An Exploratory Study on Disinformation and Fake News Associated with the U.S. 2020 Presidential Election

Ali Abbasi  
University of North Texas  
1155 Union Cir, Denton, TX, USA  
AliAbbasi@my.unt.edu

Arman Derakhti  
Universidad Católica del Norte  
1221 Larrondo, Coquimbo, Chile  
derakhti.arman@gmail.com

Abstract

With the advent of social media, the spread of misinformation is undeniable. Disinformation and more specifically fake news are types of misinformation which are devised to mislead, deceive, or confuse people intentionally. Fake news was bold in the U.S. 2016 presidential election and we expect it to grow even more at the U.S. 2020 presidential election. In this research study, we aim to answer two questions:

1. What types of disinformation/fake news are likely to be utilized in this election?
2. How will the spread of these claims affect the voting outcome?

We used text mining analysis to gain a perceivable topic model. Latent semantic analysis was applied as a text mining approach on real data collected from a well-known fact-check website. Results showed four hot topics, “Covid-19”, “Trump”, “Biden”, and “Voting Process”. Based on the results, the authors believe that the two most important topics with a strong effect on the U.S. 2020 presidential election would be “COVID-19” and “Voting Process”. These topics will play an important role on both the American voters and the election results.

Keywords

An Exploratory study on Disinformation and Fake News Associated with the U.S. 2020 Presidential Election

False stories diffuse significantly faster and more broadly than true stories, and within the false stories, political stories have the fastest spread rate. There are potential psychological explanations for this, as fake news articles tend to be more novel, or more shocking, and we as humans are attracted to such stimuli. While “misinformation” can be defined as false information, “disinformation” can be defined as misinformation in an attempt to mislead, deceive, or confuse. The distinction between misinformation and disinformation becomes especially important in political contexts, where sources may make deliberate efforts to mislead, deceive, or confuse an audience in order to promote their political objectives. Recent political events and elections have been increasingly accompanied by reports of disinformation, researchers found that people are more likely to believe stories that favored their preferred candidate (Fetzer, 2004; Torabi Asr & Taboada, 2019; Zannettou et al., 2019).

The use of technological tools and techniques, including social media platforms (Shu et al. 2017), bots (Shao et al. 2017), big data (Pearson et al. 2019), trolling (Marwick & Lewis, 2017), rumors (Koohikamali & Kim, 2015), deep-fakes (Blitz, 2018), memes (Fowler, 2020), robo-calls (Gilium & Merrill, 2020), emails and others, enables those intending to manipulate public opinion by spreading false, inaccurate, or misleading information to reach targeted audiences and create tensions in society, to further political agendas, or to delegitimize political opponents (Levish, 2019). The actors for that matter, could be social influencers (Lavorgna et al., 2018), hyperpartisan media, conspiracy theorists, fake news websites, foreign governments or politicians themselves (Tucker et al. 2018).

The news media landscape has changed dramatically over the past decades. Through digital sources, there has been a tremendous increase in the reach of journalism, social media, and public engagement. Checking for news online has become ubiquitous, people are most likely to get their news through online sources than print newspapers (Harris, 2017). Disinformation on social media spread quickly in comparison to traditional media due to lack of regulation and examination required before posting (Chen et al. 2015), hence, people are inundated with information online and consequently, exposed to tremendous amount of false information, therefore, it is vital to be able to distinguish the real information from false stories in order to decide and think more clearly.

In this research study, we focus on the U.S. 2020 presidential election and we seek to answer two fundamental questions.

1. What types of disinformation/fake news are likely to be utilized in this election?
2. How will the spread of these claims affect the voting outcome?

In order to address these questions, we study the disinformation and fake news disseminated online during a 12-weeks period (from July 1st through September 25th, 2020), then examine our data set in order to topic modeling the types of disinformation using LSA method (Latent Semantic Analysis).
Methodology
Latent semantic analysis (LSA) was introduced by Landauer et al. (1998) to aid in information retrieval and search (query) optimization. The main idea behind LSA is to collect all of the contexts within which words appear, and to establish common factors that represent underlying concepts. As compared to many other text mining techniques which only analyze textual data at the syntax level by simply counting the occurrence of particular words, LSA is able to extract the contextual-usage meaning of words and obtain approximate estimates of meaning similarities among words within the given textual data, thus providing the information at the semantic level. Therefore, LSA has a wide range of applications in natural language processing, search engine and library indexing and many other areas.

LSA is generally similar to traditional factor analysis, as its main purpose is the reduction of dimensionality of original data through singular value decomposition (SVD). In a fashion similar to principal component analysis, SVD produces simultaneous principal components for two sets of variables, the terms and the documents. As such, SVD results include two sets of factor loadings, one for the terms and one for the documents. Each latent semantic factor is associated with a set of high-loading terms and a set of high-loading documents, therefore each factor represents a word usage pattern (a theme). This study follows the well-established text mining procedures used in a prior study by Sidorova et al. (2008).

1. Data Collection
We would like to gather and study the disinformation disseminated online during July 1st, 2020 through September 25th, 2020 which is the time this paper is being written. In order to predict what the main topics would be, we facilitated factcheck.org, a reliable fact check website referred by many scientific studies, in total, 327 fake news have been determined. In order to distinguish which articles’ context is associated with the U.S. 2020 presidential election, we separated those containing the terms, “Trump”, “Biden”, “election” and “vote”, the result shows 239 fake news which accounts for roughly 73% of all fake news discovered during the time period of our study.

2. Data Analysis
The obtained data were analyzed with the help of the Rapidmining software using a text mining technique. In particular, the latent semantic analysis (LSA) technique was used for this study. LSA extracts essential tokens and identifies the different meanings of a word based on the context. Indeed, this capability has made it popular in different research areas. The LSA processing based on 239 fake news is described in the following steps.

2.1. Term Reduction
Each record is specified by one fake news, and in the first step, the links are extracted for later usage by “Extract operator”, then all records come into the “tokenize” operator in which typically each word is recognized by one space called a token, but in order to avoid losing tokens with # or @ we add other punctual marks for tokenization process. In the next step, the tokens go through the “transform case” operator changes all the letters to lowercase in order to make all tokens integrated. Next, the “stopwords” operator removes all meaningless words in the English language, such as “a,” “an,” and “is”. Since stopwords are not meaningful enough we remove them from the records. Fifth, the tokens which contain between 3 to 25 letter are considered because outside of this range, tokens do not include meaningful information. In the Sixth step, the root of words is taken into account and all suffixes are removed. For example, all “likes,” “likely,” “liked,” and “liking” tokens are considered as “lik” in the result. Finally, all tokens that only appear in a document are removed. All these procedures eventuated 976 tokens.
2.2. Term Frequency Matrix

We used the technique of calculating the relatively rare weighting called term frequency-inverse document frequency (TF-IDF). Such transformation promotes the occurrence of rare terms and decreases the impact of more common non-stopwords.

2.3. Singular Value Decomposition

Singular value decomposition (SVD) is a well-known linear algebra decomposition which makes a rectangular matrix $A$ be broken down into the product of three matrices, an orthogonal matrix $U$, a diagonal matrix $S$, and the transpose of an orthogonal matrix $V$ as follow:

$$A = U S V^T$$

SVD is the most precious section of LSA. With SVD, LSA can distinguish the different words, which have the same meaning or find one word that may mean different meanings in different contexts (Lin et al., 2017; Sidorova et al., 2008). We consider the first 100 main components that explain a variation of 95% of the total with 825 terms.

2.4. Results

According to the findings, the authors determined four topics that illustrate the best classification in the fake news related to the U.S. 2020 presidential election from a general perspective. Also, LSA helps to find the most related sources of each topic. For each topic, we created a table comprising all high-loading terms and documented and sorted them out by absolute loadings in descending order. The importance of each word based on their occurrences is shown in figure 1, moreover, table 1 shows the classification obtained and high-loading terms in each category.

Figure 1. The visualization of words based on occurrences
As we expected, Trump and Biden are our first two topic categories. A survey on fake news and rumor detection by Bondielli & Marcelloni (2019), demonstrated the advent of fake news research since 2016. It was proved that the U.S. 2016 election and Trump with the diffusion of fake news could influence the result of the election. As it is expected one of the hot topics among fake news is President Trump and he plays an important role in the rest of them. Some highlights of Trump’s category, includes a false claim by president Trump that Biden “wants to rip down the wall”, however, Biden has said only that would not build any more wall, not that he would dismantle existing fencing. Moreover, Trump during one of his speeches about Kamala Harris’ eligibility said, “I heard it today that she doesn’t meet the requirement in terms of the presidency”. In addition, fake news among players (super PAC) is a hot topic these days. The second topic explains fake news related to former vice president, Joe Biden. Chiefly, disinformation in this category is against him and more specifically his tax plan. Facebook posts wrongly claim that under Joe Biden, “tax rate on a family making 75000 dollars would go from 12% to 25%” or Trump has said “a $4 trillion tax hike on almost all-American families”. Ronna McDaniel also said, “raise taxes on 82% of American”. Despite these fake news Biden’s tax plan would raise an additional $4 trillion in taxes over the next decade, but the increases would fall mainly on very high-income earners and corporations. Another hot disinformation about Biden is associated with Ben Laden. Tom Cotton and Trump misleadingly said Biden "opposed the mission to take out Osama bin Laden". The third category describes the fake news around "COVID-19". Most of disinformation in this topic has been claimed by President Trump and his administration during conferences and speeches. He made the outdated declarations about mortality and death rates. Furthermore, he has declared a misleading comparison of the Obama administration “stopped testing” during the H1N1 pandemic repeatedly. The last topic discusses fake news around the veracity of the voting process. A viral video wrongly labeled “Registered republicans not allowed to vote”. President Trump also has declared potential fraud for the mail-in ballot or its campaign claims there’s a potential for “massive fraud” in Nevada.

**Conclusion**

This study aims to explore hot topics of fake news related to the U.S. 2020 presidential election. For this purpose, we gathered data from a reliable fact check website during a 12-weeks period and found four classes as hot topics including “Covid-19”, “Trump”, “Biden”, and “Voting Process”. Based on the results, the authors believe that the two most important topics with a strong effect on the U.S. 2020 presidential election would be “COVID-19” and “Voting Process”. These topics will play an important role on both the American voters and the election results. For instance, believing a false claim about the seriousness of COVID-19 and warnings about a potential fraud in voting process might affect a voter whether they choose to vote in-person or by mail.
References


