads/2017/06/7WONDERS_RULES_US.pdf

[3] https://boardgamegeek.com/thread/691370/some-complex-strategies-7-wonders

platform. Since the game configuration is different depending on the number of players, the data was separated into five datasets, one for each number of players, with 50 matches each [4].

B. Pre-processing and Transformation

Some players leave the game before it is finished, which prevents its sequence. When such interruption occurs in the BGA, the game is canceled and a tie is declared between all the other players. However it is kept in the game history, leading to incomplete and incoherent statistics. Since these data are not relevant, we verified all the datasets, removing the matches that finished before the time. Thus remaining only complete games with consistent data to be analyzed.

For the use of the Apriori algorithm, the data was transformed into presence matrices. To fill each matrix, and use different amount of cards, we defined parameters such as a arithmetic mean, quantile, as well as fixed values. Thus, values that are greater or equal to a parameter were converted to one, otherwise it was assigned 0.

C. Getting association rules

Using the presence matrices, we applied the APRIORI algorithm to generate the association rules. The algorithm works from two given parameters: the minimum support and the minimum confidence. Support that an attribute, or a set of attributes, A implies in an attribute, or a set of attributes, B, is given by: $Sup(A \Rightarrow B) = P(B|A) = \frac{A \cap B}{total(T)}$. Confidence that $A \Rightarrow B$ is given by: $\frac{A \cap B}{A}$ [10]. A and B are also commonly represented as Left Hand Side (LHS) and Right Hand Side (RHS).

We generated two larger sets of rules, from different matrices, one based on generated symbols, for instance, military cards may hold one to three shields; and the other based on the number of cards. In each of those sets, and for each number of players, we used as a discrimination metric the constants one and two, the first quantile, the average, and the third quantile were chosen as parameters for analysis. The constants are merely for mark the use of a card, one or two times, where the average and the quantiles seek to obtain patterns on the amount of use for each card or type of cards. For instance, if the average number of yellow cards was 3, the presence matrix will generate a 1 for every player in a match in which he/she used more than 3 cards. Using these categories, we got a total of 5796 rules with a support of 50%. However, it was not enough for some categories; for instance, the third quartile for 3 players got only 10 rules. Hence, we kept lowering the support until a set of interesting rules was reached, ending at 20%.

V. Results

A. If a war is inevitable, face it or perish

Considering victories in 3-player matches, a pattern emerges with the use of military cards. For the number of military cards played, above the average, the following rule was caught: $military \rightarrow victory$ with a support of 0.26 and confidence of 0.54. Furthermore, among the winners, 78% of them launched a high number of military cards. $victory \rightarrow military$ with a support of 0.26 and confidence of 0.78. This demonstrates that investing in military structures, in a 3-players game, may be one of the main factors leading to victory, as the player directly confronts his two opponents. This may seem obvious, yet it is not the case analyzing statistics from ordinary players (the ones not well ranked in BGA).

The confidence of this rule falls as the number of players increases. Hence, more players in the table, the less important is the use of the military cards, as the players are not directly confronting each other.

B. Coins should to be spent

TABLE I
MORE COINS THAN THE AVERAGE

Players	LHS	RHS	Support	Confidence	Lift
3 players	treasure=1	defeat	0.29	0.64	0.97
4 players	treasure=1	defeat	0.21	0.71	0.89
5 players	treasure=1	defeat	0.25	0.78	0.94
6 players	treasure=1	defeat	0.31	0.71	0.95
7 players	treasure=1	defeat	0.39	0.85	0.99

Table I shows us that have too much money is bad in the game. Considering the players' average, for those who accumulate more than half of the total value (treasure), they tend to lose. This is especially true for a seven-player game. The obvious reason is that every 3 coins score 1 point. The not so obvious reason is that fewer players, and large amounts of coins, implies in debt for the other players (point transferring), mitigating the effect. The balance is subtle, but it is more advantageous to spend with resources that are base for building cards that grant more victory points at the end of the game.

C. Trading leads to money

TABLE II
MORE COMMERCIAL CARDS THAN THE AVERAGE

Players	LHS	RHS	Support	Confidence	Lift
3 players	commercial=1	treasure=1	0.32	0.65	1.46
7 players	commercial=1	treasure=1	0.26	0.65	1.42

Table II shows strong rules regarding the relation between commercial structures and large amount of coins. Even the confidence was a match of 65% for both 3 and 7 players, which indicates that the commercial effects remains strong independent of the number of players. This rule is bonded with the previous one, which is not a good strategy since coins above the average, in most cases, leads to defeat, as seen in Section V-B. Therefore, commercial structures should be played in moderation, trying to stay in the absolute minimum for guarantee enough resources.

These rules become obvious when we analyze the points generated by the cards in the last age; therefore, depending to the height of the game, may not help as it takes a turn that should be used to play stronger cards.

[4] Our data and scripts are publicly available at https://github.com/dmag-ufsm/7wondersDataMinning

D. A science career demands sacrifices

TABLE III
MORE SCIENCE CARDS THAN THE AVERAGE

Players	LHS	RHS	Support	Confidence	Lift
3 players	scientific=1	raw materials=0	0.25	0.63	1.41
3 players	scientific=1	military=0	0.30	0.77	1.48
3 players	scientific=1	commercial=0	0.26	0.67	1.20
3 players	scientific=1	civilian=0	0.32	0.81	1.32
3 players	scientific=1	guild=0	0.31	0.79	1.18

Science cards are known for their cumulative property, where a combination of equal symbols scores 2^n if $n > 1$, 1 otherwise. Also, each set of all three symbols gives +7 points. These properties imply that a player going for a science strategy should take at least 3 of a kind to take advantage of science cards. Our rules detected that usually, successful players, in a three player game, tend to sacrifice everything else to go to science. The necessary production and raw material for playing the cards is replaced with the cards chain and commerce; even when that means sell cards for 3 coins. Table III shows these rules, which are the ones with the highest confidence in our sets. An exception to this rule is the manufactured goods, shown in Table IV. This is expected since the first scientific cards lead to the others and, each one cost one different manufactured good. This holds true for the matches with 3 to 7 players, but the strongest rules were found in 3 and 6 players matches.

TABLE IV
MORE SCIENCE AND MANUFACTURED GOODS THAN THE AVERAGE

Players	LHS	RHS	Support	Confidence	Lift
3 players	scientific=1	manufacture=1	0.30	0.76	1.19
6 players	scientific=1	manufacture=1	0.25	0.71	1.07

E. Use woods for buildings

Table V describes the action of overproduction, *i.e.,* players that create too many raw materials (above the third quantile) tend to lose. Both the support and confidence are strong for these rules. Players performing such actions to guarantee the best cards or incoming money on the trade. However, this strategy does not pay off because it is better to pay a few coins for the best cards than lose two or three cards to create raw material.

TABLE V
RAW MATERIAL ABOVE THE THIRD QUANTILE

Players	LHS	RHS	Support	Confidence	Lift
3 players	raw materials=1	defeat	0.38	0.71	1.06
5 players	raw materials=1	defeat	0.49	0.83	1.00
7 players	raw materials=1	defeat	0.43	0.86	1.00

Another case is the underproduction, which the best players know well and solve this problem with commerce (see Table VI). Otherwise, when investing in materials (near the average), players do not invest heavily on commerce, *i.e.,* an average production of raw material implies in low use of commerce. These rules hold true for all combination of players but were slightly stronger on an odd number of players. This can be explained due to the distribution of the cards.

TABLE VI
RAW MATERIAL AND COMMERCE CONSIDERING THE AVERAGE

Players	LHS	RHS	Support	Confidence	Lift
3 players	raw materials=0	commercial=1	0.23	0.51	1.16
5 players	raw materials=0	commercial=1	0.25	0.62	1.12
7 players	raw materials=0	commercial=1	0.34	0.69	1.12
3 players	raw materials=1	commercial=0	0.34	0.62	1.10
5 players	raw materials=1	commercial=0	0.29	0.49	1.10
7 players	raw materials=1	commercial=0	0.23	0.45	1.19
3 players	manufacture=1	commercial=0	0.43	0.67	1.20
5 players	manufacture=1	commercial=0	0.20	0.59	1.32

VI. FINAL REMARKS

This work described a KDD process, focused on getting association rules for the board game *7 Wonders*. After extracting matches from the best-ranked players, we were able to extract interesting and previously unknown rules. As shown in Section V, such rules comprise different types of cards, with a different number of players, allowing a strategic advantage for most *7 Wonders* scenarios. As a work in progress, our next step is to create an AI using a supervised learning algorithm, based on the set of the strongest rules we collected.

REFERENCES

[1] O. Vinyals, T. Ewalds, S. Bartunov, P. Georgiev, A. S. Vezhnevets, M. Yeo, A. Makhzani, H. Küttler, J. Agapiou, J. Schrittwieser, J. Quan, S. Gaffney, S. Petersen, K. Simonyan, T. Schaul, H. van Hasselt, D. Silver, T. P. Lillicrap, K. Calderone, P. Keet, A. Brunasso, D. Lawrence, A. Ekermo, J. Repp, and R. Tsing, "Starcraft II: A new challenge for reinforcement learning," *CoRR*, vol. abs/1708.04782, 2017.

[2] J. Rubin and I. Watson, "Case-based strategies in computer poker," *AI Commun.*, vol. 25, pp. 19–48, Jan. 2012.

[3] I. Watson and J. Rubin, "Casper: A case-based poker-bot," in *AI 2008: Advances in Artificial Intelligence* (W. Wobcke and M. Zhang, eds.), (Berlin, Heidelberg), pp. 594–600, Springer Berlin Heidelberg, 2008.

[4] P. Reuss, M. Dick, W. Termath, and K.-D. Althoff, "Case-based reasoning: potential benefits and limitations for documenting of stories in organizations," *Zeitschrift für Arbeitswissenschaft*, vol. 71, pp. 252–258, Dec 2017.

[5] E. S. Siqueira, C. D. Castanho, G. N. Rodrigues, and R. P. Jacobi, "A data analysis of player in world of warcraft using game data mining," in *2017 16th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)*, pp. 1–9, IEEE, 2017.

[6] B. A. Odierna and I. F. Silveira, "Player game data mining for player classification," in *Proceedings of SBGames*, 2018.

[7] V. da Costa Oliveira, B. J. Placides, M. d. F. O. Baffa, and A. F. da Veiga Machado, "A hybrid approach to build automatic team composition in league of legends," in *Proceedings of SBGames*, 2017.

[8] D. Robilliard, C. Fonlupt, and F. Teytaud, "Monte-carlo tree search for the game of "7 wonders"," in *Workshop on Computer Games*, pp. 64–77, Springer, 2014.

[9] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "The kdd process for extracting useful knowledge from volumes of data," *Communications of the ACM*, vol. 39, no. 11, pp. 27–34, 1996.

[10] C. Kaur, "Association rule mining using apriori algorithm: a survey," *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 2, no. 6, 2013.