ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 1, January 2019

Survey on Forecasting Events over Social Media using Chrono-Spatial Model

Ankit Kabra[1], Prof. M. P. Wankhade[2]

PG Student, Department of Computer Networks, SCOE, Pune, India[1]

Assistant Professor, Department of Computer Networks, SCOE, Pune, India[2]

ABSTRACT: It is extremely well known to identify hot topics, which can profit numerous undertakings including topic recommendations, the guidance of public opinions, etc. However, in some cases, people may want to know when to re-hot a topic, i.e., make the topic popular once more. In this project, we address this issue by presenting a spatio-Temporal User Topic Participation (UTP) model which models user's behaviors of posting messages. The UTP demonstrate considers clients' interests, Friend-circles and unexpected events in online social networks. Likewise, it considers the persistent spatio-temporal modeling of topics, since subjects are changing consistently after some time. Moreover, a weighting plan is proposed to smooth the variances in topic re-hotting prediction. At long last, trial results directed on true informational collections exhibit the effectiveness of our proposed models and topic re-hotting prediction techniques.

KEYWORDS: User Topic Participation (UTP), Topic Re-hotting Prediction, Online Social Networks (OSN)

I. INTRODUCTION

There is rapid development of data storage, information processing, and networking transmission technologies, online social networks (OSNs) have been becoming indispensable in people daily life. Everyone could freely post messages, share news, and participate in topic discussions in OSNs, e.g., Twitter (twitter.com) and Weibo (weibo.com). Along with that, many researchers have done lots of work for the convenience to analyze and use OSNs, such as topic detection, topic prediction, and topic transition. However, the phenomena of topic decay and even disappearance are inevitable. It is reported that 23% of topics have two or more hot (a.k.a. active or popular) periods.

Clearly, in many situations, after observing that a hot topic is dwindling, it is very interesting but challenging to intelligently extrapolate when this topic may be re-hot, i.e., make the topic hot again at suitable time points. It is called the problem of topic re-hotting prediction in this study, and has a lot of practical applications. This system present and formalize the problem of topic re-hotting prediction (TRP) in OSNs. It facilitates a better understanding of the topic characteristics when the focusing topics are dwindling, as well as benefits many related issues, such as topic detection and topic tracing. This system can effectively explain user's behaviors of participating in the topic discussions in OSNs.

II. LITERATURE SURVEY

In this paper authors present a novel method for clustering words in microblogs, based on the similarity of the related temporal series. This technique, named SAX , uses the Symbolic Aggregate ApproXimation algorithm to discretize the temporal series of terms into a small set of levels, leading to a string for each. This paper defines a subset of "interesting" strings, i.e. those representing patterns of collective attention. Sliding temporal windows are used to detect co-occurring clusters of tokens with the same or similar string[1].

In this paper, authors presented a new approach to detect burst novel events and predict their future popularity simultaneously. Specifically, it first detects events from online micro blogging stream by utilizing multiple types of information, i.e., term frequency, and user's social relation. Meanwhile, the popularity of detected event is predicted

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 1, January 2019

through a presented diffusion model which takes both the content and user information of the event in to account. Extensive evaluations on two real-world datasets demonstrate the effectiveness of this approach on both event detection and their popularity prediction[2].

This paper proposes a classification based approach for burst time prediction by utilizing and modeling rich knowledge in information diffusion. Particularly, Authors first propose a time window based approach to predict in which time window the burst will appear. This paves the way to transform the time prediction task to a classification problem. To address the challenge that the original time series data of the cascade popularity only are not sufficient for predicting cascades with diverse magnitudes and time spans, it explore rich information diffusion related knowledge and model them in a scale-independent manner. Extensive experiments on a Sina Weibo reposting dataset demonstrate the superior performance of the proposed approach in this paper accurately predicting the burst time of posts[3].

In this paper, authors build up a large human mobility database (GPS records of 1.6 million users over one year) and several different datasets to capture and analyze human emergency behavior and their mobility following the Great East Japan Earthquake and Fukushima nuclear accident. Based on this empirical analysis through these data, authors find that human behavior and their mobility following large-scale disaster sometimes correlate with their mobility patterns during normal times, and are also highly impacted by their social relationship, intensity of disaster, damage level, government appointed shelters, news reporting, large population flow and etc. On the basis of these findings, Authors develop a model of human behavior that takes into account these factors for accurately predicting human emergency behavior and their mobility following large-scale disaster. The experimental results and validations demonstrate the efficiency of this behavior model, and suggest that human behavior and their movements during disasters may be significantly more predictable than previously thought[4].

In this article, authors address the problem of clustering imprecise data using a finite mixture of Gaussians. They propose to estimate the parameters of the model using the fuzzy EM algorithm. This extension of the EM algorithm allows them to handle imprecise data represented by fuzzy numbers. First, authors briefly recall the principle of the fuzzy EM algorithm. Then, they provide closed-forms for the parameter estimates in the case of Gaussian fuzzy data. Authors also describe a Monte-Carlo procedure for estimating the parameter updates in the general case. Experiments carried out on synthetic and real data demonstrate the interest of this approach for taking into account attribute and label uncertainty[5].

III.PROPOSED SYSTEM

We present and formalize the problem of topic re-hotting prediction (TRP) in OSNs. It facilitates a better understanding of the topic characteristics when the focusing topics are dwindling, as well as benefits many related issues, such as topic detection and topic tracing. We propose a novel temporal model, i.e., User Topic Participation (UTP) model, for the TRP problem. UTP can effectively explain user's behaviors of participating in the topic discussions in OSNs. Also, we bring forward improved algorithms like Re-hot Topic Prediction and Topic Mining within Region and Time Interval to effectively infer the UTP model. We design a method based on the UTP model to appropriately predict the re-hotting time points for given once-hot topics, i.e., the topics which had been hot before.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 1, January 2019

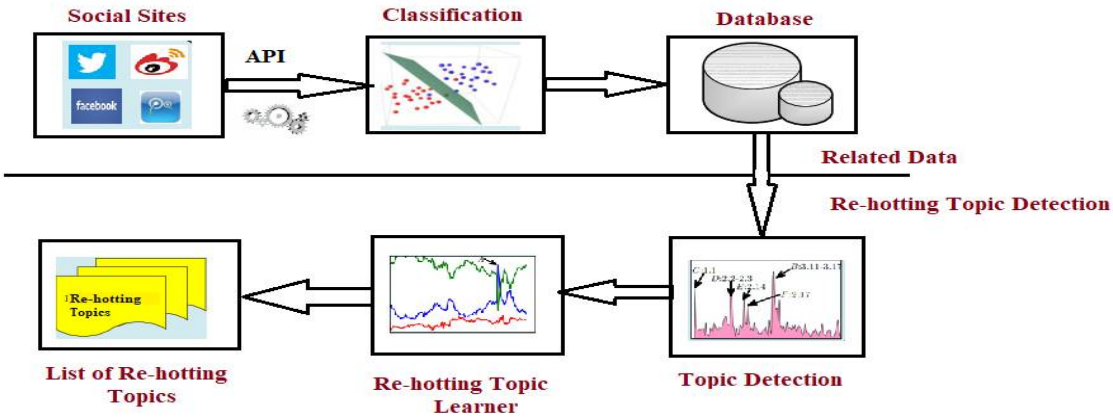

Fig 1: System Architecture

IV.CONCLUSION

In proposed system is to solve the challenging problem of topic re-hotting prediction in OSNs. By taking into account three factors, i.e., user's friend-circles, types of topics, and unexpected events, this system combines user's interests and unexpected events. Furthermore, it use re-hot topic prediction algorithm for model inference and a Topic Mining within Region and Time Interval prediction method to predict the re-hotting time points accurately. Moreover, in order to reduce the influence of slight fluctuations of topics, a weighting scheme is proposed. Finally, demonstrate the performance of the proposed methods on real-world data sets, and analyse the interesting phenomena which appear in this experiments.

REFERENCES

[1]G. Stilo and P. Velardi, "Efficient Temporal Mining of Micro-blog Texts and its Application to Event Discovery", Data Mining and Knowledge Discovery, vol. 30, no. 2, pp. 372402, 2016.
[2]X. Zhang, X. Chen, Y. Chen, S. Wang, Z. Li, and J. Xia, "Event Detection and Popularity Prediction in Microblogging", Neurocomputing, vol. 149, pp. 14691480, 2015.
[3]S. Wang, Z. Yan, X. Hu, P. S. Yu, and Z. Li, "Burst Time Prediction in Cascades", in Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015, pp. 325331.
[4]X. Song, Q. Zhang, Y. Sekimoto, and R. Shibasaki, "Prediction of Human Emergency Behavior and their Mobility following Large-scale Disaster", in Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, 2014.
[5]S. Wang, Z. Yan, X. Hu, P. S. Yu, Z. Li, and B. Wang, "CPB: A Classification based Approach for Burst Time Prediction in Cascades", Knowledge and Information Systems, vol. 49, no. 1, pp. 243271, 2016.
[6]X. Xin, C. Wang, X. Ying, and B. Wang, "Deep Community Detection in Topologically Incomplete Networks", Physica A: Statistical Mechanics and its Applications, vol. 469, pp. 342352, 2017.
[7]P. Awasthi, M. Charikar, R. Krishnaswamy, and A. K. Sinop, "Spectral Embedding of k-Cliques, Graph Partitioning and k-Means", in Proceedings of the 2016.
[8]B. Quost and T. Denoeux, "Clustering and Classification of Fuzzy Data Using the Fuzzy EM Algorithm", Journal of the Royal Statistical Society, vol. 286, no. 1, pp. 134156, 2016.
[9]Y. Yang, H. K. T. Ng, and N. Balakrishnan, "A Stochastic Expectation maximization Algorithm for the Analysis of System Lifetime Data with Known Signature", Computational Statistics, pp. 133, 2015.
[10]Y. Chen, H. Amiri, Z. Li, and T. S. Chua, "Emerging Topic Detection for Organizations from Microblogs", in Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval. ACM, 2013.