International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 7 | ISSN : 2456-3307

# Rumour Detection from Social Media : A Review

**Shital Lathiya*[1], M B Chaudhari[2]**
*[1]Student, Government engineering college, Gandhinagar, Gujarat, India
[2] Professor, Government engineering college, Gandhinagar, Gujarat, India

## ABSTRACT

In this era, social media platform are increasingly used by people to follow newsworthy events because it is fast, easy to access and cheap comparatively. Despite the increasing use of social media for information and news gathering, its nature leads to the emergence and spread of rumours i.e., information that are unverified at the time of posting, which may causes serious damage to government, markets and society. Therefore, there is necessity of effective system for detecting rumours as early as possible before they widely spread. Effective system should consist of four components: Rumour detection, rumour tracking, stance classification, and veracity classification. Lots of work has been done in later component while very less work in component rumour detection. So, now we should work on rumour detection. In this paper, we will summarise efforts done till now in this area. Most of existing methods detects a priori rumours, i.e., predefined rumours. So it is required to have automated rumour detection method which detects new emerging rumours effectively and as early as possible.

**Keywords :** Rumour Detection, Rumour Classification, Misinformation, News Events, Social Media

## I. INTRODUCTION

Today majority of people gather news online. Earlier newspaper and TV news channels were the main source of news events, but now with the increasing and easy use of internet in mobile people easily get news online faster than other sources. With use of internet in mobile, social media like twitter, Facebook, or whatsapp are the main platform used by almost all mobile users. On social networks everybody is free to obtain and share information, anywhere at any time [10]. So, breaking news spread very fast in social media. With breaking news, sometimes rumour also spread quickly in social media which may cause harm to society and government too.

The incentive for development of data mining tool for dealing with rumours increased in recent years. In data mining, there are many supervised, semi-supervised and unsupervised algorithms. Classification algorithms are supervised as they have predefined set of categories and labelled dataset. Clustering algorithms are unsupervised algorithm as data is unlabelled and no predefined set of categories is available. In semi-supervised, first clustering algorithm apply on available
Dataset and then based on clusters, classification algorithm applies.

Rumour detection is considered as binary classification task where we have predefined set of category of binary class as {Rumour, Non-Rumour} and labelled dataset is there to train classifier. Binary classification is a category of classification that classifies the events into two categories based on features. Binary classification would generally fall in the domain of supervised learning since dataset is labelled. There are various paradigms used for

learning binary classifier such as Decision trees, neural networks, Bayesian classifier or SVM [14].

## II. BACKGROUND STUDY AND RELATED WORK

To detect rumours from social media, first we need to study psychology of rumour. Then based on features and characteristics of rumour, we can make effective system that detects rumour. Here, in this part we summarize psychology of rumour in brief , general architecture of rumour classification system and introduction to existing work done to solve this problem.

### Definition: Rumour

Oxford English Dictionary defines a rumour as "a currently circulating story or report of uncertain or doubtful truth". Merriam Webster Dictionary defines it as "a statement or current report without known authority for its truth". So, basically rumour is a circulating story or message whose truth value is unverified at the time of posting. This unverified information may turn out to be true, or partly or entirely false; alternatively, it may also remain unresolved.

### Types of Rumour

Many different factors are available for classifying rumours by types as based on its veracity value (true, false, or unverified), based on credibility (low or high). Knapp et al. (1994) introduced taxonomy of three types of rumours: (1) "pipe-dream" rumours: i.e., rumours that lead to wishful thinking; (2) "bogy" rumours: i.e., those that increase anxiety or fear; and (3) "wedge-driving" rumours: i.e., those that generate hatred. With the perspective of rumour classification system, rumour can also be classified as (1) a priori rumour: It is a long standing rumour that is discussed for long period of time. (2) New emerging rumour: Rumours that emerged during breaking news event. This rumour are the one that not seen before.

Gorden et al. [8] analysed psychology of rumours. They gave a basic rule of rumour as rumour is multiplicative of importance and ambiguity. If either of these two is absent then it is not rumour. Ambiguity alone does not sustain rumour nor does importance. Rumour is set in motion and continues to travel in homogeneous social medium by virtue of the strong interest of individuals involved in transmission. Authors found that the number of details retained declines most sharply at the beginning of a series of reproductions. The number continues to decline, more slowly, in each successive version.

Zubiaga et al. [9] shows that rumours that proven to be true tends to resolve faster than false rumour. Their study revealed the importance of official announcement by a reputable person in society. The prevalent tendency of users is to support every unverified rumour. They defined follow ratio as logarithmically ratio of number of followers over number of followings. Their analysis shows that users with high follow ratios are more likely to: (1) support any rumour, irrespective of its truth value; (2) be certain about their statements and (3) attach evidence to their tweets by quoting an external source. On the other hand, users with low follow ratios are more likely to: (1) deny rumours, irrespective of their actual truth value; (2) be rather uncertain about their statements and (3) either provide no evidence in their tweets, or provide evidence on the basis of their own experience, opinions or observations. They also considered other factors to distinguish between users, such as user age, whether or not they are verified users, or the number of times they tweet, but found no significant differences.

### Architecture of Rumour classification system

Zubiaga et al. [15] defined a typical architecture of rumour classification system that includes all the components needed for a complete system. Depending on requirement, we can also omit any component. Rumour classification system generally begins with identifying information which are unverified (Rumour detection) and ends with determining its veracity value (veracity classification). The entire process consists of four components as below:

1. **Rumour Detection:** To identify whether a piece of information constitutes rumour or not. Binary classifier is used to classify stream of data into Rumour or Non-rumour.
2. **Rumour Tracking:** Once rumour is identified using rumour detection component, this will collects and filter post discussing rumour.
3. **Stance Classification:** It classifies collected related post to predefined set of stance {i.e., supporting, denying, querying, and commenting}.
4. **Veracity Classification:** It determines actual truth-value of the rumour using stance value determined in stance classification.

Lots of work has been done in later components. So, to develop a complete rumour classification system, there is need to do work in rumour detection.

Rumour detection task is to determine, from social media post, which spreading post are yet to be verified. Despite the increasing interest in analysing rumour, there has been very little work automatic rumour detection. Some of the work done by quazvinian et al.; and Hamidian and Diab but it has been limited to finding a priori rumour. This type of approach is useful for long-standing rumour only. First work that tackled the detection of new rumour is approach proposed zhao et al.[5]. Their approach based on fact that piece of information that has number of enquiry post tends to be rumourous. In contrast, zubiaga et al.[7] proposed approach based on context learned throughout the breaking news story. Their context-learning approach based on CRF (conditional random field) as a sequential classifier. Their approach improved performance over baselines zhao et al., Random forest, Naïve byes, SVM and Maximum entropy classifier. This approach achieves state-of-the-art results [15].

## III. LITERATURE SURVEY

There has been very little work done in automatic detection of new emerging rumour. Most existing method detects a priori rumour (e.g., Obama is muslim) where classifier is feed with predefined rumour, then classifier can classify post based on keyword(Obama and muslim) of predefined rumours. We study and analyse existing method to detect rumour in social media and we represent summary of all that methods in this section.

Qazvinian et al. [1] gave a general framework which predicts whether a given statement is rumour related or not and if rumour related then finds that user believe this rumour or not. In this paper, they mainly explore the effectiveness of three categories of features (1) content based, (2) network based and (3) twitter-specific memes for identifying rumours. In network based features, they focus user behaviour on twitter. They also consider user who retweets, because a tweet is more likely to be rumour if it posted or re-tweeted by user who has history of posting or re-tweeting rumour. They consider hash-tag and URL as features in twitter-specific memes category. They calculate the log likelihood ratio of each tweet. Likelihood ratio expresses how many times more likely the tweet belong to positive model than negative model. Using various features, they perform 5-fold-cross-validation. In feature analysis, they find that user history can be a good indicator of

rumour. This work is limited to a priori rumours. This approach is not effective for new emerging rumours.

Takahashi et al. [2] described how rumours spread after an earthquake. They also discussed characteristics of rumours spread after disaster. Based on characteristics, they defined a system that finds rumour candidates from twitter. They consider two rumours during earthquake disaster and analyse it thoroughly. They found that 'When people retweet a retweeted tweet, it has higher possibility as a rumour comparing with their followings' tweets'. They showed that after correcting tweet posted about a rumour, that correcting post will spread faster than rumour. They told that the high value of re-tweet ratio can be a clue to find rumour. They also find word difference in rumour and correction post. In their proposed model, they first applied **named entity recognition** to all tweets and extracted named entities which occurred more than 30 times in a day. These named entities were then used as target in further experiment. Then they filter these tweets by re-tweet ratio more than 0.80. Then they again filter by clue keyword 'false rumour' to find rumour from candidates.

Aditi gupta et al.[3] analysed fourteen high impact news events in twitter of 2011 and find its credibility. They used linear regression analysis to find content and source based features. Content based features were number of unique characters, swear words, pronouns, and emoticons in a tweet, and user based features were number of followers and length of username. They applied a supervised machine learning algorithm (SVM-Ranking) and feedback approach to rank tweets. Their performance increased when they apply re-ranking strategy (Pseudo relevance feedback). Their main limitation is that they need human annotator to obtain ground truth of each event. This model works on predefined rumours.

Suhana et al. [4] collects tweets containing false information posted during London riots 2011 from twitter and then extract content based and user based features from tweets and then also reduce features that classifies data more efficiently. They found that content based feature contributes more than user based features. They train supervised classification algorithm J48 classifier based on features and classify tweets as rumour and non-rumour and then find origin of rumour tweets but they didn't get sufficient data to test 'finding of origin' because most of the accounts which previously posted rumour has been already blocked. They get 87% weighted avg. accuracy for both rumours and non-rumours for training dataset and get 88% accuracy on reduced features.

Zhao et al. [5] detect rumours based on enquiry response from real-time data. They design some generalise regular expressions that may arise in response to a rumour post based on fact that generally more question arise in rumour more than valid news. They propose a procedure that has five steps (1) Identify signal tweets: find response tweets that match pre-defined enquiry pattern, (2)cluster signal tweets: Make cluster of all these signal tweets, (3) Detect statement : derive a statement from each cluster that represent all tweets in that cluster, (4)Capture non-signal tweets: collect non-signal tweets that doesn't match regular expression but is related to derived statement that makes candidate rumour cluster and (5)Rank candidate rumour cluster: Using statistical features of the cluster, they rank the clusters by their likelihood of really containing a disputed factual claim. This procedure works on real-time data. It is not necessary that all rum ours have enquiry response. So it has very low recall but high precision.

Jing Ma et al. [6] proposed a deep learning framework for rumour debunking. Proposed model is based on **RNN** for learning the hidden representation that

based on contextual information of relevant post over time. This RNN based model classifies microblog events into rumours and non-rumours so they detect rumours at event level not individual tweet level. They develop RNNs of three different structures tanh-RNN, single layer LSTM and GRU(LSTM-1, GRU-1) and Multi-layer GRU(GRU-2). They compare proposed model with SVM-TS, DT-Rank (zhao et al.), DTC, SVM-RBF and RFC. They showed that their proposed model outperform all the base lines on both datasets (twitter and sina weibo). Tanh-RNN achieves 82.7% accuracy on twitter data. Out of their four proposed structures, GRU-2 outperforms all other three. GRU-2 can detect rumours with accuracy 83.9% for twitter within 12-hours.

Zubiaga et al. [7] proposed a context-aware rumour detection model that uses a sequential classifier CRF to detect new rumours in new stories. They build this model on hypothesis that tweet alone may not sufficient to classify it as rumour or non-rumour, context related to that tweet is more significant. The input to CRF is Graph: G(V,E). They use two types of features, content based and social based.They analyse the performance of CRF as a sequential classifier on five twitter dataset related to five different news stories to detect new tweet that constitutes rumour. They set min retweeet ratio of each tweet as 100. Performance of proposed model is evaluated by computing precision, recall and F1-score for the target category (rumour). This model is restricted to highly retweeted tweets and when tweet is related to new event whose context is not there, then model may not perform well. CRF also suffers from cold start problem.

## IV. EVALUATION METRICS

The performance of any trained model is determined by how accurate the observation is with actual events [12]. We can evaluate any model with labeled data. So, to evaluate performance of any algorithm, we need some evaluation metrics. General evaluation metric used in any algorithm is accuracy. Apart from this, other useful metrics use in rumour detection are Precision, Recall and F1-score. While predicting values against labeled, we get four bins which are True Positive (TP), True Negative (TN), False Positive(FP), and False Negative(FN). TP is rumoured event is predicted as rumour, TN is non-rumoured event is predicted as non-rumour, FP is non-rumoured event is predicted as rumour, and FN is rumoured event is predicted as non-rumour[11][12].

**Table 1 :** Evaluation metrics with formula

| Evaluation metric | Formula |
|---|---|
| Accuracy | (TP+TN) / total events |
| Precision | TP/(TP+FP) |
| Recall | TP/(TP+FN) |
| F1-Score | 2*(Recall * Precision) / (Recall + Precision) |

## V. CONCLUSION

Generally rumours spread hatred or fear which is extremely harmful to society. So, we must take some steps to diffuse this rumour. In this paper, we summarised psychological study of rumour, existing methods to detect rumour, and evaluation matrix used to evaluate performance of method. Research in rumour detection is growing day by day as use of social media is increasing in society. As existing methods are not such capable that can efficiently process stream data and automatically detect new emerging rumours from social media, so we need a complete system that can automatically detect new emerging rumours as early as possible.

## VI. REFERENCES

[1] Vahed Qazvinian, Emily Rosengren, Dragomir R. Radev, Qiaozhu Mei “Rumor has it:

Identifying Misinformation in Microblogs" Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, pages 1589–1599, Edinburgh, Scotland, UK, July 27–31, 2011

[2] Tetsuro Takahashi, Nobuyuki Igata, "Rumor detection on twitter," SCIS-ISIS 2012, Kobe, Japan, November 20-24, 2012, IEEE.

[3] Aditi Gupta and Ponnurangam Kumaraguru "Credibility Ranking of Tweets during High Impact Events,"ACM, 2012.

[4] Sahana V P, Alwyn R Pias, Richa Shastri and Shweta Mandloi, "Automatic detection of Rumoured Tweets and finding its Origin," Intl. Conference on Computing and Network Communications (CoCoNet'15), Dec. 16-19, 2015, Trivandrum, India, Journal: IEEE.

[5] Zhe Zhao, Paul Resnick, and Qiaozhu Mei, "Enquiring minds: Early detection of rumors in social media from enquiry posts," In Proceedings of the 24th International Conference on World Wide Web. ACM, 1395–1405

[6] Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J. Jansen,Kam-FaiWong and Meeyoung Cha, "Detecting Rumors from Microblogs with Recurrent Neural Networks," Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16), 2016.

[7] Arkaitz Zubiaga, Maria Liakata and Rob Procter, "Exploiting Context for Rumour Detection in Social Media," springer, 2017

[8] GORDON W. ALLPORT AND LEO POSTMAN, "AN ANALYSIS OF RUMOR," Downloaded from http://poq.oxfordjournals.org/ at University of California, San Francisco on December 11, 2014

[9] Zubiaga A, Liakata M, Procter R, Wong Sak Hoi G, Tolmie P (2016) Analysing How People Orient to and Spread Rumours in Social Media by Looking at Conversational Threads. PLoS ONE 11(3): e0150989. doi:10.1371/journal.pone.0150989

[10] A. Friggeri, L. Adamic, D. Eckles, and J. Cheng, "Rumor Cascades," Icwsm, pp. 101–110, 2014.

[11] https://pdfs.semanticscholar.org/d051/ba0a904e4b4c45a2af145aa29b8490bbbc5c.pdf

[12] https://blogs.msdn.microsoft.com/andreasderuiter/2015/02/09/performance-measures-in-azure-ml-accuracy-precision-recall-and-f1-score/

[13] Anuradha Purohit, Deepika Atre, Payal Jaswani, and Priyanshi Asawara, "Text Classification in Data Mining," International Journal of Scientific and Research Publications, Volume 5, Issue 6, June 2015.

[14] https://www.cse.iitk.ac.in/users/se367/10/presentation_l2288888888ocal/Binary%20Classification.html

[15] ARKAITZ ZUBIAGA, AHMET AKER, KALINA BONTCHEVA, MARIA LIAKATA and ROB PROCTER, "Detection and Resolution of Rumours in Social Media:A Survey," arXiv:1704.00656v3 [cs.CL] 3 Apr 2018.