Discovering Emerging Topics in Social Streams via Link-Anomaly Detection

ABSTRACT:
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 user 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. We demonstrate our technique in several real data sets we gathered from Twitter. The experiments show that the proposed mention-anomaly-based approaches can detect new topics at least as early as text-anomaly-based approaches, and in some cases much earlier when the topic is poorly identified by the textual contents in posts.

EXISTING SYSTEM:
- A new (emerging) topic is something people feel like discussing, commenting, or forwarding the information further to their friends. Conventional approaches for topic detection have mainly been concerned with the frequencies of (textual) words.

PROBLEM DEFINITION:
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.

PROPOSED SYSTEM:

- In this paper, 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 mentionee.

ADVANTAGES OF PROPOSED SYSTEM:

- 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.
- The proposed link-anomaly-based methods performed even better than the keyword-based methods on “NASA” and “BBC” data sets.

SYSTEM REQUIREMENTS:

HARDWARE REQUIREMENTS:

- System : Pentium IV 2.4 GHz.
- Hard Disk : 40 GB.
- Floppy Drive : 44 Mb.
- Monitor : 15 VGA Colour.
- Ram : 512 Mb.
IEEE Projects 100% WORKING CODE + DOCUMENTATION+ EXPLANATION – BEST PRICE
LOW PRICE GUARANTEED

SOFTWARE REQUIREMENTS:

- Operating system : Windows XP/7.
- Coding Language : JAVA/J2EE
- IDE : Netbeans 7.4
- Database : MYSQL

REFERENCE: