

Monitoring Abnormal Behaviour of Internet Users Using Multiple Single Sign on Web Applications

A.Jayanthi¹, R.Kalaivani², L.Malar Mathi²,
Assistant Professor¹, Department of IT,
Vel Tech Hightech DR.Rangarajan DR.Sakunthala Engineering College, Chennai,
jayanthi@velhightech.com, kalaivanir21@gmail.com, malarmathiloganathan@gmail.com

Abstract

Mining of static content data works are devoted to topic modeling and the evolution of individual topics, while sequential relations of topics in successive documents published by a specific user are ignored. Hence the users activity monitoring doesn't feasibly and effectively. Monitoring individual users' activity in single web application doesn't give the effective dataset of topic extraction about the user. , we propose Sequential Topic Patterns and formulate the problem of mining User-aware Rare Sequential Topic Patterns in document streams on the Internet. Experiments on both real and synthetic datasets show that our approach can indeed discover special users and interpretable URSTPs effectively and efficiently, which significantly reflect user's characteristics

Keywords—web mining, sequential patterns, document streams, rare events.

INTRODUCTION

Data mining discovers an interesting pattern from large amounts of data. The data sources can include databases, data warehouses, the Web, other information repositories, or data that are streamed into the system dynamically. This technique, results a long process of research and product development. This evolution began when business data was first stored on computers, continued with improvements in data access, and recently, generated technologies that allow users to navigate through their data in real time. It takes this evolutionary process beyond recollected data access and navigation to prospective and proactive information delivery.

The documents which created and distributed on the Internet are ever changing in various forms. In this paper, in order to characterize and monitoring personalized and abnormal behaviors of Internet users, we propose Sequential Topic Patterns and formulate the problem of mining User-aware Rare Sequential Topic Patterns in document streams on the Internet. They are rare on the whole but relatively frequent for specific users, so can be applied in many real-life scenarios, such as real-time monitoring on abnormal user behaviors. We present a group of algorithms to solve this innovative mining problem through three phases: preprocessing to extract probabilistic topics and identify sessions for different users, generating all the STP candidates with (expected) support values for each user by pattern-growth, and selecting URSTPs by making user-aware rarity analysis on derived STPs.

EXISTING SYSTEM

In [1], the mining sequential pattern is used to uncover all sequential pattern which satisfy the user specified constraint from the given sequence database. They proposed the work which is used for extending the sequential mining approaches. They proposed the sequential pattern mining algorithm which may solves the problem of discovering the presence of frequent sequences in the given database. In this paper it is difficult to generate and examine a no of intermediate sub sequences. The apriori algorithm used can be extended to web content mining and web structure mining.

In [2], the online new event detection and tracking mainly focuses on the lack of public datasets. The privacy issues along with social network components of use hinder the availability of shared data. The existing method

shows the evaluation is being executed in event detection and mentioned the most common evaluation metrics and datasets used. The main goal of the tasks is to monitor a stream of broadcast news stories so as to determine the relationship between the stories based on the real world events that they describe.

In [3], they introduced the probabilistic frequent itemset mining in uncertain databases which uses probabilistic Fp growth algorithm and associated probabilistic Fp tree which we use to mine all probabilistic frequent itemset in uncertain transaction database without candidate generation. They demonstrated that the probabilistic pruning strategy allows us to prune non frequent itemsets early leading to a large performance gain. They proposed an order of magnitude faster than straight forward approach. The main drawback is that the data is inherently noisy such as data collected by sensors or in

In [4], the paper short and tweet experiments on recommending content from information streams for recommendation on twitter to better direct user attention. They propose to explore the other domains so as to deepen our understanding in design space to finer details and add more design options or add completely new design space. This follow a collaborative filtering method. The recommendations are rated high. The heaviest incoming tweet volume filters third party tools. The main drawback is that it may require confirmation from both sides of sender and receiver.it also needs a large volume of real time usage data.

In [5], the paper aims at microblogging during two natural hazards events that the twitter may contribute to situational awareness by harvesting the information during the crisis timing. Thus by growing ubiquitously communication rapidly and cross platform accessibility. In the existing work it may be used for extracting useful and relevant information during emergencies using information extraction technique. This paper does not used to improve the situational awareness during emergency events. They propose an idea of improving the situational awareness in emergency situations through the automatic methods such as high yield twitterness.

In [6], the paper aims at spatiotemporal social media analytics for abnormal event detection and examination using seasonal trend decomposition this existing work results in presenting a visual analytics approach that provides scalable and interactive social media data analysis and visualization. They proposes a context based analysis and to improve the current detection algorithm to allow for a faster analysis. The main advantage in finding temporal trends according to volume based importance by STL evaluation and LDA topic examination.

In [7], the effective incentive scheme is proposed to stimulate the forwarding cooperation of nodes in VANETs. In a coalitional game model, every relevant node cooperates in forwarding messages as required by the routing protocol. This scheme is extended with constrained storage space. A lightweight approach is also proposed to stimulate the cooperation.

In [8], the paper aims at a decremental approach for mining frequent itemsets from uncertain data based on the study the problem of mining frequent itemsets from *uncertain data* under some probabilistic *model*. U-Apriori and LGS-Trimming approaches are significant computational cost savings. Detrimental pruning does not require a user-specified trimming threshold values and the performance is relatively insensitive to the population of low-probability items

In [9], the paper Interactive visual analysis of text data through event identification and exploration, the interactive visual analytics system, LeadLine, used to statistically identify meaningful events in news and social media networks data and support exploration of the events. To characterize the new events, LeadLine integrates topic modeling, event detection, and named entity recognition techniques to automatically extract information regarding the investigation based on the events .In this paper they are extracting a two case studies to explore the headline to Occupy Wall Street Movement and Exploring the OWS Movement through Major Events

PROPOSED SYSTEM

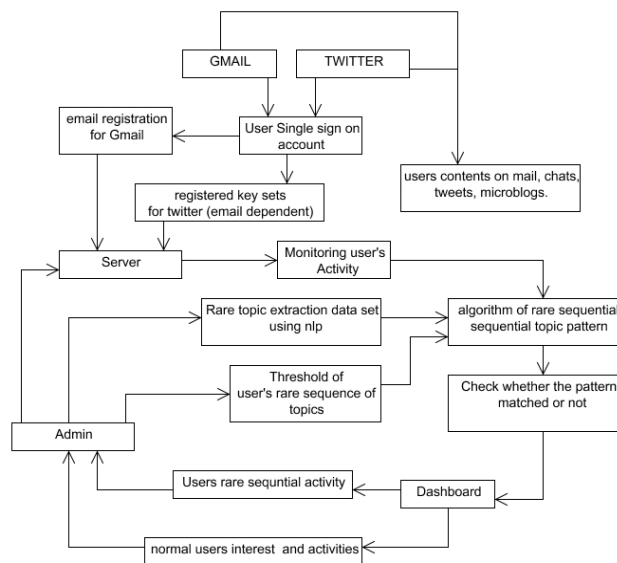
In our proposed system, Users rare and sequential activities can be monitored using sequence of document streams on multiple web application. We proposed our system to extract the user's activity on real time web application data set on Twitter and Gmail. Using our technique can monitor the user's sequential topic pattern based on their session identification on multiple applications with single sign on email id and their session id.

We used the documents of inbox and send box mail of Gmail contents and twitter's tweet and individual chats to extract the topic and mining the user's activity. We extract the topic of document stream content using Stanford Natural Language Processing. Using this NLP processing and monitoring dynamic user's different activities can be extracted and monitored effectively

ADVANTAGES:

- Suitable for inferring users' intrinsic characteristics
- Solves innovative and significant problem of mining URSTPs
- Uses multiple web application for sequential pattern
- Probabilistic description of topics helps to maintain and accumulate the uncertainty degree of individual topics.

SYSTEM ARCHITECTURE



CONCLUSION

The suggested sequential topic pattern mechanism achieves a Static and dynamic content using real and synthetic datasets Shows that our approach can indeed discover special users and interpretable URSTPs effectively and efficiently

REFERENCES

[1] R. Agrawal and R. Srikant, "Mining sequential patterns," in Proc.IEEE Int. Conf. Data Eng., 1995, pp. 3-14.

[2] J. Allan, R. Papka, and V. Lavrenko, "On-line new event detection and tracking," in Proc. 21st Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 1998, pp. 37-45

[3] T. Bernecker, H.-P. Kriegel, M. Renz, F. Verhein, and A. Zuefle, "Probabilistic frequent itemset mining in uncertain databases," in Proc. ACM SIGKDD, 2009, pp. 119-128

- [4] Jilin Chen, Rowan Nair, Les Nelson, Michael Bernstein, Ed H. Chi “Short and Tweet: Experiments on Recommending Content from Information Streams”
- [5] Sarah Viewer, Amanda Hughes, Kate Starbird & Leysia Palen “Microblogging During Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness”
- [6] Chee, D. Thom, H. Bosch, Y. Jang, R. Maciejewski, D. S. Ebert and T. Ertl, “Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition,” in Proc. IEEE Conf. Vis. Anal. Sci. Technol., 2012, pp. 143– 152
- [7] Christo Ananth, M.Muthamil Jothi, A.Nancy, V.Manjula, R.Muthu Veni, S.Kavya, “Efficient message forwarding in MANETs”, International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE), Volume 1, Issue 1, August 2015, pp:6-9
- [8] C. K. Chui and B. Kao, “A decremental approach for mining frequent itemsets from uncertain data,” in Proc. 12th Pacific-Asia Conf. Adv. Know. Discovery Data Mining, 2008, pp. 64–75.
- [9] W. Dou, X. Wang, D. Skaw, W. Ribarsky, and M. X. Zhou, “Lead Line: Interactive visual analysis of text data through event identification and exploration,” in Proc. IEEE Conf. Vis. Anal. Sci. Technol., 2012, pp. 93–102.
- [10] G. P. C. Fung, J. X. Yu, P. S. Yu, and H. Lu, “Parameter free bursty events detection in text streams,” in Proc. 31st Int. Conf. Very Large Data Bases, 2005, pp. 181–192.