International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 06 | June 2019 www.irjet.net p-ISSN: 2395-0072

# EVENT DETECTION AND TEXT SUMMARY BY DISASTER WARNING

**FATHIMA P.N[1], ASHA GEORGE[2]**

*[1]M.Tech Student Computer Science and Engineering,*
*[2]Asst.Professor, Department of Computer Science and Engineering, Indira Gandhi Institute of Engineering and Technology, Nellikuzhi, Kerala, India*

---

***Abstract -*** *Social media such as Twitter, Facebook are becoming an indispensable part in our daily lives. The rapid development of social media has attracted a large amount of research attention. In this paper we build a network of keywords based on their co-occurrence in documents. Proposed a user-interest model based event evolution model, named HEE model. The experimental results on real Twitter dataset demonstrate the efficiency and accuracy of our proposed model for both event evolution and user interest discovering. Social media such as Twitter, Facebook are becoming an indispensable part in our daily lives. The rapid development of social media has attracted a large amount of research attention. In this paper, it model a novel incremental clustering problem for comment stream summarization on SNS. It verify that IncreSTS possesses the advantages of high efficiency, high scalability, and better handling outliers, which justifies the practicability of IncreSTS on the target problem. The traditional framework of early warning systems is classified into of three phases: monitoring of precursors, forecasting of an event, and the notification of a warning. In addition of fourth phase is proposed, purpose of this phase is to recognize that there needs to be a response to the warning.*

***Key Words*****: Microblogging, event evolution, user topic, automatic clustering, user interest.**

## 1. INTRODUCTION

Twitter, Facebook are becoming a part in our daily lives. The rapid development of social media has attracted a large amount of attention. Microblogging is a type of social media where a user shares status updates and opinions in the form of short messages. The recent popularity of microblogging networks shows that microblogging is continuing to develop and rapidly attract people. The information and data being posted in microblogging is most often event-driven. This allowed microblogging to become one of the source for reporting real-world events. Social media plays an important role in people's life.

Broadcasting events in microblogging networks is an effective method of creating awareness, divulging important information. So many existing approaches the information content and discuss the event detection model and ignore the user interest. This leads to difficult in tracking the important events as they including identifying the influential spreaders. There is further complication that the influential spreaders interests will change during event evolution. The influential spreaders plays interesting role in event evolution. This has been largely ignored in traditional event detection methods.

Proposed a user-interest model-based event evolution model, named it as hot event evolution model. Model not only considers the user interest distribution but also the short text data in the social network. It is used to model the posts and the recommend methods to discover the user interests. It can solve the problem of data sparsity, as exemplified by many existing event detection methods. Improve the accuracy of event detection. A hot event automatic filtering algorithm is applied to general events, improving the quality and deficiency. Finally an automatic topic clustering algorithm is applied to arrange the short texts to the clusters with similar topics. An enhanced user-interest model is proposed to combine short texts of each cluster to form a long text document.

### 1.1 OBJECTIVE

Microblogging is a type of social media which allows people to share and disseminate real-life events. Broadcasting the social and other events in microblogging networks can be an effective method of creating awareness, divulging important information. Many existing approaches are developing at dissecting the information content. Discussing the event detection model and ignore the user interest which can be discovered during event evolution. This is very difficult in tracking the most important events as they evolve including identifying the influential spreaders. There is a complication given that the influential spreaders interests can change event evolution. The influential spreaders play a key role in event evolution and it has been largely ignored in traditional event detection methods.

It propose a model a novel incremental clustering problem and propose the algorithm IncreSTS, Which can incrementally update clustering results using with the latest incoming comments in real time. It provide the capability of comment stream summarization on SNS. With the output of IncreSTS, It will design a visualization interface consisting of basic information, key-term clouds, and representative comments. This presentation enables users to easily and rapidly get an overview understanding of a comment stream. From experimental results and a real case demonstration, it verify that IncreSTS possesses the advantages of high efficiency, high scalability, and better handling outliers,

which provides the practicability of IncreSTS on the target problem.

### 1.2 SCOPE

Proposed an improved user-interest model based event evolution model, it is named HEE (Hot Event Evolution) model. It not only considers the user's interest distribution, but also uses the short text data in microblogging network to model the data. A clustering algorithm is then used to combine the related short texts into a single text document to solve the problem of sparse data. Then according to the users in the document and the scored topics, the topics of each document is then modelled by LDA (Latent Dirichlet Allocation) topic model to get the topics of the whole document and the interests of users. To solve this problem, an automatic event and user scoring method, encompassing the authority value and minimum distance of the posts. The method can achieved on basis of the idea that the representative posts under specific event usually have higher authority. There is event difference between the posts in a microblogging network. The high popularity of SNS leads to the quantity of comments may increase at a high rate after publishing a social message. In fact that users may desire to get a brief understanding of a comment stream without reading the whole comment list. It will group comments with similar content together and generate a concise opinion summary of the message. Different users may request the summary at any moment so that the existing clustering methods cannot be directly applied and cannot meet the real-time need of this application. It model a novel incremental clustering problem for comment stream summarization. It propose IncreSTS algorithm that can incrementally update clustering results with latest incoming comments. Design a visualization interface to help users easily and rapidly get an overview summary.

## 2. RELATED WORK

Event detection methods based on topic modelling are increasing in popularity. For example, PLSA (Probabilistic Latent Semantic Analysis) and LDA are two important approaches to detect the hidden variables in microblogging. Such methods model the word occurrences with a probabilistic theory, and measure the topical similarity among the words. Existing research only discusses event detection methods but ignore the event evolution and strategies to combine messages regarding the same event. This leads to problems in dynamically tracking hot events and the identification of the influential spreaders. The traditional framework of early warning systems is classified into of three phases: monitoring of precursors, forecasting of an event, and the notification of a warning. In addition of fourth phase is proposed, purpose of this phase is to recognize that there needs to be a response to the warning and its responsibility relies on emergency response agencies.

**Table -1:** Analysis Comparison between SVM and PLNN

| Comparison Parameters | SVM Classier | PLNN Classifier |
|---|---|---|
| True Positive | 5 | 52 |
| True Negative | 920 | 905 |
| False Positive | 2 | 17 |
| False Negative | 76 | 29 |
| Precision | 71.3245 | 75.435 |
| Accuracy | 91.333 | 96.456 |
| Sensitivity | 6.45 | 64.466 |
| Specificity | 99.457 | 98.321 |

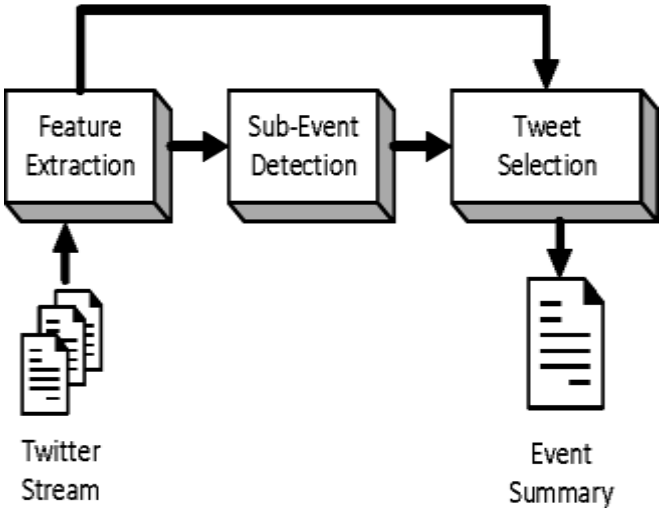


**Fig -1**: Real-time Sub-event Detection and Summarization

## 3. PROPOSED SYSTEM

Proposed a user-interest model-based event evolution model, named the hot event evolution model(HEE). This model uses the short text data in the social network to model the posts and the recommend methods to find the user interests. This can resolve the problem of data sparsity, as exemplified by many existing event detection methods, and improve the accuracy of event detection. A hot event automatic filtering algorithm is initially applied of general events, improving the quality and deficiency of mining the event. Then, an automatic topic clustering algorithm is used to arrange the short texts into clusters with similar topics. An improved user-interest model is proposed to combine the short texts into a long text document.

It explore the problem of incremental short text summarization on comment streams from social network services. It proposed the IncreSTS (standing for Incremental Short Text Summarization) algorithm to discover the top-k clusters including different groups of opinions towards one social message. For comment clusters, important and common terms will be extracted to construct a key-term

cloud. This key-term cloud provides an at glance presentation that users can easily and rapidly understand the main points of similar comments in a cluster. Our objective is that to help users get an overview understanding without reading all comments.

### 3.1 HEE MODEL

**Modules:**

➢The Event Automatic Clustering Algorithm.

➢The User Community Detection Algorithm.

➢ Personalized User Interests Discovery.

The HEE model is composed of four modules: Specifically, first, a hot event automatic filtering algorithm is proposed to remove the influence of general events. Then, an automatic topic clustering algorithm is proposed to combine all short texts with similar topics into clusters. And an improved user interest model is proposed to integrate all short texts in each cluster to form a long text document simplifying the determination of the overall topic in relation to the interest distribution of each user during the evolution of hot events. Finally, a cosine measure based event similarity detection method is proposed to judge correlation between events, detecting the process of event evolution.

### 3.2 SUMMARIZATION

Another map-like presentation system Twitter- Stand further considers incorporating geographic location of tweets to automatically obtain late breaking news. In addition, with collections of short posts on a specific topic, the authors in aim to create short summary sentences that best describe the primary gist of what users are saying about. With similar intention of, work of employs both generative model and user behaviour model to synthesize content from micro-blogging messages on the same topic into a prose description of fixed length. On the other hand, the research topic of analysing product reviews has also attracted much attention. The first step of these approaches is to obtain important idea of product features from review texts. Subsequently, in addition to traditional techniques of data mining or machine learning, natural language processing and are commonly incorporated to achieve various summarization needs.

### 3.3 EARLY WARNING SYSTEM

Million people have been killed over the last decade by disasters caused by storms, droughts, floods. EWS has framework includes the additional fourth phase. The purpose of this fourth phase is to recognize that there needs to be a response to the warning and its responsibility relies on emergency response agencies. Early warning system, EWS comprises four inter-related elements, spanning knowledge of hazards and vulnerabilities through to preparedness and capacity to respond. A weakness or failure in any one of these elements could result in failure of the whole system.

## 4. CONCLUSION

Model a novel incremental clustering problem and propose the algorithm IncreSTS, which incrementally update clustering results with latest incoming comments. Using the output of IncreSTS, a visualization interface consisting of basic information is designed, key-term clouds, and representative comments also designed. IncreSTS possess the advantages of high efficiency, high scalability, and better handling outliers, which justifies the practicability of IncreSTS. In the future work, other attributes of tweets (e.g., embedded URL) to compute the relatedness between two tweets for customizing the concepts of co-occurrences. Therefore plan to apply other types of community detection methods to better extract posy topics. Other research points are also including about how to predict the behaviour changes of influential users during the evolution. And also how to predict the popularity of hot events in the future.

## REFERENCES

[1] Q. Gao, F. Abel, G.-J. Houben, and Y. Yu, ``A comparative study of users' microblogging behavior on Sina Weibo and Twitter,'' in *User Model-ing, Adaptation, and Personalization*. Berlin, Germany: Springer, 2012, pp. 88_101.

[2] Q. Diao, J. Jiang, F. Zhu, and E.-P. Lim, ``Finding bursty topics from microblogs,'' in *Proc. 50th Annu. Meeting Assoc. Comput. Linguistics*, vol. 1. Jul. 2012, pp. 536_544.

[3] X. Zhou and L. Chen, ``Event detection over twitter social media streams,'' *VLDB J. Int. J. Very Large Data Bases*, vol. 23, no. 3, pp. 381_400, Jun. 2014.

[4] D. M. Blei, A. Y. Ng, and M. I. Jordan, ``Latent Dirichlet allocation,'' *J. Mach. Learn. Res.*, vol. 3, pp. 993_1022, Mar. 2003.

[5] K. Zhou, A. Martin, and Q. Pan, ``A similarity-based community detection method with multiple prototype representation,'' *Phys. A, Statist. Mech. Appl.*, vol. 438, pp. 519_531, Nov. 2015.

[6] T. Hofmann, ``Probabilistic latent semantic indexing,'' in *Proc. 22nd Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, vol. 8. Aug. 1999, pp. 50_57, doi: 10.1145/312624.312649.

[7] X. Sun, Y.Wu, L. Liu, and J. Panneerselvam, ``Ef_cient event detection in social media data streams,'' in *Proc. IEEE Int. Conf. Comput. Inf. Technol.,Ubiquitous Comput. Commun., Dependable, Autonomic Secur. Comput., Pervasive Intell. Comput.*, Oct. 2015, pp. 1711_1717.

[8] J. Li, Z. Tai, R. Zhang, W. Yu, and L. Liu, ``Online bursty event detection from microblog,'' in *Proc. IEEE/ACM7th Int.*

*Conf. Utility Cloud Comput., IEEE Comput. Soc.*, Dec. 2014, pp. 865_870, doi: 10.1109/UCC.2014.141.

[9] G. Salton, A. Wong, and C. S. Yang, ``A vector space model for automatic indexing,'' *Commun. ACM*, vol. 18, no. 11, pp. 273_280, Nov. 1975.

[10] J. Makkonen, ``Investigations on event evolution in TDT,'' in *Proc. Conf.North Amer. Chapter Assoc. Comput. Linguistics Human Lang. Technol Proc. HLT-NAACL Student Res. Workshop*, May/Jun. 2004, pp. 43_48.

[11] R. Nallapati, A. Feng, F. Peng, and J. Allan, ``Event threading within news topics,'' in *Proc. 13th ACM Int. Conf. Inf. Knowl. Manag.*, Nov. 2004, pp. 446_453.

[12] O. Jin, N. N. Liu, K. Zhao, Y. Yu, and Q. Yang, ``Transferring topical knowledge from auxiliary long texts for short text clustering,'' in *Proc. ACM Conf. Inf. Knowl. Manage. (CIKM)*, Glasgow, U.K., Oct. 2011, pp. 775_784.

[13] Y. Chen, X. Zhang, Z. Li, and J. P. Ng, ``Search engine reinforced semisupervised classi_cation and graph-based summarization of microblogs,'' *Neurocomputing*, vol. 152, pp. 274_286, Mar. 2015.

[14] L. Hong and B. D. Davison, ``Empirical study of topic modeling in Twitter,'' in *Proc. SIGKDDWorkshop Social Media Anal.*, Jul. 2010, pp. 80_88.

[15] W. X. Zhao *et al.*, *Comparing Twitter and Traditional Media Using Topic Models* (Lecture Notes in Computer Science), vol. 6611. Amsterdam, The Netherlands: Atlantis Press, 2011, pp. 338_349.

[16] X. Yan, J. Guo, Y. Lan, and X. Cheng, ``A biterm topic model for short texts,'' in *Proc. Int. Conf. World Wide Web*, May 2013, pp. 1445_1456.

[17] Y. Pan, J. Yin, S. Liu, and J. Li, ``A biterm-based dirichlet process topic model for short texts,'' in *Proc. 3rd Int. Conf. Comput. Sci. Service Syst. (CSSS)*, 2014.

[18] M. Rosen-Zvi, T. Grif_ths, M. Steyvers, and P. Smyth, ``The author-topic model for authors and documents,'' in *Proc. Conf. Uncertainty Artif. Intell.*, Jul. 2012, pp. 487_494.

[19] Y. Li, C. Jia, and J. Yu, ``A parameter-free community detection method based on centrality and dispersion of nodes in complex networks,'' *Phys. A, Statist. Mech. Appl.*, vol. 438, pp. 321_334, Nov. 2015. [20] A. Asuncion, M. Welling, P. Smyth, and Y. W. Teh, ``On smoothing and inference for topic models,'' in *Proc. 25th Conf. Uncertainty Artif. Intell.*, Jun. 2009, pp. 27_34.