Influence of News Media on Public Opinion

Ritu Susan Sanjay

Indiana University Bloomington

[rssanjay@iu.edu](mailto:rssanjay@iu.edu)

**ABSTRACT**

Communication channels developed first back during the Roman and Persian Empires. Media can refer to any form of information channel that delivers information i.e. news to the public – newspapers, television channels, and today social media [1]. Over the last century, media has played an important role in the rise of governments, wars and new breakthroughs in science and technology. In all these decades since the advent of modern media, we notice a common trait – the news relayed to the public has always been presented in such a ways so as to sway public opinion.

**KEYWORDS**

Text mining, R programming, Facebook, News channels, Sentiment Analysis

**1. INTRODUCTION**

Social media, including LinkedIn, Facebook, Twitter, Tumbler etc., provides great insights for business analysts, social and political scientists. Analyzing this text allows scientists to drill down into the core of the who, when, where and what. The insights derived from such studies can help shape policies as well as new marketing campaigns for ‘cherry-picking’ customers. New methods and algorithms allow us to quantitative measure large amounts of text.

**2. IMPLEMENTATION**

**2.1 DATA**

The data for this study has been obtained by principled-group selection approach, and sampled at random. News headlines and comments in their response form the main constituents of the data.

To create the sample, three current and trending topics were chosen. This was not a random selection, but based on the popularity of the news items (having greater than one million views on Google). The following three topics have been considered for the study:

1) Hong Kong Clashes

2) US 2020 Elections

3) Middle East Crisis

Once the news items were narrowed down, the next step was to identify the news channels or rather the news agencies, whose (Facebook) pages were to be parsed for headlines and comments. Five broadcasting agencies were chosen based on their relevance of viewer popularity in the respective regions:

1) British Broadcasting Corporation (BBC)

2) Cable News Network (CNN)

3) Euronews

4) Fox News

5) Channel News Asia (CNA)

The objective of this study as stated earlier is to define a relationship between the tonality of the headline and the comments in response to the same. This analysis is split into two parts. The first being the polarity of the comments directly in response to the headline of the news post, which will be hereafter referred to as direct comments. The second being the analysis of the comments which were posted in response to the direct comments, hereafter being referred to as response comments. Basically, e sum this up as the analysis of headlines posted, analysis of the direct comments to these posts and finally the analysis of the response comments.

These posts and comments were collected manually of Facebook, without making use of any automated software (like selenium, beautiful soup, ScrapeStorm etc.). In doing so, the data was stored in three files and structured with IDs to group them into specific classes. The structure of the three datasets are as described below.

The first dataset consists of the headlines (26 in total), each categorized by the news item and news channel. Additionally, each headline has a unique ID that identifies its news item and news channel. The second sheet is a culmination of all direct comments in response to the headlines. Each of these comments is given a unique ID. Finally, the third dataset consists of all comments given in response to the direct comments. These response comments are also uniquely identified by an ID. In total there are 206 direct comments and 136 response comments, giving us a sample of 342 comments across all news items and channels. The unique groupings are as given below in Table 1:

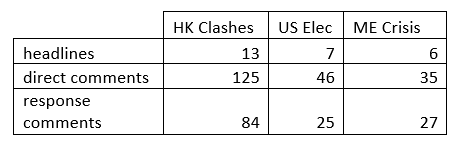


Table 1: Number of comments per news item

**2.2 ANALYSIS**

Text mining is the process of identification of trends and patterns from textual data such as those derived from articles, social media comments, customer reviews, emails, books, poetry, songs etc. [2]. The basic analysis methods include stemming, tokenization, clustering and summarizing word frequencies [3]. This can be contrasted with content analysis which involves the analysis of the context of the text under study. Text analytics also considers the structure of the text. This study makes use of sentiments analysis using R programming language to gauge the relationship between the headlines and the comments.

The first part of the analysis involves studying the content of the headlines and the comments. This will enable the use of a common ground before the sentiments analysis. Wordclouds have been used to identify the most common terms in the headlines and their comments. Note that here, the comments wordcloud includes both the direct as well as the response comments. The tm and qdap packages in R provide simple to use functions to clean the data - this involves removing any stopwords, punctuation, whitespaces and emoticons. Figure 1 below shows the wordcloud:

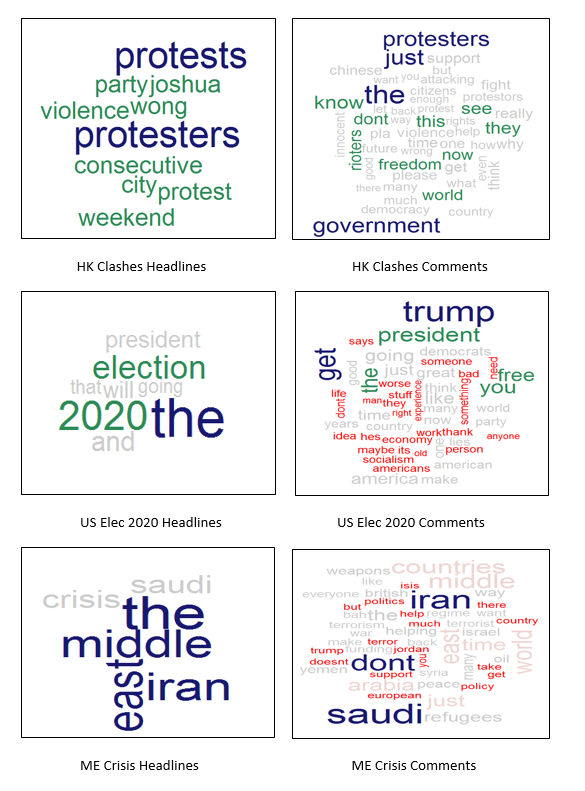


Fig 1: Wordcloud – headlines & comments

The wordclouds above, naturally invoke questions like: are most protests in Hong Kong held during the weekend or perhaps if Trump will indeed win the 2020 elections or if the tensions in the Middle East are caused by the increasing number of refugees. There are innumerable questions that may be asked. However, the objective of the study is not answering questions asked above but to explore the views of the public via the comments.

Sentiment analysis is the process of identifying the tonality, sentiment and its significance to the text inconsideration [4]. The most basic form of analysis involves classifying a simple piece of text as neutral, positive or negative. A simple technique for this classification is to find the polarity of each term in the text. Then the individual polarities are aggregated together to find the polarity of the text. A polarity value less than zero indicates that the given text is negative, while a polarity value greater than zero indicates that the text is positive. Another method of identifying the sentiment is to make use of the NRC lexicon, which includes the eight basic human emotions (anger, joy, anticipation, sadness, fear, disgust, surprise and trust) along with negative and positive classifications (this analysis may be performed using the tidytext package in R).

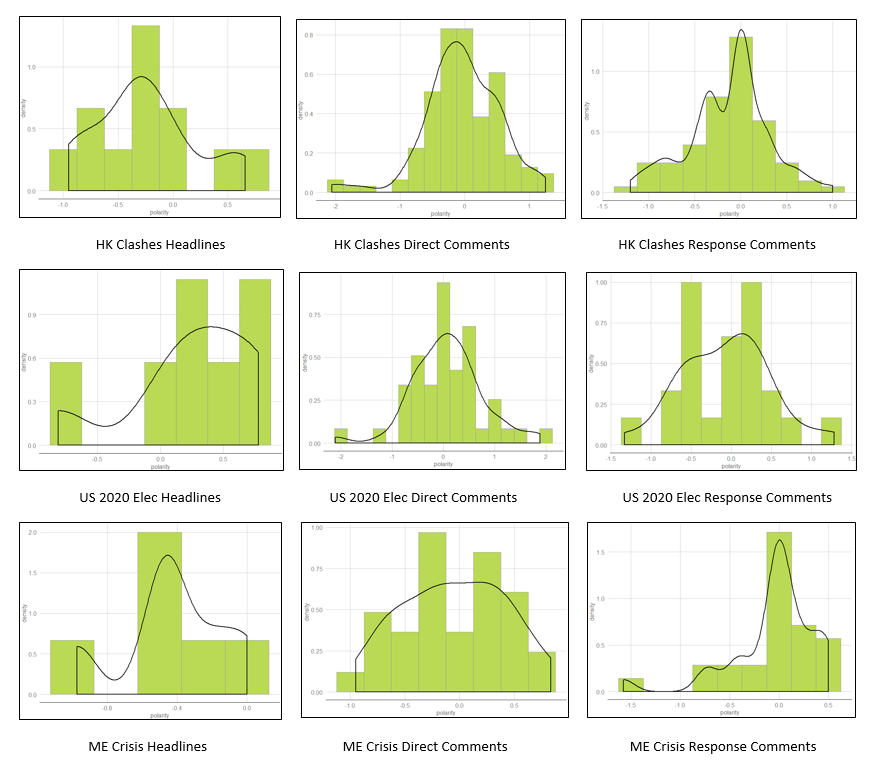


Fig 2: Density vs Polarity

Fig 2 displays the polarity of headlines and their comments (direct and response). We can examine each news item separately.

Consider the case for the Hong Kong clashes. Here, the first plot show the polarity distribution for the headlines – it is evident that a majority of the headlines (irrespective of its news channel) has made use of words that convey negative sentiments. We can verify this using the wordcloud from earlier. Coming to the direct comments, a majority of which is tending to a neutral ground with some negative sentiments. It is interesting to note that the comments in response to the above comments tend towards both neutrality as well as negativity. The news channels chosen for this news items were BBC, CNA and CNN. A closer look at this grouping shows that there is a distinctive negative quality to the comments posted in response to headlines by CNA as compared to both BBC and CNN. This could be attributed to the fact that the CNA is a channel which commonly broadcasts news relating to Southeast Asia and Singapore. It is no coincidence that the comments in this category leer closer to a negative polarity as people living in these regions are more affected.

The next news item in the study deals with the US 2020 elections. The headlines show a positive polarity. In response the direct comments express positive sentiments. However, we notice that the response comments express neutrality and negativity. Once again referring to the wordcloud, shows that most of the comments include words like ‘Trump’, ‘America’, ’Free’, ‘Great’ among others. The news channels for this item included BBC, Euronews, and Fox News. The comments data shows that there is a divided stance among the public when it comes to the next year’s elections.

The last news item to be considered is the Middle East crisis. The headlines for this news item show a distinctive negativity in their sentiments, however the comments (both direct and response) indicate a neutrality in their tones.

Table 2 summarizes the aggregated polarity for all three news items below.

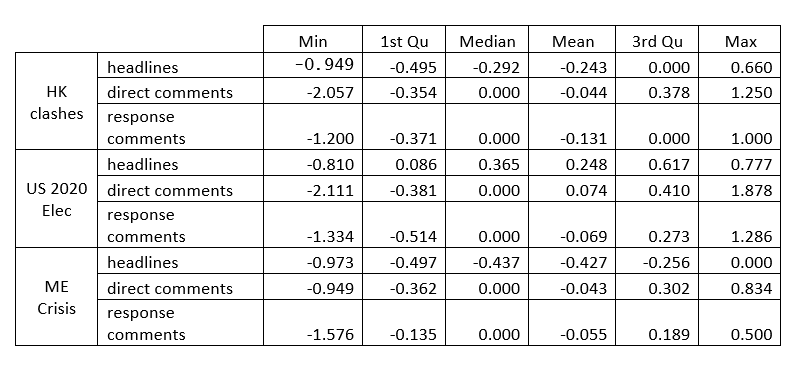


Table 2: Aggregated polarity for headlines and comments

**3. RESULTS**

The study introduces some unique and diverse topics for analysis. There is definitely a relation between how the media portrays a news item and how public opinion is influenced by the same. The wordcloud plotted in the first step of the analysis gives an indication of the trend of the topics discussed in the comments, while the sentiment analysis on the same shows the tonality of public opinion.

However, it must be noted that as the comments and the headlines were manually chosen by the author (randomly), there is a bias that could cause the results to be swayed in the direction of the study. The second point to be noted is that sample size of comments chosen is rather small to be make any form of definitive assurance with regards to the results obtained.

**4. FUTURE WORK**

To better enhance the results of this study the most important aspect is to choose a larger dataset of comments. This will help remove any of the selection bias that may be present in the data sampling. A second proposition is to make use of the NRC lexicon (having eight primary emotions) or even creating a word-sentiment network from scratch that would enable the study to analyze texts from an emotional angle rather than just classifying as positive, neutral or negative.

# **REFERENCES**

|  |  |
| --- | --- |
| [1] | Wikipedia, "Wikipedia," [Online]. Available: https://en.wikipedia.org/wiki/Media\_(communication). |
| [2] | S. Peerzada, "Digital Vidya," [Online]. Available: https://www.digitalvidya.com/blog/what-is-text-mining-guide/. |
| [3] | M. Rouse, "Search Business Anlaytics," [Online]. Available: https://searchbusinessanalytics.techtarget.com/definition/text-mining. |
| [4] | S. Gupta, "Towards Data Science," [Online]. Available: https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17. |