Rumour Detection Models & Tools for Social Networking Sites

Mohammed Mahmood Ali, Mohammad S. Qaseem, Ateeq ur Rahman

Abstract: Efficient utilization of social networking sites (SNS) had reduced communication delays, at the same time increased rumour messages. Subsequently, mischievous people started sharing of rumours via social networking sites for gaining personal benefits. This falsified information (i.e., rumour) creates misconception among the people of society influencing socio-economic losses by disrupting the routine businesses of private and government sectors. Communication of rumour information requires rigorous surveillance, before they become viral through social media platforms. Detecting these rumour words in an early stage from messaging applications needs to be predicted using robust Rumour Detection Models (RDM) and succinct tools. RDM are effectively used in detecting the rumours from social media platforms (Twitter, LinkedIn, Instagram, WhatsApp, Weibo sena and others) with the help of bag of words and machine learning approaches to a limited extent. RDM fails in detecting the emerging rumours that contains linguistic words of a specific language during the chatting session. This survey compares the various RDM strategies and Tools that were proposed earlier for identifying the rumour words in social media platforms. It is found that many of earlier RDM make use of Deep learning approaches, Machine learning, Artificial Intelligence, Fuzzy logic technique, Graph theory and Data mining techniques. Finally, an improved RDM model is proposed in Figure 2, efficiency of this proposed RDM models is improved by embedding of Pre-defined rumour rules, WordNet Ontology and NLP/machine learning approach giving the precision rate of 83.33% when compared with other state-of-art systems.

Keywords : Social Networking Sites (SNS), Rumour Detection models (RDM), Pre-defined rules, WordNet Ontology.

I. INTRODUCTION

With the use of Social media platforms there is a tremendous increase in spreading of rumours on various topics and domains. Now-a-days, these social messaging applications are excessively used in promoting of events, Advertisements, New’s channels, sharing of market data and business transactions. Sometimes, these microblogs communicate the false information which leads to misunderstanding among the group of people creating mental tensions in the society. Surveillance of falsified information (i.e., rumour) needs to be strictly monitored by e-criminal cell. The e-criminal cell is authorized to take stringent action against those culprits for sending rumours through SNS. Sending of deceitful and false information named as “rumour”, which is one of the serious cybercrimes as per the FISA Act [4]. Spreading of rumours through Websites and Social media platforms, mobile phones, laptops and vice versa may encounter various problems in the society that hinders the development by creating mental tensions among the people [5]. Specifically, many of the electronic rumours spread through mobile messaging applications is very difficult to catch at the initial stages unless it is notified by the users, and these short posts exists for short life span at the server. Similarly, microblogs communicated or shared via various interchangeable social media platform to other social mediums (i.e., WhatsApp to Facebook, Google+ to Instagram, Instagram to WhatsApp, youtube to WhatsApp, Facebook to WhatsApp and vice versa) differs in their messaging architecture and privacy restrictions of storing and retrieving policies that makes it difficult to identify the rumour words when they are encountered in microblogs [6]. Radio agencies and News channels also plays a vital role in sending of rumours through audio, video or conference communication, which becomes impossible to analyze and stop their transmissions at run-time, such contents once viewed in mobile phones are automatically auto-saved in the memory and hence, are transmitted to others at later point of time. Spying of such rumour voice communications and video recordings is still a research issue that requires rigorous surveillance at various instance of timestamps. Every post may not be a rumour, identifying factual microblogs from set of cluster of posts that are sent through social media is predicted using ranking algorithm from various enquiry patterns [7]. Twitter messaging application which is widely used by millions of people for posting, giving reply to specific tweets, forwarding of tweet to other users adversely influence on Health domains by creating mental tension in the society. To overcome, health domain problems from Twitter, few parameters are picked for evaluation such as statistics of users, sentiments of specific tweets, followers of root of tweet along with URLs and fed to classifiers for finding the rumours [3]. A new classification algorithm was proposed using statistical metrics for segregation of rumour and non-rumour twitter posts based on users frequency of interaction, structure & network establishment, temporary connectivity and linguistic features. It is concluded that linguistic features evolved to be on-top-priority with good accuracy rate in classification of rumours and non-rumours for tweets that vary for long duration [10]. Another study, suggest that rumours are detected by supervised (well-labelled datasets), unsupervised (unknown labels), and hybrid based (known and unknown keywords) approaches [5].

Revised Manuscript Received on December 15, 2019.

* Correspondence Author
Mohammed Mahmood Ali*, CSE department, Osmania University, Hyderabad, India. Email: mahmoodali4@gmail.com
Mohammed S. Qaseem, CSE department, Nawab Shah alam college of engineering, JNTUH, Hyderabad, India. Email: ms_qaseem@yahoo.com
Ateeq ur Rahman, CSE department, Shadan College of Engineering and Technology, JNTUH, Hyderabad, India. Email: mail_to_ateeq@yahoo.com

Received Number: 85465129219/201908BEIESP
DOI: 10.35940/ijeat.B3465.129219

Published By: Blue Eyes Intelligence Engineering & Sciences Publication
In this Section we illustrated the importance and its necessity to detect rumours from Social networking sites and its impact towards maintaining of Harmony and peace by avoiding unfaithful tensions in the society. In Section II, we defined the Rumour definition based on earlier studies and the various types of rumour categories. We identified the problems that are faced, if rumours are not detected in the initial stages from microblogs. The Section III, describes the thorough comparative analysis of various rumour detection models by highlighting the Features, Technique utilized, Drawbacks, and giving an asserting suggestion for improving these rumour detection models, whereas Section IV, depicts the overview of Datasets that are used in prediction of rumours from social media platforms and the currently existing tools available to predict and stop rumours from various SNS is discussed. To overcome, the necessary enhancements needed to be embedded by removing the flaws that exists in RDM with the use of Ontology and pre-defined rumour rules are elaborately discussed in Section V. The Section VI, discusses the experimental results obtained by TraceMiner, CED, and Proposed RDM approach. Section VII, concludes with the literature survey reviewed in this paper by considering the seriousness required in detection of rumours from various electronic communication channels, especially social media platforms. Further, in future the RDM can be improvised by embedding of multilingual language along with pre-defined patterns and Ontology.

II. PROBLEM STATEMENT AND RELATED WORK

A. Rumour Definition

“A rumour is defined as falsify message, intentionally created by an individual or group of individuals to gain personal benefits and it is forwarded to other individual or groups. Later, this falsify message is again forwarded by innocent individuals believing it’s a truthful message”. The spreading of rumour words through various messaging applications and creating misconceptions among the users of SNS resulted in social disaster is depicted in figure 1.

In 1944, Robert H. Knapp analysed the various categories of rumour types in his reports “A Psychology of Rumor”. He broadly categorised these rumours into three types [20]:
- Pipe dream rumors: Reflects public desires and wished for outcomes. (E.g. Japan's oil reserves were low and thus World War II would soon end.)
- Bogie or fear rumours reflects feared outcomes. (E.g. An enemy surprise attack is imminent).
- Wedge-driving rumours intend to undermine group loyalty for interpersonal relationship (E.g. German-Americans, Italian-Americans, Japanese-Americans were not loyal to the American side).

C. Problems faced due to rumours in SNS

RDM approaches that are discussed in Section I, are mediocre to detect multilingual rumour words. Further, these methodologies are deficient to predict emerging rumours that has abbreviated terms (i.e. short-forms). When the words from microblogs of social media are mapped and checked with rumour datasets the efficacy of detection rate is poor, this is due to emerging rumour words in which linguistics words are used for this reason they are not detected by RDM, and hence neglected.

The dataset needs to be updated frequently as the new trending rumour words arrive from SNS. Hence, the old rumour words (i.e., dataset) are obsolete when compared with emerging rumour words. These dataset are again categorized based on domains, obtaining dataset in specific domain and judging rumour and non-rumour posts is limited to certain extent, but collectively testing with multiple domains and comparing with each of the respective domain resulted in ambiguity outputs [3]. Many of architectures make use of Machine learning technique which is a time-consuming process that can be improvised with the use of pre-defined knowledge based rules guided with semantic ontology.

Another major drawback of rumour datasets is processing time taken to generate an alert and report to servers is delayed. In many of the rumour detection strategies it is found that machine learning approaches requires frequent training of system using the labelled dataset in case of supervised strategy. Studies suggest that, instead of machine learning or ranking algorithmic approach, the pre-defined knowledge based rules sum-up with deep learning approaches (CNN/RNN) and Natural Language processing embedded with Semantic Ontology are found to be efficient when compared to earlier approaches [16] [8]. After keen observation of various datasets of different domain it is suggested that domain specific rumour words need to be used in a particular context for better precision rate. If the Hybrid Ontology is used for finding multiple synonyms of rumours using WordNet it becomes easy in logical annotating of rumour words of a particular domain [9].

In this section we thoroughly briefed the problems that are faced in detection of rumour words from microblogs. Further, how these approaches can be improvised with the use of Knowledge-based pre-defined rumour rules, Machine learning technique and Semantic Ontology is discussed.
III. COMPARATIVE ANALYSIS OF RUMOUR DETECTION MODELS

Spreading of rumours through various means such as websites, social media, or electronic messaging systems is considered as one of the serious cybercrime [11]. To mitigate the rumours that are spread widely needs to be detected and stopped using cyber surveillance tools at the servers before it is forwarded through various means of communication channels. The technique of asserting the rumours from microblogs sent via social media is identified using Hidden Markov model and supervised classifier by deriving parameters from annotated dataset [12]. Similarly, many approaches to predict and prevent rumours from different communication channels were proposed, most of them make use of machine learning classification technique, Text mining, deep learning or SVM techniques as shown in Table I.

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Strategy/ technique</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Predicts the rumours from the reposts made via Social media. Deep neural network (i.e., CNN) is used to train the sequence of posts, for every single microblog a unique “credible detection point” is generated for better accuracy.</td>
</tr>
<tr>
<td>2.</td>
<td>Collect the available microblogs sent through communication channels over the network. Machine learning technique efficiently used for rumour detection from posts.</td>
</tr>
<tr>
<td>3.</td>
<td>Timely detection of Rumour before it spreads. Max pooling method extracts features then deep Q-learning model applied to check real rumour event, then successively train, &amp; finally minute-to-hour posts grouped and compared for real rumour.</td>
</tr>
<tr>
<td>4.</td>
<td>Cumulative dispersal of rumour patterns over time &amp; tracking the continuous changes in predictive powers of rumour features. Removal of irrelevant features using permutation, then selecting relevant features through interpretation. Next removal of redundancy from selected features. During this process Machine learning (Classifier used to learn)</td>
</tr>
<tr>
<td>5.</td>
<td>Depending upon the type of tweets sent the selection of classifier is chosen to predict the rumour and non-rumours. Rumours predicted based on context using various classifiers namely Conditional Random Fields, Maximum Entropy, Enquiry-based.</td>
</tr>
<tr>
<td>6.</td>
<td>Spying of rumour spreaders and controlling along the social networking sites is the objective. Monte Carlo is used to simulate rumour propaganda. The high degree and low degree at the network nodes is identified by intuitive ranking and K-Shell decomposition method using Degree centrality.</td>
</tr>
<tr>
<td>7.</td>
<td>Detects long lasting and newly emerging rumours spawned across the network. Machine Learning approaches and NLP techniques used for Rumour Classification by rumour tracking, stance &amp; Veracity classification.</td>
</tr>
<tr>
<td>8.</td>
<td>Proposed a technique to identify trending rumours, which are defined as a topic that includes disputed factual claims. Performs four major tasks: i) Identify tweets that has Questions, ii) perform clusters of those questions, iii) annotate by detecting the rumour in the tweet, iv) Rank clusters using statistical features.</td>
</tr>
</tbody>
</table>

IV. RUMOUR DATASETS AND DETECTION TOOLS

A. Rumour Datasets

Collection of datasets by considering historical belongings is trivial as discussed on websites snopes.com and urbanlegends.about.com where the events that were occurred in 1876 were suddenly re-appeared in 1974 showing old collective materials which were never existed earlier, but at the same time few contents such as Address and the shops used to exist till date, which becomes extremely difficult to believe. For, such events, tweets were widely discussed by different people with respect to current situation which will inadvertently gives rise to propaganda of rumours.
Detected of short rumour tweets such as “Obama is a muslim guy”, this does not require excessive learning for this authenticated books to be checked and a conclusion can be derived within seconds using a pre-defined rule based classifier [13]. In literature, most of the researchers had picked datasets of their own considering a particular scenario and few have taken from previously existed rumour datasets [3] [13] [16]. Many of the researchers had picked rumour dataset of microblogs from Twitter, Facebook, Weibo and Instagram to predict the category of rumours [1] [18].

The proposed approach of Rumour detection model after improvising by embedding of features namely Linguistics, Pre-defined rumour rules and Machine learning approach. The proposed approach of Rumour detection model after embedding the above features is shown in Figure 2.

B. Rumour Detection Tools

Many tools for detecting of Fake rumours was proposed earlier, one among them is TraceMiner that detects the Fake news or rumors generated using the social media linkages in various social networking communities which is a static approach [15]. Rumour detection projects aimed at developing of specialized tools namely Emergent, PHEME, RumorLens, TwitterTrails, RumourFlow, COSMIC, SUPER, Hoaxy, REVEAL, InVID, CrossCeck, Decodex, Check, ClaimBuster, Una Hakika, Seriously Rapid Source Review, TweetCred were successfully deployed for detection of rumours from microblogs [18].

VI. PROPOSED RUMOUR DETECTION MODEL

After analyzing, various Rumour detection tools and problems faced by RDM in Section 3. These RDM can be improvised by embedding of features namely Linguistics, Pre-defined rumour rules and Machine learning approach. The proposed approach of Rumour detection model after embedding the above features is shown in Figure 2.

The WordNet, logically establishes the links with set of synonym words from the posts that are sent by the users via SNS [2] [22]. This WordNet lexical database comprises of 155,327 words forming a synsets word-pairs of 207,016.

In this proposed Rumour detection approach, the messages that are sent via SNS are picked and then unnecessary words are filtered. After this, filtered message along with linguistic words is fed to NLP or Machine learning Tool, which then tries to detect the Rumour words by computing with pre-defined rumour rules which is assisted with WordNet Ontology. On detection of rumour words the domain of Rumour type is identified and reported to E-crime department cell. This e-crime cell will remove the rumour words and then forwards the filtered message without rumour words to other users of SNS. Subsequently, a warning message and an alert is generated and forwarded to the concerned user of SNS.

VI. EXPERIMENTAL RESULTS & ANALYSIS

Based on stance and veracity these rumour messages communicated through Social Networking sites are classified into Four (4) types:

i) Rumour messages persist for long durations and continue to exist forming an infinite linked-chain.

ii) Rumour messages that exist temporarily for short-duration and then disappear after certain period.

iii) Emerging rumour messages are those that changes with respect to time, context, necessity and situation.

iv) Use of diverse and versatile language style in chatting session that constitute of rumour words.

First and Second type of rumour messages are detected using rumour detection methodologies which are discussed in Section IV, they are dependent on historical datasets for categorizing posts as rumour or non-rumour using machine learning approaches which is a time-consuming process. Whereas Third and Fourth types efficiently predicts rumours from SNS using NLP guided with pre-defined rules shown in Table II instead of historical datasets as discussed in Section V.

<table>
<thead>
<tr>
<th>Type of rumors (Domain)</th>
<th>Rumour words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wish rumors</td>
<td>Imagining, day mare, illusion, mirage, hallucination, pipe dream, gossip, vision, fantasy, fiction, dream, visualization,...</td>
</tr>
<tr>
<td>Dread rumors</td>
<td>Bone chilling, startling, threatening, weird, horrid, alarming, ghastly, fearsome, terror, panic,...</td>
</tr>
<tr>
<td>Wedge Driving rumors</td>
<td>Hateful, loathsome, distasteful, bold, determined, insolent, dominating, fierce, vicious, savage, fighting,...</td>
</tr>
<tr>
<td>Accident</td>
<td>Crash, disaster, tragedy, collision, survive, deathblow</td>
</tr>
</tbody>
</table>

A. Rumour Scenario in Social Networking Sites

In this section, an example is chosen that illustrates, how the rumour words are propogated through online-chatting session by group of users in SNS, which is shown in Table III.
Table-III: Depicts the domain of ‘Accident’ rumour words communicated between the users

<table>
<thead>
<tr>
<th>Domain of Rumour</th>
<th>User 1</th>
<th>User 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accident</td>
<td>“Did u hear the latest news of a car crash? It turned out to be a disaster.”</td>
<td>“Did anyone survive?”</td>
</tr>
<tr>
<td></td>
<td>“the collision took place at the intersection where the driver fell asleep”</td>
<td>“what a tragedy, how did it happen”</td>
</tr>
<tr>
<td></td>
<td>“no, it was a deathblow.”</td>
<td></td>
</tr>
</tbody>
</table>

Rumour words to be detected: Crash, disaster, tragedy, collision, survive, deathblow

In Table III, the ‘user 1’ propagates the fake information of rumour words pertaining to ‘Accident’ domain to another ‘user 2’. In this chatting session the rumour words like crash, disaster, tragedy, collision, survive and deathblow are used which needs to be detected. The above chatting session is tested with Rumour Detection strategies TraceMiner[15], CED[16] and the proposed RDM. The precision rate that are obtained are shown in Table IV.

Table-IV: Rumour words detected from chatting session by TraceMiner, CED, & Proposed RDM strategy

<table>
<thead>
<tr>
<th>Word Detected</th>
<th>TraceMiner</th>
<th>CED</th>
<th>Proposed RDM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crash, Disaster</td>
<td>Collision, Tragedy, Disaster</td>
<td>Crash, disaster, tragedy, collision, deathblow</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Count of Words</th>
<th>2</th>
<th>3</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision rate</td>
<td>33.33%</td>
<td>50.0%</td>
<td>83.33%</td>
</tr>
</tbody>
</table>

B. Analysis of RDM model

TraceMiner has detected only Two (2) rumour words namely ‘Crash’ and ‘Disaster’ with a precision of 33.33%, TraceMiner initially forms a cluster of messages from the users of Social media with the aid of Gaussian distribution and segregates the collected messages into two sets one is training and another is validation, then these sets are fed to RNN for training purpose for a certain period of time. Thus this training for TraceMiner using reinforcement learning resulting in poor detection of rumour [15]. Another rumour detection approach named as CED succeeded in detecting of rumours with a precision rate of 50%. The words ‘Collision’, ‘Tragedy’ and ‘Disaster’ are words identified from Table II. The CED uses checkpoint for every single post sent by the user, these posts are fed to CNN and RNN for training of CED, during this process the weights changes as the post increases. This CED approach is better when compared to TraceMiner, because it uses checkpoints for every post.

In our proposed RDM approach, the precision rate is 83.33%. This approach initially, filters unnecessary words from messages, parses the text using NLP approach, identifies the Entity relationships between the words. After finding the entities, it starts mapping with the pre-defined rumour rules which is shown in Table II. During this process, WordNet ontology is efficiently used that logically finds the exact word by mapping with the words from various posts sent by users. On detection of matched words the domain ‘Accident’ along with set of rumour words is predicted by our proposed RDM model.

VII. CONCLUSION AND FUTURE WORK

This paper reviews different ways by which rumours are propagated by criminals through various means of social media platforms. Further, how these falsified rumours are picked and communicated to other social users or groups that resulted into a serious cybercrime and disturbing of mental peace in the society. Interestingly, the various rumour detection models developed till date are discussed in this paper in section III, it is found that, many of these approaches of rumour detection methodologies could able to predict and prevent rumours from different communication channels (i.e. social media platforms) most of these RDM make use of Classifiers, machine learning, Text mining, deep learning, SVM techniques, and statistical approaches. Whereas, in section IV, we discussed the various datasets that are used for analysis for rumour detection, which are chosen from various sources of SNS sites namely, Facebook, Twitter, LinkedIn, WhatsApp, Sina weibo and Instagram. Subsequently, we have listed the rumour detection tools that are available for detection of rumours.

Lastly, we conclude that most of these detection models fail to predict emerging long-term twitter posts and multilingual rumours. To improve the efficacy of rumour detection rate the trending use of NLP/machine learning technique along with semantic ontology guided with pre-defined knowledge based rumour rules shown in Table II, is integrated into the existing rumour detection approach as discussed in Section V. Ultimately, we propose RDM approach that aid in prediction of rumours from microblogs and generating a report that will be submitted to e-crime department to take appropriate action as per the law. Apart from this RDM will pop-up an alert with a warning message for sending false information to the users of SNS [21].

REFERENCES


AUTHORS PROFILE

Dr. Mohammed Mahmood Ali, completed his PhD in 2017, from Department of CSE, Osmania University, Hyderabad, Telangana, India. He received his Masters degree from Department of CSE, JNTU (Hyderabad), India in 2007. His current research interests include data mining, social web applications, ontology-based information extraction, decision support systems and cyber security. He has presented and published a total of 24 research papers in reputed International journals and International Conferences of IEEE, ACM, Springer and Inderscience. He is an active reviewer of reputed journals like IEE Transactions on Systems, Man, and Cybernetics: Systems, Springer Journal of The Institution of Engineers (India): Series B, International Journal of Electrical and Computer Engineering and many more journals. He is a lifetime member of Institution of Engineers (India).

Dr. Mohammed S. Qaseem, completed his Ph.D.(CSE), from Acharya Nagarjuna University in January 2015, M.Tech in 2007. His current research interests include Data mining, Ontology based information extraction, social web applications, etc. He has presented and published a total of 18 research papers in various International journals and International Conferences of IEEE, ACM and Springer. He is a lifetime member of ISTE and IETE. With more than 27 years of experience in Academics, research, he headed and successfully executed the prestigious TEQIP-II project, sponsored by the World Bank, through the MHRD’s NPIU and SPFU at the State govt. level worth Rs. 4 crores in 2017 while working at NIET. At the current place of work, NSAKCET, apart from heading the CSE dept. and serving as Vice principal (Acad), he is heading the accreditation works for NAAC and NBA as the IQAC coordinator.

Dr. Ateeq ur Rahman, completed his PhD in 2014, from Department of CSE, JNTUH, Hyderabad, Telangana, India. He received his Masters degree from Department of CSE, VTU, Karnataka, India in 2005. His current research interests include data mining, Image processing, Cloud computing. He has published more than 30 research paper in various International Journals and International conferences of IEEE, ACM and Springer. He holds a Life time Fellow membership from Institution of Engineers (India).