Extracting and Analyzing User Comments to Expose How Americans Respond to Gun Violence

Erika Fox '21
Professor Alicia Simmons, Professor Michael Hay
Department of Sociology, Department of Computer Science
Colgate University, Summer 2019

Background

- *Framing* is the process of selecting some aspects of reality and making them salient, while directing attention away from other aspects. These word choices made by journalists can have implications on public opinion regarding major issues in our current political climate. Mass shootings and their associated issues such as gun control and mental health are no exception.
- The New Media age has drastically changed the way that Americans consume news, in that the internet enables news consumers to bounce from source to source with ease and to leave comments in the meantime. These comment sections are vast pools of data that have the potential to answer some of the burning questions that media sociologists have had for decades regarding just how much frames in media coverage affect the takeaway readers receive from articles.
- Natural Language Processing is a subfield of computer science that is concerned with the development of software that is capable of processing large amounts of human language data. This study requires research of this field in order to determine which natural language processing methods, if any, would be the most fruitful in the attempt to analyze thousands of user comments and their articles frames.

Research Question

1. How are the implications of news framing reflected in the comments of New York Times articles on the Las Vegas Massacre and the Parkland MSD High School Shooting?
   a) What is the nature of users' comments about online news articles about mass shootings? What topics are discussed?
   b) How does the population of comments vary across case characteristics? How do discussion topics differ between the Las Vegas Massacre and the Parkland Shooting?

Methods

Article Content Analysis

- A sample of 20 New York Times articles was collected, including 10 articles from the Las Vegas Massacre of October 1st, 2017 (dated from 10/2/2017-10/3/2017) and 10 articles from the Parkland MSD High School Shooting of February 14th, 2018 (dated from 2/15/2018-2/17/2018)
- The entire sample of articles was coded in MAXQDA to get a sense of the frames that were present in the articles themselves. (See table: "Overview of Frames")
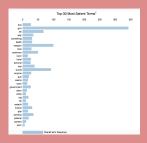
Data Extraction/Organization

- The comments had to be extracted from the NYT articles someway, and the NYT's API came in handy for that. The API worked by performing html calls with the article's NYT URL as an input, and would return 25 comments per call.
- I wrote python code to automate this process and extract over 28,000 comments across the sample of 20 articles. This code not only extracted the data but also placed them each into separate json files for each article's corpus of comments, each of which were uploaded to GitHub in preparation for use by a natural language processing algorithm.

Natural Language Processing (NLP)

- After lots of experimentation with different NLP techniques and sample code, the most success seemed to come with something called Latent Dirichlet Allocation (LDA) Topic Modeling. This statistical modeling tool allowed me to discover abstract "topics" that occur in my sample of articles and their corresponding comment sections. My LDA topic model used Python code and many of its packages, including Gensim, Pandas, Numpy and Spacy to return five of these topics, consisting of 10 keywords each, along with charts that displayed the top 30 most salient terms. Below is an example of what the LDA model would return.

Overview of Frames:

Code type		Las Vegas	Parkland
Attribution of Responsibility	Stephen Paddock/ Nikolas Cruz (Perpetrator)	5	6
	Lack of Gun Control	4	5
	Assault Rifles	6	6
	FBI/Police Negligence	3	3
Conflict	Right to Bear Arms vs. Gun Control	2	5
	Americans vs. NRA	3	2
	Democrats vs Republicans	3	2
	Mental Health Issues	2	1
	Concert/Hotel/ School Security	5	1
Human Interest	Accounts of Grievance/Trauma	4	4
	Humanization of Perpetrator	3	4
Morality	Thoughts and Prayers	3	1
	Unity for Action	4	2
	Demoralization	1	5

Results and Analysis

- Overwhelmingly, "gun" was the most used term across all the comment sections in my sample. The table below shows the salience of the term in the comments in comparison to the salience of the term in the articles themselves.

Las Vegas Articles						
Article ID	"Gun" in top 30 salient terms (Article)	"Gun" is #1 most salient term (Article)	"Gun" appears as topic keyword (Article)	"Gun" in top 30 salient terms (Comments)	**"Gun" is #1 most salient term (Comments)**	"Gun" appears as topic keyword (Comments)
1	Yes	No	No	Yes	**Yes**	Yes
2	No	No	Yes	Yes	**Yes**	Yes
3	No	No	No	No	**No**	No
4	No	No	Yes	Yes	**Yes**	Yes
5	No	No	Yes	Yes	**Yes**	Yes
6	No	No	Yes	No	**No**	Yes
7	Yes	No	Yes	Yes	**Yes**	Yes
8	No	No	No	Yes	**No**	Yes
9	Yes	Yes	Yes	Yes	**Yes**	Yes
10	Yes	No	Yes	Yes	**Yes**	Yes
Parkland Articles						
Article ID	"Gun" in top 30 salient terms (Article)	"Gun" is #1 most salient term (Article)	"Gun" appears as topic keyword (Article)	"Gun" in top 30 salient terms (Comments)	**"Gun" is #1 most salient term (Comments)**	"Gun" appears as topic keyword (Comments)
11	Yes	No	Yes	Yes	**Yes**	Yes
12	Yes	No	No	Yes	**Yes**	Yes
13	Yes	No	Yes	Yes	**Yes**	Yes
14	No	No	Yes	Yes	**Yes**	Yes
15	No	No	No	No	**No**	Yes
16	No	No	No	Yes	**Yes**	Yes
17	No	No	No	Yes	**No**	Yes
18	Yes	Yes	Yes	Yes	**Yes**	Yes
19	No	No	Yes	Yes	**Yes**	Yes
20	Yes	Yes	Yes	Yes	**Yes**	Yes

- 93% of the time, if an article mentions "gun", the term will dominate the comment section as the most salient term. Even articles that don't include "gun" still have the term dominate the comments more often than not.
- More focused articles (containing only one frame category) are likely to have more comments that don't include "gun".
- Other issues include Mental Health and the Disputed Power of the NRA, and Hotel/School Security. 17 out of 20 comment sections significantly discuss the NRA (from both Las Vegas and Parkland articles) and 2 out of 20 comment sections significantly discuss mental health (only Parkland articles). In the articles, hotel security (Las Vegas) is more salient than school security (Parkland), but in the comments the opposite is true.
- According to Sentiment Analysis testing, the majority of the comments are negative.

Relative salience of terms among the aggregate article and comment data sets:

Article Topics

Comment Topics

Conclusion

- This research looks closely at how the media's routine for covering mass shootings impacts the way Americans come to understand and remember them. Although there are several issues that are commonly associated with American mass shootings, this study proves that the issues that center around "gun" terms are the most central to the cognitive processes of the American public. This frequent discussion of the gun debate is not a bad thing, especially since several of the comment sections that discussed this particular topic contained productive conversations that could even inspire readers to take action to make a change of how guns are handled in the USA. However, this study doesn't only show that "guns" are the most important topic to the public, it also shows just how hard the topic is to avoid amongst readers in response to mass shooting articles. Because of this, if a journalist ever wanted to start a conversation in response to a shooting that doesn't center around guns, they would have to be very smart about how they framed their article. In other words, studies such as this one can begin to inform journalists on how to frame their articles in order to have more control over what message their audience receives from their writing. According to this study's findings, more focused articles are the most productive in starting target conversations (ex: an article that focuses only on school security is more likely to start conversations only about school security, meanwhile if an article mentions "gun" there's a 93% chance that "gun" will dominate the comment section). Finally, this research should make the public more aware of how news coverage impacts their understanding of gun violence. Perhaps the study of the implications of news coverage can someday get us all on the same page, to the point where we can address issues most efficiently.

References
https://developer.nytimes.com/, https://www.nytimes.com/2017/10/02/us/las-vegas-shooting-live-updates.html, https://www.nytimes.com/2017/10/02/us/las-vegas-shooting.html, https://www.nytimes.com/2017/10/02/business/las-vegas-shooting-fake-news.html, https://www.nytimes.com/2017/10/03/opinion/vegas-gun-control-shooting.html, https://www.nytimes.com/2017/10/02/opinion/mass-shooting-vegas.html, https://www.nytimes.com/2017/10/03/opinion/vegas-shooting-gun-control-paddock.html, https://www.nytimes.com/2017/10/03/us/las-vegas-gunman.html, https://www.nytimes.com/2017/10/02/business/hotel-security-las-vegas.html, https://www.nytimes.com/2017/10/02/opinion/gun-control-vegas-shooting.html, https://www.nytimes.com/2017/10/03/us/las-vegas-shooting-live-updates.html, https://www.nytimes.com/2018/02/15/us/florida-school-shooting-updates.html, https://www.nytimes.com/2018/02/15/us/florida-shooting-nikolas-cruz-florida.html, https://www.nytimes.com/2018/02/16/us/florida-mass-shooting.html, https://www.nytimes.com/2018/02/14/us/parkland-school-shooting.html, https://www.nytimes.com/2018/02/14/us/fbi-nikolas-cruz-shooting.html, https://www.nytimes.com/2018/02/17/opinion/sunday/gun-laws-school-shootings.html, https://www.nytimes.com/2018/02/17/opinion/sunday/appeasing-the-trigger-gods.html, https://www.nytimes.com/2018/02/16/opinion/repeal-repeal-second-amendment.html, https://www.machineleikrngplus.com/nlp/topic-modeling-gensim-python/, Arendholz, J. 2013. (In)appropriate online behavior: A pragmatic analysis of message board relations. Amsterdam/Philadelphia; John Benjamins Publishing Company., E. H. Kerr, S. 2018. Gun violence prevention? The politics behind policy responses to school shootings in the united states. Cham: Springer International Publishing, Pablo, Carolyn R. 2016. "Social Media and News Coverage as Vicarious Exposure." The Wiley Handbook of the Psychology of Mass Shootings. by Laura Wilson, Wiley-Blackwell 136–152, Highfield, Tim. 2016. Social Media and Everyday Politics. Polity Press. Kinder, Donald R. 1998. "Communication and Opinion." Annual Review of Political Science 1:167-197., Papacharissi, Z. 2002. The virtual sphere: The internet as a public sphere. New Media & Society, 4(1), 9-27. , Schildkraut, Jaclyn, and H. Jaymi Elsass. 2016. "The Influence of Media on Public Attitudes." The Wiley Handbook of the Psychology of Mass Shootings, by Laura Wilson, Wiley-Blackwell 117-135., Tewksbury, David and Jason Rittenberg. 2012. News on the Internet: Information and Citizenship in the 21st Century. New York: Oxford University Press., Wojcieszak, M. & Mutz, D. 2009. Online groups and political discourse: Do online discussion spaces facilitate exposure to political disagreement? Journal of Communication, 59(1), 40-69. Zhou, Yuqiong and Patricia May. 2007. "Parsing Framing Processes: The Interplay Between Online Public Opinion and Media Coverage." Journal of Communication 57(1):79-98.