

Exploiting Network Structure to Detect Fake News

Meghana Rao
Department of Statistics
Stanford University
mvrao@stanford.edu

Neel Ramachandran
Department of Computer Science
Stanford University
neelr@stanford.edu

Anika Raghuvanshi
Department of Computer Science
Stanford University
anikar@stanford.edu

Abstract—Fake news classification has mostly been limited to text classification. We analyze the process of news propagation through social media by looking at the way in which articles are shared and spread from person to person. We use the FakeNewsNet depository, which contains data on several Twitter articles. It contains data on each article’s text as well as data on the social context (how users share articles and who they follow). We leverage both text and social context data to improve purely text-based models. Our results show that using a combination of features surrounding the article improves detection. We also demonstrate the effectiveness of purely social context features as inputs to a logistic regression model and neural network, as well, demonstrating the power of leveraging user community interactions for fake news detection.

I. INTRODUCTION

The spread of fake news through social media has been a dominant topic since the 2016 elections. Most work on fake news detection relies purely on text-based models to determine article validity. For example, a lot of research has been done to train NLP models to classify articles and social media posts as fake news. We aim to analyze the network structure of news propagation and leverage relationships between users and articles in order to improve on text-based models to detect fake news.

A new approach to combating fake news is to study the spread of it from a network perspective, where relationships between users and articles, as well as between users and other users, are important. It is quite possible that the manner in which the news was shared could be indicative of how authentic the news is. Perhaps if a user shares more often, this means that they have a higher percentage of shared articles that are fake. Maybe if a user has more followers, that this suggests that they are more trustworthy. It could be possible that articles shared multiple times are more likely to be fake. These are all potential hypotheses focused on how the article is shared rather than the actual content of the article.

It could also be true that the content of the article influences how it was shared. Perhaps users who more

often believe that a fake news article is real are drawn to similar uses of language or writing style. Because there is likely intersection in how the text and social features interact with each other, we decided to run models which took into account both types of features.

For our models, we experimented with three different types of inputs. Some models used purely text-based features (termed ‘Article Context’) for input. Some of the models use features purely based on social network relationships (termed ‘Social Context’). Some models used a hybrid of article context and social context features for input (termed ‘Hybrid’).

Our baseline model was a Naive Bayes Classifier using only Article Context as inputs. We also implemented a Logistic Regression model that was tested on Article Context, Social Context, and on Hybrid inputs. We further improved the Hybrid Logistic Regression model through Iterative Classification. We also ran a shallow neural network model on Article Context, Social Context and Hybrid. The output of all of these models were the fake/real predictions for the articles in our test dataset.

II. RELATED WORK

Jang et al. showed a comparison between fake and real news propagation and found that real news reached a wider audience, but fake news spreads through echo chambers and goes through a much higher number of iterations. [1] The study was limited in scope (looking only at the spread of 60 stories through tweets) and provides threads for further work. Previous research has also experimented with simpler models such as Naive Bayes classifiers. For example, a research group from Vinnitsia National Technical University achieved 74% accuracy in classifying news articles that had been shared on Facebook [2]. MIT’s Computer Science and Artificial Intelligence Lab recently published a paper on determining the validity of a news source using article and social context as features. Their inputs consisted of a sample of articles from the news medium, information from its Wikipedia page, information from its Twitter account,

the structure of its URL, and information about its Web traffic, with the article body being the best performing feature type overall [3]. Researchers from the University of California detected fake from real articles shared on Facebook purely from who liked the posts, with 99% accuracy [4]. Research has also looked at utilizing user’s perceptions of fake news with Facebook’s new “flagging” feature and understanding how to leverage community signals to identify misinformation [5].

III. DATA AND FEATURES

A. Data

We use the FakeNewsNet data repository released in August 2018 by Shu et al. The dataset contains 422 articles that were posted on Twitter in 2016, with 211 articles labeled real and 211 labeled fake. These ground truth labels are drawn from both BuzzFeed and PolitiFacts article fact-checking services.

Each article in the dataset includes the headline, text, source, images, and other metadata. In this project, we focus only on the headline and text of the article. We choose to ignore the source because we wanted the fake news detection task to be more difficult than simply learning which sources tend to produce fake and real articles (i.e. articles from a source such as nytimes.com are almost definitely real news). However, we see images and metadata as legitimate features to incorporate in future work.

Abbreviated examples of fake and real articles are shown below.

Real Article:

Source: *washingtonpost.com*

Date: *September 19, 2016*

Title: *France becomes the first country to ban plastic plates and cutlery*

Text: *France has apparently become the first country in the world to ban plastic plates, cups and utensils, passing a law that will go into effect...*

Fake Article: Source: *ihavethetruth.com*

Date: *December 3, 2016*

Title: *Obama: I bet when you die youll be happy to pee on my grave*

Text: *This is a true story. General Stanley McChrystal was the Commander of US Forces in Afghanistan and he had frequent disagreements on the conduct of the war...*

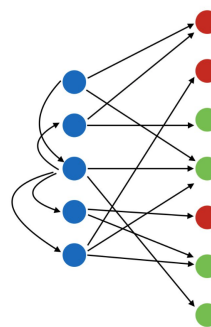


Fig. 1. A graph visualization of a small subset of the dataset. User nodes (left column) are shown in the blue, while article nodes (right column) are shown in red for fake articles and green for real articles. Edges between users indicate follower-followee relationships, and edges between users and articles indicate articles shared by a user.

The dataset includes what we term the Social Context of the articles being posted on Twitter. The dataset includes the (anonymized) user IDs of 39,122 Twitter users who interacted with the given articles. For each user, the dataset provides the set of articles that user shared on Twitter (ie, user-article interactions) as well as how many times they shared each article. It also provides each users set of followers (ie, user-user interactions). Thus, the social content portion of the dataset forms a graph where nodes are users and articles with edges between them. A visualization of a subset of the dataset is shown in Fig. 1.

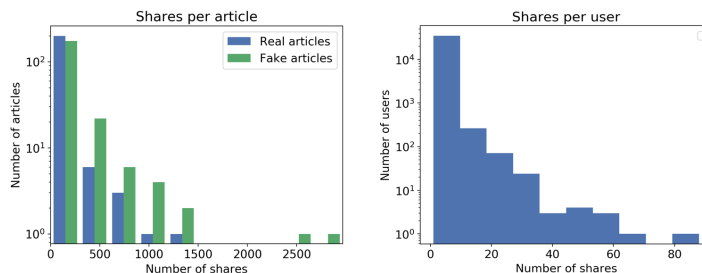


Fig. 2. Shares per article (left) and shares per user (right) histograms generated from the dataset. The y-axis is on a logarithmic scale.

Preliminary analysis of social context in Fig 2. shows both expected and surprising behaviors. Firstly, we find that the distribution of shares per user is roughly exponential. A large number of users have a low number of article shares (ie 0-20), and a small number of users have a large number of article shares (20+). The same holds for shares per article – most articles have a low number of shares, and a small number of articles have a large number of shares. Perhaps surprisingly, we also observe

that on average fake articles have much higher user share counts than real articles. This is an indicator that there are substantial differences in the social propagation of real and fake articles.

B. Features

We construct the three types of feature vectors we test across our models as follows:

1) *Article Context*: For article context, we generate features based purely on the title and text body of the article. Here, we take a simple bag-of-words approach where the feature vector is approximately the size of the training set vocabulary, and feature x_i in the feature vector x is 1 if the i_{th} word in the vocabulary is present in the article and 0 otherwise. We include a few other simple features, such as the number of misspelled and capitalized words, and the length of the title and text body.

2) *Social Context*: For social context, we made use of certain information for each article including (a) how many times the article was shared, (b) users who shared the article, (c) users who viewed the article and (d) how many articles the sharer of the article has shared.

Each article has one feature for each user, where the value at the index of user i is a certain weight if the user has shared the article and a certain lesser weight if the user has been exposed to the article (i.e. is a follower of someone who shared the article). The weights we ended up using in our final model were 2 for sharers and 1 for followers.

3) *Hybrid*: The hybrid features are essentially the Article Context and Social Context features appended to each other.

IV. METHODS

A. Baseline Approaches: Naive Bayes, Logistic Regression

As a baseline, we first decided to implement a Naive Bayes method using the text features (Article Context). This relies on the Naive Bayes assumption. Given a feature vector x where x_i indicates the presence of the i_{th} word in the article, we assume:

$$P(x|y) = P(x_1, x_2 \dots x_n|y) = \prod_i P(x_i|y)$$

which, by Bayes' Theorem, allows us to predict labels by:

$$P(y|x) = P(x|y)P(y)$$

where we compute $P(y|x)$ for both $y = 0$ and $y = 1$ and pick the label which maximizes the quantity.

Another baseline model we implemented was logistic regression model for our simplistic baseline approach. In logistic regression, we make predictions as follows:

$$P(y|x) = h_{\theta}(x) = g(\theta^T x)$$

where g is the sigmoid function. We predict a value of 1 if $P(y|x) > 0.5$ and 0 otherwise.

Observing the top words (especially those identified by Naive Bayes), there appears to be a stark difference in topics and content of the fake and real news stories in the dataset.

B. Iterative Classification

To improve on the baseline, we incorporate the network structure into our logistic regression classifier using iterative classification techniques suggested by Neville and Jensen [2]. In iterative classification, we train two classifiers, which we call the *regular* and *dynamic* models. The models both use logistic regression with slightly different feature vectors, detailed below:

1) *Regular model*: The regular simply uses the hybrid feature vector described in section III.

C. Dynamic model

The dynamic model uses the hybrid feature vector, and adds features that capture the relationships between articles, which are not explicitly provided in the dataset. Given an article x , we define *related articles* as the set R of all articles shared by users that also shared x . We add features for the number of articles in R that are real and fake, as well as the ratio of real to fake articles. In the training data, we can easily compute these extra features and train our model by checking the ground-truth labels of articles in the set R . However, at testing time we cannot check the labels of R – thus, we first use the regular model to compute predictions, or 'soft assignments' for the labels of the articles, and then use these assignments to construct the feature vector for our dynamic model. When we then make predictions on the test data, our initial soft assignments may change, which means we can recompute the feature vector and make predictions again, repeating this process until convergence (hence the name iterative classification).

D. Article-Context Neural Network

We implemented a neural network in PyTorch, using the words from an article and its weighted title as input

features. The network has an initial linear layer that maps from all input features to the number of the hidden units (10). Then, the network has an activation layer which uses a rectified linear unit (ReLU) that outputs the maximum value of the input and 0. The output of the ReLU activation is fed to a second linear layer that maps from the 10 hidden units to a single scalar value. Ultimately, a sigmoid output layer calculates the probability that an article is fake. If the output probability is greater than 0.5, then the article is classified as being fake, otherwise it is classified as being real. The NN has 10 hidden units with binary cross entropy loss and stochastic gradient descent with momentum. All NN models had a learning rate of 0.001 and were trained for 1000 epochs.

E. Social-Context Neural Network

We ran a shallow neural network on the social context features described in the Data and Features section. The hyperparameters and setup were the same as for the article-context neural network.

F. Hybrid Neural Network

The hybrid method uses the same setup as the previous two methods. It combines the article context and social context features and runs the model on this updated feature array.

V. EXPERIMENT AND RESULTS

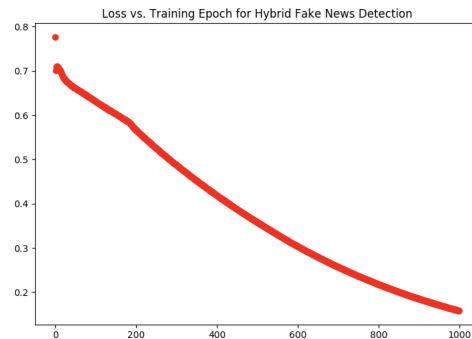
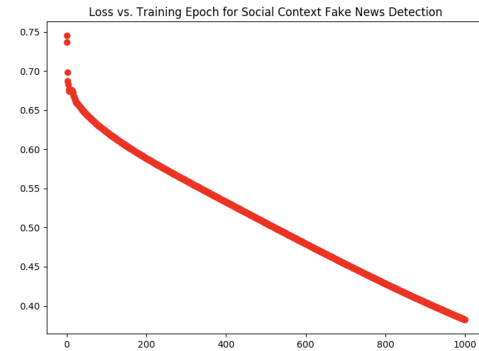
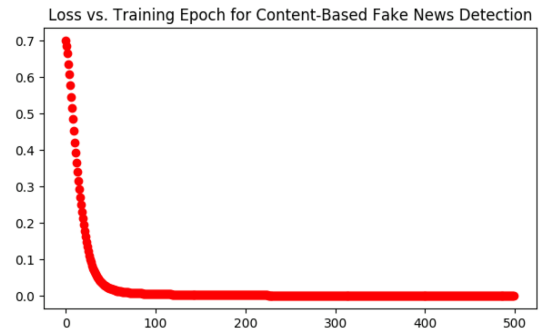
A. Figures and Tables

The following table shows the results of the top 5 words found through Naive Bayes and Logistic Regression. We observed these words in order to see if there was a stark contrast in the language that is used in fake versus real news articles.

| Model | Indicator | Top 5 Words |
|---------------------|-----------|--|
| Naive Bayes | Fake News | Fake, Marijuana, Murder, Sex, Prison |
| | Real News | Debate, Voting, Candidates, Keith, September |
| Logistic Regression | Fake News | President, Your, Were, Comments, White |
| | Real News | Story, Black, Debate, Trump, September |

Fig. 3 shows the training and test accuracy for all of the models that we implemented.

Additionally, the following graphs show the loss over the training epochs for the article context, social context and hybrid neural networks. For the hyperparameters, we ended up using a learning rate of 0.001 and 1000 epochs.



VI. DISCUSSION

Iterative classification with logistic regression achieved the lowest test error. We can see that logistic regression with hybrid features achieves a test accuracy of 83.26% and logistic regression with iterative classification achieves an accuracy of 87.7%; thus, the iterative classification framework accounts for over a 4% increase in accuracy.

While its performance was not the highest, a neural network with purely social context identified fake news articles with 63% accuracy. One of the main challenges of the project was understanding how to appropriately incorporate user-user interactions into neural networks. The training and testing were done on a mixed dataset of articles fact-checked by BuzzFeed and Politifact, creating a sparse matrix of user interactions between the sets of articles. The hybrid neural network was created by combining the input features of the social context and

| Model Type | Input Features | Training Accuracy | Test Accuracy |
|----------------------------------|-----------------|-------------------|---------------|
| Naive Bayes | Article Context | 98.6% | 71.5% |
| Logistic Regression (LR) | Article Context | 100% | 81.5% |
| | Social Context | 100% | 81.5% |
| | Hybrid | 100% | 83.26% |
| LR with Iterative Classification | Hybrid | 100% | 87.7% |
| Neural Network | Article Context | 99.9% | 72.1% |
| | Social Context | 85.1% | 65.1% |
| | Hybrid | 99% | 81% |

Fig. 3. Accuracy of different models and input features on training and test sets.

article context NN, but the expected lift in performance was not seen. This may be caused by redundancies in input features and the sparseness of the input matrix.

VII. FURTHER WORK

Future work includes improving the Social Context and Article Context neural network performance through additional feature engineering and additional tuning of hyperparameters. We tried several different combinations of neural network layers, but it would be beneficial to complete a more in-depth analysis on which layers are suited for this particular problem.

Additionally, it would be beneficial if we could get access to a larger dataset to test our models on. With the relatively FakeNewsNet dataset, the models are more subject to overfitting, and a large dataset would prevent this as well as provide an even more accurate representations of the social trends that are actually being seen in practice.

VIII. CONTRIBUTIONS

We believe we have done a good job of distributing tasks on this project. Neel implemented the logistic regression baseline and iterative classification. Meghana implemented the Naive Bayes baseline and the neural

network structure. Anika implemented the social context features after doing a preliminary analysis on user relations, along with the social context neural network. The team together implemented the hybrid methods and spent time reading up on previous work and debugging challenges of the models together.

IX. LINK TO CODE

<https://github.com/nee1r11/cs229>

REFERENCES

- [1] S. Mo Jang, et al. A computational approach for examining the roots and spreading patterns of fake news: Evolution tree analysis Computers in Human Behavior: Volume 84, 2018. P. 103-113, ISSN 0747-5632,
- [2] M. Granik, V. Mesyura, "Fake news detection using naive Bayes classifier," 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), Kiev, 2017, pp. 900-903.
- [3] R. Baly, et al. "Predicting Factuality of Reporting and Bias of News Media Sources," 2018 Conference on Empirical Methods in Natural Language Processing, 2018.
- [4] E. Tacchini, et al. "Some like it hoax: Automated fake news detection in social networks." arXiv preprint arXiv:1704.07506. 2017.
- [5] S. Tschitschek, et al. "Fake News Detection in Social Networks via Crowd Signals." 2018. In Companion Proceedings of The Web Conference. 517524. Alternate Track on Journalism, Misinformation, and Fact-checking.