International Journal of Advance Research in Science and Engineering
Vol. No.5, Issue No. 05 , May 2016
www.ijarse.com
IJARSE
ISSN 2319 - 8354

RECOGNIZING THE PROCEEDING THEME IN POPULAR STREAM BY MEANS OF ASSOCIATION-INCONSISTENCY AUTHENTICATION

D. Kiran Kumar[1] , Ambala Srinivas[2] , Yadaiah Madgula[3]

[1,2,3] *Research Scholar, Sunrise University, Alwar Mewar University, RJ Dept of CSE, GNITC ,(India)*

ABSTRACT

Detection of emerging topics is now receiving renewed interest motivated by the rapid growth of social networks. Conventional-term-frequency-based approaches may not be appropriate in this context, because the information exchanged in social-network posts include not only text but also images, URLs, and videos. We focus on emergence of topics signalled by social aspects of these networks. Specifically, we focus on mentions of user links between users that are generated dynamically (intentionally or unintentionally) through replies, mentions, and retweets. We propose a probability model of the mentioning behaviour of a social network user, and propose to detect the emergence of a new topic from the anomalies measured through the model. Aggregating anomaly scores from hundreds of users, we show that we can detect emerging topics only based on the reply/mention relationships in social-network posts. We demonstrate our technique in several real data sets we gathered from Twitter. The experiments show that the proposed mention-anomaly-based approaches can detect new topics at least as early as text-anomaly-based approaches, and in some cases much earlier when the topic is poorly identified by the textual contents in posts.

Keywords: Authentication, aggregation, knowledge representation, segmentation, streaming media.

I. INTRODUCTION

Communication over social networks, such as Face book and Twitter, is gaining its importance in our daily life. Since the information exchanged over social networks are not only texts but also URLs, images, and videos, they are challenging test beds for the study of data mining. In particular, we are interested in the problem of detecting emerging topics from social streams, which can be used to create automated “breaking news”, or discover hidden market needs or underground political movements. Compared to conventional media, social media are able to capture the earliest, unedited voice of ordinary people. Therefore, the challenge is to detect the emergence of a topic as early as possible at a moderate number of false positives. Another difference that makes social media social is the existence of mentions. Here, we mean by mentions links to other users of the same social network in the form of message-to, reply-to, retweet-of, or explicitly in the text. One post may contain a number of mentions. Some users may include mentions in their posts rarely; other users may be mentioning their friends all the time. Some users (like celebrities) may receive mentions every minute; for others, being mentioned might be a rare occasion. In this sense, mention is like a language with the number of words equal to

the number of users in a social network. We are interested in detecting emerging topics from social network streams based on monitoring the mentioning behavior of users. Our basic assumption is that a new (emerging) topic is something people feel like discussing, commenting, or forwarding the information further to their friends. Conventional approaches for topic detection have mainly been concerned with the frequencies of (textual) words. A term-frequency-based approach could suffer from the ambiguity caused by synonyms or homonyms. It may also require complicated preprocessing (e.g., segmentation) depending on the target language. Moreover, it cannot be applied when the contents of the messages are mostly non textual information. On the other hand, the "words" formed by mentions are unique, require little preprocessing to obtain (the information is often separated from the contents), and are available regardless of the nature of the contents. System shows an example of the emergence of a topic through posts on social networks. The first post by Bob contains mentions to Alice and John, which are both probably friends of Bob, so there is nothing unusual here. The second post by John is a reply to Bob but it is also visible to many friends of John that are not direct friends of Bob. Then in the third post, Dave, one of John's friends, forwards (called retweet in Twitter) the information further down to his own friends. It is worth mentioning that it is not clear what the topic of this conversation is about from the textual information, because they are talking about something (a new gadget, car, or jewelry) that is shown as a link in the text.

II. RELATED WORK

In this system, we propose a probability model that can capture the normal mentioning behavior of a user, which consists of both the number of mentions per post and the frequency of users occurring in the mentions. Then this model is used to measure the anomaly of future user behavior. Using the proposed probability model, we can quantitatively measure the novelty or possible impact of a post reflected in the mentioning behavior of the user. We aggregate the anomaly scores obtained in this way over hundreds of users and apply a recently proposed change point detection technique based on the sequentially discounting normalized maximum-likelihood (SDNML) coding. This technique can detect a change in the statistical dependence structure in the time series of aggregated anomaly scores, and pinpoint where the topic emergence is; see in this system. The effectiveness of the proposed approach is demonstrated on four data sets we have collected from Twitter. We show that our mention-anomaly-based approaches can detect the emergence of a new topic at least as fast as text-anomaly-based counterparts. Furthermore, we show that in three out of four data sets, the proposed mention-anomaly-based methods can detect the emergence of topics much earlier than the text-anomaly-based methods, which can be explained by the keyword ambiguity we mentioned above. A new (emerging) topic is something people feel like discussing, commenting, or forwarding the information further to their friends. Conventional approaches for topic detection have mainly been concerned with the frequencies of (textual) words. A term-frequency-based approach could suffer from the ambiguity caused by synonyms or homonyms. It may also require complicated preprocessing (e.g., segmentation) depending on the target language. Moreover, it cannot be applied when the contents of the messages are mostly nontextual information. On the other hand, the "words" formed by mentions are unique, require little preprocessing to obtain (the information is often separated from the contents), and are available regardless of the nature of the contents. In this paper, we have proposed a new approach to detect the

emergence of topics in a social network stream. The basic idea of our approach is to focus on the social aspect of the posts reflected in the mentioning behaviour of users instead of the textual contents. We have proposed a probability model that captures both the number of mentions per post and the frequency of mentioned. The proposed method does not rely on the textual contents of social network posts, it is robust to rephrasing and it can be applied to the case where topics are concerned with information other than texts, such as images, video, audio, and so on. The proposed link-anomaly-based methods performed even better than the keyword-based methods on "NASA" and "BBC" data sets.

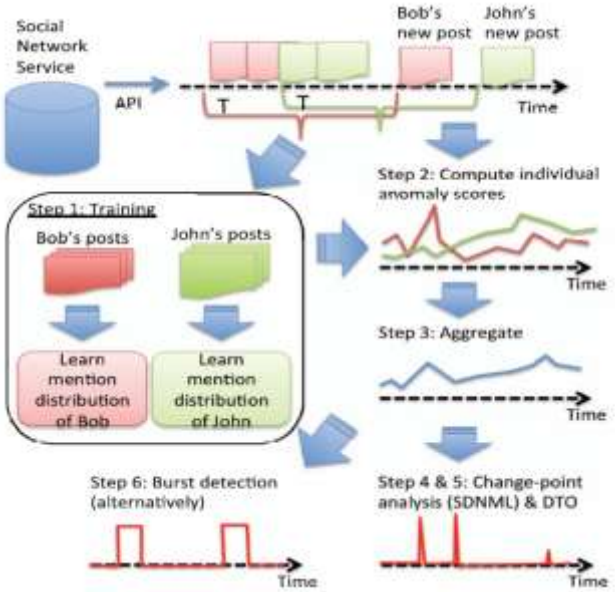

Fig 1 : System Architecture:

III. SYSTEM PREELIRIES

A. Admin

In this module, the Admin has to login by using valid user name and password. After login successful he can do some operations such as search history, view users, request & response, all topic messages and topics.

B. Search History

This is controlled by admin; the admin can view the search history details. If he clicks on search history button, it will show the list of searched user details with their tags such as user name, searched user, time and date.

C. Users

In user's module, the admin can view the list of users and list of mobile users. Mobile user means android application users.

D. Request & Response

In this module, the admin can view the all the friend request and response. Here all the request and response will be stored with their tags such as Id, requested user photo, requested user name, user name

request to, status and time & date. If the user accepts the request then status is accepted or else the status is waiting.

E. Topic Messages

In this module, the admin can view the messages such as emerging topic messages and Anomaly emerging topic messages. Emerging topic messages means we can send a message to particular user. Anomaly emerging topic message means we can send message on a particular topic to all users.

F. User

In this module, there are n numbers of users are present. User should register before doing some operations. And register user details are stored in user module. After registration successful he has to login by using authorized user name and password. Login successful he will do some operations like view or search users, send friend request, view messages, send messages, anomaly messages and followers.

G. Search Users

The user can search the users based on users and the server will give response to the user like User name, user image, E mail id, phone number and date of birth. If you want send friend request to particular receiver then click on follow, then request will send to the user.

H. Messages

User can view the messages, send messages and send anomaly messages to users. User can send messages based on topic to the particular user, after sending a message that topic rank will be increased. Then again another user will also re-tweet the particular topic then that topic rank will increases. The anomaly message means user wants send a message to all users.

I. Followers

In this module, we can view the followers' details with their tags such as user name, user image, date of birth, E mail ID, phone number and ranks.

J. Android Test Book

We can easily use this application. This application user has to install in a mobile. Before using this application user should register, after registration he should login by using authorized user name and password. After login successful he will do some operations such as view emerging topic messages, view anomaly topic messages, view all users, view request and response.

IV. CONCLUSION

In this system, we have proposed a new approach to detect the emergence of topics in a social network stream. The basic idea of our approach is to focus on the social aspect of the posts reflected in the mentioning behavior of users instead of the textual contents. We have proposed a

probability model that captures both the number of mentions per post and the frequency of mentionee. We have combined the proposed mention model with the SDNML change-point detection algorithm and Kleinberg's burst-detection model to pinpoint the emergence of a topic. Since the proposed method does not rely on the textual contents of social network posts, it is robust to rephrasing and it can be applied to the case where topics

are concerned with information other than texts, such as images, video, audio, and so on. We have applied the proposed approach to four real data sets that we have collected from Twitter. The four data sets included a wide-spread discussion about a controversial topic ("Job hunting" data set), a quick propagation of news about a video leaked on Youtube ("Youtube" data set), a rumor about the upcoming press conference by NASA ("NASA" data set), and an angry response to a foreign TV show ("BBC" data set). In all the data sets, our proposed approach showed promising performance. In three out of four data sets, the detection by the proposed link-anomalybased methods was earlier than the text-anomaly-based counterparts. Furthermore, for "NASA" and "BBC" data sets, in which the keyword that defines the topic is more ambiguous than the first two data sets, the proposed link-anomaly-based approaches have detected the emergence of the topics even earlier than the keyword-based approaches that use hand-chosen keywords.

All the analysis presented in this paper was conducted offline, but the framework itself can be applied online. We are planning to scale up the proposed approach to handle social streams in real time. It would also be interesting to combine the proposed link-anomaly model with text-based approaches, because the proposed link-anomaly model does not immediately tell what the anomaly is. Combination of the word-based approach with the link-anomaly model would benefit both from the performance of the mention model and the intuitiveness of the word-based approach.

REFERENCES

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

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

[3] Y. Urabe, K. Yamanishi, R. Tomioka, and H. Iwai, "Real-Time Change-Point Detection Using Sequentially Discounting Normalized Maximum Likelihood Coding," Proc. 15th Pacific-Asia Conf. Advances in Knowledge Discovery and Data Mining (PAKDD' 11), 2011.

[4] S.Morinaga and K. Yamanishi, "Tracking Dynamics of Topic Trends Using a Finite Mixture Model," Proc. 10th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 811-816, 2004.

[5] Q. Mei and C. Zhai, "Discovering Evolutionary Theme Patterns from Text: An Exploration of Temporal Text Mining," Proc. 11th ACM SIGKDD Int'l Conf. Knowledge Discovery in Data Mining, pp. 198-207, 2005.

[6] A. Krause, J. Leskovec, and C. Guestrin, "Data Association for Topic Intensity Tracking," Proc. 23rd Int'l Conf. Machine Learning (ICML' 06), pp. 497-504, 2006.

[7] D. He and D.S. Parker, "Topic Dynamics: An Alternative Model of Bursts in Streams of Topics," Proc. 16th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 443-452, 2010.

[8] H. Small, "Visualizing Science by Citation Mapping," J. Am. Soc. Information Science, vol. 50, no. 9, pp. 799-813, 1999.

[9] D.Aldous, "Exchangeability and Related Topics," _ Ecole d' _ Ete´ de Probabilite´s de Saint-Flour XIII—1983, pp. 1-198, Springer, 1985.

[10] Y. Teh, M. Jordan, M. Beal, and D. Blei, "Hierarchical Dirichlet Processes," J. Am. Statistical Assoc., vol. 101, no. 476, pp. 1566-1581, 2006.

[11] D. Lewis, "Naive (Bayes) at Forty: The Independence Assumption in Information Retrieval," Proc. 10th European Conf. Machine Learning (ECML' 98), pp. 4-15, 1998.

[12] K.Yamanishi and J. Takeuchi, "A Unifying Framework for Detecting Outliers and Change Points from Non-Stationary Time Series Data," Proc. Eighth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, 2002.

[13] 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.

[14] J. Rissanen, "Strong Optimality of the Normalized ML Models as Universal Codes and Information in Data," IEEE Trans. Information Theory, vol. 47, no. 5, pp. 1712-1717, July 2001.

[15] T.Roos and J. Rissanen, "On Sequentially Normalized Maximum Likelihood Models," Proc. Workshop Information Theoretic Methods in Science and Eng., 2008.

[16] J. Rissanen, T. Roos, and P. Myllyma¨ki, "Model Selection by Sequentially Normalized Least Squares," J. Multivariate Analysis, vol. 101, no. 4, pp. 839-849, 2010.

[17] C. Giurc_aneanu, S. Razavi, and A. Liski, "Variable Selection in Linear Regression: Several Approaches Based on Normalized Maximum Likelihood," Signal Processing, vol. 91, pp. 1671-1692, 2011.

[18] C. Giurc_aneanu and S. Razavi, "AR Order Selection in the Case When the Model Parameters Are Estimated by Forgetting Factor Least-Squares Algorithms," Signal Processing, vol. 90, no. 2, pp. 451-466, 2010.

[19] K. Yamanishi and Y. Maruyama, "Dynamic Syslog Mining for Network Failure Monitoring," Proc. 11th ACM SIGKDD Int'l Conf. Knowledge Discovery in Data Mining, pp. 499-508, 2005.