INFO 5601 Final Project
Analyzing Reddit Commenter's Opinions on Net Neutrality

Luke Meszar

1 Introduction

Over the last few years, discussions around net neutrality have become more prevalent on the internet, including Reddit. The project aimed to study the public opinion of net neutrality on Reddit. In particular, to understand how opinions differed across different subreddits. In particular, subreddits that likely have different types of users. I choose a thread from five different subreddits and performed both a qualitative coding and sentiment analysis on the comments in each of these threads.

2 Net Neutrality

Net neutrality has been a hotly debated topic recently with recent policy changes by the FCC.

3 Methodology

The first choice that had to be made was which subreddits and which threads to consider. A few factors went into choosing this. First, a thread had to have a decent number of comments. A thread with only a handful of comments might not be the most representative of a subreddits overall opinions due to small sample size. The subreddits chosen to pick threads from was based on getting a diverse set of subreddits.

The five that were chosen were /r/askreddit, /r/technology, /r/books, /r/the_donald, and /r/democrats. /r/askreddit was chosen because it is one of the largest subreddits, and likely has one of the most diverse user bases of any subreddit. The thread chosen was titled "*[Serious] Americans that *DO NOT* support net neutrality, why?*" This article in particular was chosen because it asks what I assumed to be for the opposite opinion than the majority of Reddit users hold. I thought this would make for an interesting discussion from both sides. /r/technology was chosen because net neutrality is directly related to topic of the subreddit. The article chosen was "*Net neutrality day of action update: Twitter, Soundcloud, and Medium, have joined. Reddit, This could be as big as SOPA*" primarily because it was one of the threads with the most comments. /r/books was chosen because it was a seemingly neutral subreddit. It is neither directly political or technology related. The thread chosen was "*Join the Battle for Net Neutrality!! We need to stop them from allowing ISPs to charge us extra fees to access ebooks, games or anything else!*" The subreddits /r/the_donald and /r/democrats were chosen for their clear political leanings. I realized they aren't exactly a fair comparison since one is about a specific president and the other is about a whole political party. However, because the thread chose for /r/democrats, "*This is President Barack Obama. He did not sell Americans out to the telecom lobby, but instead called upon on the FCC to take up the strongest possible rules to protect net neutrality, which they did at his instruction in 2015*" is very much about Obama. The thread chosen from /r/the_donald was "*One Year Later, 'Net Neutrality' Zealots Proved Dead Wrong. 'net neutrality' zealots warned that its repeal would spell doom for a 'free and open" internet.*'"

Once the threads were chosen, comment data was collected using the Reddit API and the Python wrapper praw. After this, open coding was done on 50 comments, with 10 coming from each post. Two types of codes where chosen to generally represent people's opinions on net neutrality: their stance on net neutrality and their reason for their stance. The following tables list the codes with a short description:

Codes for Stance on Net Neutrality

Code	Description
For NN	It is clear that the comment is for NN
Against NN	It is clear that the comment is for NN
Probably for NN	The comment's stance on NN is ambiguous but context says they are likely for NN
Probably against NN	The comment's stance on NN is ambiguous but context says they are likely against NN
Ambiguous	It is entirely unclear the comment's position on NN

Code for Reason for Stance

Code	Description
Political Reasons	Primary reason stems from opinions of politicians, or general political views
Economic Reasons	Primary reason is driven by economic ideas like markets, monopolies
Personal Reasons	Primary reason is how it affects their lives individually
Other	The other categories don't fit well

I originally tried a form of topic modeling known as Latent Dirichlet Allocation to find which comments were on topic. However, this proved ineffective since all the topics were too broad. I decided to write my own crude system to determine which comments were on topic. I created a dictionary containing terms like "net neutrality, Comcast, ISPs" etc. Using this dictionary, I filtered out comments that contained these words in them.

Then, I built a simple Python script that allowed me to go through each comment, choose whether to accept or reject it, and then choose a numerical code for each of the two codes presented above. For all the threads except /r/democrats, I coded 41 comments. For /r/democrats, I only coded 38 because so many of the comments where off-topic. Using these codes, I computed simple percentage statistics for all the codes and combinations of codes in each of the threads which is the basis of my results. Finally, I ran the pretrained sentiment analysis known as VADER [HG14] on all the comments I coded. For each comment, a compound score was given in the range $(-1, 1)$ where a more negative number is a negative sentiment and a larger positive number is a more positive comment. I considered any comment with a compound score greater than 0.25 to be positive, anything with a compound score less than -0.25 negative, and everything else is considered neutral. From here, percentages for each category were calculated. I also ran the sentiment analysis on the thread titles.

4   Results

The complete results of all the percentages from the coding are long so they can be found in section 9. Here, only the top two codes from each category for each thread will be shown.

/r/askreddit		Stance		Reason
	Probably against NN	21.9%	Economic Reasons	68.3%
	Probably for NN	21.8%	Political Reasons	24.4%
/r/technology				
	For NN	48.8%	Political Reasons	31.7%
	Probably for NN	36.6%	Economic Reasons	29.3%
/r/books				
	Probably for NN	48.8%	Political Reasons	34.1%
	For NN	29.3%	Other	26.8%
/r/democrats				
	Probably for NN	47.2%	Political Reasons	66.7%
	Ambiguous	33.3%	Economic Reasons	16.7%
/r/the_donald				
	Probably against NN	65.9%	Political Reasons	63.4%
	Against NN	26.8%	Other	19.5%

I am choosing not to report the most popular combinations of stance and reasons since they can almost always be inferred from the topic reason and stance separately. The sentiment analysis for the titles resulted in:

	Compound Score	Sentiment
/r/askreddit	0.4019	Positive
/r/technology	0.0	Neutral
/r/books	-0.5386	Negative
/r/democrats	0.8074	Positive
/r/the_donald	-0.8885	Negative

The results for the sentiment analysis are as follows:

	Positive	Neutral	Negative
/r/askreddit	53.7%	17.1%	29.3%
/r/technology	46.3%	24.4%	29.3%
/r/books	46.3%	19.5%	34.1%
/r/democrats	33.3%	11.1%	55.6%
/r/the_donald	29.3%	26.8%	43.9%

5 Analysis

6 Potential Biases

7 Future Work

8 Conclusion

9 Appendix

Full results

References

[HG14] Clayton J Hutto and Eric Gilbert, *Vader: A parsimonious rule-based model for sentiment analysis of social media text*, Eighth international aaai conference on weblogs and social media, 2014.