Credibility Verification of Social Media Users for Detecting Fake News

S. Sivasankari
Faculty of Engineering and Technology, Department of Information Technology, SRM Institute of Science and Technology, Kattankulathur, India.
E-mail: sivasans2@srmist.edu.in

Dr.G. Vadivu
Faculty of Engineering and Technology, Department of Information Technology, SRM Institute of Science and Technology, Kattankulathur, India.
E-mail: vadivug@srmist.edu.in

Received November 28, 2020; Accepted December 30, 2020
ISSN: 1735-188X
DOI: 10.14704/WEB/V18SI03/WEB18040

Abstract

Fake news contains wrong information's and mostly it spreads through social media. This is mostly done to impose some ideas and is implemented with reasons. These news containing false claims, may end up with viralized. The role of coordinated users in social media is high as they try to give fake reviews to promote or remove some YouTube videos. These coordinated users also can promote unworthy products for sale. With respect to politics, they can even change the scenario by giving negative votes. This paper implemented to verify the user credibility in social media based on their similarity measures. If the user is incredible then the content posted by them is also assumed to be incredible. It is time consuming and includes lot of difficulty in verifying fake content in social media. So we have tried to segregate incredible users that indirectly helps us to identify fake content posted by them.

Keywords

Fake News, Credibility, Similarity, Social Media, Measures.

Introduction

Nowadays, along with the growing technology, the use of social media amongst all the different age groups is at its apex. With the enhancement in the use of the social media, there has been an increase in the rumours on the social media platforms. In order to identify whether the rumour is appropriate, certain methods and techniques are employed namely naïve bayes based on probability value and using Recurrent neural network for classification. This paper is based on how to resolve the rumours and check into with their liability based on user information rather than content of news.

There are basically two types of rumours which have been elaborated i.e. long standing rumours and emerging rumours. The stepwise techniques employed that are detecting the rumours, tracking of rumours, stance classification, and veracity classification of rumours which makes a rumour classification system. We all are aware of the fact that if on one side social media has n number of benefits, on the other hand it has m number of disadvantages. One amongst them is the spread of the rumours which may include societal rumours, celebrity rumours, political rumours etc. A rumour classification system is thus employed in order to be pitch sure whether a particular fact is legitimate. This system employs us with the methodology to get into the resources of the info, veracity, all going through a sense making process.

A rumour has been defined in different ways on different platform. Rumour is defined as "a currently circulating story or report of uncertain or doubtful truth". Also it can be defined as "a statement or report without known authority for its truth". We can exaggerate both the definitions and define rumour as "An unofficial interesting story or piece of news that might be true or invented, and quickly spreads from one person to another". The Long Standing Rumours are basically the rumours that have been discussed for a prolonged period of time. These rumours will circulate for a prolonged period of time without any verification and liability. These rumours are in a much more demand to be clarified due to their longevity and huge interests of the audience whereas Emerging rumours are sort of rumours occurs in breaking news. The occurrence of rumours in the context of breaking news are most probably the one that have not been spotted before.

Studying the Rumours

Rumours have been a part in the society since a long time. Exceptionally talented people have been through these. Several people have several opinions on these rumours. Traditionally, it consumed a lot of time, in order to identify whether a particular was a rumour or not. And social media after all plays an important role in this as social media is a source for verifying rumours which have gained a lot of importance, because social media nowadays is such a platform which has huge fan base and along with the fan base, it has huge number of datasets. Taking an example of Twitter, which has been handling the so called rumours from a long time. It not only supports the comments section but also has a voting feature i.e. how many people liked the post and how many didn't can be seen.

Scope and Organisation

Due to the need of the society for this sensitive topic, this paper was implemented. This particular paper is on the bases of increase in the use of numerous social media platforms namely Facebook, Instagram, and Twitter etc. We are very well known to the fact that if on one hand these social media platforms have numerous advantages, on the other hand it has plenty disadvantages too. Advantages like getting connected to family and friends, getting the news from the whole world, time pass (common phrase used by youngsters nowadays) and disadvantages like fake news regarding politicians and celebrities, religion and caste issues, cyber bullying and so many things which make it unhealthy. The biggest problem which prevails on these social media platforms is the ease of uploading the content i.e. anyone anywhere in the world having an access to the internet can upload anything on the social media. That content which has been uploaded remains unverified. This paper is thus implemented to ensure the techniques which can be employed in order to resolve the problem of rumours through identifying incredible users.

This paper includes the challenges imposed on and by the rumours and how these rumours can be diffused. Social Media being an Information source we have factors which promote the use of social media in News Gathering, Emergencies and Crises, Public Opinion and Financial and Stock Markets.

Related Work

Zhao et al. 2015 idea was built on considering that rumours usually give rise to tweets from users who inquire the conformity to facts. The authors have created a manual list of regular expressions (for example “is (this/it/that) true”) that were used to identify about the inquire tweets. These inquiring tweets are further clustered with respect to their similarity, each cluster have been finally regarded as candidate rumour.

In contrast, Zubiaga et al. 2016, 2017 suggested a method that learns the circumstances that form the setting for an event through a breaking news to verify whether the relevant tweet establish to a rumour or not. They were built on the assumption that having a tweet alone was not adequate to verify if its primary information is a rumour or not, since the context information is not enough. Further, they avoid the dependence on inquiring tweets, as all rumours does not trigger and for this reason it leads to low recall as rumours that do not stimulate inquiring tweets might be fail to notice. Their approach depend on conditional random fields that is used as sequential classifier and it learns from the reported dynamics throughout an event, so that classifier can identify for each incoming tweet, whether it is rumour or not a rumour based on what have been observed so far during the event. This approach leads to improvement in performance than the fundamental classifier by Zhao et al., 2015.

Tolosi et al., 2016 used feature based analysis on the rumours across various events have determined the difficulty to categorize rumours with non-rumours since their feature changes continuously across various events.

Mc Creadie et al., 2015, analysed the convenience of using a crowd sourced platform to detect rumours and non-rumours, identifying that the annotators will be able to achieve higher inter annotator agreement. Here they have categorized rumours into 6 types namely unsubstantiated message, disinformation, dispute message, report, opinionated and the linked dispute.

Zubiaga et al., 2017, delivered an approach leverages the context obtained from previous posts for a specific event to identify whether a tweet posted is a rumour or not.

Sardar et al., 2015, tried supervised machine learning approach in order to determine rumor. They used the guidelines annotation as label and use different technique for classification. They are limited to finding the rumor that are known prior. They work on the limited data that are labelled.

Hareesh et al., 2016 basically deals with the detection of the rumor based on the stance. The also provide good efficiency but not as good as unsupervised learning technique. They also rely on the label which cannot process the unlisted data.

Bajaj, 2017, have applied numerous deep learning methodologies on dataset that consists of various news article(fake) from Kaggle. And also real news articles from the dataset related to media news and analyzed the classifiers based on GRU, LSTM and Bi-directional LSTM has performed well compared to the classifiers that works on convolutional neural network.

Ruchansky et al., 2017 uses social media dataset and they designed a hybrid model which has accuracy of 0.89 for twitter data and 0.95 on weibo data. Listed both the articles temporal behavior and learning source behavior of the users that are essential for detecting fake news.

Y. Long et al., 2017 has worked using liar dataset. They proposed a hybrid LSTM model for fake news detection task, which performed better than hybrid CNN model with accuracy of 14.5%.

Sanjay KS et al., 2019 has identified the fake reviews on products using SVM Classifier and the proposed methodology has 94% accuracy compared to manual detection of fake reviews.

Adnan Hussein et al., 2019 used content and context based features of news from twitter. This model is graph based approach to identify fake news. It is believed that context based will give high accuracy model.

System Overview

Data Set Description

This dataset consists of health tweets from bbc health. The data includes 3929 tweets. The dataset contains the user id, tweet id, timestamp, and tweet.

Data Processing

This step allows us to remove all the unnecessary words that are basically not required in our detection process and it just reduces the accuracy. These words are the noise or the outliers. The word like stop words like “want” and all the punctuation marks should be removed “SpaCy” library in python.

Proposed Methodology

Information Credibility is classified into 3 parts as Message credibility, Source credibility and Media credibility. In this paper we have ranked credibility of social media users in order to measure the credibility of information. This can be implemented by identifying the coordinated users (clique). And clustering the users with similar behavior. Then we calculate the cluster weight for every member in the cluster. Low credibility weight is assigned for users involved in coordinated behavior.

In order to measure similarity of users we have many methods namely TF-IDF (Term Frequency – Inverse Document Frequency), Edit distance and Jaccard’s Coefficient. In our proposed methodology we make use of jaccard’s coefficient to calculate behavior similarity. (Eq. 1)

$$J(B_i,B_j) = (B_i \cap B_j)/(B_i \cup B_j) \quad - (1)$$

Where B_i is behavior of user at timestamp - ‘t’ and
B_j is behavior of user at timestamp - ‘t’.

Algorithm: Credibility Algorithm

Step 1: The similarity between users is measured based on their behavior of every pair (U_i & U_j) – User i and j.

Step 2: Define a threshold value Ɵ.

Step 3: If the user's similarity value exceeds Ɵ then cluster them.

Step 4: The cluster with many users is assigned with lower credible weight, w. {Weights assigned in descending order based on increasing of cluster size}.

Experiments and Results

BBC Health dataset contains 3929 tweets. We first applied pre-processing of tweet data like removing tweet id, time stamp. Removed the URL's and Hashtags. To measure similarity in user behaviour, we used jaccard distance. Based on the distance, we applied k means clustering algorithm to cluster user into groups. Then the number of cluster size is increased as k =1, 2, 3… based on SSE (Sum of Squares of distances from centroid).

Iteration 1 for k = 3
1: 1493 tweets
2: 1210 tweets
3: 1226 tweets
SSE: 3416.7526775028173

Iteration 2 for k = 4
1: 1061 tweets
2: 693 tweets
3: 1362 tweets
4: 813 tweets
SSE: 3363.525962316692

Iteration 3 for k = 5
1: 1107 tweets
2: 586 tweets
3: 672 tweets
4: 983 tweets
5: 581 tweets
SSE: 3338.4759064277573

Iteration 4 for k = 6
1: 473 tweets

2: 639 tweets
3: 458 tweets
4: 1272 tweets
5: 610 tweets
6: 477 tweets
SSE: 3313.7227328125387

Iteration 5 for k =7
1: 864 tweets
2: 466 tweets
3: 343 tweets
4: 492 tweets
5: 531 tweets
6: 841 tweets
7: 392 tweets
SSE: 3272.117103752349

As we can see from the Table 1, the SSE value decreases with respect to increase cluster size, k.

For K = 7, it is very clear that cluster 1 and cluster 6 has greater number of similar tweets 864 and 841 respectively when compared to other clusters. According to credibility algorithm, the cluster of users with highest similarity in behaviour is considered as incredible. Thus the content posted by user id in these groups are not to be considered.

Table 1 Cluster Size Vs SSE

Cluster Size(k)	SSE from centroid
3	3416
4	3363
5	3338
6	3313
7	3272

Conclusion and Future Work

We took health tweet dataset of BBC Health and filtered the tweet id, URL, Hashtag and timestamp. Based on user id and tweet content, applied credibility algorithm to identify similar pattern or behaviour users who involve in promoting fake news and clustered them using k means algorithm. These users are finally considered incredible based on cluster group size. The paper proves if the user is incredible then the information posted by them is also considered as not credible or fake.

Here we have increased the number of iterations for increasing cluster size till we reach a converging point based on SSE value. This can be modified and we can use Elbow method to find the optimal cluster size and go ahead with single iteration and hence reduces the time of trial in range of k.

References

Zhao, Z., Resnick, P., & Mei, Q. (2015). Enquiring minds: Early detection of rumors in social media from enquiry posts. *In Proceedings of the 24th international conference on world wide web*, 1395-1405.

Arkaitz, Z., Maria, L., & Rob, P. (2016b). Learning reporting dynamics during breaking news for rumour detection in social media. *arXiv:1610.07363* (2016).

Zubiaga, A., Liakata, M., & Procter, R. (2017). Exploiting context for rumour detection in social media. *In International Conference on Social Informatics*, *Springer, Cham,* 109-123.

Tolosi, L., Tagarev, A., & Georgiev, G. (2016). An analysis of event-agnostic features for rumour classification in twitter. *In Proceedings of the International AAAI Conference on Web and Social Media*, *10*(1).

McCreadie, R., Macdonald, C., & Ounis, I. (2015). Crowdsourced rumour identification during emergencies. *In Proceedings of the 24th International Conference on World Wide Web*, 965-970.

Sardar, H., & Mona, D. (2015). Rumor Detection and Classification for Twitter Data. *In proceedings of the 5th International Conference on Social Media Technologies, Communication, and Informatics*, 71-77.

Bahuleyan, H., & Vechtomova, O. (2017). UWaterloo at SemEval-2017 Task 8: Detecting stance towards rumours with topic independent features. *In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017),* 461-464.

Samir Bajaj. Fake News Detection Using Deep Learning. https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1174/reports/2710385.pdf

Ruchansky, N., Seo, S., & Liu, Y. (2017). Csi: A hybrid deep model for fake news detection. *In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 797-806.

Yunfei, L. Qin, L., Rong, X., Minglei, L., & Chu, R.H. (2017). Fake News Detection through Multi-Perspective Speaker Profiles. *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers) IJCNLP*, 252-256.

KS, S., & Danti, A. (2019). Online fake review identification based on decision rules. *International Journal of Advanced Trends in Computer Science and Engineering*, *8*(2), 140-143.

Adnan, H., Farzana, K.A., & Siti, S.K. (2019). Content-Social based Features for Fake News Detection Model from Twitter. *International Journal of Advanced Trends in Computer Science and Engineering*, *8*(6), 2806-2810.