Sentiment Analysis of Internet Jargon / Slang

Jeremy Singh

Abstract

The internet has changed the way people communicate with one another and has led to the creation of new words and changed the way some words are used, and the way people view these words and how they react to their use. The sentiment behind these words can change month to month, year to year so categorizing how the sentiment these words have changed over time can be a worthwhile endeavor.

Motivation

The motivation for this problem was to categorize in what contexts slang is used and how it changes over time. I also wanted to see if there is a correlation between how people react to different comments, and if the use of different slangs affected their reactions. There is a very good dataset of Reddit comments compiled by Reddit user u/Stuck_In_the_Matrix which inspired me to look at how jargon changed over the timespan of a few years. I will be using data from 2008 to 2010 simply because it takes too much time to download the entire ~160GB dataset.

Method

Firstly, I want to create a list of the various slangs that appear in the corpus. To accomplish this, is used a list of slang words compiled using a list of slangs from Liang Wu, Fred Morstatter, and Huan Liu. This is a compilation of slang words from various sources such as Urban Dictionary and UD. These words have already been marked with a sentiment which will not be used to generate sentiments but will be used as a comparison on the effectiveness of my methodology. One problem with that is that it could miss jargon that are in the corpus but not in the dictionary. I will simply be doing analysis on the words in the dictionary. Once I have a list of jargon, I now need to get the sentiments from corpus. Reddit has a feature where every user generated body of text can be upvoted or downvoted. Upvotes generally mean agreement with the sentiment of the post and downvotes, the opposite. The difference between upvotes and downvotes is called the score, and every Reddit comment in this corpus has its own score. I can then use Naïve Bayes Sentiment Analysis on the different sets of data from different years with three classifiers, positive, negative, and neutral. Words will generally receive a positive sentiment if they occur in more sentences that have a positive upvote to downvote ratio, negative if the opposite. Words that occur most often in comments that have no score will generally receive a neutral sentiment. Words that do not occur in the corpus will receive a neutral sentiment.

The first ~1 million comments from each year, 2008, 2009, 2010, were used to generate a list counts for the words that I decided doing analysis on. After assembling the list of vocabulary and getting the counts for whenever these words occur in an upvoted, downvoted, or neutral comment using the files ‘read2008.py’, ‘read2009.py’, and ‘read2010.py’, I was able to generate the marginal probabilities for each word, and the total count and marginal probabilities for each classifier, positive (1), negative (-1) and neutral (0). With these, I was able to generate the probability for each word to be any sentiment using Naïve Bayes, and decided that the classification of the word is the highest of these probabilities all done with ‘analyze.py’. I then compared the classifications that I came up with to the classifications that were generated from SlangSD, the dataset used to create the vocabulary and compared the change in sentiment between the years using ‘compareSentiments.py’. I believe that using the user given score of a comment can be useful for tagging the sentiment of slang words. I believe that words with higher or lower sentiments will occur more often in comments with the same score. Even if the slang word does not have a significant sentiment on its own. When it is used with other words in certain contexts, it can take on their sentiments and be used in the same way as more pronounced slang.

Analysis

The counts for the vocabulary as they occurred in the corpus can be found in the <year> <sentiment>. pickle files or in the 'Sentiments' excel sheet. The purpose of this experiment was to assign a sentiment to

those changes, it does not look like many words changed in their sentiment.

Change 2008 to 2009: 0.9404709667983033
Change 2009 to 2010: 0.9389270797783303
Change 2008 to 2010: 0.9351892480457641

Table 1: Change in sentiments between years

It looks to be that roughly 6 ~ 7 % of all words changed their sentiment from year to year, but there is a problem with this analysis. There are many words that were in the vocabulary, but not in the corpus. This lead to many of the vocabulary words receiving a sentiment of zero, neutral. If the word showed up even once in another year, it is very likely that it would receive another sentiment. I believe that this is the most likely scenario for words to change using this model. Comparing my sentiments to the sentiments generated from SlangSD,

Comparison 2008 to Slang: 0.33707116506588996
Comparison 2009 to Slang: 0.33346334487185736
Comparison 2010 to Slang: 0.33213072660198595

Table 2: Comparison of my sentiments to SlangSD

We can see that only a third of the generated sentiments in this model match the sentiments from SlangSD. This could simply be due to the size of the corpus that was used to generate the counts. A million comments from each year may not have been enough to get the counts needed to generate the proper sentiments. The other likely issue is that using the score generated by upvotes and downvotes may not be a realistic measurement of sentiment of individual words. It could be that using an n-gram model with an n greater than 1 where certain words used in conjunction with others are the main drivers for higher or lower scores.

While I have not done a full analysis of dataset, from the analysis done on the 3 million comments for this experiment, it appears that negatively scored comments occur less frequently than positively or neutrally scored comments or that negatively scored comments tended to be shorter than the others.

Total 2010 Positive	Total 2010 Negative	Total 2010 Neutral
9302343	476443	524013

Table 3: Counts for Positive and Negative Words 2010

This is not the counts for every year, only 2010, but it is reprehensive of the findings for the other years. Positively scored comments occurred the most frequently, neutral the second most, and negative the least. Looking at how these counts affected the sentiment score was very surprising.

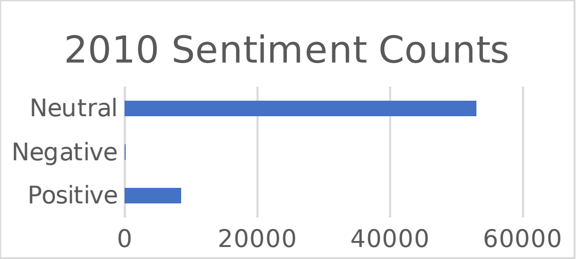

Figure 1: Counts for sentiments for 2010

Despite there being less neutral comments, most of the words were classified as neutral and almost none of them were negative. What this shows is that neutral comments may have tended to us a larger array of words, and that negative comments may have been unusually short. This of course, ignores the fact that many words did not occur in the corpus, so that may skew the results. Since there is such a skew between negatively scored comments and negative sentiments, it calls into question using score as the main decider for word sentiments. I believe that score can be used as a parameter among many and using an algorithm to learn the weights for the different parameters.

Another possibility for this discrepancy could be bias of Reddit comments. Reddit comments are not solely focused on the use of slang, and the slang across the entire website may be radically different due to the nature of the site being made up of millions of smaller subcommunities with possibly different and more emotionally charged slang.

Relation and Comparison to other Works

In recent years, there have been a few other people working towards creating sentiments for words and slang on social media sites like Twitter. SlangSD, the slang dictionary used to create the vocabulary for this experiment is one that used slang compiled from Urban Dictionary and other sources. SlangSD used a method for slowly building up sentiments from hand annotated slang words to assigning sentiments based on association with known slang with sentiments. Instead of using a simple count-based model, they used a weighted model where the associated sentiment is weaker the farther away it is from the known slang in any given text.

Finn ˚Arup Nielsen is another person that attempted to attach sentiment to various words using a pre-annotated list of words to begin with. He compiled his vocabulary from various sources and generated sentiments from the annotated list and learned his model on Twitter comments like from Liang Wu, Fred Morstatter, and Huan Liu. He went farther and used his model to compare against several different annotated dictionaries containing slang.

My model takes the SlangSD dictionary, and instead of using the sentiments given to each slang, I have opted to use the reactions of individuals on Reddit to the comments to determine the sentiment of a word given its occurrences in those sentences and the score given to those sentences.

Limitations

Since some jargon are just normal English words, some words used in a normal context can be miscategorized as jargon and vice versa. Simply using the words as they occur in these comments does not account for the context in which these words occur and ignores their parts of speech and differences in singular and plural occurrences. I also only did my analysis on words that did not have spaces in them. This limits the scope of the words that I can analyze. This ignores compound words, phrases, etc. There may be more profound analysis if I decide to create a model that takes them into account. In my analysis, I noticed that many of the words that I chose to analyze never showed up in any of the sentences. This lead to many of them having a sentiment of 0 which may not be true for many of these words, but since they did not show up in the corpus, no data was available to create a real sentiment. I believe that this was due to the relatively small size of the corpus that I analyzed and the small year range as well.

Using only one parameter to determine the sentiment of a slang word may have been misguided as the difference in my sentiments and SlangSD shows. The score the users of Reddit give comments may be a good for determining the overall agreement with a certain set of ideas, score may very well just give a random score for different words as certain words with negative sentiment may very well have a higher score than a comment with words with positive sentiment. This could be due to the community and if the community encourages the use of these words or if the people in these communities create comments and score them in a specific frame on mind.

Follow-up

In the future, if I were to come back to this problem, I may make a few changes to it. Firstly, if I wanted to continue with the same methodology of using the scores to find the sentimentality of certain words and slang, I would try to use more of the corpus than I did during this experiment. I believed that using ~3 million comments would be enough to capture all the slang that I wanted to analyze, but I was wrong in that assumption as many of the words did not have a single occurrence. Using more of the corpus could help in this regard. Another issue that I encountered was entirely my decision to only use words that had no spaces in them and separating each sentence by spaces to parse them. Because of this, I may have been unable to observe radical changes in the sentimentality of some phrases, compound words, etc.

Something unique about websites like Reddit is that these sites allow for smaller communities and subcultures to gather in one place to talk amongst themselves. Concerns about echo chambers, political polarization, and radicalization aside, this gives a special opportunity to see how different communities change over time and how emotionally charged their words and language is. This is like how different cultures that may speak the same language have different words for the same object or idea and that some words that may be innocent in one culture are offensive in another. An interesting idea would be to find two are more different communities that exist in opposition to one another and determine the difference in sentiment with the same list of vocabulary.

Due to the nature of slang, it may be necessary to use another dataset for slang on Reddit, or to manually generate a list of slang for Reddit or any other social media website to get more counts for the selected vocabulary which could create more accurate sentiments for each word. I could also do the same thing as others and compile a list of known slang from various sources to capture as much slang as possible increasing the ability to apply sentiments to other words by association with the known list of annotated slang.

I could also follow the methodology of others and use the annotated sentiments of various slang words and use their associations via some n-gram model with weights given for the distance from the slang to determine the slang of different words. I may have also used a smaller time range than is necessary to observe changes in the sentimentality of certain slang. For some slang, 3 years is a significant time where it experiences a rapidly changing definition or sentiment. For other words, they may have changed significantly later which would not have been captured by my model. This could have been true for most of the words in my vocabulary especially those that did not have a single occurrence in the corpus. Those words may have only appeared after my time frame or may have become commonly used after my time frame. In the future, I would like to increase the number of comments I use and to increase the range of years for these comments.

If these changes listed here are effective in determining the sentiment of given slang words, then that model could be used for other social media sites and micro-blogs like Twitter, Tumblr, Snapchat, etc. A problem there is that the different demographics and cultures of these websites may lead to some slang words needing to be dropped or added from the vocabulary. While there are many databases for Twitter comments, there are less for other social media sites. If someone should compile a list of comments for other social media platforms, then I could use those. Alternatively, I could use this experience in combing through data to compile my own list of comments from other social media platforms.

Conclusion

While there are merits to using a community’s reaction to comments to be indicative of the sentiment behind the words used in those comments, correct care of the context of those words and sentences is also necessary. The time, place, and audience all need to be taken into consideration. Those same words can change over time, but to properly analyze those changes, it seems that a large body of data across a longer timespan are both necessary for proper analysis. This model that I have made did not consider those things and performed poorly as a result. In the future, I believe that revisiting this experiment with the proper changes will lead to better results.

Resources

u/Stuck_In_the_Matrix. “r/Datasets - I Have Every Publicly Available Reddit Comment for Research. ~ 1.7 Billion Comments @ 250 GB Compressed. Any Interest in This?” *Reddit*, July 2015, www.reddit.com/r/datasets/comments/3bxlg7/i_have_every_publicly_available_reddit_comment/.

Wu, Liang, Morstatter, Fred , Liu, Huan. “SlangSD: Building and Using a Sentiment Dictionary of Slang Words for Short-Text Sentiment Classification” Arizona State University, Aug 2016, https://arxiv.org/abs/1608.05129

Nielsen, Finn ˚Arup “A new ANEW: Evaluation of a word list for sentiment analysis in microblogs” DTU Informatics, Technical University of Denmark, Lyngby, Denmark, Mar, 2011