Synopsis

Fake News Detection

Abstract

In recent years, due to the booming development of online social networks, fake news for various commercial and political purposes has been appearing in large numbers and widespread in the online world. With deceptive words, online social network users can get infected by these online fake news easily, which has brought about tremendous effects on the offline society already. An important goal in improving the trustworthiness of information in online social networks is to identify the fake news timely. This paper aims at investigating the principles, methodologies and algorithms for detecting fake news articles, creators and subjects from online social networks and evaluating the corresponding performance. This project addresses the challenges introduced by the unknown characteristics of fake news and diverse connections among news articles, creators and subjects. This project introduces a novel automatic fake news credibility inference model, namely FAKE DETECTOR. This project reviews various fake news detection methods involving feature extraction methods like Count vectorizer, TF-IDF Vectorizer, Word Embedding and also, different classification algorithms like SVM, Logistic Regression and Gradient Boosting, Random Forest, Decision Trees, KNN and XG-Boost.

Introduction

Fake news denotes a type of yellow press which inten- tionally presents misinformation or hoaxes spreading through both traditional print news media and recent online social media. Fake news has been existing for a long time, since the “Great moon hoax” published in 1835 . In recent years, due to the booming developments of online social networks, fake news for various commercial and political purposes has been appearing

in large numbers and widespread in the online world. With deceptive words, online social network users can get infected by these online fake news easily, which has brought about tremendous effects on the offline society already. During the 2016 US presidential election, various kinds of fake news about the candidates were widely spread in the online social networks, which may have a significant effect on the election results. According to a post-election statistical report , online social networks account for more than 41.8% of the fake news data traffic in the election, which is much greater than the data traffic shares of both traditional TV/radio/print medium and online search engines respectively. An important goal in improving the trustworthiness of information in online social networks is to identify the fake news timely, which will be the main tasks studied in this paper.

Fake news has significant differences compared with tra- ditional suspicious information, like spams , in various aspects: (1) *impact on society*: spams usually exist in personal emails or specific review websites and merely have a local impact on a small number of audiences, while the impact fake news in online social networks can be tremendous due to the massive user numbers globally, which is further boosted by the extensive information sharing and propagation among these users (2) *audiences' initiative*: instead of receiving spam emails passively, users in online social networks may seek for, receive and share news infor- mation actively with no sense about its correctness; and (3) *identification difficulty*: via comparisons with abundant regular messages (in emails or review websites), spams are usually easier to be distinguished; meanwhile, identifying fake news with erroneous information is incredibly challenging, since it requires both tedious evidence-collecting and careful fact- checking due to the lack of other comparative news articles available.

These characteristics aforementioned of fake news pose new challenges on the detection task. Besides detecting fake news articles, identifying the fake news creators and subjects will actually be more important, which will help completely eradicate a large number of fake news from the origins in online social networks. Generally, for the news creators, besides the articles written by them, we are also able to retrieve his/her profile information from either the social network website or external

knowledge libraries, e.g., Wikipedia or government-internal database, which will provide fundamental complementary information for his/her background check. Meanwhile, for the news subjects, we can also obtain its textual descriptions or other related information, which can be used as the foundations for news subject credibility inference. From a higher-level perspective, the tasks of fake news article, creator and subject detection are highly correlated, since the articles written from a trustworthy person should have a higher credibility, while the person who frequently posting unauthentic information will have a lower credibility on the other hand. Similar correlations can also be observed between news articles and news subjects. In the following part of this paper, without clear specifications, we will use the general *fake news* term to denote the *fake news articles*, *creators* and *subjects* by default.

Problem Studied: In this project, we propose to study the fake news detection (including the articles, creators and subjects) problem in online social networks. Based on various types of heterogeneous information sources, including both textual contents/profile/descriptions and the authorship and article- subject relationships among them, we aim at identifying fake news from the online social networks simultaneously. We formulate the fake news detection problem as a credibility inference problem, where the real ones will have a higher credibility while unauthentic ones will have a lower one instead.

The fake news detection problem is not easy to address due to the following reasons:

- *ProblemFormulation*:The Fake News Detection Problem studied in this paper is a new research problem, and a formal definition and formulation of the problem is required and necessary before studying the problem.
- *Textual Information Usage*: For the news articles, creators and subjects, a set of their textual information about their contents, profiles and descriptions can be collected from the online social media. To capture signals revealing their credibility, an effective feature extraction and learning model will be needed.
- *Heterogeneous Information Fusion*: In addition, as men- tioned before, the credibility labels of news articles, creators and subjects

have very strong correlations, which can be indicated by the authorship and article-subject relationships between them. An effective incorporation of such correlations in the framework learning will be helpful for more precise credibility inference results of fake news.

Literature Review

Mykhailo Granik et. al. in their paper shows a simple approach for fake news detection using naive Bayes classifier. This approach was implemented as a software system and tested against a data set of Facebook news posts. They were collected from three large Facebook pages each from the right and from the left, as well s three large mainstream political news pages (Politico, CNN, ABC News). They achieved classification accuracy of approximately 74%. Classification accuracy for fake news is slightly worse. This may be caused by the skewness of the dataset: only 4.9% of it is fake news.

Himank Gupta et. al. gave a framework based on different machine learning approach that deals with various problems including accuracy shortage, time lag (BotMaker) and high processing time to handle thousands of tweets in 1 sec. Firstly, they have collected 400,000 tweets from HSpam14 dataset. Then they further characterize the 150,000 spam tweets and 250,000 non- spam tweets. They also derived some lightweight features along with the Top-30 words that are providing highest information gain from Bag-of- Words model. 4. They were able to achieve an accuracy of 91.65% and surpassed the existing solution by approximately18%.

Marco L. Della Vedova et. al. first proposed a novel ML fake news detection method which, by combining news content and social context features, outperforms existing methods in the literature, increasing its accuracy up to 78.8%. Second, they implemented their method within a Facebook Messenger Chabot and validate it with a real-world application, obtaining a fake news detection accuracy of 81.7%. Their goal was to classify a news item as reliable or fake; they first described the datasets they used for their test, then presented the content-based approach they implemented and the method they proposed to combine it

with a social-based approach available in the literature. The resulting dataset is composed of 15,500 posts, coming from 32 pages (14 conspiracy pages, 18 scientific pages), with more than 2, 300, 00 likes by 900,000+ users. 8,923 (57.6%)

posts are hoaxes and 6,577 (42.4%) are non-hoaxes.

Cody Buntain et. al. develops a method for automating fake news detection on Twitter by learning to predict accuracy assessments in two credibility- focused Twitter datasets: CREDBANK, a crowd sourced dataset of accuracy assessments for events in Twitter, and PHEME, a dataset of potential rumours in Twitter and journalistic assessments of their accuracies. They apply this method to Twitter content sourced from BuzzFeeds fake news dataset. A feature analysis identifies features that are most predictive for crowd sourced and journalistic accuracy assessments, results of which are consistent with prior work. They rely on identifying highly retweeted threads of conversation and use the features of these threads to classify stories, limiting this works applicability only to the set of popular tweets. Since the majority of tweets are rarely retweeted, this method therefore is only usable on a minority of Twitter conversation threads.

In his paper, **Shivam B. Parikh** et. al. aims to present an insight of characterization of news story in the modern diaspora combined with the differential content types of news story and its impact on readers. Subsequently, we dive into existing fake news detection approaches that are heavily based on text- based analysis, and also describe popular fake news datasets. We conclude the paper by identifying 4 key open research challenges that can guide future research. It is a theoretical Approach which gives Illustrations of fake news detection by analysing the psychological factors.

Proposed Work

The fake news detection model is built using the following steps:

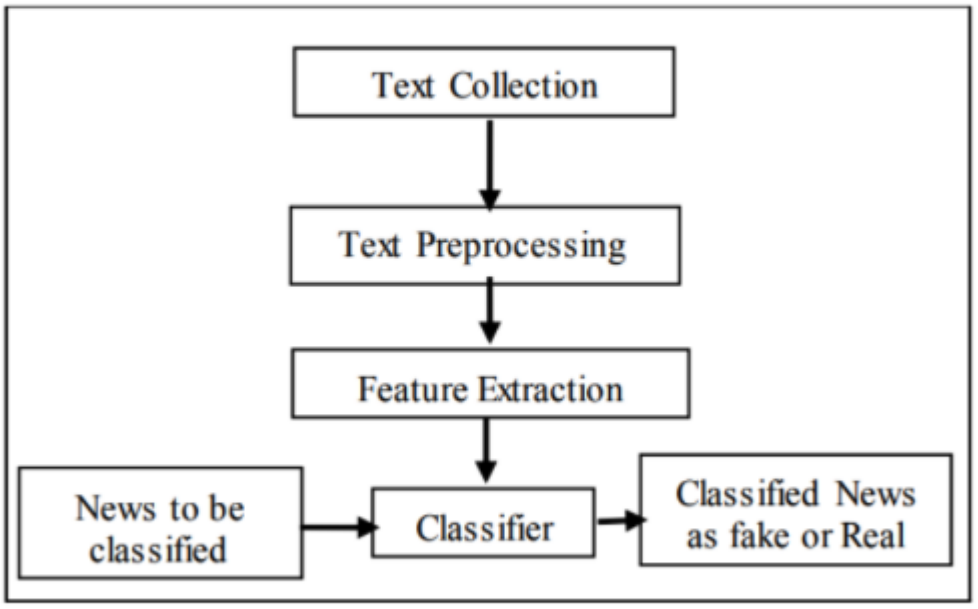

Fig 1: Model architecture

Methodology

TABLE III: ALGORITHM FOR MODEL BUILDING

Algorithm	
Input:	News Content
1.	Convert text to lowercase
2.	Remove punctuations, digits, stop words from text
3.	Repeat: Input: Receive each news article Calculate count vector for it Append the count vector to count_feature vector Until the end of news article
4.	Repeat Input: Receive each news article Calculate the TF-IDF vector for it Append the count vector to tfidf_feature vector Until the end of news article
5.	Repeat Input: Receive each news article Calculate the Spacy vector for it Append the spacy vector to Spacy_feature vector Until the end of news article
6.	Parse count_feature vector, tfidf_feature vector and spacy_feature vector into classifier Return feature vector gives us highest accuracy
7.	Build model with the feature vector
Output:	Predict label of news - Fake or Real

TABLE V: CLASSIFICATION METRICS WITH TF-IDF VECTORIZER

Classifiers	Performance evaluation			
	Accuracy	*Precision*	*Recall*	*Fscore*
SVM Linear	0.94	0.93	0.93	0.93
Logistic Regression	0.93	0.93	0.91	0.92
Decision Tree	0.82	0.79	0.80	0.80
Random Forest	0.90	0.94	0.83	0.88
XG-Boost	0.89	0.89	0.88	0.88
Gradient Boosting	0.90	0.89	0.88	0.88
Neural Network	0.93	0.93	0.91	0.92

CONCLUSION

In recent years, deceptive content has grown in popularity, and its influence on online users has increased. The authors of this paper have surveyed on fake news detection methods involving three different feature extraction methods like Count vectorizer, TF-IDF Vectorizer, Word Embedding and also, different classification algorithms. The greatest accuracy attained by classification techniques is by using SVM Linear classification algorithm with TF-IDF feature extraction with 0.94 accuracy, as shown in TABLE IV, V, VI. Even though both Neural Network and Countvectorizer achieve the same accuracy, the Neural Network takes longer to train and is more sophisticated. So, in the proposed system, Linear SVM is used, which is less difficult and takes less time to compute.

FUTURE WORK

Deep learning methods and sentiment analysis to categorise news with high accuracy might be considered in the future, and more useful text such as the publication of the news, URL domain, and so on, could be extracted for the process.

For more accuracy, a dataset with a higher number of articles from various sources can be employed, as it includes more jargon and notable material.