International Journal of Advanced Trends in Computer Science and Engineering

Content-Social Based Features for Fake News Detection Model from Twitter

Adnan Hussein [1], Farzana Kabir Ahmad [2] Siti Sakira Kamaruddin[3]

[1]Universiti Utara Malaysia, Malaysia, taibh_310@hotmail.com
[2]Universiti Utara Malaysia, Malaysia, farzana58@uum.edu.my
[2]Universiti Utara Malaysia, Malaysia, sakira@uum.edu.my

ABSTRACT

The advancement of social networks has facilitated the sharing and spread of news among people over the world. With the growth of these networks and growth of the volume of the news shared daily, the phenomena of fake news become more stronger and widely spread. In this paper, content-social based features for fake news detection model from Twitter data has been proposed. This work aims to analyze content-based features of news content including of linguistic features, writing style features, semantic features and sentiment features along with social-context based features of news diffusion over the social network including user-based features and network-based features to detect fake news from Twitter news posts. With using of unsupervised graph-based clustering approach, no labelled data is required and this make the proposed model more practical to detect online fake news.

Key words : Content features, Fake news detection, Social features, Unsupervised learning.

1. INTRODUCTION

With the quick revolution and advancement in online social networks, the majority of people tends to use it as one of the main source for consuming information [1]. Twitter is one of the biggest social networks that has been used by a wide range of people to share and spread their important news and events [2]. Every day, there is a huge amount of information and news generated online on the social networks. Many people post, share and discuss the daily news and events online through the social networks as a normal routine of daily life. As a result, big volumes of data is increasing on the web by time and this made the online users face several problems such as the searching and retrieving of specific accurate information [1]. On the other side, these big volumes make suitable environment to spread and diffuse of the misleading false information and fake news. [2].

Using of social networks as a source to consume news and other information is a double-edged sword. On one side, people can reach news on social networks easily, more cheaply than the traditional news media, and they can diffuse their information faster [3], [4]. On the other edge, the online social networks provide a perfect environment for generating and disseminate of fake news. Most of the people consumed news from these networks regardless of their credibility and truthfulness [5]. This in turn caused many troubles for individuals, organizations and even governments. Fake news can lead to big negative changes among society, like what has happened in the Arabic world during 'Arabic Spring' where the fake news was widespread in social networks and was used as a weapon of Psychological war against individuals and governments [6]. In this article, a conceptual model is proposed for fake news detection from Twitter using several features of news content including of linguistic features, writing style features, semantic features and sentiment features as well as social context-based features including user-based features and network-based features. Moreover, this study propose to use unsupervised approach using graph-based clustering to detect fake and misleading information. The rest of the paper discuss on the general view of the model with an explanation on the feature extraction stage of different news features. Then gives details on the detection analysis of fake news. Description of unsupervised approach are presented. At the last, study ends with a conclusion.

2. THE CONCEPTUAL MODEL

Figure 1 shows the general conceptual model of fake news detection from twitter. Data collection is the first stage where the twitter posts (tweets) are collected and saved in one dataset. This dataset pass through many stages of processing and analysis to detect the fake news that may be conveyed inside tweets.

Pre-processing of the dataset is an essential stage. The collected tweets usually have noisy data such as URLs, special characters, stop words and other unrelated text like advertisements [7]. In this stage, the tweets go through some text pre-processing mechanisms to prepare text for next stage of analysis. This includes text tokenization where each tweet

is broke down into its individual words. Normalization is another mechanism of text pre-processing where the long words that may have some redundant letters normalized to their original word form. For example, some users tend to write (gooooood / Thanksssss) to express of their deep and strong emotion and sentiment regarding to something. The normalization here will remove the redundant characters and transform the word to the original form (good /Thanks). In addition, stemming task can be applied to the tokenized words to remove the suffixes and prefixes and it is an important task to reduce the dimension of words volume. Moreover, stop words removal mechanism also can be applied on the cleaned tweets to remove the stop words such as "the", "a", "an", "in" [8].

After the data is cleaned and prepared, it passed through to the next stage, which is the feature engineering stage. Features engineering has two main components: feature extraction and feature selection. In feature extraction, several features of news are extracted. There are two main categories
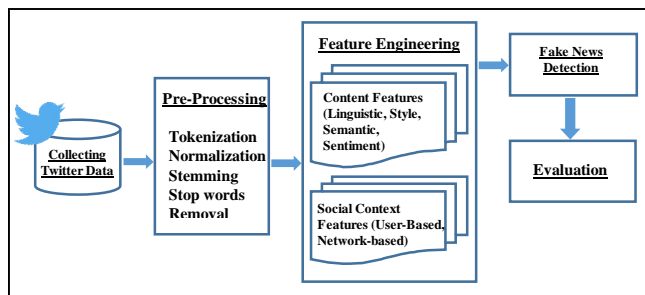
Figure 1: Conceptual Fake News Detection Model

of news features, which will be discovered in the next subsections of this paper. Feature selection is a process in which the irrelevant and noisy features are removed from the extracted news features [9]. This process is applied by choosing a subset of features based on specific criteria [9]. Feature selection is an essential and important task to reduce the features space and help to perform detection process with more high accuracy [9]. Fake news detection process primarily depend on the analysis of news features. As mentioned above, there are two main categories of these features: features related to the news content called content-based and features related to the diffusion context called social-based features.

3. CONTENT-BASED FEATURES ANALYSIS

News content has several important features that can be used to detect fake news. These features including explicit and latent features of the news content. Basically, content-based features includes 1) Linguistic features. 2) Writing style features. 3) Semantic features. 4) Sentiment features. 5) Visual-Based features.
Linguistic features indicating the basic language element analysis and sentence structure of the news content language.

The linguistic features of the text are good indicators and cues for detection of suspicious news, which are purposely written to mislead the readers [1].
Writing style features analysis discovers the different authoring and writing styles for the author of fake news and reveal its special writing attributes [1]. In spite of the fact, fake news creators aim to imitate the same way of writing of true news authors to mislead the readers and encourage them to believe their claims but, there are several distinctions are still exist. These distinctions can help to recognize and differentiate the fake news authors and true news authors [3] such differences including keystrokes usage and several syntactical structure features which provides clarity on structure syntax. However, writing style features analysis does not include the semantic of sentences, which is crucial in detecting fake news.
Semantic features are very important features that can help to detect and expose the fake news. There is a lack in utilizing semantic features in fake news detection, only few studies have utilized these features in their works [10]. The results of those works have showed that using semantic features to detect fake news is very important and effective and need more investigation [11].
Sentiment features analysis is an important approach to detect fake news and expose the suspicious and fake user accounts [12]. The fake news authors can overdraw the facts and mislead the online readers. Sentiment analysis of the news contents and Psychological keywords can reveal the emotions and judgment that are carried within the news content and it is a helpful method for false information detection [13] , [14]. There are several methods have been proposed to utilize the sentiment of news. Such methods: arousal valence dominance score, happiness detection and analysis, emotion analysis and analysis of polarization and strength [1], [15]. Sentiment analysis has been proved as a sufficient method to detect fake news by combining several sentiment scores and can detect and distinguish between the fake bot accounts and real human accounts [12].
The visual-based features analysis indicates to feature analysis of visual content included in news content such as the number of image and videos, multi-image frequency, popular image frequency and so on. Images and videos features are significant cues for detecting misleading information and several fabricated news [3].

4. SOCIAL CONTEXT-BASED FEATURES ANALYSIS

Beside content-based features of news there are other important features called social-based features which are related to the analysis of news distribution and diffusion patterns through the different social networks, analyze how fast and widely the fake news is spread and how the different users interact with these news and how they share it between them. Social context-based features are critical features that

cannot be ignored. However, most of the recent fake news detection approaches are depend widely on the analysis of news content-based features only. There are only few researches have used the social context-based features to analyze and detect of the fake news. [1], [3]. Using social context-based features analysis consider as a way to complement content-based features analysis that rely on analyzing of news language and leakage cues to predict fake news [3]. There are two basic categories of social context-based features: user-based features and network-based features.

4.1 User-based features analysis

The analysis of these features aims to analyze the online users features, which they create, share and spread the fake news. By utilizing the unique characteristics of the fake news authors, the fake or bot accounts can be exposed and differentiated from the true accounts [1]. The user-based features analysis approach can be divided into 1) user profiling features analysis, 2) posting behavior and temporal features analysis and 3) credibility features analysis.

User profiling features represents the fundamental user data involves the language that is used in his account, the geographic location, the profile creation time, verification of the account, how many posts have been authored, etc.[14]. The analysis of profile details of online social user define and reveal how the user is active or suspicious on the social network. This analysis is helpful for the identification of fake and unreal accounts in social networks [14].

News creation and posting behaviour indicates to the behaviour of the online social account posting patterns at certain time period [1]. These patterns and features includes the time pattern between two successive posts, sharing, mentioning, and so on. Usually, the suspicious accounts such as bots and cyborgs follow specific time posting pattern when they post the fake news [1].

The credibility level of the news can be detected and predicted through the interaction of the online users with this news. For example, if the piece of news has been "liked" or "commented" widely by unreliable users and anomalous accounts then this news more probably to be a fake news [5]. it is approved that the trustiness of a piece of news can be specified by the truthfulness of the users who interact with that news [1].

4.2 Network-based features analysis

Recently, the number of active users of online social networks exceeds one billion. Each of them writing, posting and sharing different news and information with their friends and followers. User-based features analysis indicating analysis of the users of the social network and how they interact with each other, whereas in the network-based features analysis the diffusion of data in social network and the spreading paths are studied and analyzed [14]. Study of network diffusion and

spreading can give a value information to mitigate and stop of fake news and misleading information distribution [16].

Two distinct networks can represent the interaction between online users and news posts: 1) Homogeneous. and 2) Heterogeneous. Each of them has it is own characteristics and behavior. In homogeneous, nodes and edge that have same type form the network. There are three main types of such network:

1) Friendship Networks indicates the structure of following and follower users who post related tweets or posts [3].
2) Diffusion Networks which tracks the news spreading path [3], in this network each user is represented as a separated node. These nodes are connected by edges (links) which represent the different paths of the diffusion of the information between users. Assume that there are two users Ux and Uy, the diffusion path is exist between those users if and only if: (1) Uy follows Ux, and (2) Uy post (or re-tweet) the information only after Ux did so. After the analysis and draw of the users and the paths between them, the diffusion network can be constructed well [16].
3) Credibility Networks this kind of homogeneous networks can be represented as undirected graph Gc=(V , Ec , s). Where V indicating the set of user posts with corresponding score s of the post credibility. E indicating the edge (link) type such as: support or agree relation between the two of nodes.

Heterogeneous Networks are different from homogeneous. The set of nodes are different and the edges (links) types are different. The main characteristic of heterogeneous networks is it can represent and encode the data and links of relations from different perspectives. Heterogeneous networks can be divided into three types:

4) Knowledge Networks is a graph-based network Gk=(I , Ei , R). Where I indicating the graph node set, Ei represents the link (relation) between graph nodes, and R represents the relation sets.
5) Stance Networks consist of set of nodes that indication the online posts and set of edges representing the weight of stances. [3].
6) Interaction Networks are heterogeneous network Gi=((P,U,V),Ei). Where P represents the publisher of information, U indicating the user set, V represents the information and Ei represents the interaction between the previous elements. For example, (p → v) represent that the information (V) is published by the publisher (P).Machine Learning Based Fake news Detection.

5. FAKE NEWS DETECTION

Many studies have considered the problem of fake news detection on social media as a classification problem. Basically, for classification approach there is a need of labelled data which is used as training data to train the classifier before the classification task is performed [16].

However, there are two basic approaches of machine learning which are used in detecting fake news: supervised learning and unsupervised learning approaches.

In the previous studies of fake news detection and other hoaxes and frauds detection studies, supervised machine learning algorithms are widely used. As mentioned above, some researchers consider the detection of fake news as a supervised learning classification task. many of supervised machine leaning algorithms were used such as: Support Vector Machine (SVM), K-nearest Neighbour (KNN), Decision Tree (DT), Logistic Regression (LR), Random Forest (RF) and so on. [17].

However, for the evaluation of the performance of these algorithms, many criteria for evaluation are used. The most widely used criteria are confusion matrix (True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN)). In addition, there are some rates can be computed for the classifier evaluation like: Precision, Recall, False Positive Rate, False Negative Rate, F-score and Accuracy [3].

The quality and accuracy of the classification in supervised learning deeply depend on the quality of labelled datasets. However, it is more hard and cost to build a fake news dataset with high quality and big size due to some reasons: first, the online-generated data is usually has missing information, has many noises, unlabeled and unstructured data. It needs a lot of effort and labor to organize and clean [3]. Second, the big produced volume of fake information with variety of features and characteristics is created every day in social networks [14]. Moreover, the labelling process of data with the truthfulness labels is time consuming and hard task. Thus, the future directions refer to the unsupervised learning approach as more practical and appropriate approach to solve these problem [1].

The unsupervised leaning approach is a machine learning approach. In this approach, there is no need for training classifiers to perform classification task. Instead, the data is clustered and grouped according to similarities and differences. There are five types of analysis for this approach which are clustering analysis, unsupervised news embedding, outlier analysis, semantic similarity analysis and information diffusion analysis. For information diffusion analysis, there are two approaches followed be the literature to analysis the diffusion of information. Graph-based approach and non-graph based approach. There are only few studies used unsupervised approach to detect fake news and it needs more investigation in future studies of fake news [1], [3].

6. CONCLUSION

Fake news spreading is a phenomenon that have emerged recently among the social networking users. It is an important to detect fake news and stop its negative effects on individuals and society. In this paper, fake news detection model was introduced to detect fake news by utilizing variety of several content-based features and social context-based features of diffused news in Twitter. Moreover, this model used graph-based unsupervised approach to detect fake news. It is believed that analyzing of different content-based features as well as social context-based features can give high accuracy and the use of the unsupervised approach makes this model more practical to detect fake news on social media.

REFERENCES

1. Zhang, X. and A.A. Ghorbani, **An overview of online fake news: Characterization, detection, and discussion**. *Information Processing & Management*, 2019.
   https://doi.org/10.1016/j.ipm.2019.03.004
2. Aphiwongsophon, S. and P. Chongstitvatana. **Detecting Fake News with Machine Learning Method**. in *2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*. 2018. IEEE.
3. Shu, K., et al., **Fake news detection on social media: A data mining perspective**. *ACM SIGKDD Explorations Newsletter*, 2017. 19(1): p. 22-36.
   https://doi.org/10.1145/3137597.3137600
4. Ab Aziz, A., et al. **Designing a robot-assisted therapy for individuals with anxiety traits and states**. in *2015 International Symposium on Agents, Multi-Agent Systems and Robotics (ISAMSR)*. 2015. IEEE.
5. Castillo, C., M. Mendoza, and B. Poblete. **Information credibility on twitter**. in *Proceedings of the 20th international conference on World wide web*. 2011. ACM.
6. Hammad, M. and E. Hemayed. **Automating credibility assessment of arabic news**. in *International Conference on Social Informatics*. 2013. Springer.
   https://doi.org/10.1007/978-3-319-03260-3_13
7. Yang, S., et al. **A Topic Detection Method Based on KeyGraph and Community Partition**. in *Proceedings of the 2018 International Conference on Computing and Artificial Intelligence*. 2018. ACM.
8. Alotaibi, S.S., **Sentiment analysis in the Arabic language using machine learning**. 2015, Colorado State University. Libraries.
9. Agarwal, S., P. Ranjan, and R. Rajesh, **Dimensionality reduction methods classical and recent trends: a survey**. *IJCTA*, 2016. 9(10): p. 4801-4808.
10. Jia, S., et al. **Fake reviews detection based on LDA**. in *2018 4th International Conference on Information Management (ICIM)*. 2018. IEEE.
    https://doi.org/10.1109/INFOMAN.2018.8392850
11. Xu, K., et al., **Detecting fake news over online social media via domain reputations and content understanding**. *Tsinghua Science and Technology*, 2019. 25(1): p. 20-27.

12. Dickerson, J.P., V. Kagan, and V. Subrahmanian. **Using sentiment to detect bots on twitter: Are humans more opinionated than bots?** in *Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. 2014. IEEE Press. https://doi.org/10.1109/ASONAM.2014.6921650

13. Thakur, P. and D.R. Shrivastava, **A Review on Text Based Emotion Recognition System**. *International Journal of Advanced Trends in Computer Science and Engineering*, 2018. 7(5). https://doi.org/10.30534/ijatcse/2018/01752018

14. Zhao, J., et al., **# fluxflow: Visual analysis of anomalous information spreading on social media**. *IEEE transactions on visualization and computer graphics*, 2014. 20(12): p. 1773-1782. https://doi.org/10.1109/TVCG.2014.2346922

15. Oueslati, O., A.I.S. Khalil, and H. Ounelli, **Sentiment Analysis for Helpful Reviews Prediction**. *International Journal*, 2018. 7(3). https://doi.org/10.30534/ijatcse/2018/02732018

16. Shu, K., H.R. Bernard, and H. Liu, **Studying fake news via network analysis: detection and mitigation**, in *Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining*. 2019, Springer. p. 43-65. https://doi.org/10.1007/978-3-319-94105-9_3

17. Mustaffa, Z., Y. Yusof, and S. Kamaruddin. **Application of LSSVM by ABC in energy commodity price forecasting**. in *2014 IEEE 8th International Power Engineering and Optimization Conference (PEOCO2014)*. 2014. IEEE. https://doi.org/10.1109/PEOCO.2014.6814406