I J C T A, 9(40), 2016, pp. 581-585 ISSN: 0974-5572

Time Series Analysis of Rumors Propagation on the Social Networks

Upendra Barthwal[1], Shubham Dobhal[2], Shilpam Rawat[3], Vijay Singh[4,*] and Bhasker Pant[5]

ABSTRACT

According to Social Sciences, rumors are statements or conversation which is not confirmed, circulating from person to person about the specific object or specific event. Rumors are considered as "Misinformation" or "Disinformation". Social Networks works as a catalyst to the instant propagation of events, it leads Government machinery in the big trouble. In this article time series rumor propagation is analyzed using the Twitter. In this research article the post rumor analysis is done. A special case of "Salt shortage rumor" on Nov 11, 2016, is considered, and the results show that most of the tweets done on Nov 11, at around 11:00 pm.

Keywords: Twitter, Rumor, Time Series Analysis, Social Network.

1. INTRODUCTION

Rumors are "misinformation" which propagate from one person to another person about the specific event or specific object. In the era of Social Network, there are a huge chances rumors can affect many population. Sometimes Rumors can create the troublesome problem for the society and difficult to control the situation for the administration. So, online rumor transmission is an open challenge for the Social Network researchers. If the rumors can be detected at the early stage, then many disastrous consequences can be avoided. The similar kind of situation occur when many users uploaded various fake images about the specific events, these fake images and videos are becoming new age rumor. Which could be handled timely and effectively.The Rumor's truth value is unverified or generally false[1]. Emilio et al., [3] proposed a work in which they din the agent-based novel simulation modeling for rumor propagation in Twitter. For this they analyzed and studied two rumor datasets. Vijay Singh et al., [4] proposed a mechanism for event detection and the location detection where the event occur. They proposed event detection algorithm on twitter dataset. Rudra M. Tripathy et al., [5] focused on their work about the strategies on Social Network to control the Rumors. In this article our main concern is about how to detect the initial point from where the rumor started and the points where it is largely affected. Zhe Zhao et al., [6] proposed a very interesting method to early detect of the rumor, they made a cluster of those where, there is some confusion about the truthfulness of the tweets and the cluster of normal tweets. Rumor has shown that rumor analysis creates the pathway for information exchange in the situation like emergency and disastrous condition [7,8,9]. Some researchers worked on attributes of the social network graph[10,11] and their distribution of the rumor popularities [12]. But there is a problem associated with these theories, due to lack of meaningful information it is cumbersome to decide whether to follow power-low or lognormal distribution [13]. Previous studies on rumor suggest rumor behavior with the notion of volume [14,15,16]. The volume can be defined as rumor related tweet in each discrete interval of time. The volume V of the i^{th} rumor r_i at the specific time t, as the number of tweets (rumor-related), m, observed at time t^2[17] shown in equation 1.

[1-5] Department of Computer Science and Engineering, Graphic Era University, 566/6 bell road, Clement Town, Dehradun, Uttarakhand, INDIA, *Emails: upendrabarthwalymail@gmail.com, Shubhamdobhal10@gmail.com, shilpam18081997@gmail.com, vijaysingh_agra@hotmail.com, pantbhaskar2@gmail.com*

$$V_{rit} = m_{rit} \tag{1}$$

Linear Filtering of Time Series: Time series models are generally considered as one series is the filtered form of another series.Linear Filtering methods are heavily used in statistical analysis, data analysis, bi-technology and mechanical engineering.One of the main concept of traditionally used time series is the decomposition of series X_t in a pattern T_t, S_t Seasonal component and the remainder e. The general procedure for obtaining the pattern is to use linear filter on given time series [2], shown in equation 2 and 3:

$$T_t = \sum_{i=-\infty}^{\infty} \lambda_i X_{t+i} \tag{2}$$

The Basic change of the linear filters are moving averages with equal weights:

$$T_t = \frac{1}{2a+1} \sum_{i=-a}^{a} X_{t+i} \tag{3}$$

The rest of the paper is organized as follows.Data Collection and Preprocessing in section II. Proposed Methodologyin section III. Experimental results are presented in section IV. Concluding remarks are given in section IV.

2. DATA COLLECTION AND PREPROCESSING

Data is generated from the twitter using the R program and Twitter Archiver using the various filters related the Salt shortage rumor. Twitter is the very useful source for collecting information, and Twitter Streaming API is Utilized. The main reason behind the used of twitter dataset is the diversity of its user. Around twenty thousand are collected, containing meaningful information as well as some noise. Then apply basic noise removal technique for pre-processing of data.

Sample Tweets are shown in Table 1.

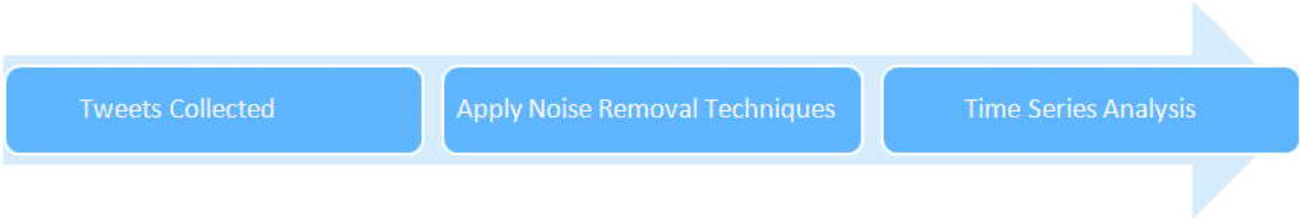

Table 1
Sample Tweets related salt shortage rumor.

Sample Tweets (Before pre-processing)
þ@IndianExpress Nov 11 Salt shortage rumour creates panic in Western UP
Seven Sisters þ@7SistersProjekt 19 Nov 2013#Impact- We knew the reason behind the sudden escalation in salt prices - http://bit.ly/185IBgJ #SaltRumor #FoodSuppy
while #India is a proven argumentative country, its rumor-mongering is peaking. seems to be market share or profit motived #SaltRumor
Ashish Srivastava þ@ashish_7790 Nov 11 Faizabad, India It's a rumor do not spread it .. Make law and order ..and maintain peace ... #SaltRumor
@IndianExpress the same rumours was there couple of years ago in our city Imphal spreading from up & I was one of fools buying @120pkg

Sample Tweets (After pre-processing)

Salt shortage rumour creates panic in Western UP
We knew the reason behind the sudden escalation in salt prices
India is a proven argumentative country, its rumor-mongering is peaking. seems to be market share or profit motived
India It's a rumor do not spread it . Make law and order ..and maintain peace
the same rumours was there couple of years ago in our city Imphal spreading from up I was one of fools buying 120pkg

3. PROPOSED METHODOLOGY

Time Series analyis of salt shortage rumor –

After the demonetization move by the government, salt shortage rumor panic the people on India at various geographical locations. The analysis is done on the tweets collected related to salt shortage rumor. The main dimensions of the analysis are:

- Starting Point of the rumor.
- Density point on the Map where Rumor affect the huge number of people.
- The highest propagation time.

At the time of tweet collation, filters are applied to extract other attributes of the tweets, like Timestamp, Location of tweet and the content. Timestamp is used to find when the numbers of tweets are increased exponentially, in this specific case it is around 11p.m. on 11, Nov shown in the figure 1. Location and Timestamp is used to detect the initial location from where the rumor started on salt shortage, in this case it is Western UP. For mapping the Location Google Fusion tables are used. Google fusion tables can be used to visualize the data and compatible with the Google Map as well. Mapping with the data is used to find the density points where the rumor is propagated again and again, shown through heatmap in figure 3.

4. EXPERIMENT AND RESULT

For Demonstration purpose Twitter Dataset is used. R and Python program are used to preprocessing and time series analysis. The PC for experiment is equipped with an Intel P4 2.4GHz Personal laptop and 4GB memory

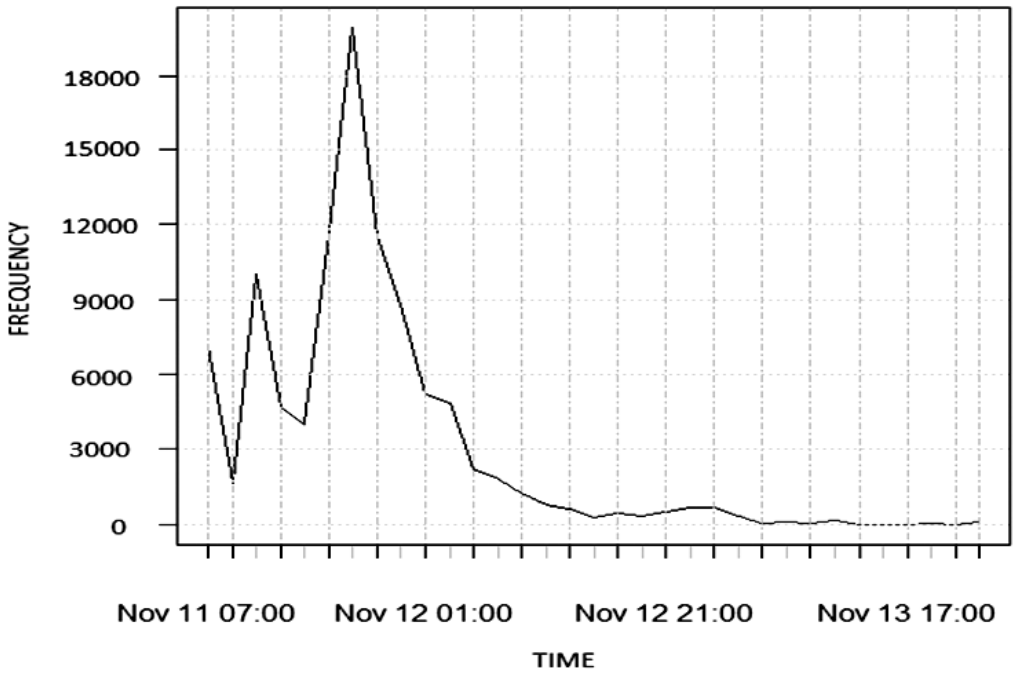

Figure 1: Time Vs. Frequency of Tweets

Figure 2: Locations on the map where rumors are detected.

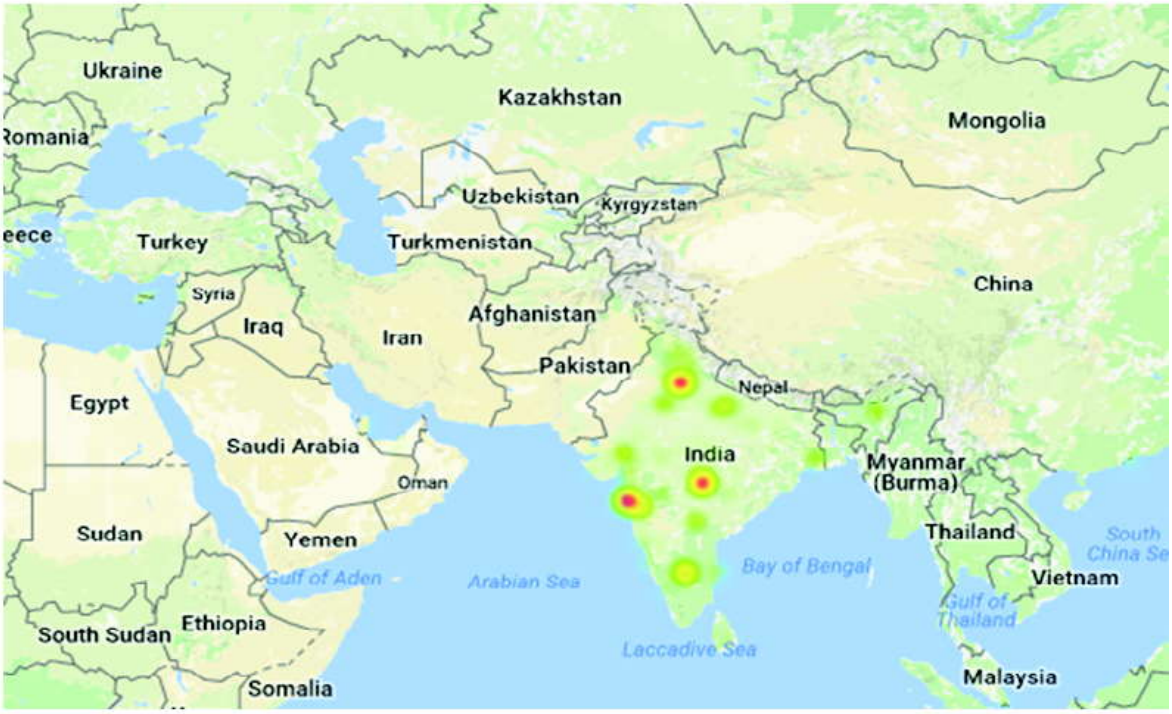

Figure 3: Heatmap representing Highest Density point of Rumor

From the experiment results, we reached to the conclusion that this proposed method can be used to identify the density points where the tweets are propagated again and again, initial point of the rumor. This analysis can be performed to other similar kind of applications.

4. CONCLUSION

In this research the basic time series analysis is done for examining the various points about the rumor. The main focused area of the research are starting point of the rumor, the density area of the rumor and the time interval where most tweets are done. In the future work , would try to develop a system for detecting the rumor and automatic post rumor analysis.

REFERENCES

[1] Xiaomo Liu, Armineh Nourbakhsh, Quanzhi Li, Rui Fang, Sameena Shah, "Real-time Rumor Debunking on Twitter",CIKM'15 International Conference on Information and Knowledge Management, ACM, pp. 1867-1870.

[2] Walter Zucchini, lain L.MacDonald, "Hidden Markov Models for Time Series: An Introduction Using R", Monographs on Statistics and Applies Probability, CRC Press.

[3] Emilio Serrano, Carlos Angel Iglesias and Mercedes Garijo, "A Novel Agent-Based Rumor Spreading Model in Twitter", International Conference on World Wide Web, ACM, pp. 811-814.

[4] Vijay Singh, Bhasker Pant and Devesh Pratap Singh, "Identifying the Event type and Event Location using Sentiment Analysis on Tweets", International Journal of Control Theory and Applications, 9(19), pp.9101-9105.

[5] Rudra M. Tripathy, Amitabha Bagchi and Sameep Mehta,"A study of Rumor control startegies on Social Networks", International Conference of Information and Knowledge Management, ACM, pp. 1817-1820.

[6] Zhe Zhao, Paul Resnick and Qiaozhu Mei, "Enquiring Minds: Early Detection of Rumors in Social Media from Enquiry Posts", International Conference on World Wide Web, ACM, pp. 1395-1405.

[7] Scanlon, T.Post-disaster rumor chains: A case study. Mass Emergencies 2, 126 (1977), 22-27.

[8] Erickson, B., Nosanchuk, T., Mostacci, L., and Dalrymple, C. , "The flow of crisis information as a probe of work relation", Canadian Journal of Sociology (1978), pp.71-87.

[9] Starbird, K., and Palen, L.,"Pass It on?:Retweeting in Mass emergency", International ISCRM Conference (2010). pp.1-10.

[10] R. Ghose and B.A.Huberman,"Ultrametricity of Information cascades" CoRR, abs/1310.2619, 2013.

[11] M.Barthelemy, A.Barrat, and A. Vespignami, "The role of geography and traffic in the structure of Complex networks".Advances in Computer Systems, 10(1):5-28,2007.

[12] J.L. Iribarren and E.Moro, "Branching dynamics of Viral information spreading". CoRR, abs/1110.1884, 2011.

[13] Andrzej Pacuk, Piotr Sankowski and Karol Wegrzycki, "There is Something Beyond the Twitter Network",ACM Conference and Social Media, pages 279-284.

[14] Kwon Kwon, Meeyoung Cha, Kyomin Jung, Wei Chen and Yajun Wang, "Prominent features of rumor propagation in online social media", IEEE 13th International Conference on Data Mining(ICDM) 1103-1108 (2013).

[15] Jim Maddock, Kate Starbird, Haneen Al-Hassani, Daniel E. Sandoval, Mania Orand, and Robert M. Mason, "Characterizing Online Rumorning Behaviour Using Multi_dimensional Signatures", 18th ACM Computer Supported Cooperative Work and Social Computing(CSCW'15).

[16] Marcelo Mendoza, Barbara Poblete, and Carlos Castillo, "Twitter Under Crisis: can we trust what we RT?", workshop on Social Media Analytics, 9, 2010.

[17] Ahmer Arif, Kelley Shanahan, Fang-Ju Chou, Yoanna Dosouto, Kate Starbird, Emma S. Spiro, "How Information Snowballs: Exploring the role of Exposure in Online Rumor Propogation" , ACM Conference on Computer-Supported Cooperative Work and Social Computing, pp. 466-477.