ISSN:0975-9646

Detecting of Materialize Topical Trend Topic from Social Networks

Tamilmani.G
Assistant Professor
Department of CSE
Vel Tech University
Chennai, India

Rajathi. K
Assistant Professor
Department of CSE
Vel Tech University
Chennai, India

Ponnaruvi. P
Assistant Professor
Department of CSE
Vel Tech High Tech Engg College
Chennai, India

INTRODUCTION

In this, we focus on social networks, such as Facebook and Twitter, which gaining more importance in our daily life. Since the information exchanged over social networks 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

We are 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. All the above-mentioned studies make use of textual content of the documents, but not the social content of the documents. The social content (links) has been utilized here. However, citation networks are often analyzed in a stationary setting. The novelty of the current paper lies in focusing on the social content of the documents (posts) and in combining this with a change-point analysis.

PROJECT DESCRITION

Product Perspective:

In this paper, 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 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

Product Features:

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 signaled by social aspects of theses networks. Specifically, we focus on mentions of users—links between users that are generated dynamically (intentionally or unintentionally) through replies, mentions, and retweets. We propose a probability model of the mentioning behavior 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. Our mention-anomaly-based approaches can detect the emergence of a new topic at least as fast as text-anomaly-based counterparts. the proposed mention-anomaly-based methods can detect the emergence of topics much earlier than the text-anomaly-based methods

Design and Implementation Constraints

Constraints in Analysis

- Constraints as Informal Text
- Constraints as Operational Restrictions
- Constraints Integrated in Existing Model Concepts
- Constraints as a Separate Concept
- Constraints Implied by the Model Structure

Constraints in Design

- Determination of the Involved Classes
- Determination of the Involved Objects
- Determination of the Involved Actions
- Determination of the Require Clauses
- Global actions and Constraint Realization

Constraints in Implementation

A hierarchical structuring of relations may result in more classes and a more complicated structure to implement. Therefore it is advisable to transform the hierarchical relation structure to a simpler structure such as a classical flat one. It is rather straightforward to transform the developed hierarchical model into a bipartite, flat model, consisting of classes on the one hand and flat relations on the other. Flat relations are preferred at the design level for reasons of simplicity and implementation ease. A flat relation corresponds with the relation concept of entity-relationship modeling and many object oriented methods.

System Features

In this project, BECAN Scheme is a feature of the implementation of this project. The Proposed scheme is a life cycle of the filtering the false data. The filtering the false data is maintained using the MAC concept. The MAC is used to find the Injected data in the Sensor node. If the Injected data was found the Sensor node will be filtered the data.

EXISTING SYSTEM

In the existing system 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.

Disadvantages of Existing System

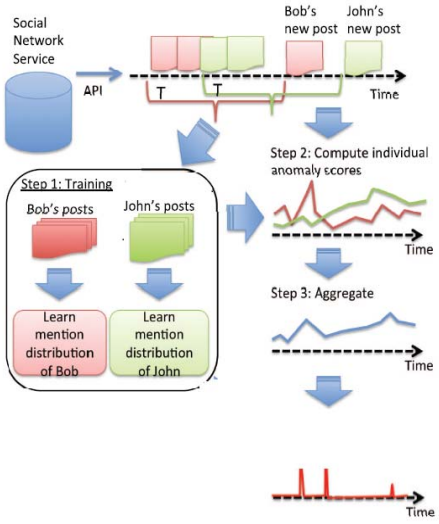
1. Existing system using conventional Media
2. All finished incidents and things will come as data in conventional media
3. From that data only we find Emerging Topics
4. Term-frequency method used to identify Emerging Topics

Data part only we are considering

PROPOSED SYSTEM

We focus on emergence of topics signaled by social aspects of networks. Specifically, we focus on mentions of users—link between users that are generated dynamically through replies, mentions, and retweets. We propose a probability model of the mentioning behavior 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 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. 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.

BLOCK DIAGRAM:



Advantages of Proposed System

1. We are using Social Networking Sites to find Emerging Topics
2. All finished incidents, things will come as data.
3. Each person's thoughts also will come as data (post, comment etc…)
4. From that data we are identifying Emerging topics
5. Link-Anomaly-Detection method we are using
6. Data part + Social part (social connection) we are using.
7. Social part is the Link (like, comment, share, mentions etc…)

(APPN) and advanced program-to-program communications (APPC).

MODULES

- **Social Network Service implementation**
- **Data Collection And Training**
- **Link Anomaly score calculation**

MODULE EXPLANATION:

Social Network Service implementation:

First, we create the social network service and implement its functionalities such as posts, comments; online users, profiles and every information will be stored in the repository. Here we take the data, that mean s the post and comments, mentions, retweets that is required for our process from the social networking stream that are used as input. First we collect the post of each and everyone who are registered with the social network stream. We deal with the comments also that are posted against each post.

Data Collection and Training:

Secondly we categorize each individual's posts and find the mentioning distribution. We will find how many number of mention s in a post as well as comments and also the frequency with which each user is mentioned. There are two types of infinity we have to take into account here. The first is the number k of users mentioned in a post. The second type of infinity is the number of users one can possibly mention. First we will calculate the predictive distribution with the number of mentions. Then we calculate the predictive distribution with number of users.

Link Anomaly score calculation:

The tradeoff between the security and detection cost, iTrust introduce periodically available Trust Authority (TA), which could launch the probabilistic detection for the target node and judge it by collecting the forwarding history evidence from its upstream and downstream nodes. Then TA could punish or compensate the node based on its behaviors. To further improve the performance of the proposed probabilistic inspection scheme, we introduce a reputation system, in which the inspection probability could vary along with the target node's reputation. Under the reputation system, a node with a good reputation will be checked with a lower probability while a bad reputation node could be checked with a higher probability. We model iTrust as the Inspection Game and use game theoretical analysis to demonstrate that TA could ensure the security of DTN routing at a reduced cost via choosing an appropriate investigation probability.

IMPLEMENTATION

CLIENT SIDE VALIDATION

Various client side validations are used to ensure on the client side that only valid data is entered. Client side validation saves server time and load to handle invalid data. Some checks imposed are:

- JAVASCRIPT in used to ensure those required fields are filled with suitable data only. Maximum lengths of the fields of the forms are appropriately defined.
- Forms cannot be submitted without filling up the mandatory data so that manual mistakes of submitting empty fields that are mandatory can be sorted out at the client side to save the server time and load.
- Tab-indexes are set according to the need and taking into account the ease of user while working with the system.

SERVER SIDE VALIDATION

Some checks cannot be applied at client side. Server side checks are necessary to save the system from failing and intimating the user that some invalid operation has been performed or the performed operation is restricted. Some of the server side checks imposed is:

- Server side constraint has been imposed to check for the validity of primary key and foreign key. A primary key value cannot be duplicated. Any attempt to duplicate the primary value results into a message intimating the user about those values through the forms using foreign key can be updated only of the existing foreign key values.
- User is intimating through appropriate messages about the successful operations or exceptions occurring at server side.
- Various Access Control Mechanisms have been built so that one user may not agitate uponanother. Access permissions to various types of users are controlled according to the organizational structure.

CONCLUSION

In this project, 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 mentioned. We have combined the proposed mention model with the SDNML change-point detection algorithm [3] and Kleinberg's burst-detection model [2] 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.

FUTURE ENHANCEMENT

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-anomaly based 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] B. Parno, A. Perrig, and V. Gligor, "Distributed detection of node replication attacks in sensor networks," in *Proc. IEEE Symp. Security Privacy*, 2005, pp. 49–63.

[2] H. Balakrishnan, M. F. Kaashoek, D. Karger, R. Morris, and I. Stoica, "Looking up data in P2P systems," *Commun.ACM*, vol. 46, no. 2, pp. 43–48, 2003.

[3] Y. Zhang,W. Liu,W. Lou, andY. Fang, "Location-based compromise tolerant security mechanisms for wireless sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 2, pp. 247–260, Feb. 2006.

[4] S. Zhu, S. Setia, and S. Jajodia, "LEAP: Efficient security mechanisms for large-scale distributed sensor networks," in *Proc. 10th ACM CCS*, Washington, DC, 2003, pp. 62–72.

[5] R. Anderson, H. Chan, and A. Perrig, "Key infection: Smart trust for smart dust," in *Proc. 12th IEEE ICNP*, 2004, pp. 206–215.

[6] M. Conti, R. D. Pietro, L. V. Mancini, and A. Mei, "A randomized, efficient, and distributed protocol for the detection of node replication attacks inwireless sensor networks," in *Proc. 8thACMMobiHoc*,Montreal, QC, Canada, 2007, pp. 80–89.

[7] B. Zhu, V. G. K. Addada, S. Setia, S. Jajodia, and S. Roy, "Efficient distributed detection of node replication attacks in sensor networks," in *Proc. 23rd ACSAC*, 2007, pp. 257–267.