Identify Misinformation on Twitter with Machine Learning

Supanya Aphiwongsophon, Prabhas Chongstitvatana

Department of Computer Engineering, Faculty of Engineering
Chulalongkorn University, Bangkok, Thailand
supanya.a@student.chula.ac.th, prabhas.c@chula.ac.th

ABSTRACT

Misinformation on Twitter became the national agenda which many parties are interested in fixing it. This work proposed two methods of machine learning to deal with this issue. The data set, 948,373 messages, was collected from Twitter in 2017. The data contained both the truth and the misinformation. This data is used to train a model to identify misinformation. Two machine learning methods are tested: Naïve Bayes and Support Vector Machine. The result of the experiment shows that the accuracy of Naïve Bayes is 96.08% and Support Vector Machine is 99.89%.

Keywords: Misinformation, Twitter, Online Social Network, Machine Learning

1. INTRODUCTION

Thailand's growth in the use of the internet via wireless devices was very high [1]. Rapid news reporting and the news forwarding can propagate both truth and misinformation.

Misinformation may be a notice rumours that there is no evidence to confirm its truth and reliability. It may be an incident in which the relevant parties have issued a statement confirming from the source of news content that has been distorted or it is not true, as the claiming in the news story [2]. To investigate the truth the work requires that the topic of the news content be specified in order to be processed into the credibility of the specified news topic [3] [4] [5]. The research in [6] verified the truth and processed the algorithms for the news reliability by considering the source of news or a person with specific expertise or who can give facts without errors.

The truth could spread so far and rapidly but the misinformation may change the content and spread out as well [7]. In addition, [8] analyses factors related to messages that affect negative attitudes and incorrect believes based on information received. The research in [9] offered advice on how to effectively manage the risks that occur as well as improving the method of acknowledging news, understanding, and solving long-term problems of counterfeit news at the event.

2. EXPERIMENTS

The experiment consists of three steps. The first step is the news collection. The topics from Twitter were collect and the text was pre-processed. The raw data are stored in an unstructured form and it is transformed to the structure data. The unstructured data is not suitable for use in the machine learning process [10]. The second step is the process of managing the data format by normalization and elimination of the duplicate data. The last step is the machine learning process. The details of the overall process are shown in the Fig.1.

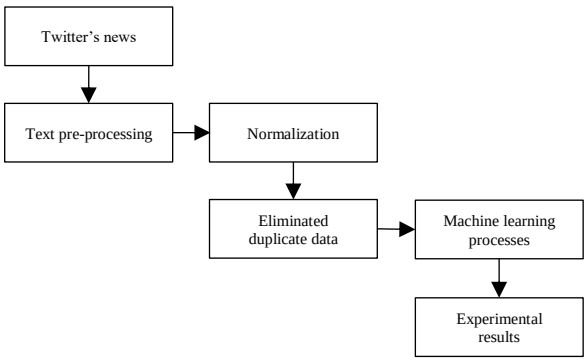

Fig.1: *Overview of the experiment*

The 22 attributes on Twitter API are used. These attributes are similar to [11]. The selected attributes are Id, Name, IsVerified, ProfileImageUrl, FollowersCount, FriendsCount, FavouritesCount, StatusesCount, Description, Location, TimeZone, UserCreatedDate, Status, Url, Mentions, Number of Mentions, HashTags, Number of HashTags, RetweetCount, TweetCreatedDate, MessageText and MessageImage.

The data was collected between October and November 2017 on the selected of news topics. The data set is described in Table 1. After the raw data (948,373 messages) are normalized to all numbers and eliminate the duplication of data, there were leaving only 327,784 messages with unique values.

Table 1: *Topic from Twitter*

Sample Topic	Quantities of topic information			Percentage of information	
	Amount	Truth	Misinformation	Truth	Misinformation
ดอกไม้จันทน์, ดอกไม้เพื่อพ่อ, ส่งเสด็จสู่สวรรคาลัย, พระราชพิธีถวายพระเพลิง, รัชกาลที่ 9, สถิตในดวงใจนิรันดร์	363,639	357,485	6,154	98.31	1.69
น้ำท่วม, เขื่อนแตก, ฝนตก, พายุ, ไล่ฝุ่น, อากาศหนาว, แผ่นดินไหว, ก้าวคนละก้าว, อุบัติเหตุ, มาร์ค เยือนไทย, ร้านสะดวกซื้อ ขายเบียร์สด, เพิ่มเงินสมทบประกันสังคม, ลดหย่อนภาษี	584,734	469,588	115,146	80.31	19.69
Total	**948,373**	**827,073**	**121,300**	**87.21**	**12.79**

The experiment compares two machine learning methods: Naïve Bayes and Support Vector Machine. 10-fold cross validation is used for the performance testing of the model. The results from the experiment are shown in Table 2. The F-measure used to measure the overall efficiency of the model calculated from the average between the precision and recall.

Table 2: *Experimental result*

	F-Measure	Precision	Recall (True Positive Rate)	True Negative Rate	False Positive Rate	False Negative Rate	Accuracy
Naïve Bayes	0.9779	0.9909	0.9652	0.9232	0.0768	0.0348	96.08%
Support Vector Machine	0.9994	0.9997	0.9992	0.9971	0.0029	0.0008	99.89%

3. DISCUSSION

An accuracy is a measure of the accuracy of the prediction of models that can predict both truth and misinformation from the total amount of data. The percentage of accuracy obtained from Naïve Bayes is 96.08%. The percentage of accuracy from Support Vector Machine is 99.89%.

Observations made about the misinformation in Twitter found that it occurs mostly in a short period of time and disappears when the truth appears. The time spent in publishing the truth and misinformation are significantly differences. The average duration of published misinformation is 5 days, 1 hour 19 minutes and the average time that the truth is published 7 days 7 hours 13 minutes. It can be said that the lifecycle of the misinformation is shorter than the truth. Misinformation has been spreading for a long time until the truth appears and it will disappear silently. However, the problem is that as long as the truth is not revealed, the misinformation may damage people and society.

If there are a lot of unknown information, the model will not be able to correctly identify the data. Therefore, it is necessary to gather a lot of information and more diverse news for training data. In addition, the misinformation was happened in the limited time because when the truth appears the effect of misinformation will be lost.

4. REFERENCES

[1] S. Kemp. (2019). *Digital in 2019 Golbal digital Overview*. Available: https://wearesocial.com/global-digital-report-2019

[2] S. Aphiwongsophon 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, pp. 528-531.

[3] A. Ehsanfar and M. Mansouri, "Incentivizing the dissemination of truth versus fake news in social networks," in *2017 12th System of Systems Engineering Conference (SoSE)*, 2017, pp. 1-6.

[4] D. G. Young, K. H. Jamieson, S. Poulsen, and A. Goldring, "Fact-Checking Effectiveness as a Function of Format and Tone: Evaluating FactCheck.org and FlackCheck.org," *Journalism & Mass Communication Quarterly*, vol. 95, no. 1, pp. 49-75, 2018/03/01 2017.

[5] R. El Ballouli, W. El-Hajj, A. Ghandour, S. Elbassuoni, H. Hajj, and K. Shaban, "CAT: Credibility Analysis of Arabic Content on Twitter," Valencia, Spain, 2017, pp. 62-71: Association for Computational Linguistics.

[6] Á. Figueira and L. Oliveira, "The current state of fake news: challenges and opportunities," *Procedia Computer Science*, vol. 121, pp. 817-825, 2017/01/01/ 2017.

[7] S. M. Jang *et al.*, "A computational approach for examining the roots and spreading patterns of fake news: Evolution tree analysis," *Computers in Human Behavior*, vol. 84, pp. 103-113, 2018/07/01/ 2018.

[8] M.-p. S. Chan, C. R. Jones, K. Hall Jamieson, and D. Albarracín, "Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation," *Psychological Science*, vol. 28, no. 11, pp. 1531-1546, 2017/11/01 2017.

[9] K. Niklewicz, "Weeding out fake news: an approach to social media regulation," *European View*, vol. 16, no. 2, pp. 335-335, 2017/12/01 2017.

[10] S. R. Guruvayur and R. Suchithra, "A detailed study on machine learning techniques for data mining," in *2017 International Conference on Trends in Electronics and Informatics (ICEI)*, 2017, pp. 1187-1192.

[11] M. AlRubaian, M. Al-Qurishi, M. Al-Rakhami, S. M. M. Rahman, and A. Alamri, "A Multistage Credibility Analysis Model for Microblogs," presented at the Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015, Paris, France, 2015.