Fake News Detection

Si Hong Long and Mohd Pouzi Bin Hamzah

Abstract Everyday people receive a lot of information through social media and online news portals. To distinguish whether the information is fake or true is a big problem. An algorithm has been developed to distinguish fake news and true news by searching the relevant news from reliable news website based on the news given. This results in the similarity percentage between news and the relevant news. The algorithm has been tested with the dataset collected by Dr. Victoria L. Rubin that consists of 180 true news and 180 fake news from several American and Canadian news websites. The precision of 69.44% has been achieved with the dataset.

Keywords Fake news · Cosine similarity · Natural language processing

1 Introduction

The meaning of fake news is false stories that appear to be news, spread on the internet or using other media, usually created to influence political views or as a joke. Nowadays the amount of fake news keeps increasing especially news regarding Covid-19, but people cannot distinguish between the true and fake news. In large part, a deeper concern that the prevalence of “fake news” has increased political polarization, decreased trust in public institutions, and undermined democracy [1]. An example of fake news such as in 2016, a Facebook post about nationwide order banning The Pledge of Allegiance in schools in the United States that had been signed by President Obama was shared and commented upon a total of 2.2 million times on Facebook [2]. There are two examples of fake news that happened in Malaysia that have been reported in New Straits Times. First, the news regarding PT3 papers for...
subject English, Mathematics, Science and Integrated Living Skills reported leaked [3]. Second, the news regarding October beer festival in Terengganu [4].

The type of fake news can be divided into four types which are actual “fake news”, satire news, poorly reported news and misleading news. The actual “fake news” are stories that are completely made up and do not happen in the world. Satire news are fake articles that are meant for humour and without prove. Poorly reported news is news that are reported badly but not completely made up. The misleading news are news that try to change the perspective of readers toward a topic [5].

People are receiving information every day, but they do not have the ability to recognize the fake news. An exclusive Ipsos poll conducted for BuzzFeed News found that 75% of American adults who always use Facebook as the source of news are likely to believe fake news headline than those who do not use Facebook as source for news [6]. According to estimate made by police and by local community leaders, 86 to 238 Berom ethnic minority were killed in Gashish between 22 and 24 June 2018, just because a fake Facebook news posted by a man in United Kingdom that Fulani Muslims were killing Christians [7].

In order to overcome the problem stated above, there is a need to develop an algorithm that can help people to distinguish fake news and true news by comparing the news from user with several reliable news website and give the related news from reliable news website as references to the user.

2 Related Work

Sirajudeen et al. [8] proposed three-phase method to detect online fake news that use java programming language. The first phase is checking IP validity. The second phase is checking the content of the information of online news such as article, title, author and background information of the article with a database that contain the verification information. The third phase is to determine the status of fake news based on result from two previous phases. Another approach proposed by Gahirwal et al. [9] is by comparing the headlines and compare news article with top search. Feyza and Bilal [10] proposed a two-step method for identifying fake news in social media and the method was tested on three real data sets in terms of different evaluation metrics.

Granik and Mesyura [1] proposed an algorithm making use of naïve Bayes classifier. This approach achieved accuracy approximately 74% on test set. They found that spam messages and fake news article have common properties like a lot of grammatical mistakes, emotionally coloured, often use same set of word and affect reader’s opinion on some topic in manipulative way. The main idea is to treat each word of the news article independently. Wei and Wan [11] introduced a method that uses class sequential rules (CSR) and basic features (body-independent features) extracted from headline to train support vector machine (SVM) classifier. They also add body dependent features such as Informality, Sentiment, InformalGap, sentiGap,
Similarity, Recognizing Textual Entailment (RTE) to train SVM. The SVM toolkit is from the scikit-learn.

Ahmed [12] introduced feature extraction using term frequency (TF) and term frequency-inverted document frequency (TF-IDF). Other features are keystroke such as editing patterns and timespan, n-grams features and semantic similarity. The experiments involve six different machine learning algorithms which are Stochastic Gradient Descent (SGD), K-Nearest Neighbour (KNN), Logistic Regression (LR), Decision Tree (DT), Support Vector Machines (SVM) and Linear Support Vector Machines (LSVM). The experiment also studies the impact of n-grams size on performance. Total of four experiment were carried out. From the first experiment, Linear-based classifiers such as Linear SVM, Logistic regression and SDG yield better result than nonlinear methods. The accuracy increases as number of feature values increase. As the size of n-grams increases, the accuracy will decrease. The performance for TF-IDF is better than TF. KNN achieve lowest accuracy which is 47.2% with 4-g word and 50,000 feature values. From the second experiment, they found that Linear-based classifier is still better than nonlinear as Linear SVM achieved accuracy as 92%. The performance of Linear SVM is not affected by number of feature values, but as size of n-gram increase the accuracy decrease. From the third experiment, they found that keystrokes feature with n-gram yielded better accuracy. From the fourth experiment, they found that as the percentage of change increase the semantic measurement decrease.

Most of the researchers use artificial intelligence in online fake news detection. In our work, we propose a new approach by comparing news article from the user and the articles from reliable news sources.

3 Methodology

Figure 1 shows the algorithm of fake news detection system. The headline, article or URL are provided by the user as an input to the system. Then the article and the headline will be extracted from the webpage. Next, all related article will be retrieved from the reliable news website based on the headline. Then data pre-processing such as lemmatization, stop word removal will be carried out. After the pre-processing, the algorithm proceeds with the calculation of the Term Frequency-Inverted Document Frequency (TF-IDF) and cosine similarity between news. Finally, percentage of similarity will be displayed.

3.1 Lemmatization

Lemmatization is to improve the performance of natural language processing by generating the word into its root word. For example, playing, plays and played after
1) Get article, headline or URL from user.
2) If user use headline and article to check news, then use headline as keywords to find similar news.
3) If user use article only to check news, if the length of article longer than 20 word, use first 20 word as keywords to find similar news.
4) If user use URL to check news, extract headline and article from given URL, use headline as keywords to find similar news.
5) Remove all the punctuation from keywords and declare article as String.
6) Declare three list all_news_article, reference_list and all_news_url.
7) Pre-process article and add to all_news-article list.
8) Replace all the space in keywords with symbol “+”.
9) Find all the URL link that relate with keyword from reliable news website with web scraping and add into all_news_URL list.
10) Extract all article and headline from all the URL in all_news_URL list, then add pre-process article into all_news_article, and add article, headline and URL into reference_list.
11) Calculate Term Frequency-Inverted Document Frequency.
12) Calculate cosine similarity.
13) Convert cosine similarity into percentage with 2 decimal places.
14) Add similarity to reference_list and sort list in descending order base on similarity.
15) Find the similarity of first news in reference list if more than or equal to 70, then delete the rest of news in reference_list.
16) If first news in reference_list not equal to 0, check the news with 0 percent similarity and delete it.
17) Calculate average similarity from reference_list.
18) If average similarity greater or equal to 70, then status is “The news is true”.
19) If average similarity smaller then 70, then status is “The news is not reliable”.
20) Display the average similarity, status and news in reference_list to user.

Fig. 1 Fake news detection system algorithm

lemmatization will become play. All the lemmas are the English words that can be found in dictionaries.

3.2 Stop Words and Punctuation Removal

Stop words are the words that do not provide important information to document and common to most documents. Stop words will decrease the performance in natural language processing. The example of stop words are ‘as’, ‘the’, ‘be’, ‘are’ and etc.
Like the stop words, the punctuation is not important in natural language processing and it will affect the performance. The example of punctuations are (?, !, ;, :; , ‘, “).

3.3 Term Frequency-Inverted Document Frequency (TF-IDF)

In information retrieval, TF-IDF is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus (Fig. 2).

In mathematical representation, TF-IDF = TF * IDF which is term frequency multiply with inverted document frequency. The formula for Term frequency is the number of a word appear in a document divide the total number of words in document. The formula for Inverse document frequency is the total number of documents in corpus by document frequency for each term and apply logarithmic scaling on the result.

\[
TF - IDF = \frac{A}{B} \times \left(1 + \log \frac{C}{1 + D}\right)
\]  

where A represents the number of a word appear in a document, B represents the total number of words in a document, C represents total number of documents in corpus and D represents number of documents a word appears.

3.4 Cosine Similarity

Cosine similarity is used to measure the cosine angle between two terms as they represented in their vectorized forms and non-zero positive vectors in an inner product space. The term vectors are close to each other and in the same direction, thus the score is closer to 1 (cos 0°), mean they are similar. If the term vectors score close to 0 (cos 90°), mean they are not similar. Term vectors score close to −1 (cos 180°), mean they are unrelated, and they are in opposite direction to each other.

Cosine similarity are dot product of the two term vectors \( u \) and \( v \), divided by the product of their L2 norms. The mathematically representation of dot product between two vectors as shown in Fig. 3.
Fig. 3 The formula for cosine similarity. Source [13]

\[
\cos(u, v) = \frac{u \cdot v}{\|u\| \|v\|} = \frac{\sum_{i=1}^{n} u_i v_i}{\sqrt{\sum_{i=1}^{n} u_i^2} \sqrt{\sum_{i=1}^{n} v_i^2}}
\]

\[u \cdot v = |u||v| \cos(0)\]  \hspace{1cm} (2)

The cosine similarity can be derived from the above formula where \(u_i\) represents the various features of term vector \(u\), \(v_i\) represents the various features of term vector \(v\) and \(n\) represents the total number of features.

3.5 Dataset

The dataset used in this research was collected by the Language and Information Technology Research Lab directed by Dr. Vitoria Rubin, Western University, London, Ontario, Canada [14]. The dataset consists of 360 news from several United State and Canada news websites. The dataset is divided into two sets.

Set 1 consists of 240 articles in which 120 out of 240 articles are legitimate, and the remaining articles are satirical. The legitimate news set is collected from The New York Times of United State and The Toronto Star of Canada. The satirical news set is collected from The Onion of United State and The Beaverton of Canada.

Set 2 consists 120 articles, half of it is legitimate and the others are satirical. The legitimate news set is collected from USA Today, The Wall Street Journal, The Los Angeles Times, The New York Post, Newsday, The Denver Post of United State and The Globe and Mail, The Vancouver Sun, The Calgary Herald, The National Post, The Edmonton Journal, The Hamilton Spectator of Canada. The satirical news set is collected from The Daily Currant, The Spoof, The National Report, The People’s Cube, World News Daily Report, Urban Anomie of United State and CBC’s Punchline, The Codfish, The Lapine, The Syrup Trap, Sage News of Canada.

3.6 Evaluation

To evaluate the performance, precision has been used as a measurement. The calculation of precision is using \(#(true\_positive)\) divided by \(#(true\_positive, false\_positive)\). \(#(true\_positive)\) is the number of news correctly classified by the approach. \(#(true\_positive, false\_positive)\) is the total number of news in the dataset. The formula is as shown below
Precision = \frac{\#(true \ positive)}{\#(true \ positive, \ false \ positive)} \tag{3}

4 Results and Discussion

Dataset consists of 360 news; 180 news are legitimate (true news) and 180 news are satirical (fake news). After testing with the dataset, the algorithm is able to classify correctly all the fake news but can only classify correctly 70 out of 180 for true news. An average precision of 69.44% is achieved with the dataset.

5 Prototype

The interfaces of prototype are as shown in Figs. 4 and 5. Figure 4 is the input interface for the prototype. There are two section and one button in this interface which are checking news by using and related information about the news. The user can choose either by using URL only, article only or both headline and article for checking the news.

The result of news checking is as shown in Fig. 5. The result interface will show six information. The similarity between the news provided by user with the news from pre-defined reliable news sources. The comment either the news is fake or true. The headline, article, URL and similarity of news from reliable news sources

![Fig. 4 The input interface](image_url)
related to user news if available. If no related news from reliable news sources, then the prototype will show an alert box with content ‘Sorry that we cannot find news related to your topic. Your news has high probability of FAKE NEWS’.

6 Conclusion

The fake news tries to change the reader perspective toward a specific topic. We had proposed an algorithm to detect fake news by comparing headline and article with several reliable news website. Our algorithm can achieve a precision of 69.44%. There are three suggestions to improve the algorithm. First, is to increase number of reliable news website, so that comparison can be made with higher accuracy. Second, the algorithm can detect the place and from the place name fake news detection can be localized. This can limit the news sources and speed up the time needed for checking. For example, if Kuala Lumpur is stated in the news article, the algorithm will identify the country as Malaysia and will then retrieve the reliable news from websites in Malaysia such as The Star Online to classify the news provided by the user.
References

1. Granik M, Mesyura V (2017) Fake news detection using naive Bayes classifier. In: Proceedings of IEEE 1st Ukraine conference on electrical and computer engineering, UKRCON 2017, pp 900–903
2. Roberts H (2016) This is what fake news actually looks like—we ranked 11 election stories that went viral on Facebook. Retrieved from https://www.businessinsider.com/fake-president-ial-election-news-viral-facebook-trump-clinton201611/?IR=T
3. Abas A (2018) PT3 papers not leaked, rumours untrue. New Straits Times, 9 Oct 2018. Retrieved from https://www.nst.com.my/news/nation/2018/10/419352/pt3-papers-not-leaked-rumours-untrue
4. Nizam F (2018) No Oktoberfest in Terengganu. New Straits Times, 12 Oct 2018. Retrieved from https://www.nst.com.my/news/nation/2018/10/420604/no-oktoberfest-terengganu
5. Ashe S (2017) The 4 types of ‘Fake News’. Observer, 1 Apr 2017. Retrieved from https://observer.com/2017/01/fake-news-russia-hacking-clinton-loss/
6. Silverman C, Singer-Vine J (2016) Most Americans who see fake news believe it, new survey says. Retrieved from https://www.buzzfeednews.com/article/craigsilverman/fake-news-survey
7. Adegoke Y, BBC Africa Eye (2018). Nigerian say “fake news” on Facebook is killing people. Retrieved from https://www.bbc.co.uk/news/resources/idtsh/nigeria_fake_news
8. Sirajudeen SM, Azmi NFA, Abubakar AI (2017) Online fake news detection algorithm. J Theor Appl Inf Technol 95(17):4114–4122
9. Gahirwal M, Moghe S, Kulkarni T, Khakhar D, Bhatia J (2018) Fake news detection. Int J Adv Res Ideas Innov Technol 4(1):817–819
10. Feyza AO, Bilal A (2020) Fake news detection within online social media using supervised artificial intelligence algorithms. Phys A Stat Mech Appl 540. https://doi.org/10.1016/j.physa.2019.123174
11. Wei W, Wan X (2017) Learning to identify ambiguous and misleading news headlines. In: IJCAI International joint conference on artificial intelligence, pp 4172–4178. https://doi.org/10.24963/ijcai.2017/583
12. Ahmed H (2017) Detecting opinion spam and fake news using N-gram analysis and semantic similarity by. MSc thesis, University of Victoria
13. Sarkar D (2016) Text analytics with Python: a practical real-world approach to gaining actionable insight from your data. New York
14. Rubin VR, Conroy NJ, Chen Y, Cornwell S (2016) Fake news or truth? Using satirical cues to detect potentially misleading news. In: Proceedings of the workshop on computational approaches to deception detection at the 15th annual conference of the North American chapter of the association for computational linguistics: human language technologies (NAACL-CADD2016), San Diego, California, 17 June 2016