Fei Guo
Victor Li
Jonathan Lin
Xinyu Zheng

Text Analysis of Video Game Reviews

Introduction and background:
The video game marketplace in North America has grown up tremendously from its birth in early 1980s. What was once a very small market competing for space with pinball machines in an arcade has grown to a multimillion dollar industry that employs tens of thousands of people to develop and create new games. This rise of a new entertainment industry coupled with the advent of the internet allowed for publications and websites of all sizes to begin covering the industry. This has resulted in a wealth of coverage on video games through previews and reviews.

Because the games journalism industry is a pretty niche market, prominent reviewers and commentators tend to write many reviewers over the course of their careers. This means that their preferences on certain game genres, types, and even companies can be tracked through their review scores. The volume of these reviews make for an exciting platform for studying the unseen biases and perceptions reviewers may have about games in their critical assessments.

So what:
Critical reviews of a game cannot serve the entirety of an audience because the reviewer has intrinsic biases towards certain genres and styles that may subconsciously color their review. If those biases could be revealed to the reader, the reader could go into the review knowing that his or her taste either matches closely with or does not match closely with the reviewer's taste. With this knowledge, the reader knows whether or not to give more or less credence to the reviewer when reading through their assessment of games.

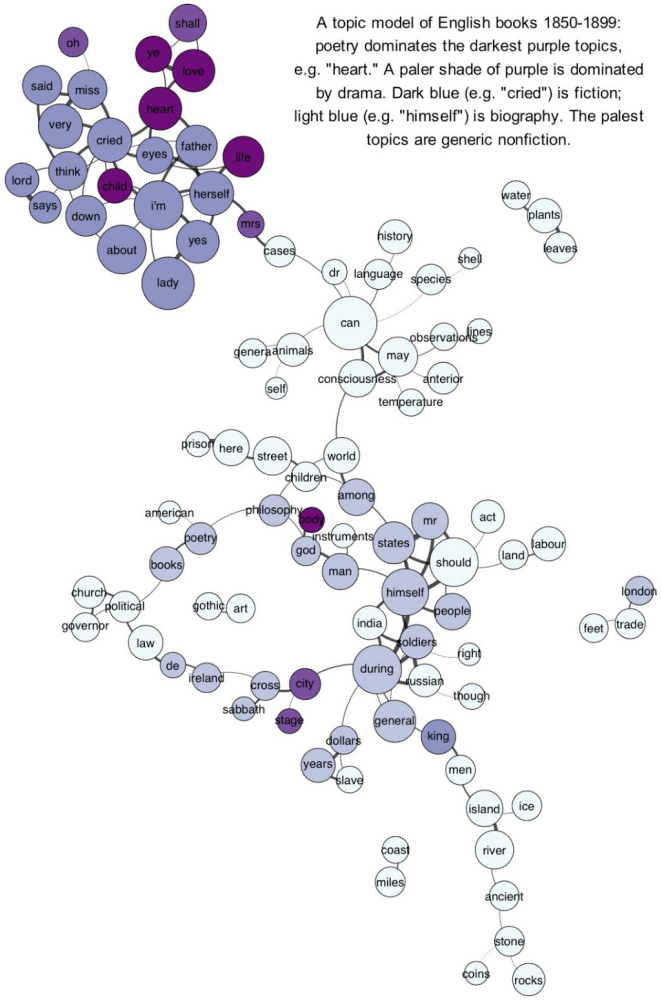
Question:
Can the writing tendencies and habits of a video game reviewer indicate their disposition towards certain genres and types of video games? Alternatively stated, can certain phrases and word choices in a review indicate a higher or lower score for a given product?

Data:
For this analysis, full text data of reviewers will need to be gathered. We will use web scraping to collect reviews data from game review websites such as giantbomb.com and gamespot.com. If website scraping proves to be too difficult, we may attempt to do transcriptions of video reviews for games from the websites which will take longer than the scraping. We could also try scraping the metacritic database to obtain links to published reviews. There is an unofficial API at

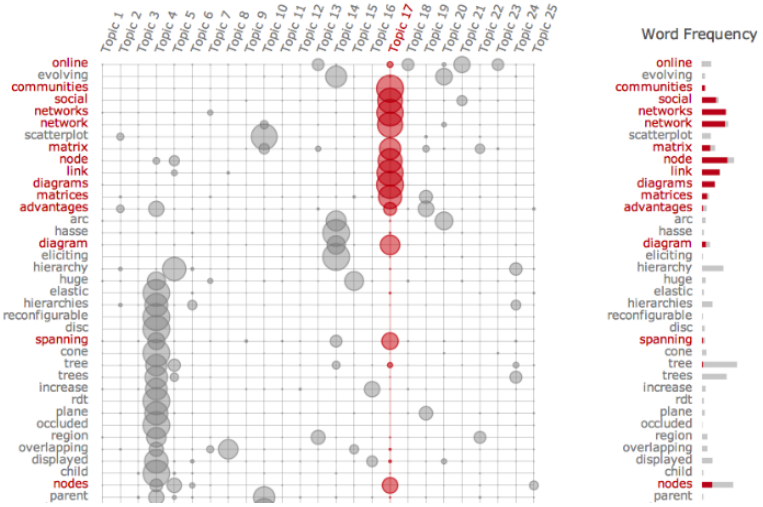
https://www.mashape.com/byroredux/metacritic-v2#get-critic-reviews that can make scraping review links easier. The larger problem would be pulling text from the differently structured review websites.

Analysis:
After scraping and cleaning text data, we plan to apply LDA model to review analysis using Python. First of all, we will divide games by types. For each type of games, we will identify phrases and words that appear more frequently in higher-score reviews and lower-score reviews, respectively. By comparing these phrases and words across game types and review scores, we will be able to infer whether they can indicate a higher or lower score for a given game.

Figures:
Some common text visualization techniques includes nodes-and-lines graph such as network paragraphs. As shown below:



The Stanford Vis Group, now UW Interactive Data Lab developed a tool called Termite, which “is a tabular view displays term-topic distributions for an LDA topic model”, and a “bar chart shows the marginal probability of each term”. Termite “can filter the display to show the most probable or salient terms. Users can choose between 10 and 250 terms”. Eventually Termite helps researchers to “identify coherent and significant themes.”



[Termite: Visualization Techniques for Assessing Textual Topic Models](#)
[Jason Chuang](#), Christopher D. Manning, [Jeffrey Heer](#), *Advanced Visual Interfaces*, 2012

Another great inspiration from the UW Data Lab is from the paper “[Interpretation and Trust: Designing Model-Driven Visualizations for Text Analysis](#)”. The paper puts individual texts, their case, PhD papers on a scatter plot and used the distance from the center of the group as a proxy for interpreting in many different ways. We favorite is this below:

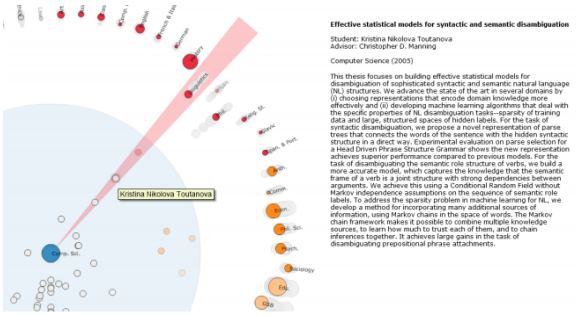


Figure 4. The Thesis View shows individual dissertations as small circles placed between the focus department and the next most similar department. Reading the original text of the dissertation enables experts to evaluate observed dept-dept similarities, and confirm the placement of three computational linguistics Ph.D.s that graduated in 2005.

[Interpretation and Trust: Designing Model-Driven Visualizations for Text Analysis](#)
[Jason Chuang](#), Daniel Ramage, Christopher D. Manning, [Jeffrey Heer](#)
ACM Human Factors in Computing Systems (CHI), 2012

Contingencies:
Scraping for text reviews across a variety of gaming websites could be tougher than initially planned. We will need to explore the complexity behind the websites prior to scraping and select which websites we will use. The text and natural language processing may also return results that may not be what we are expecting. Making adjustments to the text processing to ensure all documents are successfully analyzed may require some tinkering.