Social Monitoring System for Dynamically Evolving Anomalies over Text Streams

Pranali Ratnaparkhi[1], Rohini Jadhav [2], Prakash Kshirsagar [3]

[1, 2] Student, Computer Dept., ZCOER, Savitribai Phule Pune University, India.

[3] Asst.Prof, Computer Dept., ZCOER, Savitribai Phule Pune University, India.

Abstract – A blog having a small group of members with their discussion. In a blog that is smaller than a traditional blog and contains very short entries is called as microblog. Using this real-time diffusion of information is going on. So in that so many abnormal discussion on microblog, this things are trending on the social network and be able to monitor their evolution and find related anomalies. So in this paper we proposed a system RING (real-time emerging anomaly monitoring system over microblog text streams). To find out and monitor this abnormal discussion, RING proposed a graph analytic approach such having some advantages like 1)RING is detect current trends emerging anomalies topics, 2)RING is among the first to discover emerging anomalies correlations in a streaming fashion, 3)RING works on the real-time data that work on the minutes to months data. RING is able to process big data to the entire Weibo or Twitter text stream with linear horizontal scalability.

Index Terms — RING, Twitter text stream, blog and microblogs, emerging, anomalies.

1. INTRODUCTION

Social media are interactive Web-2.0 based applications like Facebook and Twitter etc. We are the social media having media having positive and negative impacts. Social media can help to improve individuals' sense of connectedness with real or online communities and social media can be an effective communication (or marketing) tool for corporations, entrepreneurs, non-profit organizations, including advocacy groups and political parties and governments. A popular component and feature of Twitter is retweeting. Twitter allows other people to keep up with important events, stay connected with their peers, and can contribute in various ways throughout social media. Retweeting is beneficial strategy, which notifies individuals on Twitter about popular trends, posts, and events. On the basis of these popular trends some abnormal things are happen so in this paper we are introducing RING System, that managing Real-Time Emerging Anomaly Monitoring System.

Now a day's we seen the trends of social media, while using the social media like Facebook, twitter, user post their opinion. The business class people used social media like Twitter, in that they trying to do business, so in that each and every twits or post is important. In this we are monitoring the online incoming text data, and find out the Emerging topics with abnormal things.

Microblog platforms have been extremely popular in the big data era due to the real-time nature and viral diffusion of information. Big data is a term for data sets that are so large or complex that traditional data processing application software is inadequate to deal with them. Big data challenges include capturing data, data storage data analysis. Search, sharing, transfer, visualization, querying, updating and information privacy.

2. RELATED WORK

Using keyword co-occurrence, we review related work about topic detection and emerging event detection. It has long been recognized that modeling topics or events based on keyword co-occurrence is an effective approach. For term clustering and keyword removal from documents Co-occurrence information has been used. A short text topic model that directly models the generation of word co-occurrences pattern has been proposed. In contrast, our approach adopts a unified graph processing framework through each phase and meets all the listed semantic requirements. Detecting emerging events basically requires recomputing clustering from scratch while the efficiency of largely exploits incremental computation. For Twitter, this method would generate many trends that only contain a single keyword, which is hard to understand.

In this paper we are using twitter dataset having more than 100000 user's twits.

2.1. Existing System

We are going to finds emerging trends, So RING monitoring on abnormal trend going on currently. RING system a graph analytic approach such have some step to find the abnormal step, RING is detect current trends emerging anomalies topics, RING is among the first to discover emerging anomalies correlations in a streaming fashion, RING works on the real-time data that work on the minutes to months data. RING is able to process big data to the entire Weibo or Twitter text stream with linear horizontal scalability.

2.2 Methods

The new user register on the social network like twitter and then login. The user here has to post their comment on the social network and be able to monitor their evolution and find

related anomalies and the system then retrieves the anomaly thing of that particular post from database which the user has posted and these posts are displayed in the list. And detect anomaly event in trends which store into database. So, for every post there is a different list and there is also facility for user to view all the post. Here the user can see the post in their system.
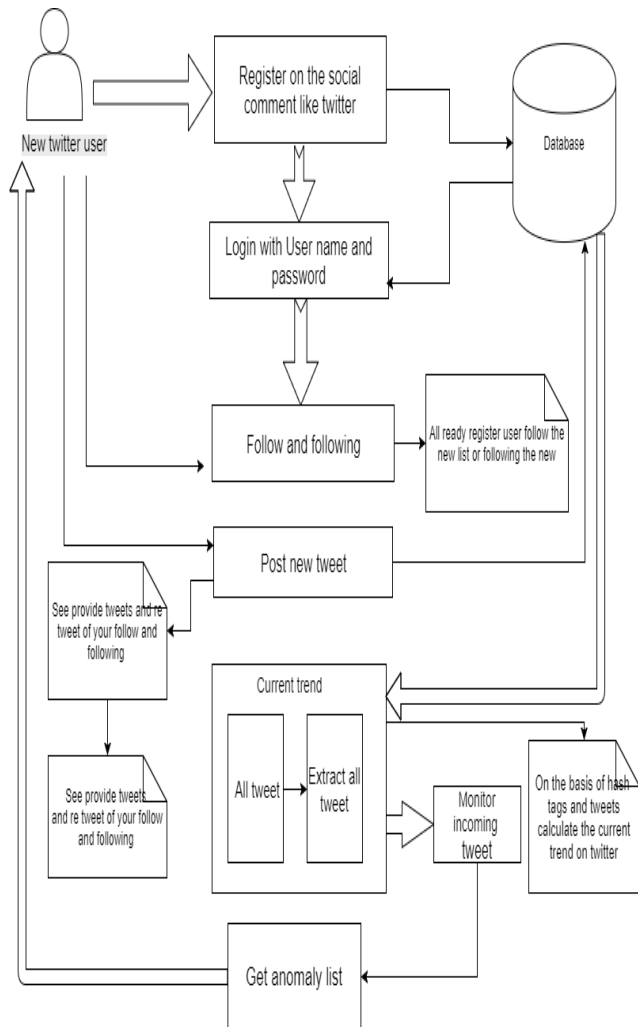
3. PORPOSED MODELLING

Fig: System Framework

3.1 Product Functionality

The prime focus of this system is to analyze the post to their tweets on social network. We are applying our system on social media like Twitter, Twitter have some current trends with his hash tags and current topics. And we will use those current topics and hashtags to detect trending topics which will have anomalies. Using this real-time diffusion of information is going on. So, in that so many abnormal discussions on microblog, this thing are trending on the social network and be

able to monitor their evolution and find related anomalies. So, in this paper we proposed a system RING (real-time emerging anomaly monitoring system over microblog text streams).

3.2 System Workflow

The new user register on the social network like twitter and then login. The user here has to post their comment on the social network and be able to monitor their evolution and find related anomalies and the system then retrieves the anomaly thing of that particular post from database which the user has posted and these posts are displayed in the list. And detect anomaly event in trends which store into database. So, for every post there is a different list and there is also facility for user to view all the post. Here the user can see the post in their system.

3.3 Hardware Interface

Processor – Pentium –III

Speed – 2.4 GHz

RAM - 256 MB (min)

Hard Disk - 20 GB

3.4 Software Interface

Operating System: Windows 95/98/2000/XP

Frontend: HTML, JSP, CSS

Backend: MySQL

JDK 1.8

3.5 Statement of Scope

This application can be used to detect anomalous event whenever some user posts any tweets related to their trending topic. In this project we are going to form graph that are related to all tweets distributed with an efficient and linear scalable properties. The proposed RING system is more useful to find out real time incoming tweets (text stream) to detect anomaly.

4. RESULTS AND DISCUSSIONS

4.1 Equations

Step 1: To form undirectional graph of G(t),

$G(t) = (N(t),A(t))\ldots 1$

Where, N(t)- Set of all distinct nodes in G(t) at time t,

A(t)- As the sequence of edges received so far.

Step 2: Decay weight: To calculated arrival time of incoming data

Time $t = 1/\lambda \ldots 2$

Step 3: Weighted Frequency: Accumulated decay weights over all instances of its arrivals till time t.

$$F(i,j,t) = \sum_{k=1}^{n_{ij}^t} N(i,j,T(i,j,k)) \cdot 2^{-\lambda \cdot (t-T(i,j,k))}$$

….3

Step 4: Node Activity Frequency: Sum of the edge frequencies.

$$\alpha(i,t) = \sum_{k=j_1^i(t)}^{j_{|S(i,t)|}^i(t)} F(i,k,t)$$

….4

Step 5: Denoised Keyword Graph): Is a directed weighted graph

GE(t) = (VE(t),EE(t)), VE(t)….5

Where, EE(t)- Contain edges among all keywords with their weighted frequency.

Step 6: Trending Keyword Graph: A binary graph

GT (t) = (VT (t), ET (t)), ….6

where VT (t) - contains only trending keywords.

 ET (t) - Edges between trending keywords.

Step 7: Trending Event: To calculate the current trending event, we used scalability and event denoising.

4.2 Naive Bayes

Naive Bayes is applied to differentiate news, ads and wisdom words among detected trending events, where the latter two classes are major types of spams on Weibo.

The classifier is trained with manually labeled data depends on features of content, users and temporal information.

4.3 Web Based GUI

Server will be web based application and this module will be responsible to take inputs from admin. The GUI is developed in HTML and Java input will be taken through this GUI where proper validations are supported. This includes new topic and post upload, etc.

4.4 Methodology

The hardware design of the system includes designing the hardware units and the interface between those units.

A social networking service is an online platform which people use to build social relation with other people who share similar personal or career interests, activities, backgrounds or real-life connections. Social networking services are Internet-based applications. In our system it's important to know what anomalous events are trending on the social network and be able to monitor their evolution and find related anomalies. People use the hashtag symbol (#) before a relevant keyword or phrase in their tweet to categorize those Tweets and help them show more easily in Twitter search. Once you have followers on Twitter, you should try to build a positive relationship with them to grow your business. For this, you should make an effort to post updates and tweets that your customers find interesting and useful. A Twitter message can be retweeted any number of times, which means any useful or interesting message you post can be passed on to a number of Twitter users in very little time. But sometimes some user post something which is irrelevant to that topic of discussion. In such a situation we consider that event as an anomaly.

- Input: Tweets

- Output: Get anomaly list.

    1. User posts any tweets related to their trending topic.

    2. Find the anomaly things in current trends

    3. And then generate the graph of that anomaly things

    4. Get the list of that anomaly things.

4.5 Features

1. User Registration with interested topic user can easily post and read the comment related to their topic.

2. Login with user name and password.

3. Admin Login

4. Admin registration

5. Post new tweet on related topic and see provide tweets and retweet of your follower and following

6. Already registered user follow the new user in the list or following the new user.

7. Monitoring incoming tweets.

8. On the basis of hashtags and tweets we calculate the current trend on twitter and then   detect abnormal things.

5. CONCLUSION

In this paper, the advantages RING system is using the existing and proposed system using the monitoring the abnormal emerging event task. Further, RING's infrastructure is equipped with customized optimization on its full-text search

engine and distributed graph processing engine to perform event monitoring more efficiently.

5.1 Applications

This system varies post of user for detecting the anomalies. This system provides a user-name and password for verification of application user. The system provides tweets to the user's topic or post. It also verifies the post generated by the system manually. This application run on web environment, when we are host our application on server the internet is required. The main input of system is tweets that are come at real time. We want to manage this tweets and fetch from database like MySQL. For verification we are using mail API so internet is required. The system RING is more useful to finding real time incoming tweets (text stream) to detect anomaly. In that we are forming the module like current trending detection and the real-time event detection on the basis of this things follows detect procedure.

REFERENCES

[1] T.Sakaki, M. Okazaki, and Y.Matsuo, "Earthquake shakes twitter users: real-time event detection by social sensors," in www,2010.

[2] R. McCreadie, C. Macdonald, I. Ounis, M. Osborne, and S. Petrovic, "Scalable distributed event detection for twitter," in IEEE BigData, 2013.

[3] Y. Chen, H. Amiri, Z. Li, and T.-S. Chua, "Emerging topic detection for organizations from microblogs," in SIGIR, 2013.

[4] W. Xie, F. Zhu, J. Jiang, E.-P. Lim, and K.Wang, "Topicsketch: Realtime bursty topic detection from twitter," in ICDM, 2013.

[5] E. Schubert, M. Weiler, and H.-P. Kriegel, "Signitrend: scalable detection of emerging topics in textual streams by hashed significance.

[6] F. Wei, S. Liu, Y. Song, S. Pan, M. X. Zhou, W. Qian, L. Shi, L. Tan, and Q. Zhang, "Tiara: a visual exploratory text analytic system," in KDD, 2010.

[7] P. Lee, L. V. Lakshmanan, and E. E. Milios, "Incremental cluster evolution tracking from highly dynamic network data," in IEEE International Conference on Data Engineering (ICDE), 2014, pp. 3–14.

[8] C. Li, A. Sun, and A. Datta, "Twevent: segment-based event detection from tweets," in CIKM, 2012.

[9] D. Metzler, C. Cai, and E. Hovy, "Structured event retrieval over microblog archives," in HLT-NAACL, 2012.

[10] C. Budak, T. Georgiou, and D. A. A. El Abbadi, "Geoscope: Online detection of geo-correlated information trends in social networks," PVLDB, vol. 7, no. 4, 2013.

[11] J. Allan, R. Papka, and V. Lavrenko, "On-line new event detection and tracking," in SIGIR, 1998.

[12] T. Hofmann, "Probabilistic latent semantic analysis," in UAI, 1999.

[13] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," the Journal of machine Learning research, vol. 3, pp. 993–1022, 2003.

[14] D. M. Blei and J. D. Lafferty, "Dynamic topic models," in ICML, 2006.

[15] X. Yan, J. Guo, Y. Lan, and X. Cheng, "A biterm topic model for short texts," in WWW, 2013.

Authors

Pranali Ratnaparkhi Student of Zeal College of Engineering and Research & study in BE computer.

Rohini Jadhav Student of Zeal College of Engineering and Research & study in BE computer.

Prof. Prakash Kshirsagar Professor of Zeal College of Engineering and Research with 8.6 yrs. Teaching Experience.