UC Berkeley
CS294-10 Visualization
Fall 2014

Visualizing Crowdfunding

Alex Chao
http://bit.ly/1zhjJet
Can be found on Github
github.com/alexchao56

Problem

Work has been done in predicting the success of crowdfunding campaigns.

However, it is a largely untapped data source for visualization. Most visualizations that exist just show the accuracy of prediction algorithms, but do not look at the data itself. Crowdfunding data is mostly textual.

No tools thus currently exist to visually explore crowdfunding data.

Motivation

With the rise of websites like Indiegogo and Kickstarter, crowdfunding has become an increasing technological phenomenon that will only scale with time. With access to scraped data from campaign pages on Indiegogo, the opportunity was ripe to do exploratory data analysis and visualizations.

Figure 1: A word cloud showing the frequency of terms from user comments of unsuccessful campaigns in the business/tech category. It can be compared to the chart below showing a word cloud for comments for successful campaigns in the same category.

Figure 2: One can see that both successful and unsuccessful campaigns in the business/tech category share similar words. It appears as a whole that people use positive words in their comments on campaigns.

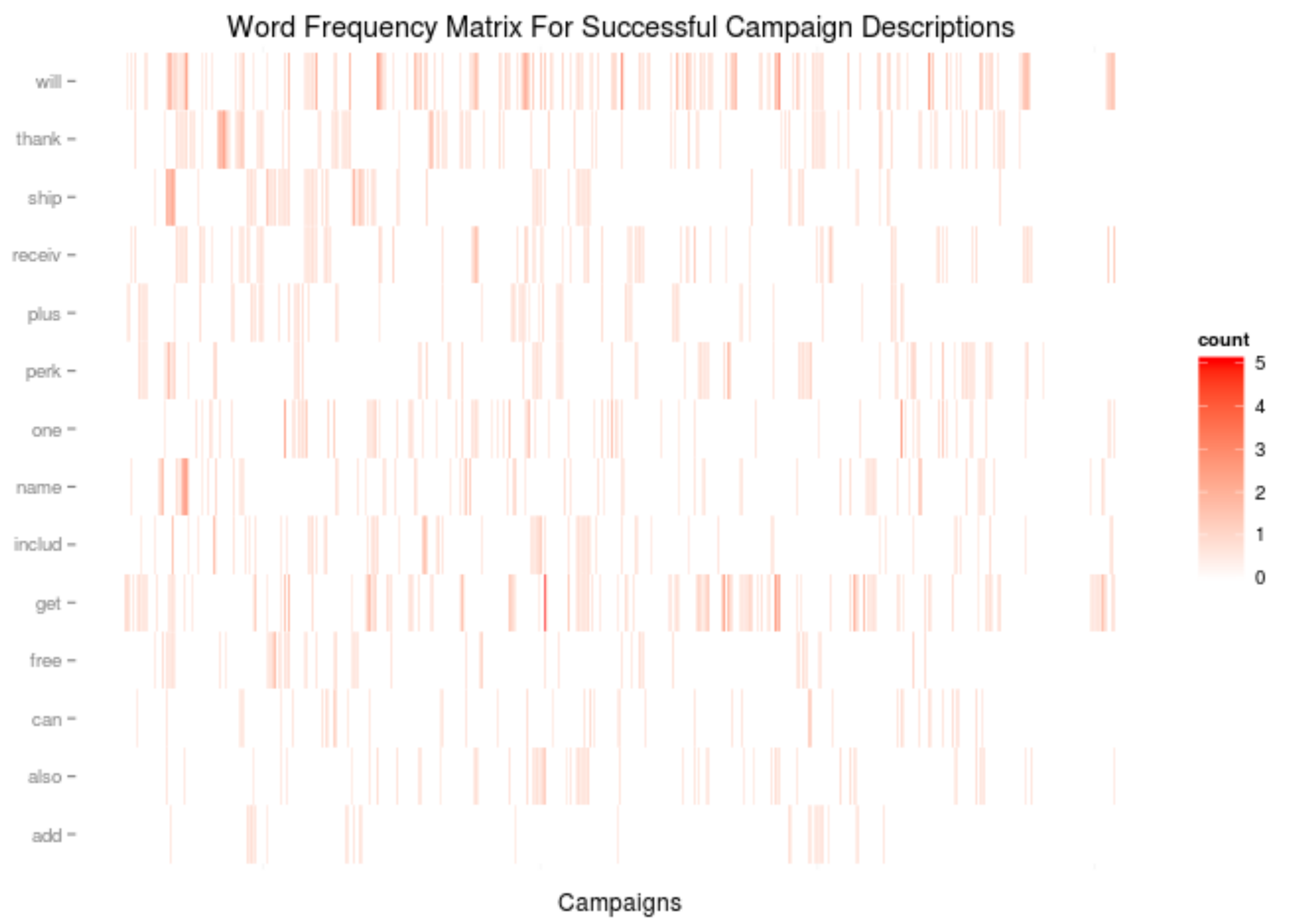

Figure 3: The following plot is a word frequency matrix for successful campaign perk descriptions under the business/tech category. Essentially it is trying to capture the terms that occur most frequently across all "documents in our corpus" or campaigns in the same category. Ideally from this, we should be able to construct a vocabulary of words that are characteristic of successful campaigns. Some of these are words are consistent in plots for campaigns of other categories and even unsuccessful ones, so a more interesting table (captured in the source code) displays some of the other more unique words that show up in campaigns. Words like "photo", "family", and "level" show up across campaigns.

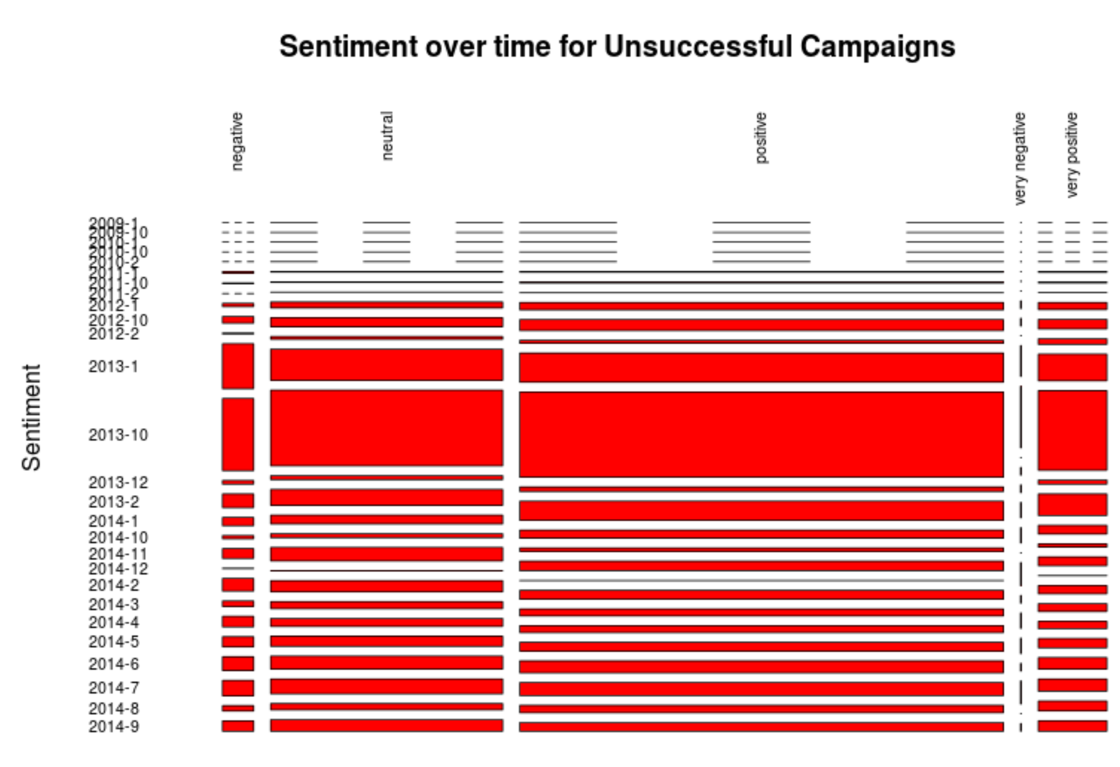

Figure 4

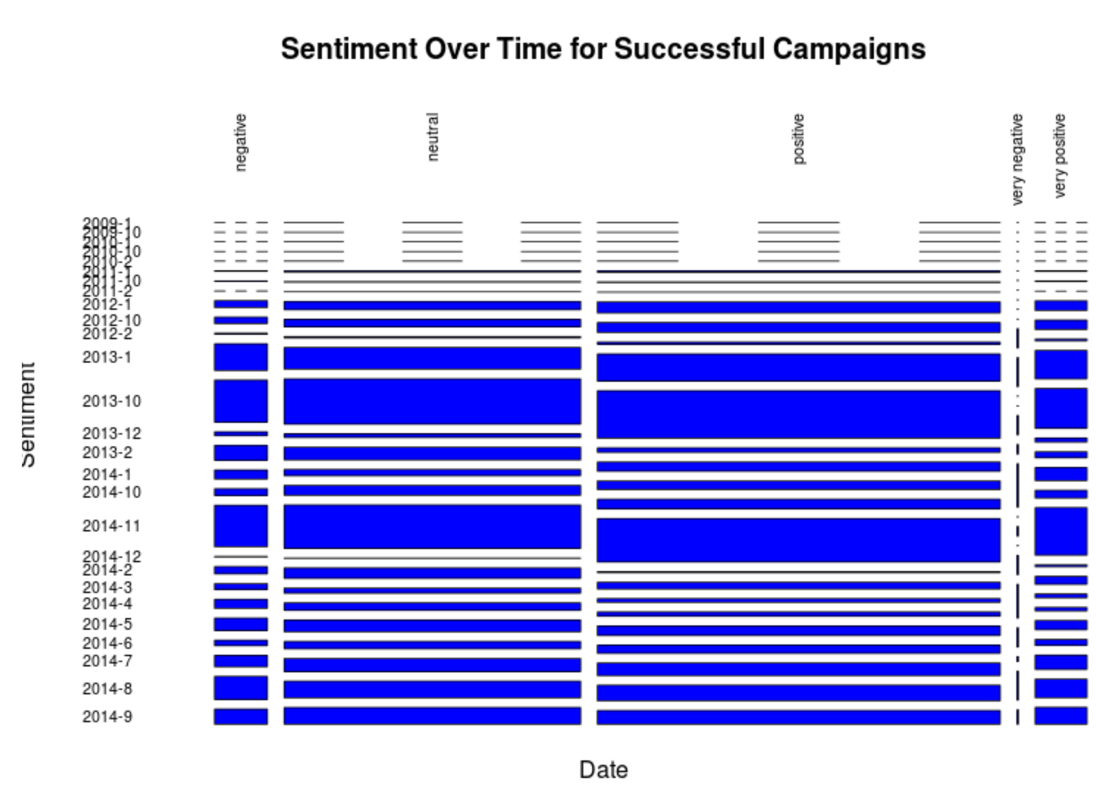

Figure 5

Approach

Using scraped data, primarily utilized R to clean and munge the data to an actionable form. From there, used R's several text mining and clustering libraries to create data structures like a Term Document Matrix and visualizations like dendrograms, word-frequency matrices, and word clouds.

Results

What I was able to accomplish was to provide particular views that could be helpful in exploratory analysis of the dataset. Because of limitations in the data (dates were difficult to work with)

Future Work

Given such a rich and descriptive data set, one area that would be extremely beneficial as a visualization tool would be something that can provide selection and interactivity so that people can identify particular campaigns easier given a visualization from exploratory analysis.

Figure 4: A plot tracking the overall sentiment over time of campaigns that failed to meet their goal in the business/technology category. Makes use of comment data: the content of the comment and the relative date of each comment ("posted over a year ago").

Figure 5: Same type of plot but this time looking at the overall sentiment of successful business/tech campaigns. One can see that there were more neutral and negative comments for successful campaigns than unsuccessful ones, an interesting result.