



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 10, October 2014

Hot Topics Perception in Social Network

Kayalvizhi P¹, Anoor Selvi C²

P.G. Student, Department of Computer Science and Engineering, V.S.B Engineering College, Karur, Tamil Nadu, India¹

Assistant Professor, Department of Computer Science and Engineering, V.S.B Engineering College, Karur, Tamil Nadu, India²

ABSTRACT: Social Network have become a very big network where many people are discussing about the hot trends, events and everyday activities which they considered as important with their friends, families and many unknown people. Compare to media like Newspaper, Television the social network is spreading the latest news in a very fast manner and the original talk of the people is spread over the network. The proposed model is to identify the hot topics which is discussed in the social network. The previous methods of finding the hot topics in social network has some limitations like the performance of the model is low, the system detect the topic with high false positive rate. In the proposed model the posts that is shared by the people in social network with their friends by forwarding the posts is considered. The particular post which is forwarded by many people and irregular forwarding nature will be monitored and using change point detection technique and comparing the contents of the posts the perception of hot topic is done.

KEYWORDS: Social Network, Topic Detection, Change Point Detection

I. INTRODUCTION

The Communication between people is increasing day by day with other people through many ways of communication media such as phone, internet news channels, newspaper, etc. The topic discussed may be a hot topic or their personal day today activities. In this work, we propose a probability model that can capture the normal sharing behavior of a user. The sharing of information through the social network consists of both the number of links that is created while sharing per post and the frequency of users occurring in that sharing. Then this model predict the future user behavior. Using the proposed probability model, we can quantitatively measure the novelty or possible impact of a post reflected in the sharing behavior of the user. The previous method like Link based anomaly detection and topic tracking in social network may have some disadvantages. That can be avoided in the proposed model by considering the sharing of the news and the content of the topic is also analyzed whether the shared posts contains text message or images etc.

II. RELATED WORK

In [1] the authors have used the mentioning behavior of the user only. There by using Sequential Change point detection and Burst detection method the emerging topics are identified. 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. Drawbacks of the link based detection is The quality of anomaly detection is lower than the other system this approach is doesn't well handle social streams in real time applications. This system gives the lower accuracy rate. Time complexity rate of the system is highest with lower accuracy. In [2] the keyword based topic detection uses keyword based analysis in [2] user Topic Detection and Tracking (TDT), this method may suffer from confusions due to texts that is considered for the analysis. The text may be written in different languages and meaning of the words may differ from one user's perspective to another user's perspective. The disadvantage is Online detection cannot yet be performed reliably. Substantial work is needed to reduce the errors to manageable numbers.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 10, October 2014

III. PROPOSED WORK

The proposed model works by analysing the content of the message which may be text, image or video and calculating the outlier score which is find from the sequence of scores which is generated while sharing the posts. The data set from the social network is obtained by the social network API such as Facebook API for Facebook. Using the individual unique ID generated in the social network, the names of the users involved in shared posts and the content of the posts are retrived for some time period. With that outlier score is calculated and summation of all the users who also shared posts are considered. From that scores the change point detection and burst detection is done. The content of the message is analyzed using semantic information without any delay the hot topics is finalized.

IV. IMPLEMENTATION

The proposed work is described with following System Design which contains the architecture diagram and System Modules.

A. System Design

- The dataset from the social network is obtained using API
- The normal sharing pattern of the user is analyzed
- The prediction of the sharing of topic is done using Change Point Detection
- The content of the message shared is analyzed

B. System Modules

The proposed model works by analysing which post is forwarded by many people to many of their friends. If a particular post is getting anomaly score (as outlier) then its content is analysed using WordNet tool for text message and for messages with images and videos the future work can be done by extracting the image feature such that colour, texture etc. similarly for videos can be done and the content of the message can be analysed. The proposed model has the following modules.

1. Training Phase
2. Change Point Detection
3. Analyzing the content of post.

1. Training Phase

First step in the proposed model is the training phase. In the training phase the past behavior of the user is considered, the posts that are shared with their friends are extracted from social network dataset using a social network API for analyzing the forwarding behavior of the user. Here number of user k who are mentioned in the post and IDs (Names of the user mentioned in the post) is taken as set V . Here the number of users who are mentioned in the post is limited by geometric distribution internally. With k and V , we are calculating the joint probability distribution to predict the probability of each user mentioned in forwarding list.

$$s'_j = \frac{1}{\tau} \sum_{t_i \in [\tau(j-1), \tau j]} s(x_i) \quad (1)$$

2. Sequentially Discounting Normalized Maximum Likelihood - Change Point detection

The Sequentially Discounting Normalized Maximum Likelihood Coding method [5] is used to find the change point from the sequence of anomaly score for all the post, this process is done through two layers of processes. In the first layer, from the collection of aggregated anomaly score which is calculated in specific time period (2), the outliers is detected by using the density function. In the Second layer from the outliers which is detected in first layer is used again the change point is detected.

Let $x^{j-1} = \{x_1, \dots, x_{j-1}\}$ be the aggregated anomaly score from time period 1 to $j-1$. The outlier is detected using the density function,

$$y_j = \frac{1}{k} \sum_{j'=j-k+1}^j (-\log p_{SDNML}(x_j | x^{j-1})) \quad (2)$$



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 10, October 2014

Finally using the Dynamic Threshold Optimization algorithm, the change point which is calculated (5) is converted into a binary alarm. It is raised by dynamically adjusting the threshold over a long period of time.

For a variable $x = x(t)$ in the discrete time series $x = \{x_t | t = 0, 1, \dots\}$

$$EMA(n)[x]_t = \sum_{k=0}^n \alpha (1 - \alpha)^k x_{t-k} \quad (3)$$

here n is the window size. The difference of its t_1 and t_2 moving averages:

$$\text{Moving Average}(t_1, t_2) = EMA(t_1) - EMA(t_2) \quad (4)$$

The histogram gives the difference between the moving average, this difference gives the burst in the outlier score.

3. Analyzing the content of post

After identifying the change point in the aggregated score the post using the above two techniques the post can be confirmed as the dynamic post which carries the hot topic, but the content of the post should be analyzed since the anomaly score is calculated based only on the link that is generated while forwarding. The content of the post is not considered till now for confirming the dynamic topic. To confirm the dynamic topic we need to analyze the content of the posts also. If the content of the posts is text message then it can be analyzed using the WordNet tool [6]. WordNet is a lexical database for English which has Nouns, Verbs, adjectives are grouped as synonym sets called synsets. The words are linked as synsets according to lexical and conceptual relations. All synsets are connected to other synsets via the semantic relations. With WordNet the similarity between words can be determined, this can be done by using algorithms that measures the distance between the words and forms the WordNet graph structure by counting the number of edges among the synsets. After the analysis of content of the post, the post which carries the dynamic topic can be identified.

V. CONCLUSION

The proposed work is to detect the dynamic topic that is discussed in the social network, by considering the past posts that is discussed just before the current post and predicting the future behaviour of the user. With the training set the anomaly score is calculated for the current post and users in the post and the aggregated anomaly score is calculated. Using the change point detection and burst detection the change point in the forwarding behaviour of post is detected and also the content of the message is analysed for checking whether the same topic is discussed in all the change point analysed post. The dynamic topic is finalized which is expected to detect the topic before the conventional media finds the hot topic.

REFERENCES

- [1] Toshimitsu Takahashi, Ryota Tomioka, and Kenji Yamanishi, Member, IEEE, "Discovering Emerging Topics in Social Streams via Link-Anomaly Detection", IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 1, January 2014.
- [2] J. Allan et al., "Topic Detection and Tracking Pilot Study: Final Report," Proc. DARPA Broadcast News Transcription and Understanding Workshop, 1998.
- [3] Daniel C. Berrios, Richard M. Keller, "Semantic Analysis of Email Using Domain Ontologies and WordNet", Source of Acquisition NASA Ames Research Center.