OR/MS Today – Pre-print version
Published February 2020, Volume 47, Number 1, Pages 34-39

Fighting Fake News during Disasters

Authors: Kyle Hunt, Puneet Agarwal, Ridwan Al Aziz, Dr. Jun Zhuang

Affiliation: Department of Industrial and Systems Engineering, University at Buffalo

Article overview: Over the last decade, fake news has plagued social media platforms during disasters, causing unneeded chaos and confusion. In order to create a safer online environment, major agencies often debunk misinformation to provide the public with updated and correct news. Such tasks demand human time and effort in order to identify and subsequently monitor the misinformation that spreads during crisis events. Given this, we study the use of machine learning to support the efficient monitoring and resolution of fake news. In order to aid agencies in their decisions related to debunking false news, we develop game theoretic models that identify the optimal strategies for major agencies and accounts.

---------- Start of Article----------

Introduction

With billions of daily users, social media platforms prove to play an important role in the dissemination of news, opinions, comments, and even personal updates. On sites such as Twitter and Facebook, users can find up to date information on just about any topic of interest. Given the extreme speeds at which information can travel across these platforms, they play a vital role in the dissemination of timely news, such as political updates and risk communications. Although there are significant benefits of using social media to release timely news and updates, the unmoderated nature of sites such as Twitter can create a rumor mill among users. When false information is maliciously or mistakenly spread on social media sites it can result in a millions of misinformed users. Over the last decade, there has been an overwhelming amount of cases where false news spreads during crisis events, such as natural disasters and terrorist attacks. This misinformation proves to be very dangerous during these menacing situations, as the integrity and accuracy of emergency communications is of the utmost importance. Given that emergency communications are often needed in order for affected populations to make informed decisions regarding planning and evacuation, it is vital that misinformation is controlled in an effective and efficient manner.

The spread of fake news on Twitter during crisis events

In order to understand the magnitude of fake news dissemination during crisis events, we turn to real cases that occurred within the last decade.

In October of 2012, Hurricane Sandy caused destruction across the Caribbean before impacting the entire East Coast of the United States. The storm caused significant devastation throughout New Jersey and New York. As a result of the storm's surge, New York City was victim to flooded streets, subway lines, and tunnels. On October 28th, 2012, a meteorologist incorrectly informed CNN that the floor of the New York Stock Exchange (NYSE) was flooded. An invalid tweet was posted, which read "BREAKING: Confirmed flooding on NYSE. The trading floor is flooded under more than 3 feet of water." After this false information was disseminated throughout the news and social media, a NYSE official denied that there was any flooding, and CNN issued a correction. This misinformation caused a significant panic among the public in New York City

and around the world, as damage to the floor of the NYSE could have resulted in an economic hit, as well as the loss of a landmark building.

Every year, the Boston Marathon attracts around 30,000 contestants, and even more fans supporting the event. During the 2013 race, a bombing near the finish line killed three people, and injured many more. This act of terrorism led to a panic throughout Boston, and law enforcement evacuated the area and halted flights leaving Boston. During the online dismay which followed, many false rumors were spread. One of the most widespread rumors stated that a young girl, who was running in remembrance of the Sandy Hook victims, was killed in the bombing. A tweet, which read "R.I.P. to the 8-year old girl who died in Boston's explosions, while running for the Sandy Hook kids. #prayforboston," received over 33,000 retweets, proving to reach millions of users across Twitter's network. This false information diverted attention away from other critical news following this extreme act of terrorism.

During Hurricane Harvey in August of 2017, many homes and buildings in Houston were destroyed as a result of the high winds and overwhelming flooding. When residents were evacuating their homes in an effort to find safety, a malicious false rumor was spread across Twitter which stated that undocumented immigrants could not seek safety in Texas shelters. The claim was that shelters were checking IDs at their doors, and therefore undocumented immigrants would not be able to enter. This proved to be extremely dangerous to the lives of over 500,000 undocumented immigrants in the Houston area, as their opportunities for reaching safety were unfavorable.

On October 1, 2017, the United States was struck by an act of terrorism as a shooter opened fire at an outdoor concert in Las Vegas. The shooting resulted in 58 deaths, and over 400 injuries from gun shots. Many concert attendees rushed to find safety and evacuate the area, while many emergency responders and volunteers showed their heroism as they remained in the high-risk area to tend to the wounded people. In the chaos that ensued, false information was spread both online and offline, leading to skewed beliefs and unneeded misguidance. One of the most prominent pieces of misinformation, which was spread by an emergency dispatcher, claimed that the University Medical Center was completely out of beds and no more patients could be serviced. This information quickly contaminated Twitter's network, and led to the falsehood being extensively spread. False news such as this can cause significant harm, as the injured people needed prompt medical attention.

Although there are many cases where misinformation has been spread during disaster situations, these selected cases show the panic and distress that false news can cause during threatening situations. In order to resolve misinformation that is spread during disasters, major accounts such as government organizations, news agencies, public figures, and emergency managers often post to social media in order to dispel the falsehoods and offer corrected information.

Correcting fake news

The threat of misinformation propagating throughout social media brings the need for timely and valid debunking posts from trustworthy accounts. Without such posts, Twitter users may remain misinformed, resulting in a wider reach of the incorrect news. In many of the cases where false information spreads during disaster events, government accounts and other major accounts turn to

social media to supply the public with updated information. Sometimes this false news is even posted on webpages to expose the risks.

Following the Boston Marathon bombing and the Las Vegas shooting, false news was addressed on websites such as The New York Times, The Washington Post, USA Today, and The Wall Street Journal. Additionally, many accounts posted to Twitter in order to disprove the inaccuracies. When rumors were spread during Hurricane Sandy, the Federal Emergency Management Agency (FEMA) released a Hurricane Sandy "Rumor Control" page, where the false rumors were debunked. Many other accounts took to social media to correct the falsehoods. In the case of the NYSE flooding rumor, the account who originally spread the misinformation ended up posting a new tweet to admit to the wrongdoing, and the account also released a public apology.

When false news was spread during Hurricane Harvey, a huge debunking effort was exhibited by many different agencies. Similar to Hurricane Sandy, FEMA released a Rumor Control page which listed all of the false rumors. Aside from this, many governmental and news agencies, including Immigration and Customs Enforcement (@ICEgov), Customs and Border Protection (@CBP), the City of Houston (@HoustonTX), CNN (@CNN), and The Hill (@thehill), posted to Twitter to clarify the misinformation. A tweet from @HoustonTX, which read "WE WILL NOT ASK FOR IMMIGRATION STATUS OR PAPERS AT ANY SHELTER. No vamos a pedir documentos ni estatus migratorio en ningun albergue," received over 100,000 retweets, proving to be an extremely effective debunking effort.

From our conversations with emergency managers, and the historical evidence showing the many rumor debunking strategies, the process of identifying and monitoring misinformation can require a significant amount of time. In order to correct false news, agencies must first identify the case, and subsequently track the online and/or offline activity to understand the coverage of the topic. If the story has reached a large number of people, and no other major accounts have clarified the misinformation, then the agency will likely choose to expend their human resources to debunk. If the case is not widespread, or if other accounts have made significant debunking efforts, then agencies will likely choose to remain out of the conversation. Tracking tweets in order to monitor the impact and depth of the information spreading takes significant efforts and human resources, and through our research, we have shown how machine learning can be used to limit the need for human intervention in monitoring misinformation.

A machine learning approach to monitor misinformation

In order to automatically monitor misinformation during disasters, machine learning proves to be an effective tool. By training algorithms on a small portion of incoming data related to a misinformation case, the models can be deployed to predict the veracity of newly emerging tweets. After tweets are collected from Twitter's API, and labeled based on the veracity of their content (as either true, false, or neutral), our framework is put to work in order to learn the labeled tweets, and subsequently predict the veracity of new tweets. To learn twitter data, our methods vectorize the text of tweets using term frequency-inverse document frequency (tf-idf), and also take into account numerical features such as the number of retweets and likes. The extracted features are used to train the models using k-fold cross-validation, and then the models are tested on new data. The results of this study show that out of eight different algorithms, Support Vector Machine performed the best in monitoring misinformed tweets, achieving a predictive performance of over

87%. This research was driven by the need for efficient methods to monitor misinformation during crisis events.

So, how can this work, and how does it prove to be more effective than traditional methods?

When misinformation is identified on social media, agencies can easily collect some of the tweets using the Twitter API. After collection, the agencies then label the tweets as either true, false, or neutral. The tweets are then fed into our framework to train the models, and then deploy them on newly emerging tweets. From the results of the framework, analysts can easily and automatically monitor the misinformation to see if users are continuing to post false news related to the misinformation, or are beginning to share the truth. The analysts can also see *who* the accounts are that are spreading or debunking the misinformation. Depending on the number and nature of the accounts posting about the misinformation, the analysts can make informed decisions on whether or not to debunk. Traditionally, misinformation is monitored manually, requiring human resources to read and track the cases. With this automated approach, agencies can quickly quantify the misinformation's spread.

Using game theory to support debunking decisions

When multiple rumors propagate, official agencies must choose the specific rumor case(s) to debunk in order to effectively use their resources. Using their available information, they can react fast to minimize the spread of these rumors. If this information is imprecise, it can cause widespread confusion, and people may continue to trust misinformation and spread it unknowingly. If agencies choose to invest time to completely learn and verify the details of a possible misinformation case, this can allow false news to spread with full force, making the process of strategic decision-making very challenging.

We apply game theory to develop mathematical frameworks that can model the strategic interactions between official agencies and social media users during false rumor propagation. In game theory, each player considers the strategies of the other player(s) in formulating her own strategy in order to maximize her expected payoff. In a misinformation debunking-spreading game, the official agencies have a set of objectives, that is, to minimize the costs of debunking and the impact of false news transmission. On the other hand, social media users have a different set of objectives, such as maximizing their influence and credibility rating in the social networks. Based on these objectives, game-theoretic models serve as decision support tools for the emergency agencies to make critical decisions regarding the rumor case(s) which need to be debunked; subsequently releasing correct information to the public by effectively utilizing available resources. These models can also be utilized to determine the optimal debunking strategies for the government agencies so that they can minimize the spread of misinformation during crisis events by addressing the trade-offs between reacting fast with partial/incomplete information and reacting at a later stage with complete information.

Acknowledgement

This research was partially supported by the National Science Foundation under Award No. 1762807. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

OR/MS Today – Pre-print version
Published February 2020, Volume 47, Number 1, Pages 34-39

----------End of Article----------

Kyle Hunt Bio: Mr. Kyle Hunt is a Ph.D. student in the Department of Industrial and Systems Engineering at the University at Buffalo, focusing his studies in operations research. Hunt's research has been funded twice by National Science Foundation REU awards, and three times from the University at Buffalo's Experiential Learning Network. Hunt has received the School of Engineering and Applied Sciences (SEAS) Dean's Excellence Scholarship (November 2019), the SEAS Dean's Achievement Award (May 2019), and the University at Buffalo's Research and Scholarship Award of Distinction (April 2019). His research interests are in emergency management, national defense, and sports analytics, and his research has been widely disseminated through publications, news coverage, and conference presentations.

Puneet Agarwal Bio: Mr. Puneet Agarwal is a Ph.D. student specializing in operations research in the Department of Industrial and Systems Engineering at University at Buffalo. His research interest lies in the field of disaster risk management and strategic decision-making. His collaborative efforts have received many acknowledgements, and his team's research has been widely covered in publications and by media agencies. In 2019, he received the Geohazards Research Award from the University at Buffalo's Center for Geohazards Studies to support his research on misinformation diffusion during disasters.

Ridwan Al Aziz Bio: Mr. Ridwan Al Aziz is currently a Presidential Fellow and a Ph.D. student in the Department of Industrial and Systems Engineering at the University at Buffalo. He received his B.Sc. degree in 2014 and M.Sc. degree in 2017 in Industrial and Production Engineering (IPE) from Bangladesh University of Engineering and Technology (BUET). He served as a faculty member in the Department of IPE in BUET for 3 years. His research interest includes leveraging data analysis, machine learning, and game theory for analyzing information diffusion, debunking rumors in social media, and securing the U.S. border.

Jun Zhuang Bio: Dr. Jun Zhuang is a Professor, Department of Industrial and Systems Engineering, School of Engineering and Applied Sciences (SEAS) at the University at Buffalo (UB). Dr. Zhuang received a Ph.D. in Industrial Engineering in 2008 from the University of Wisconsin-Madison. Dr. Zhuang's long-term research goal is to integrate operations research, big data analytics, game theory, and decision analysis in order to improve mitigation, preparedness, response, and recovery for natural and man-made disasters. Other areas of interest include applications to health care, sports, transportation, supply chain management, sustainability, and architecture. Dr. Zhuang's research has been supported by the U.S. National Science Foundation (NSF), the U.S. Department of Homeland Security (DHS), the U.S. Department of Energy, the U.S. Air Force Office of Scientific Research (AFOSR), and the National Fire Protection Association. Dr. Zhuang has published 100+ peer-reviewed journal articles in Operations Research, IISE Transactions, Risk Analysis, Decision Analysis, and European Journal of Operational Research, among others.