RESEARCH ARTICLE OPEN ACCESS

Detection and Analysis of Twitter Trending Topics via Link-Anomaly Detection

Chandan M G*, Chandra Naik**
*(PG Scholar, Department of Computer Science Engineering, NMAMIT, Nitte, Udupi-574 110, India)
** (Department of Computer Science Engineering, NMAMIT, Nitte, Udupi-574 110, India)

ABSTRACT
This paper involves two approaches for finding the trending topics in social networks that is key-based approach and link-based approach. In conventional key-based approach for topics detection have mainly focus on frequencies of (textual) words. We propose a link-based approach which focuses on posts reflected in the mentioning behavior of hundreds users. The anomaly detection in the twitter data set is carried out by retrieving the trend topics from the twitter in a sequential manner by using some API and corresponding user for training, then computed anomaly score is aggregated from different users. Further the aggregated anomaly score will be feed into change-point analysis or burst detection at the pinpoint, in order to detect the emerging topics. We have used the real time twitter account, so results are vary according to the tweet trends made. The experiment shows that proposed link-based approach performs even better than the keyword-based approach.
Keywords **-** anomaly-detection, social network, change-point analysis, burst detection.

I. INTRODUCTION

A. Introduction to Social network and Twitter:

Nowadays Social network has become one of the most important aspects in our daily life. Social network is a network of social interactions and personal relationships. And also dedicated website or application which enables users to create and information exchanged over social networks is not only texts but also URLs, images, and videos. Some of the biggest social networks used today like Face book, Twitter, Google+, LinkedIn etc.[1] …

Another type of information (i.e., intentionally or unintentionally) exchanged through social networks: mentions. Mentions means links to other users of the same social network stream in the form of message-to, reply-to, retweets-of.

Twitter: Twitter is an online social networking service that enables users to create application, it sends and read 140-character messages called "tweets". Only the registered users can read tweets and post tweets, but the unregistered users can only read tweets. Users may subscribe to other users known as following and subscribers are known as followers. In twitter, Replies and Mentions are two ways for twitter user to exchange ideas [2].

B. Anomaly-detection:

Anomaly Detection is a pattern in the data that does not conform to the expected normal behaviour. And also referred to as outliers, exceptions, peculiarities, surprise etc. some applications of anomaly detection [3].

- Cyber intrusions
- Credit card fraud

Cyber intrusion is the unauthorized act of spying, snooping, and stealing information through cyber space. Credit card fraud means purpose may be to obtain goods without paying, or to obtain unauthorized funds from an account.

II. RELATED WORK AND ITS LIMITATIONS

A new (trending) topics is something people feel like discussing, commenting the information further to their friends. Conventional key-based approaches for topic detection have mainly focus on frequencies of (textual) words [4].

A key-based approach could suffer from the ambiguity caused by synonyms. It may also require complicated pre-processing (e.g., segmentation). It cannot be applied when the contents of the messages are mostly no textual information. Another way, words formed by mentions are unique, require little pre-processing to obtain.

III. PROPOSED SYSTEM AND ITS ADVANTAGES

The anomaly detection in the twitter data set is carried out by retrieving the trend topics from the twitter in a sequential manner by using some API and corresponding user for training. Then computed anomaly score is aggregated from different users. Further the anomaly score will be feed into change-point analysis or burst detection at the pinpoint, in order to detect the emerging topics [5]. System architecture for detecting emerging topics in twitter is

as shown in Fig.1. In this architecture the detecting topics involves five modules.

Twitter trends: In this module use Twitter REST (**Re**presentational **S**tate **T**ransfer) API, this API is to collect the top 10 trending topics for specific WOEID (Where on Earth ID), if trending information is available for twitter. Then trending topics are in the form of JSON format. The twitter API calls Trends=https://api.twitter.com/1.1/trends/place.json?id=23424848

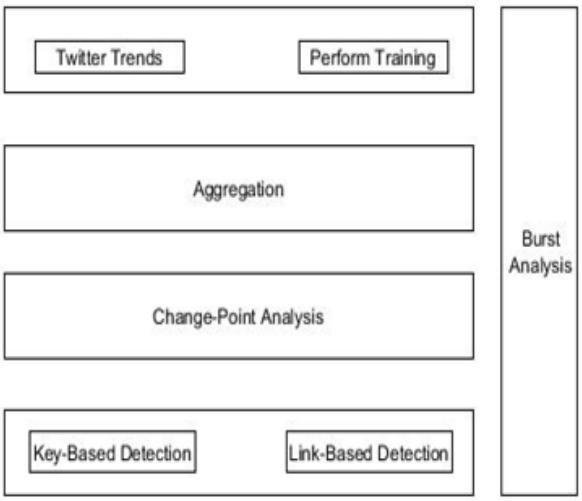

Fig.1 Detecting System Architecture.

Perform training*:* In this module use Twitter Search API is the part of Twitter's v1.1 REST (**Re**presentational **S**tate **T**ransfer) API. It allows queries against the recent or most popular Tweets and behaves similarly. It's important note that the Search API is focused on relevance and not completeness. This means that some Tweets and users may be missing from search results. The twitter search API call.
search=https://api.twitter.com/1.1/search/tweets.json?q

Parameters:

q (Required) ***The search query to run against people search.***

Example Values: q=Bangalore Blast.

Aggregation*:* In this module, to combine the anomaly scores from different of hundreds users. From the previous data, the anomaly score is computed for each user depending on the current post of user u and his/her past behaviour. We propose a aggregated anomaly scores obtained for posts X1,X2, X3.....XN using window size $\lambda>0$ [5].

In aggregation of key-based approach to find out time in and time out, whereas link-based approach to find out time in and time out.

Change-Point analysis*:* In this module, using previous aggregated anomaly score is applied to change-point detection. It detects change point from the aggregated anomaly scores [6].

Burst detection: The burst-detection method is based on a probabilistic automaton model. The algorithm aims to analyze documents to find features that have high intensity over finite/limited durations of time periods [7].

Advantages*:* The proposed approach does not confidence on the textual contents of social streams posts, it is a strong to rephrasing and it can be applied where topics are focused with information other than texts, such as images, videos, URLs audio, and so on. The proposed link-based approach performs even better than the keyword-based approach.

IV. RESULTS

To detect the emerging topics in twitter. We have used the real time twitter account, so results are vary according to the tweet trends made. Our experiment compare that the proposed link-based detection approach against a key-based detection approach. We have considered real twitter trends topics for compare the results which one best performance. Fig 2 the result of #MSD2011 trend topics, proposed link-based approach detect much earlier than keyword-based approach. So proposed link-based approach is best. Fig 3 the result of 14th Indian trend topics, proposed link-based approach detect much earlier than keyword-based approach. So proposed link-based approach is best. Fig 4 the result of 8th ODI trend topics, proposed link-based approach detect much earlier than keyword-based approach. So proposed link-based approach is best. To observe the above three possible results proposed link-based approach performs even better than the keyword-based approach. And the proposed approach is highly scalable.

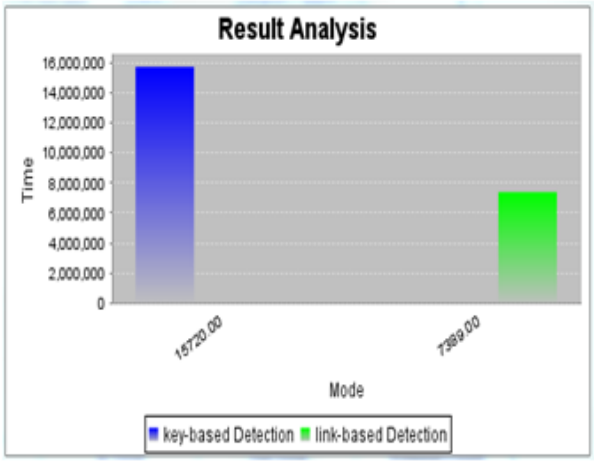

Fig 2: #MSD2011 trend topics

V. CONCLUSIONS

In this paper, we have proposed link-based approach to detect the trending of topics in twitter social network. The basic idea of link-based approach is to focus on the post reflected in the mentioning behaviour of hundreds user instead of the textual contents. The anomaly detection in the twitter data set is carried out by retrieving the trend topics from the twitter in a sequential manner by using some API and corresponding user for training. Then computed anomaly score is aggregated from different users. Further the anomaly score will be feed into change-point analysis or burst detection at the pinpoint, in order to detect the emerging topics. We have collect data from real time twitter account, so results vary according to the tweet trends made. Finally, to gives the better performance results when compared with the existing system.

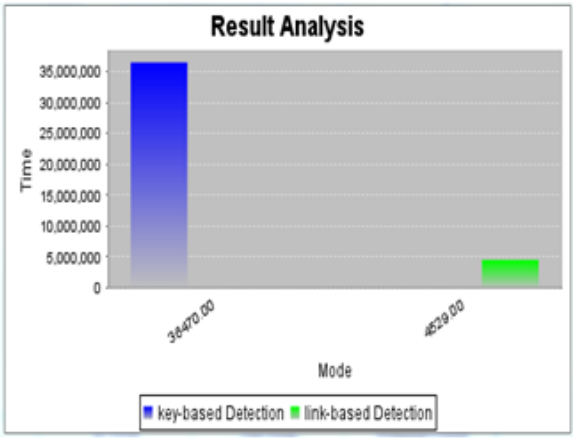

Fig 3: 14th Indian trend topics

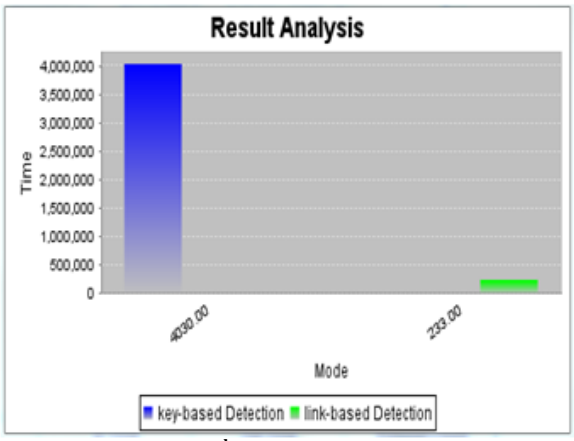

Fig 4: 8th ODI trends topics

REFERENCES

[1] D. Boyd and N. B. Ellison, "*Social Network Sites: Definition, History, and Scholarship*," Journal Computer-Mediated Communication , vol.13, no. 1-2, Nov. 2007.

[2] Gabriel Weimann, "*Terror on Facebook, Twitter, and Youtube*," The Brown Journal of World Affairs, volume 16, issue 2, 2010.

[3] Yang Li, Bin-Xing Fang, "*A Lightweight Online Network Anomaly Detection Scheme Based on Data Mining Methods*," International Conference on Network Protocols, pp.340-341, 16-19 Oct. 2007.

[4] J. Allan et al., "*Topic Detection and Tracking Pilot Study: Final Report*," Proc. DARPA Broadcast News Transcription and Understanding Workshop, 1998.

[5] T. Takahashi, R. Tomioka, and K. Yamanishi, "*Discovering emerging topics in social streams via link anomaly detection*," arXiv: 1110.2899v1 [stat.ML], Tech. Rep., 2011.

[6] J. Takeuchi and K. Yamanishi, "*A Unifying Framework for Detecting Outliers and Change Points from Time Series*," IEEE Trans. Knowledge Data Eng., vol. 18, no. 4, pp. 482-492, Apr. 2006.

[7] J. Kleinberg, "*Bursty and Hierarchical Structure in Streams*," Data Mining Knowledge Discovery, vol. 7, no. 4, pp. 373-397, 2003.