

A Novel Technique for Detection of Fake News on Social Media - A Review

¹Moulya M, ²K S Veda, ³Chandana N S, ⁴Amulya

Department of ISE,
VTU/AIET, Mangalore, Karnataka, India

Abstract: Automatic identification of fake news in social media like Facebook has attracted attention for a few days now. The main challenge is to acquire sufficiently large learning data. Here we also address about weakly supervised method, which can automatically gather large-scale training dataset of hundreds of thousands of tweets. During the collection process, we can automatically identify the source as trustworthy or untrustworthy sources and train the classifier on these datasets.

Keywords: Fake news, Weakly Supervised.

I. INTRODUCTION

First let us know what is Fake news, Fake news is a neologism that is often used to refer to manufactured news. Now, the Fake news has become a well-recognized topic and much more important to identify them. Many methods of identifying fake news in society have therefore gained some attention.

It's all about trustworthy and untrustworthy classification of the dataset. We compile a dataset of large scale with tags of low quality. The source's self-confidence instead of the tweet.

False news can introduce false positive sources of untrustworthy spread a mixture of real and fake news. We show you that by gathering such a dataset, which can be done automatically, the lack of hand-labelled information can be solved. This paper assumes the secrecy of the source of tweet, rather than the tweet reality itself. And it even shows the results of the high quality identifying a tweet as fake news or real news.

Objectives

The main objective is to detect the fake news, which with a straight forward proposal is a classic text classification problem. Building a model that can distinguish between Real News and Fake News is needed.

II. Literature Review

False identification of news in social media now gets equal publicity. Fake news identification has been performed with flux of research work and publications. There are the following types of fake news.

First the Satire or Parody – sites like the Onion and Daily Mash post fake news stories as satirical efforts to satirize the media, but when posted out of context they have the ability to mislead.

The fake news that is kind of accurate but used in the wrong context – selectively selected real facts that are published to gain attention but appear to misrepresent scientific research.

Thirdly, The Sloppy article which suits the agenda – news containing some grains of truth not completely checked and used to support a particular opinion or view.

Forth, The Misleading news that is not based on facts but promotes an overarching narrative – news that has no foundation for evidence, often where views or beliefs conflict and unconscious biases come into play. Theories of conspiracy tend to fall down here!

Five the intentionally disappointing – news intentionally created to either make money through a number of clicks or cause confusion or dissatisfaction or as sensational propaganda.

In detecting fake news, machine learning plays a vital role. The algorithm such as Naive Bayes, Support Vector Machine, etc. Helps us to spot fake news and provide us with better results.

Naive Bayes Theorem can be used to infer that any news from a large or small dataset can be categorized in less time as fake or real news by comparing it with previous data set values, which in turn helps users believe in specific news. Now using this theorem on the test set, which is a decent result, they had achieved classification accuracy of about 74 %.

Support Vector Machine is a supervised machine learning algorithm that can be used to classify and regress. Using the support vector machine algorithm, 92% of the accuracy can be achieved.

Natural Language Processing (NLP) is a tool that deals with the communication using natural language between computers and humans. Natural language processing's ultimate goal is to learn, translate, comprehend and understand human languages in a meaningful manner. It segment sentences, word tokenizes, predicts parts of speech for each token, lemmatizes text, identifies stop words, parses dependence, finds noun phrases and recognizes the entity. Hence is enabled to detect fake news resulting almost 99% of accuracy.

Recently, they achieved F1 score results of 0.77 and up to 0.9 by using weakly supervised approach. Through allowing a certain amount of tag noise that can still be well done, the issue of collecting large-scale training data sets for fake news identification can be solved.

III. RELATED WORK

Typically, in the following, we will examine some of the published work. False news can be classified into three groups. The fake news which is totally fake, created by the articles authors. The satirical satire news that is fake news with the main purpose of making humor accessible to the readers. Fourth as mostly fake news, which is poorly written news articles and includes some real news but not completely accurate. In short, this kind of news aims at pushing some agenda or biased opinion.

IV. Datasets

The principle of machine learning is used to identify fake news, collect twitter information on a large scale. We also collected a small, hand marked data set for evaluation purposes.

A Large Scale Training Dataset: they compile the training dataset with trusted and untrustworthy sources. All articles from the trusted source are listed as actual news, and all tweets from an untrusted source are labelled as fake news.

It doesn't mean that all the tweets in the untrustworthy sources are fake news, at times it may also be real. If the tweet in the trusted and untrustworthy source has the same tweet as the actual news.

While it can be done automatically on a large scale, this labeling may not be perfect. Most of the time, untrustworthy sources mix fake and real news, and accidentally fake news can be spread by trustworthy sources. But we assume that this may be rare in the event.

We would collect list of different web pages and even web catalogs open sources to collect fake news source.

We could use a copy of the latest DMOZ shutdown catalogue to collect real news sources, including those news sites listed as trustworthy in open sources and filtered the web to those with an active twitter stream. Next, you can use the Twitter API to access original tweet.

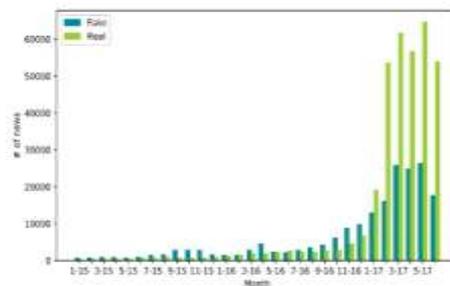


Fig. 1. Distribution of tweets labeled as real and fake news in the training dataset

The calculation reveals that the class of fake news consists of only 15% fake news tweets, 40% real news tweets, while the remainder are either no news or undecidable. Furthermore, since the study involves both real and fake news tweets, we can also consider real news as non-fake news and because the actual news classification is higher by factor of three, the classifier is more likely to identify them as real news.

A Small Scale Analysis Dataset: 116 tweets from the homepage of politifact identified by reporters as fake news. To generate negative examples, We may select those 116 tweets in a positive class based on TF-IDF and cosine similarity that were the nearest to fake news tweets and delete those 116 tweets from the dataset before training the classification models.

An Evaluation Possibilities: There are two possibilities 1 for tweets only and 2 for tweets that include user account data from which the message was sent.

V. Approach

Since it's a growing, noisy dataset, we need to use different algorithms for Machine Learning. Five different features were used here:

User-Level Features- Collecting all features from the twitter API instance number of followers. The rate of tweets was generated to create additional metrics to describe the user's instance of actions.

Tweets-Level Features- Information available directly from the Twitter API (ex: no re-tweets) and meta-data information can be added (ex: weekday and time). Also numerical counting on contents such as word count and question and exclamation mark ratio can be achieved. We may automatically identify new tweets for real-time scenarios.

Text Features: The Bag of Words (BOW) model is trained on the corpus using TF-IDF vectors and a Doc2vec neural model. DM and DBOW were used in later genesis and train models with dimensions of 100,200 and 300.

Subject Features: Modelling topics can be used to build tweet features. Both the latent model of Dirichlet Allocation (LDA) and the entire dataset model of Hierarchical Dirichlet Process (HDP) are educated. It requires a number of subjects to be picked.

The Features of Sentiment Analysis: Senti Word Net is used to measure tweet polarity in terms of positive, negative and neutral phrases. Text Blob library is used to measure tweet subjectivity.

Feature Scaling and Selection: The results and strategies of all feature set are very large. We therefore expect performance gains from the choice of features.

Training algorithm and parameter optimization: Naive Bayes, Decision Trees, Support Vector Machine and Neural Networks can be used as the training algorithm as a basic classification algorithm.

VI. Evaluation

In the first phase, we perform cross-validation of our noisy training set. Secondly, we train and validate training models against a manually generated gold standard. The explanation for this is to model two use cases in which a tweet from an established user account is evaluated and a tweet from a new user is evaluated.

Although the data set training has been categorized by source and not by tweet, the first setting tests how well the model can identify fake news tweets.

Training Dataset Cross-validation: It is necessary to perform the cross-validation first on the training dataset. The effect on the incorrectly labelled training data set is expected to overestimate the actual performance. As noted, XGBoost achieves the best results, culminating in F1 score on the 0.78 and 0.94 fake news category respectively.

Gold standard validation: This system validates the solution with a manually annotated gold standard. The gold standard is obtained separately on the training data set and is not used to prepare, pick potential parameters or optimize them. In the cross validation settings, future selection, parameter settings can be used to set up the work. Neural network learner performs the best possible scenario for cross-validation.

VII. Conclusion

The realistic approach to detecting fake news with machine learning algorithms is shown in this paper. It is a challenge to collect the data set of appropriate sizes. The use of a large dataset gives us good results.

By using machine learning algorithms, fake news can be found and prevented. As the research goes on, the accuracy of the fake news is also enhanced to be identified and avoided by the approaches described above. This will motivate us to stop spreading false news worldwide.

References

- [1] J. Ratkiewicz, M. Conover, M. R. Meiss, B. Gonçalves, A. Flammini and F. Menczer, "Detecting and tracking political abuse in social media". 2011
- [2] Z.H. Zhou, "A brief introduction to weakly supervised learning", National Science Review, 2017.
- [3] A. Deshwal and S. K. Sharma, "Twitter sentiment analysis using various classification algorithms", in 2016 5th International Conference on Reliability, Infocom Technologies and Optimization.
- [4] L. C. Molina, L. Belanche, and A. Nebot, "Feature selection algorithms: A survey and experimental evaluation", in Data Mining, 2002.
- [5] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on twitter", in Proceedings of the 20th International Conference on world wide web, ser. WWW'11.
- [6] Stefan Helmstetter, Heiko Paulheim, "Weakly supervised learning for fake news detection on twitter", in preceding of IEEE, International conference, 2018.