Terrence Neumann

Is It Something I Said?

Correlating Rankings Generated from Unsupervised Sentiment Analysis with Media Polls to Assess Character Likeability and Favorite Seasons from the Show *Friends*

Summary

In this paper, I use an unsupervised technique for assessing the sentiment polarity of every sentence uttered by one of the six main *Friends* characters over the 10 seasons of the show. I then generate aggregate statistics from this polarity dataset, such as max sentiment, average sentiment, and variance sentiment. Next, I rank the characters in order of their generated statistic, and see how this correlates to several online polls about the “likeability” of *Friends* characters. I repeat this analysis by aggregating to the season level. The results reveal that statistics generated from sentiment polarity are a good predictor for fan’s favorite characters ($p < 0.06$), but a poor predictor for the likeability of a season.

Introduction

Sentiment polarity is the magnitude of positive or negative emotion elicited from a given piece of text, and in this case, it is determined from a pre-defined lexicon of positive and negative words and a series of heuristic rules.

In this paper, I explore the relationship between the mined sentiment and likeability of characters and seasons of the show *Friends*. I was interested initially to see if the sentiment polarity from text alone could predict the likeability of a character on a show. But in addition, I was also interested to see exactly what feature of the text sentiment made a character likeable.

You could imagine a character being likeable in a sitcom for a variety of reasons. For instance, they might be likeable because they have an overall positive sentiment and they generally bring positivity to any scene that they are in. This could be represented as either the sum or mean of sentiment polarity across all episodes of a show. In addition, you could imagine that a character could be likeable (or perhaps memorable would be a better word) due to some measure of variance in their sentiment. If they are very volatile, then they would likely have a very big impact on every scene that they are in. This could be measured by looking at the variance, the range, or the maximum levels of sentiment polarity of a character. Similarly, you might imagine that the same attributes might lead to a particular season of a show being preferred over another.

I generated four statistics from the sentiment polarity of each sentence uttered by one of the six main characters in the show: average polarity, maximum positive polarity, polarity variance, polarity range. These statistics were then rank ordered and the results were compared to ranking polls posted online by media outlets about favorite characters and seasons using Kendall and Spearman rank correlation. While these polls are inherently subjective, I provide an explanation about some significant results.

In the results, we also see that the three female characters have the highest variance in sentiment polarity, which indicates that they were written as being more “unstable” than the male characters, which research shows is a common theme in the portrayal of women in the media. However, the men have the highest range of sentiment polarity, which is a common feature of male characters “bottling up” and having “explosions” of emotion at specific times.

Methodology

Gathering and Cleaning the Text

I began by gathering 214 out of the 236 scripts from the ten seasons of *Friends* from a website[1]. Some scripts were not properly formatted, and thus did not get pulled into memory correctly. I considered this acceptable, since such a high proportion (>90%) were able to be stored and parsed. I stored them in a Python dictionary with seasons, episodes, and characters as nested keys for easy manipulation later. I only kept text from the six main characters, as I did not wish to analyze any other characters in this study.

I then preprocessed the data by tokenizing, spell-checking based on the WordNet lexicon, part-of-speech tagging, and splitting into sentences. Stop word removal was unnecessary for my given sentiment analysis technique, but in retrospect could have increased the speed at which the algorithm performed by having to look up fewer words.

Unsupervised Sentiment Analysis using Lexicons and Rules

After gathering the data, the next step was to generate a polarity score for each character for each episode. This was achieved by using an unsupervised method detailed in Wiegand and Klakow[2].

Polarity is determined from a rule-based classifier. There are two polarity lexicons (positive and negative), and any word that isn't included in this lexicon is discarded.

1. Positive words are assigned a score of +1 and negative words are assigned a score of -1.
2. Part-of-speech tagging is used to disambiguate words. For instance, if "like" is used as a verb (*I like you*), it will receive a score of +1. If "like" is used as a preposition (*You look like her*), then it will be discarded.
3. Negations are also taken into consideration. The authors choose to reverse the polarity of given words if one of five negating words is found directly preceding the word in the polarity lexicon.

---

[1] http://www.livesinabox.com/friends/scripts.shtml
[2] Weigand, M., & Klakow, D. (2013). Bootstrapping Supervised Machine-learning Polarity Classifiers with Rule-based Classification. *Language Resources and Evaluation*.

4. Finally, heuristic weighting is applied. Words with an intensifier (“very”, “extremely”, “madly”, etc) preceding are assigned a 2x multiplier, and adjectives included in either polarity lexicon are assigned a 2x multiplier as well. This is because if an adjective has an associated polarity, the prior probability of the sentence being polar is much higher than that of nouns and verbs.
5. All of the scores for each word in a sentence are summed, delivering a polarity score for each sentence. Then the polarity score of each sentence is summed by character to give a polarity score for each episode.

This method is domain independent, which works well for this show. *Friends* is not a show that focuses on a particular occupation or idea; it is largely about emotions and relationships. Thus, we can make the assumption that a generalized lexicon dictionary should capture a lot of the underlying sentiment evoked by characters.

One concern that I have about the domain is that *Friends* contains quite a bit of sarcasm, and the audience often responds well to this. Sarcasm is something that would be very difficult to detect with an unsupervised technique. Thus, those moments will largely go unnoticed, or perhaps will receive the opposite polarity score of their intended effect on the audience. This could have an impact on my “likeability” scores detailed in the next section.

Results

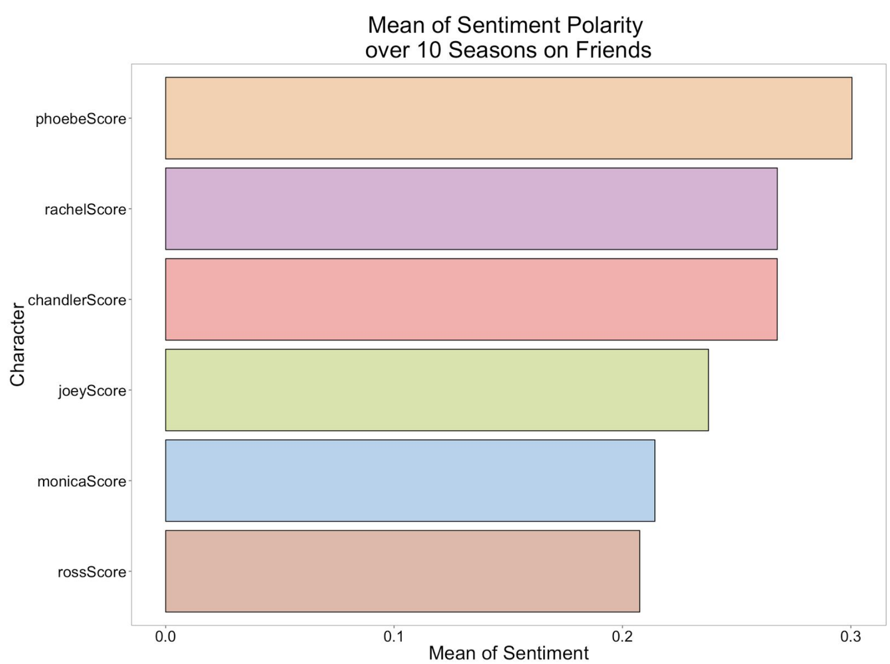
After running the unsupervised, rule-based polarity score technique for each character for each episode, I aggregated the results to different levels to generate key statistics. The first set of results I will show are for characters, followed by season. I then correlated these rankings generated from the aggregate statistics with three polls for the characters using Kendall's rank correlation and Spearman's rank correlation.

Kendall's correlation is a measure for determining similarity in rank orders, and only takes positional ties into account. Spearman's correlation is very similar, but it takes into account non-tied points by looking at the distance of ranked items from one another.

I then use a hypothesis test with the null hypothesis being that the two lists are uncorrelated to determine whether or not a significant relationship exists between the two.

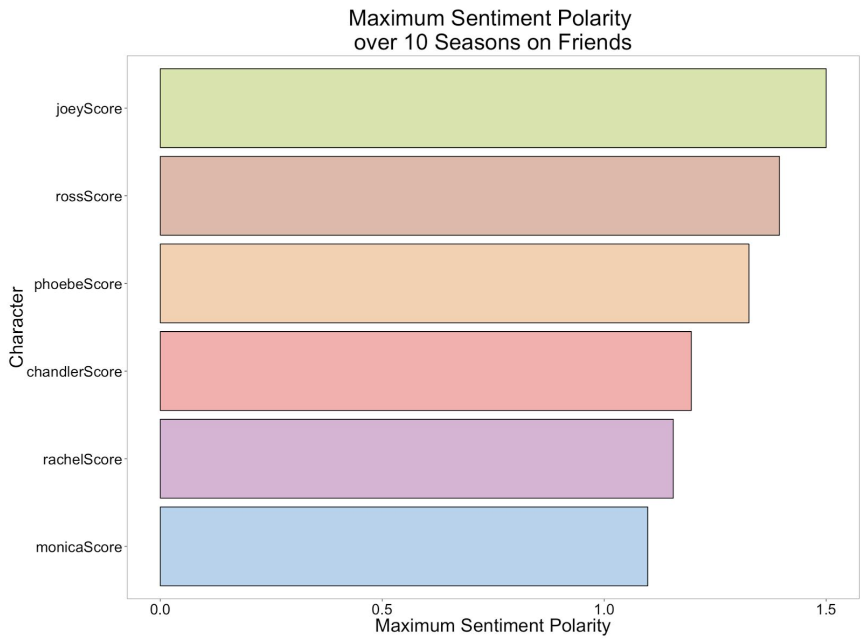
Analysis of Character Results

1. Mean Sentiment by Character



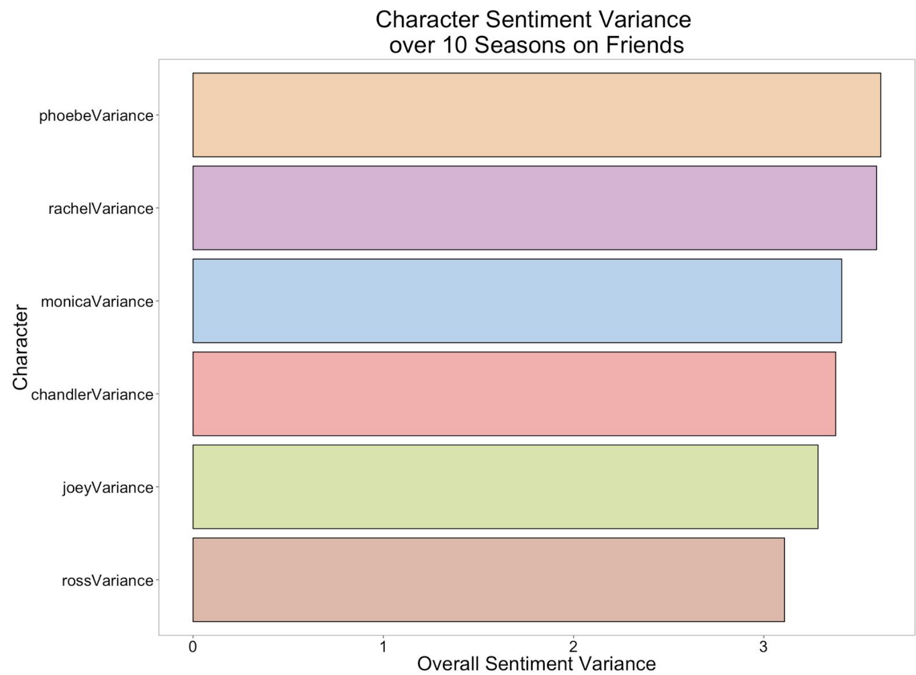
Here, we see that Phoebe takes the lead with the highest mean sentiment across the 10 seasons, followed by Rachel, Chandler, Joey, Monica, and Ross. This makes a lot of sense, as Phoebe is known for bringing a quirky positivity to most scenes that she is in, while Monica and Ross tend to be neurotic and negative.

2. Maximum Sentiment Polarity By Character



Looking at maximum positive sentiment polarity, we see that Joey takes the lead here. Joey is known for being one of the more erratic characters in the series, and due to his job instability as an actor, is constantly waiting for a “big break”. This huge leap in sentiment came in an episode when he landed a recurring part in a sitcom “Mac & C.H.E.E.S.E”, so this could have had something to do with him celebrating this part.

3. Variance in Polarity by Character



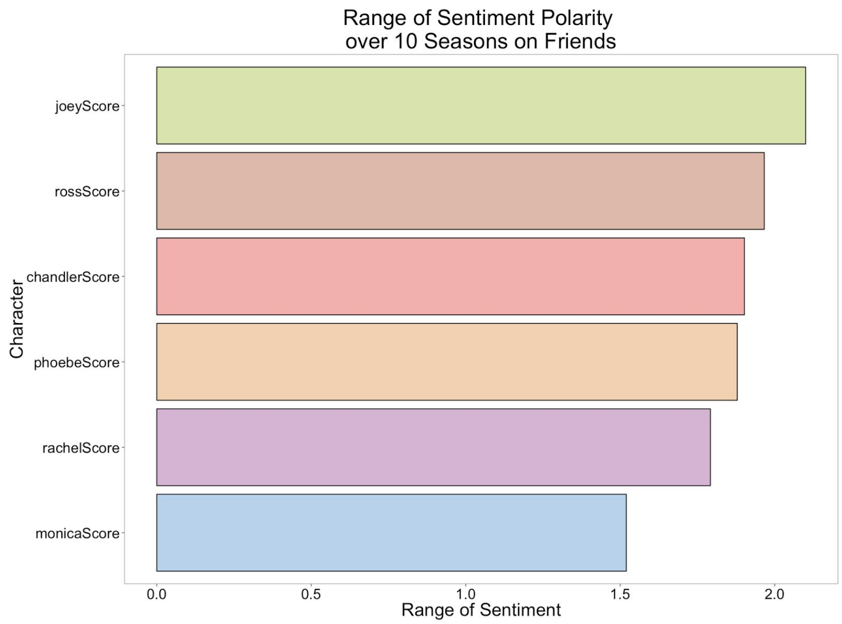
Here we see a particularly interesting result. The women characters have landed in the top 3 for variance in sentiment over the 10 seasons. I decided to calculate how likely this was to occur probabilistically. Using a simple brute force approach computationally, I found:

$$\#\, Number\ of\ times\ females\ land\ in\ top\ three\ spots \div Total\ Permutations\ Possible\ =\ 36 / 720\ =\ 0.05$$

Thus, the probability of this happening by chance is 0.05. Women are often portrayed as more emotional than men in the media, often due to historical biases and under-representation by women in Hollywood. According to a study put together by the USC Annenberg School of Communication, we see that representation in writing and directing roles is also largely unbalanced, and could be to blame for these stereotypical results. In fact, only 10.8% of Hollywood film writers, 31.6% for broadcast TV writers, 28.5% cable TV writers, and 25.2% streaming writers are women[3], which shows that the emotions of women are likely oversimplified or written from a male perspective.

[3] Smith, S. L., Choueiti, M., & Pieper, K. (2016, February 22). Inclusion or Invisibility? Comprehensive Annenberg Report on Diversity in Entertainment. *USC Annenberg External Memo.*

4. Range of Sentiment by Character



Here we have a similarly interesting result to the variance statistic generated above. We see that the men land in the top 3 spots for total range of sentiment expressed throughout the course of the show. This also conforms to typical media stereotypes of men "bottling up" and having emotional "explosions" only rarely. Men are often portrayed as being more stoic and thus would have lower variance, but high outliers in sentiment polarity, as we see above.

Correlating Results with Media Polls

I found 3 media polls online that had ranked "likeability" of Friends in slightly different ways. First of all, it is clear that these rankings are highly subjective, and in 2 of the 3 cases, are based on the opinion of one person - albeit a *Friends* expert.

The first poll[4] assesses each character's "tolerability" by asking the questions "Could we tolerate them in real life? How nice are they? Which would make the best friend?". Using this criteria, the author ranks them and provides a brief explanation. The second poll[5] was conducted via Twitter by Comedy Central UK in 2016, where users directly voted for their favorite friends character (only 1). Then the list was generated by

[4] http://www.puckermob.com/entertainment/ranking-the-friends-characters-from-least-to-most-tolerable
[5] http://metro.co.uk/2016/10/10/the-uks-favourite-friends-character-has-been-revealed-but-is-it-the-one-you-like-best-6183589/

frequency of votes, and is targeting character memorability. The third and final poll[6] was similar to the first in that it ranked the characters by thinking of, overall, what were their best and worst moments.

In each of the cases, the characters were ranked on a 1 (best) to 6 (worst) scale, and they were correlated to my results using Kendall's and Pearson's rank correlation. Then, a simple hypothesis test was run to determine if the correlation was significant. The results can be seen below.

	Poll 1: "Tolerability"			
	Kendall	**P(corr = 0)**	**Spearman**	**P(corr = 0)**
Average Polarity	0.73	0.055	0.83	0.058
Maximum Polarity	0.2	0.719	0.14	0.82
Variance Polarity	0.47	0.272	0.54	0.298
Range Polarity	0.06	1	0.09	0.919

	Poll 2: "Memorability"			
	Kendall	**P(corr = 0)**	**Spearman**	**P(corr = 0)**
Average Polarity	-0.2	0.719	-0.31	0.564
Maximum Polarity	0.6	0.136	0.71	0.136
Variance Polarity	-0.47	0.272	-0.71	0.136
Range Polarity	0.73	0.055	0.83	0.058

	Poll 3: "Best and Worst Moments"			
	Kendall	**P(corr = 0)**	**Spearman**	**P(corr = 0)**
Average Polarity	0.6	0.136	0.77	0.108
Maximum Polarity	-0.2	0.719	-0.26	0.658
Variance Polarity	0.33	0.469	0.49	0.355

[6] https://www.joe.co.uk/entertainment/the-six-main-characters-of-friends-ranked/78420

Range Polarity	-0.06	1	-0.14	0.802

These results are very interesting. For clarification, Spearman and Kendall correlation both measure rank-order correlation, but Spearman takes into account proximity in order, while Kendall only looks at ties in position. I have highlighted the most significant results from each poll. We can see that polls 1 and 3, which were most interested in examining overall “goodness” of a character, were most significantly correlated to the average sentiment statistic. Further, the correlation coefficient was positive, indicating that the higher the average sentiment of a character, the higher the likeability ranking.

This intuitively makes sense. If you want to determine the overall reliability or quality of a friend, you would look at how they behave or express themselves on average. We see that for poll 1, the probability of the null hypothesis being true -- that the actual correlation between these results is 0 -- is less than 0.06 for both Spearman’s and Kendall’s correlation. This is a very good result, considering it is incredibly hard to reject the null hypothesis with n = 6 observations. For poll 3, the probability that the null hypothesis is true is slightly higher, 0.108, but still within the range of significance.

For the second poll, we get a similarly interesting result. Due to the structure of the poll, users could only vote for one character. Thus, one could hypothesize that they are most likely to vote for the one who was most memorable across the seasons - that is a character with extreme high and low points. And indeed, that is what the data reveals. The range polarity statistic is an extremely good predictor of the memorability of a character. We see that the probability of the null hypothesis being true for both Kendall and Spearman’s correlation coefficients is less than 0.06. In addition, we see that the correlation coefficient is positive, indicating that characters with a higher range are more likely to be remembered.

Analysis of Seasons Results

I conducted an analysis on each season as an aggregate by creating the same aggregate statistics from the character section and correlating them to media lists of “Best Seasons of *Friends*”. In each case, the polls were written by a single person and provide analysis of episodes within a given season. Seasons are generally considered worse if they were deemed “inconsistent”, or contained episodes that either didn’t move the plot along, were too ridiculous, or not very funny. Instead of presenting the rankings from all the aggregated statistics, I will just product the summary tables to show that the statistics did not perform well at all in predicting the “likeability” of a season.

	Poll 1: www.collider.com			
	Kendall	**P(corr = 0)**	**Spearman**	**P(corr = 0)**
Average Polarity	-0.07	0.862	-0.02	0.973
Maximum Polarity	0.11	0.727	0.1	0.785
Variance Polarity	0.02	1	0	1
Range Polarity	0.24	0.38	0.3	0.407

	Poll 2: www.digitalspy.com			
	Kendall	**P(corr = 0)**	**Spearman**	**P(corr = 0)**
Average Polarity	0.29	0.29	0.33	0.348
Maximum Polarity	0.11	0.727	0.3	0.407
Variance Polarity	0.02	1	0.06	0.865
Range Polarity	0.24	0.38	0.35	0.33

	Poll 3: www.buzzfeed.com			
	Kendall	**P(corr = 0)**	**Spearman**	**P(corr = 0)**
Average Polarity	-0.11	0.727	-0.18	0.632
Maximum Polarity	0.15	0.6	0.09	0.814
Variance Polarity	-0.2	0.484	-0.33	0.349
Range Polarity	0.02	1	0	1

As you can see, none of the results even approach significance. There could be many possible reasons for this. The most obvious one that comes to mind is that none of these variables captures what makes a season of a sitcom good. Although you might think that range of sentiment polarity would be a good indicator, it is perhaps more about the interaction between characters that makes a season great.

In addition, sentiment analysis cannot capture what is really going on in an episode. Scenes from a show might be very memorable due to the imagery or facial expressions of certain cast members. None of this is incorporated into my analysis. Creating features about how many times a certain cast member is mentioned in an episode might be a better predictor for the likeability of a season

Conclusions

Unsupervised, rule-based sentiment analysis worked very well for predicting “likeability” of characters from the show *Friends*. Using various statistics, we saw that memorability was correlated with the range of sentiment, and general likeability was correlated to the average sentiment over the course of the show. Unsupervised sentiment analysis worked particularly well for *Friends* because it is not a context-dependent show. Almost all conversations are about relationships, and thus a basic lexicon of positive and negative words was able to work very well. A season’s likeability was not correlated with the statistics generated from the rule-based analysis. This is likely due to the more complicated nature of what makes a season of shows enjoyable. We also saw that the female characters were written with more emotional variance than the males, while the males had more explosive moments.

Note on software

I used Python software developed by Nikolaos Pappas, which was kindly made available on his github at the following location:

https://github.com/nik0spapp/usent