

Assignment 1

Bartosz Walusiak

February 2021

1 Abstract

GOG.com is an online store selling DRM-free video games. Because the games are DRM-free, new games made by today's Game Dev giants rarely opt to sell their games through GOG.com as that would make them easy to pirate. The catch is that GOG.com also takes time to make sure that each game, going as far back as 1980, works well with today's PCs, hence the name of the service: Good Old Games. Having launched in 2008, 28 years after the release date of the oldest game available in their store, and 13 years prior to time of writing, GOG's reviews are in a good position to be analysed year by year, to see which words are being used to review new vs old games both positively and negatively and how the popularity of different games over the ages is viewed from a modern perspective.

2 Data

The data available for analysis includes all games on the platform and their reviews. Relevant attributes of the games include but are not limited to: name, price, player rating, genres, OSes, age rating, release date, developer and publisher, languages and reviews - which are described by their author's handle, rating, time of posting, title and review description.

The main focus of this project is to correlate which words used to describe games over the entire period from 1980 - 2021 are most unique to those years. For example: it is expected that today's review of old games will likely talk about the gameplay, soundtrack and nostalgic feelings rather than groundbreaking graphics and effects.

3 Visualisation

An easy way to visualise word occurrence frequency is a word cloud. A word cloud displays all the words clumped together, increasing the font size according to the amount of occurrences. The user's first interaction with the data will be an overview which will display one big word cloud with all the unfiltered words in the center. A button underneath will invite the user to interact with the data. Once pressed the button will be replaced with a slider underneath, which ranges from 1980 to 2021. Moving the slider will allow the user to zoom in on a certain year, causing the word cloud to morph to represent the words most relevant that year. Check boxes will allow the user to filter the display to only show the words based on reviews of certain genres or publishers, or to remove non-key words like "the" or "and". A radio button will allow the user to split the word cloud into two different word clouds for positive and negative reviews. Clicking on the word cloud will show a window which displays the details and numeric values associated with the visualised data. Another option the user will have will allow them relate two different years and show which words are common between the two.

4 Case Study

4.1 Paper

In 2018 researchers performed sentiment analysis on Twitter posts about the game PlayerUnknown's Battlegrounds over a period of 8 months [1]. Their results, mainly concerned the different sentiment analysis methods and

algorithms and their accuracy, which were visualised using scatter plots and bar charts. Their algorithms were trained using Amazon's reviews, labeling reviews from 1 to 2.49 stars as negative, and from 2.5 to 5 as positive. The results of this training and the weight each word had on the sentiment was not published. Instead, the general subjects of the reviews were compared on a scatter plot (e.g. gameplay vs. graphics).

4.2 Article

Daniel McNihol, a Data Scientist working at the Philadelphia Inquirer, argues against word clouds in his article on towardsdatascience.com [2]. In his article, Daniel mentions multiple gripes with word clouds. To begin with, scale is absent from word clouds. The most frequent word appearing in a sample text is displayed, but not the number associated with it. The difference in frequency of occurrence is also not shown. Everything has to be approximated.

The data in a word cloud is also not ordered. The graph does not suggest where the user should start looking, where they should finish and no easy method of keeping track of what has already been viewed. The article then proceeds to explain a better approach. More formal and quantitative. Despite some valid points about the downsides of the word cloud, the focus on the project is to create a fun and interactive experience rather than clear and academic data.

4.3 describingwords.io

describingwords.io [3] is an online service that provides related or describing words for nouns. It was initially created based on crunching through Project Gutenberg - an online library of over 60000 free ebooks. This has since been expanded to include over 100 Gigabytes of text. This visualisation is a variance of the word cloud, ordering the words most used from left to right; top to bottom and denoting them with color. This is less engaging and does not immediately highlight the relevant words, however, the purpose seems to be writing assistance and not data visualisation, and this visualisation

method allows the user to easily read all the proposed words.

popular	bad	handheld	favorite	childish	clean	built-in
latest	fancy	illegal	perfect	interactive	classic	real-life
simple	global	online	violent	impossible	no-win	
endless no-win	black-and-white	portable	lousy	successful		
noisy	progressive	sole	hand-held	slow-motion	well-designed	
best	gigantic	various	sophisticated	modern	hush-hush	
asinine	expensive	standard	savage	high-end	eternal	specific
freelance	bizarre	top-of-the-line	recent	occasional	pointless	
bloody, pointless	bloody	frantic	dreamlike	nascent		
home-grown	pornographic	unimaginative pornographic	cosmic			
fossilized						

Figure 1: describingwords.io result for "Video Game"

5 Conclusion

The visualisation technique that I have chosen to represent this data misses out on the opportunity to create a formal and informative visualisation, choosing instead to make a fun and interactive site which approximates the information. A similar project was not found anywhere on the internet and this might be a nice addition to the world of data visualisation.

References

- [1] S. Chakraborty, I. Mobin, A. Roy, and M. H. Khan. Rating generation of video games using sentiment analysis and contextual polarity from microblog, 2018.
- [2] Daniel McNihol. Rip wordclouds, long live chatterplots. <https://towardsdatascience.com>, 2018.
- [3] Describing words for video game. <https://describingwords.io>.