UNIVERSITÀ DEGLI STUDI DI MILANO

Information Retrieval Project

Study on the complexity of boardgames from the analysis of rulebooks

Giovanni Avati

April 2023

Contents

1 Project outline

Playing boardgames is a wonderful hobby that is growing in popularity in the last years. Since 2000, the website *BoardGameGeek* (BGG) provides a complete database about boardgames and users playing boardgames around the world. Users provide also stats and ratings that evaluate the popularity of each game according to several criteria, including the game complexity (called weight) (see for example the stats for the game Gloomhaven). For many games, either on the producer website or on some community forum (sometimes on the BGG page as well), the Rulebook is also available. See for example the games from GMT games that usually provide a pdf rulebook downloadable for each game. Goal of the project is to automatically analyze the rulebooks in order to define appropriate metrics and strategies for automatically evaluate the complexity score of the game on the basis of the rules. This score should then compared against the BGG weight score in order to discuss similarity and differences between the two metrics.

2 Building dataset

Due to the specificity of the task, I could not find a dataset that fit my needs. In particular, I was looking for a dataset with some information about games and for each of them the text of the rulebook. Consequently, I had to build the dataset by myself.

To build the dataset I use the *Esoteric Order of Gamers site*. From this website I could find a list of about 400 selected boardgames and for each of

them the website provide some pdfs including also the rulebook. Moreover, all the rulebooks are formatted in the same way by the website owners and this helped a lot the process of extracting the needed information.

2.1 Downloading rulebooks

So, the first step was to build a python script to automatically download all the rulebooks featured on the site. This was done by starting from this *page* and using the *Beautiful Soup library* to download only the pdfs containing the rules.

2.2 Extracting information from pdfs

After having downloaded all the rulebooks in the website, I built another script to open each rulebook and extract the needed information. The information that I retrieved from each boardgame were *name*, *year of publication*, *publisher* and the *text of the rules*. The rules are saved in two formats: one computed to be in lower case and without multiple blank spaces and line break, instead, the other one was saved in a raw form, exactly as it was extracted from the pdf. It will be clear later the reason for this distinction.

From this data a first Pandas dataframe is created.

2.3 Link dataset to BGG

The dataframe built in the previous step then has to be linked to the BGG information. Using the BGG API it was possible to retrieve the BGG id of each game starting from the name extracted from the pdf. Then, having the id of a game it is easy to get all the information in the BGG site about that game, such as the weight.

2.4 Final dataset

The final dataset, the one then used in the core script of the project, has the following columns:

- **ID** - The id that unequivocally identifies the game in the BGG database.
- **NAME** - The name of the game
- **YEAR** - The year of publication
- **PUBLISHER** - The company that published the game
- **RULES** - The text of the rules formatted
- **RAW_RULES** - The text of the rules in the raw form

After all the preprocessing and filtering done in order to have a clean dataset, the final dataset is of about 300 rows.

3 Custom weighting system

With the dataset ready, the goal is to work on the text of the rulebooks to compute a series of metrics and choose one, or more, among them that give us an idea of the complexity of the game and then compare it with the BGG weight.

3.1 Metrics

The metrics computed in the project are explained below.

3.1.1 Syntactic complexity metrics

The following 3 metrics are metrics used to measure the syntactic complexity of a text.

- **Gunning Fog Index**
  In linguistics, the Gunning Fog Index is a readability test for English writing. The index estimates the years of formal education a person needs to understand the text on the first reading. The index is calculated as:

$$0.4\left[\left(\frac{words}{sentences}\right)+100\left(\frac{complex\ words}{words}\right)\right]$$

- **Flesch–Kincaid grade level**
  The Flesch–Kincaid readability tests are readability tests designed to indicate how difficult a passage in English is to understand. The grade level is calculated as:

$$0.39\left(\frac{\text{total words}}{\text{total sentences}}\right)+11.8\left(\frac{\text{total syllables}}{\text{total words}}\right)-15.59$$

- **Automated Readability Index**
  The automated readability index (ARI) is a readability test for English texts, designed to gauge the understandability of a text. The index is calculated as:

$$4.71\left(\frac{\text{characters}}{\text{words}}\right)+0.5\left(\frac{\text{words}}{\text{sentences}}\right)-21.43$$

All these 3 indexes are similar and give us an idea on the syntactic complexity of a text. But the syntactic complexity of the text seems to not be a good indicator for the complexity of a boardgame. This because there a lot of games that are easy to understand but, at the same time, that are really complex to play. The opposite is also sometimes true.

3.1.2 Quantitative Metrics

Other metrics that I computed are more quantitative, such as:

- The **number of words**
- The **number of sentences**
- The **number of distinct words**

These metrics are based on the idea that the more complex a boardgame is, the longer its rules are likely to be. This could be a good assumption but these metrics are too simple to be significant.

3.1.3 Number of "*rules*"

The last metric is the one chose as the most significant to give us an idea on the complexity of a game.

I want to identify the number of "rules" that each game has, where for "rule" is meant an action or a group of actions that a player needs to do. This because, enriching the assumption stated above, if we can correctly identify the number of "rules" that a game has, then the more complex a boardgame is, the higher number of "rules" it is likely to have.

To identify a "rule", I started by identifying a list of words that represent an *action* that a player can do during a game. The list of words that I chose is {*'play', 'count', 'move', 'go', 'return', 'draw', 'reveal', 'control', 'place', 'gain', 'lose', 'win', 'roll', 'discard', 'use', 'shuffle', 'attack', 'defend', 'produce', 'receive'*}.

From this list, a **word2vec** model is trained with the rulebooks in the dataset in order to enrich the list of *action* words. I took the 10 most similar words that the word2vec model returned for each words, only if the similarity is greater than 0.5. Having now a much longer list of words, I removed from this list the word that are not verbs because I want only words that represent an action and so these have to be verbs.

With the list of *action words* ready, a possible approach could be to simply count these words, but based on the fact that a typical "rule" in a boardgame can contain more than one of these words, I decided to count the **number of "rules" as the number of paragraphs where some of these words appear more than 2 times**.

3.2 Custom weight

After having established what a "rule" is, it was possible to count them for each of the boardgame in the dataset. The distribution of the weight over the number of rules could be seen in Figure 1 below.

To get a reasonable comparison between the *rules count* metric and the BGG weight I compute a custom weight based on the rules count. This **custom weight** will work as follow:

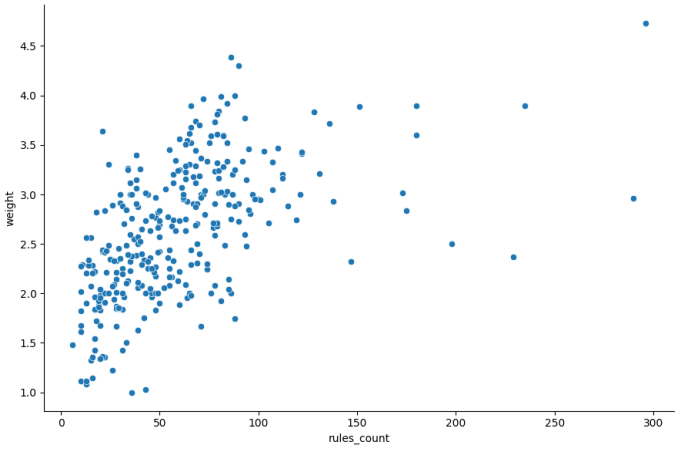

Figure 1: Relative plot between BGG weight and rules count

- As BGG weight, it goes from 1 to 5
- A *rules count* of 0 is equal to the lowest weight: 1
- A *rules count* of 160 is equal to the lowest weight: 5
- If a rulebook has more than 160 rules, the weight is capped to 5

All the rules above were decided empirically by the observation of the Figure 1.

According to the rules just mentioned it was possible to plot a graph with the custom weight and the actual BGG weight (Figure 2).

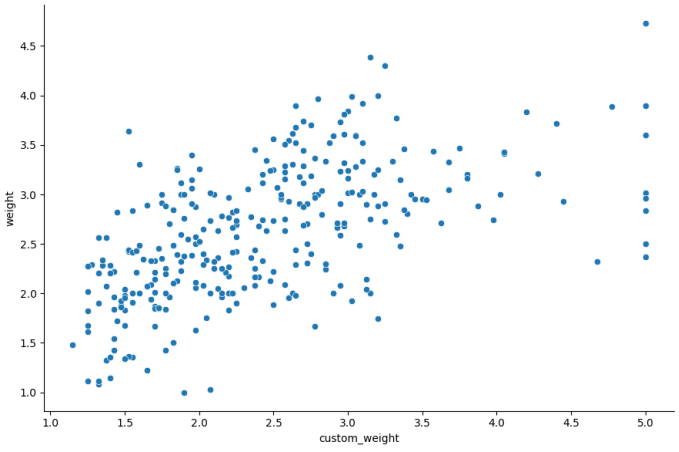

Figure 2: Relative plot between BGG weight and custom weight

3.3 Complexity labels

To simplify the comparison between the weight and the custom weight, the weight scores that were going from 1 to 5 were labelled as follow:

- **EASY** - For weights from 1 to 2.33
- **MEDIUM** - For weights from 2.34 to 3.66
- **HARD** - For weights from 3.67 to 5

In that way it is easier to understand how the complexity of a boardgame is classified by the two systems.

4 Results and Criticism

Based on the complexity labels identified above, it was possible to compare the labels from the two systems and understand how much this two type of weight differ. The label from the BGG weight and from custom system were compared in order to understand how many times the two label match, and the results of this analysis could be seen below.

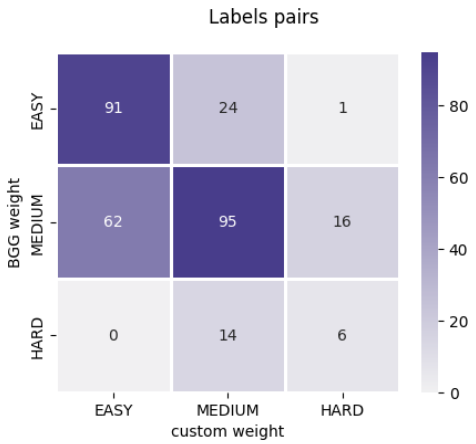

Figure 3: Heatmap of the labels combination from the two systems

As it is visible from the Figure 3, the results seem to be reasonable. In fact, the majority of pairs match on the EASY and MEDIUM labels. There is only a pair for the EASY-HARD or HARD-EASY match, that is a good result because means that is really rare for a game to be classified as EASY from a system and HARD from the other.

For the pairs that differ by only one label (e.g. MEDIUM-EASY, EASY-MEDIUM, etc.), the pair with most occurrences is the MEDIUM-EASY and this could be explained by the fact that a lot of games may have short rulebook

but they are complex for other reasons that are not the "number of rules", the parameter calculated by the custom system.

At the end, after the study of the rulebooks of boardgames it is clear that the syntactic complexity of a rulebook is not an indicator of the game's complexity. Instead, the metric choose to build the custom weighting system seems to be a decent indicator. The system, however, is very simple and so it has to be intended as a starting point.

Because of the fact that the text analysis of rulebooks alone can not give us a complete idea on the complexity of the game, in further implementation, could be interesting to add in the analysis some parameters that are not strictly related to the text of rulebooks. One of these could be, for example, the *average time of a game* but it could be possible to incorporate a lot of other metrics.