FakeNewsTracker A Tool for Fake News Collection, Detection, and Visualization

FakeNewsTracker is an end-to-end system developed in Arizona State University for fake news pieces collection, detection, and visualization. FakeNewsTracker uses fact-checking websites like PolitiFact and GossipCop to collect news articles related to fake and real news and their relevant social media data from Twitter. FakeNewsTracker provides deep learning based model that uses news content and social context to detect a news article as fake or real. It also provides a web-based interface for analysis and understanding of features from the data collected.

Challenges of Fake News Detection

Fake news detection is an important and technically challenging problem. In an attempt to tackle the growing misinformation, several fact-checking websites have been deployed to expose the fake news. These websites play a crucial role in clarifying fake news, but they require expert analysis which is time-consuming. Numerous articles and blogs are written in order to distinguish fake news from the true news. However, they are not from authority sources and may be biased, which make the labels not fully reliable and convincing. Due to the volume and diversity of the social media, it is almost impossible to manually label the fake news and true news. Also, the information in social media is spread at an alarming rate and hence a framework is required to detect fake news in early stages and avoid dissemination. Thus, to solve these challenges, we present a system FakeNewsTracker, to facilitate the community for studying fake news.

Access FakeNewsTracker using `http://blogtrackers.fulton.asu.edu:3000/#/about`

FakeNewsNet dataset is available at `https://github.com/KaiDMML/FakeNewsNet`

For details or qeuries contact

Kai Shu
kai.shu@asu.edu

Deepak Mahudeswaran
dmahudes@asu.edu

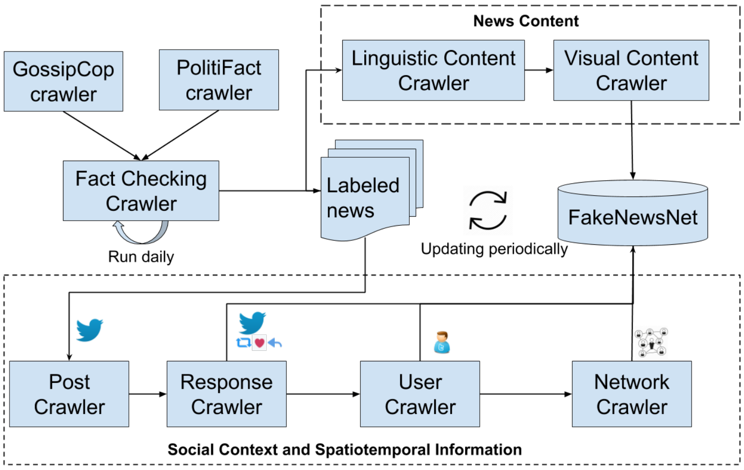
Huan Liu
huan.liu@asu.edu

The major components of FakeNewsTracker (Shu, Mahudeswaran, & Liu, 2019) system are as follows,

- *Fake News Collection*: collecting news contents and social context automatically, which provides valuable datasets for the study of fake news
- *Fake News Detection*: extracting useful features and build various machine learning models to detect fake news;
- *Fake News Visualization*: presenting the characteristics of fake news dissemination in social media through effective visualization techniques.

As part of the demo, we will discuss data collection process, FakeNewsNet (Shu, Mahudeswaran, Wang, Lee, & Liu, 2018) Github repository usage and FakeNewsTracker interface for understanding dissemination process. This is an ongoing project and we plan to provide several other functionalities in the future.

Data Collection



The flowchart of dataset integration process for FakeNewsNet. It mainly describes the collection of news content, social context and spatiotemporal information.

In this section, we demonstrate how we can collect news contents with reliable ground truth labels, how we obtain additional social context information, and how we can dynamically updating our repository in a periodical manner. The flowchart of the data collection process is shown in above figure, which mainly consists of the collection of news contents, social context, and dynamic information. In these fact-checking sites, fake news articles are labeled by the trusted professional authors and relevant claims are made by the authors on why the mentioned news is not true. The FakeNewsTracker system collects multiple dimensions of data including text, image, and social information. The news content crawler collects the news article's text contents and its metadata including a list of urls of images in the article, publisher information, etc. The social context crawler makes use of Twitter API to gather tweets related to fake/real news that spread on Twitter. Additionally, the different social engagements including replies of tweets, retweets of tweets and favorites of tweets are collected for each tweet. Comments to tweets provide information about user's emotion towards a news post on Twitter and can provide useful signals in fake news detection.

Using the user and network crawler, all the user profiles and the social network of the users involved in engagements are collected. This provides rich information to analyze news articles from the social perspective. The dataset is updated in a dynamic manner.

FakeNewsNet dataset

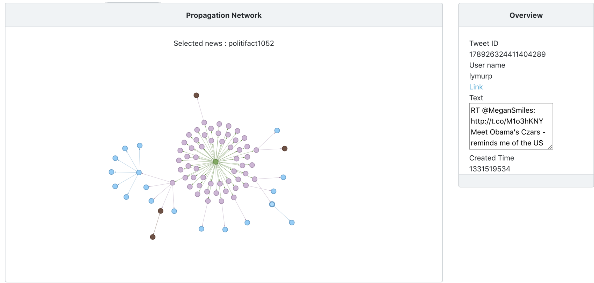
We construct and publicize a multi-dimensional data repository *FakeNewsNet*, which currently contains two datasets with news content, social context, and spatiotemporal information. The constructed FakeNewsNet repository has the potential to boost the study of various open research problems related to fake news study. First, the rich set of features in the datasets provides an opportunity to experiment with different approaches for fake new detection, understand the diffusion of fake news in social network and intervene in it. Second, the temporal information enables the study of early fake news detection by generating synthetic user engagements from historical temporal user engagement patterns in the dataset. Third, one can investigate the fake news diffusion process by identifying provenances, persuaders, and developing better fake news intervention strategies. Our data repository can serve as a starting point for many exploratory studies for fake news, and provide a better, shared insight into disinformation tactics. Because of Twitter's policy, we cannot disclose the entire dataset and hence we provide API's in the Github repository to collect the entire dataset.

Trend of tweets posting fake and real news in Twitter

Fake News Detection

We have proposed Social Article Fusion (SAF) model that make use of news article's text contents and social context to classify a news article as fake or real. The features from the news article are extracted using auto-encoder that uses reconstruction loss to learn its representation. The temporal pattern of social engagements is captured using recurrent neural networks(Ruchansky, Seo, & Liu, 2017). The model combines the features generated by the auto-encoder and social context recurrent neural network and concatenates them together to form a single concatenated feature vector for the classification. Then weights are learned to map the combined features to classified labels. The softmax layer is used to classify the news articles in the output layer. We train both the feature learning and classification tasks together so that the features are learned relative to the detection task rather than capturing plain linguistic differences and social engagements.

Fake News Visualization



Propagation network and its details from Twitter

We have developed a fake news visualization as shown in the figures for the developing insights on the data (Morstatter, Kumar, Liu, & Maciejewski, 2013). We have developed various interfaces for visualizing the different dimensions of our dataset. For identifying the differences in the news content of the real and fake news, we have used word cloud for analyzing the topical distributions. The dashboard also allows one to observe the trend of tweets related to fake and real in a specified interval of time. Further, we have provided a comparison of feature significance and model performance as part of this dashboard. Using the social network of the users posting and engaging with tweets, we visualize the social network to identify the differences between the users who interact with the fake news and the true news. Additionally, to understand the propagation of information on Twitter, we have visualized hierarchical propagation network(Shu, Mahudeswaran, Wang, & Liu, 2019) which involves both the retweet network and reply network. This interface allows to visualize the patterns and contents of information propagation.

Conclusion and Future work

Through this demo, we introduce FakeNewsTracker system, which provides general solutions for data collection, interactive visualization, and analytical modeling towards fake news detection. Further, we provide guidelines to use API's from FakeNewsNet repository to collect the complete dataset. The rich dimensions of dataset provides way for several interesting future works like using understanding propagation networks for fake news detection, early fake news detection using publisher bias and spatiotemporal features.

References

Morstatter, F., Kumar, S., Liu, H., & Maciejewski, R. (2013). Understanding twitter data with tweetxplorer. In *Proceedings of the 19th acm sigkdd international conference on knowledge discovery and data mining* (pp. 1482–1485). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/2487575.2487703 doi: 10.1145/2487575.2487703

Ruchansky, N., Seo, S., & Liu, Y. (2017). Csi: A hybrid deep model for fake news detection. In *Proceedings of the 2017 acm on conference on information and knowledge management* (pp. 797–806). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/3132847.3132877 doi: 10.1145/3132847.3132877

Shu, K., Mahudeswaran, D., & Liu, H. (2019, Mar 01). Fakenewstracker: a tool for fake news collection, detection, and visualization. *Computational and Mathematical Organization Theory*, 25(1), 60–71. Retrieved from https://doi.org/10.1007/s10588-018-09280-3 doi: 10.1007/s10588-018-09280-3

Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2018). Fakenewsnet: A data repository with news content, social context and dynamic information for studying fake news on social media. *CoRR, abs/1809.01286*. Retrieved from http://arxiv.org/abs/1809.01286

Shu, K., Mahudeswaran, D., Wang, S., & Liu, H. (2019). Hierarchical propagation networks for fake news detection: Investigation and exploitation. *CoRR, abs/1903.09196*. Retrieved from http://arxiv.org/abs/1903.09196

This material is based upon work supported by, or in part by, the ONR grant N00014-16-1-2257, ARO (W911NF-15-1-0328) and ONR N000141310835.