Predicting the Influence of Fake and Real News Spreaders (Student Abstract)

Amy Zhang[1], Aaron Brookhouse[2], Daniel Hammer[3], Francesca Spezzano[4], Liljana Babinkostova[4]

[1]University of Minnesota - Twin Cities
[2]Michigan State University
[3]Western Carolina University
[4]Boise State University
zhan7007@umn.edu, brookho8@msu.edu, dlhammer2@catamount.wcu.edu,
{francescaspezzano,liljanababinkostova}@boisestate.edu

Abstract

We study the problem of predicting the influence of a user in spreading fake (or real) news on social media. We propose a new model to address this problem which takes into account both user and tweet characteristics. We show that our model achieves an F1 score of 0.853, resp. 0.931, at predicting the influence of fake, resp. real, news spreaders, and outperforms existing baselines. We also investigate important features at predicting the influence of real vs. fake news spreaders.

Introduction

The spread of misinformation across social media is one of the biggest national security threats in the 21st century. Previous research has been successful at identifying misinformation spreaders on Twitter based on user demographics and past tweet history (Shu et al. 2019), and others have been relatively successful at predicting the number of retweets of a given tweet (Nesi et al. 2018; Raj, T Vinayaka 2020). However, the problem of predicting the number of retweets of news articles tweeted by a specific user (either a fake or real news spreader) has not yet been tackled, which determines the influence of the user in the community. We use data from FakeNewsNet, containing a list of 43K known fake news spreaders and 135K real news spreaders, and the past 500 tweets of each user to build profiles of each user to predict the number of retweets their latest news article tweet will receive. We then address the problem of predicting the user influence as a multi-class classification problem and present a Random Forest classifier that categorizes the number of retweets a news tweet will receive into five ranges using user profile characteristics, emotion and sentiment analysis of tweets, and information about past tweets. This classifier results in a 0.931 and 0.853 weighted F1 score for real and fake news spreaders, respectively, performing better than other existing models, which resulted in a 0.928 weighted F1 score for real news spreaders and 0.832 F1 score for fake news spreaders at best. By comparing important features for predicting the influence of real and fake news spreaders, we show that an established account and the utilization of sources better characterize the influence of real news spreaders, while user interaction on Twitter is more important to determine the influence of fake news spreaders.

Dataset

We used the Politifact dataset from the FakeNewsNet repository (Shu et al. 2020). [1] This dataset consists of 429 fake news articles and 619 real news articles labeled by the politifact.com fact-checking website and Twitter IDs of users who shared those articles on Twitter. In total, 70,655 users had posted at least one fake news article listed in the dataset and 189,951 had posted only real news articles in the dataset. For each user, we used the web-scraper Twint[2] to extract the text, posted time and date, number of likes, and number of retweets of up to 500 tweets the user had posted before the date of the news article tweet. We then excluded users with less than 100 tweets in their timeline and those who tweeted both real and fake news. This reduced our datasets to 43,119 fake news spreaders (FNS dataset) and 135,234 real news spreaders (RNS dataset). In both datasets, each user's ground truth (influence) is related to the number of retweets their latest tweeted news article received. More specifically, by following the problem statement proposed by Nesi et al. (Nesi et al. 2018), we grouped the number of retweets into five classes: 0 retweets, 1-10 retweets, 11-100 retweets, 101-1000 retweets, and 1000+ retweets.

Features

We propose the following three categories of features to train machine learning models for predicting the influence (retweets) of users in the FNS and RNS datasets.

User Profile Features: These features consist of the number of followers, number of following, follower to following ratio, account age in days at the time of the news tweet, number of statuses (total tweets of the user), listed count (the number of people that added the author to a list), and number of posts favorites by the user.

History-based User Features: These features were calculated from each tweet collected from Twint before the date of the target tweet and averaged by the number of valid tweets that were retrieved. These consisted of the average number of replies, average number of likes, average number of retweets, average number of mentions, average number of hashtags, average number of URLs, average time be-

---

[1]The GossipCop dataset is not used in this study as it deals with a different form of misinformation.

[2]https://github.com/twintproject/twint

Model	RNS (F1 Score)	FNS (F1 Score)
(Nesi et al. 2018)	0.928	0.832
(Raj, T Vinayaka 2020)	0.900	0.821
GCN	0.523	0.634
GAT	0.336	0.444
SAGE_GCN	0.562	0.641
Our Approach	**0.931**	**0.853**

Table 1: Weighted F1 Scores: Real News Spreaders (RNS) and Fake News Spreaders (FNS) Datasets. Best scores are in bold.

tween tweets in seconds, average emotions present (presence of keywords in tweet history indicative of 8 emotions and 2 polarities computed using the NRC Emotion Lexicon (Mohammad and Turney 2013), normalized to sum up to 1), and average sentiments present (analysis of collected post history using VADER[3] sentiment analysis).

Tweet-based Features: These features are directly related to the tweet sharing a news article. These features consist of the number of hashtags, number of mentions, number of URLs, average emotions present, average sentiments present, the cosine similarity of average sentiments to the sentiments in a user's history, and the cosine similarity of average emotions to the emotions in a user's history.

Experimental Results

We tested our features for the classification task by using 10-fold cross-validation with several machine learning models, namely K-Nearest Neighbors, Multi-layer Perceptron, Random Forest, and Ada Boosting. We used class weighting to deal with class imbalance. Due to multi-class and data imbalance, F1 score was used as evaluation measure. Due to space limitations, we report F1 scores only for the best performing model (Random Forest) which achieved a weighted F1 score of 0.931 on the real news spreaders dataset and 0.853 on the real news spreaders dataset (cf. Table 1).

Comparison with Baselines. We compared our approach with the following baselines. As network properties in the follower network may also characterize real vs. fake news spreaders (Vosoughi, Roy, and Aral 2018), we considered three types of graph neural networks (GNN) as baselines, namely Graph Convolutional Network (GCN), Graph Attention Network (GAT), and SAGE GCN (Hamilton, Ying, and Leskovec 2017). The node data for each user was the set of user-based and tweet-based features we proposed in this paper. We also compare with the approach proposed by Nesi et al. (Nesi et al. 2018) for predicting retweets of general tweets (also based on feature engineering plus classification with Random Forest) and the first place solution to the CIKM AnalytiCup 2020: COVID-19 Retweet Prediction Challenge, which proposed a deep-learning-based solution with personalized attention (Raj, T Vinayaka 2020).

[3]https://pypi.org/project/vaderSentiment/

As we can see from Table 1, GNN-based baselines perform consistently worse than the other two considered, although they were more effective on the fake news spreaders than the real news spreaders. (Nesi et al. 2018) resulted in the best baseline method with an F1 score of 0.928 for real news spreaders and 0.832 for fake news spreaders. However, our proposed model achieves better performances in both datasets and for each individual class. We also observe that predicting the influence of fake news spreaders resulted, in general, more challenging than for real news spreaders.

Feature Importance. We ran feature importances with a forest of trees on both datasets and, for each dataset, normalized the feature importance scores to sum up to 1 to allow comparison. We then selected the top-5 features with the highest absolute value of the difference between the importance score in the RNS and FNS datasets. We found out that the number of URLs in the tweet and the user account age were more important at predicting the influence of real news spreaders than fake news spreaders (positive difference between importance scores). This indicates that an established account and the utilization of sources contributed more to characterize the influence of real news spreaders. On the other hand, the average number of replies, the average number of users the author is following, and the number of statuses were more important at predicting the influence of fake news spreaders than real news spreaders (negative score difference). This indicates that user interaction is more important to determine the influence of fake news spreaders.

Acknowledgements

References

Hamilton, W. L.; Ying, Z.; and Leskovec, J. 2017. Inductive Representation Learning on Large Graphs. In *NIPS*, 1024–1034.

Mohammad, S. M.; and Turney, P. D. 2013. Crowdsourcing a Word-Emotion Association Lexicon. 29(3): 436–465.

Nesi, P.; Pantaleo, G.; Paoli, I.; and Zaza, I. 2018. Assessing the ReTweet Proneness of Tweets: Predictive Models for Retweeting. *Multimedia Tools Appl.*, 77(20): 26371–26396.

Raj, T Vinayaka. 2020. CIKM AnalytiCup 2020: COVID-19 Retweet Prediction with Personalized Attention. *CEUR Workshop Proceedings*, 2881.

Shu, K.; Mahudeswaran, D.; Wang, S.; Lee, D.; and Liu, H. 2020. FakeNewsNet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media. *Big Data*, 8(3): 171–188.

Shu, K.; Zhou, X.; Wang, S.; Zafarani, R.; and Liu, H. 2019. The role of user profiles for fake news detection. In *ASONAM*, 436–439.

Vosoughi, S.; Roy, D.; and Aral, S. 2018. The spread of true and false news online. *Science*, 359(6380): 1146–1151.