European Journal of Molecular & Clinical Medicine
ISSN 2515-8260 Volume 07, Issue 07, 2020

A Survey Of Event Detection Techniques In Online Social Media Networks

Ms. S. Chandra Kala[1], Dr..S.Albert Antony Raj[2]

[1,2]*Associate Professor, Department of Computer Applications, SRM Institute of Science and Technology Kattankulathur – 603 203, India*

ABSTRACT - The online social media networks have become an important source for detecting real-world happenings in this modern digital world. Event detection has become an essential research topic nowadays because it illustrates different scenarios during crisis or events as it contains significant information about them. Several approaches exist for detecting and analyzing events which provide valuable information for various applications like disaster and crisis management, detecting disease outbreak, health monitoring, traffic management and opinion mining. Traditional event detection methods that are proposed for handling large, formal and structured documents are not suitable for social media networks due to its unique characteristics such as restricted length, unstructured phrases, time sensitivity and massive information. This article provides a survey of methods used for detecting event in online social media and also gives an outline of challenges that are faced in handling social media post.
Keywords: Event Detection, Online Social media networks, Twitter.

1. INTRODUCTION

One of the most defining phenomena of the present times reshaping the world is the worldwide accessibility to the internet. According to Statista, the widespread use of social media is such that the number of global users is estimated to reach nearly some 3.02 billion monthly active social media users around 2021, nearly a third of world population. By 2022, it is expected nearly 750 million of these users are from China and nearly a third of a billion from India. On-line social media networks, comes in many forms, including blogs, web forums, newswire, photo-sharing platforms, social gaming and micro-blogs. They form a powerful means of communication for people looking to share and exchange information on a wide variety of real-world event as they unfold. With the increasing number of real-world events that are originated and discussed over social networks, Researchers are showing much interest in automatically detecting real world happening, to analysis and describe the happenings of an event like pandemic disease outbreak, flood disasters, monitoring traffic control, news events like electoral polls, fires, etc. However, the traditional text mining techniques to event detection on large text streams are not applicable to online social media data streams because of the restricted size of message, the huge number of spelling and grammatical mistakes, and the frequent usages of informal and multiple languages. This paper presents an overview of challenges that are faced in handling social media data and describes various approaches used for detecting events in online social media networks

2. CHALLENGES

Detecting events in social media networks requires considerably different approaches when compared to traditional media. Social media data is generated at huge volumes and speed than traditional media sources. Moreover, social media post is made up of short, noisy and

unstructured data that essentially requires different methods to solve information retrieval problem.

Voluminous data.

Larger volume of data demands high powered computing algorithm and huge storage space for storing, accessing and processing all data within restricted time period. For instance, every day Twitter platform generates millions of tweets. So, to process such large voluminous data, event detection techniques must have some dynamism and suitable execution environment so that it runs without any interruption even after rapid increase in social media post due to the happening of some bursty events.

Short Length.

Majority of social media platforms restrict the size of posts. For instance, Twitter restricts the users to post tweets that should not exceed 140 characters. Similarly, Picasa comments are also restricted to 512 characters and Windows Live Messenger restricts their user status messages to 128 characters. When compared to standard text which contains lots of words and their resulting statistics, short messages have few phrases or sentences. They may not provide useful context information for efficient similarity measure which forms the base of many text processing techniques.

Unstructured Phrases.

Traditional media outlets are well-structured, and edited news releases, whereas social media posts has large amounts of meaningless messages,polluted and informal content, irregular, andabbreviated words, large number of spelling andgrammatical errors, and improper sentence structuresand mixed languages. In addition, in social networks,the distribution of content quality has high variancefrom very high-quality items to low-quality, sometimesabusive content, which negatively affect theperformance of the detection algorithms.

Abundant Information.

In general, most social media platform contains a rich amount of information sharing options in addition to the content itself. For instance, Users of Twitter platform use "#" symbol, called hashtag, to present topics in a Tweet; an image normally contains many labels which are characterized by different regions in the image; social media users share information with others (link information). Previous textanalytics sources most often appear as <user, content>structure, while the text analytics in social media is ableto analyze data from different aspects such as user, content, link, tag, timestamps and others.

Diversified data sources.

Online social media networks have given tremendous contribution to huge outburst of diversified data composed of unstructured textual data, audio, images, video, multivariate records and spatiotemporal data. Thus, one of the problems encountered in real time event detection is to determine what data are really useful for the event detection under study and which technique must be applied to process the data from selected sources.

Authenticity and missing data.

The inaccuracy and incomplete nature of social media data should be addressed in event detection algorithms. For instance, location or position information given as longitude and latitude is most likely to be missing or incorrrect. The information about environmental activity may not have expected confidence level. Hence, event detection techniques, while detecting the events in social media platforms must take into consideration the problems such as confidence level, incompleteness and inaccuracies.

3. EVENT DETECTION IN TRADITIONAL MEDIA

Document-Pivot and Feature-Pivot techniques are most widely used method to detect event in traditional media data sources. The following sections describe the working of these techniques.

3.1 Document-Pivot Techniques:

Document-pivot techniques cluster the documents by applying their textual similarity to detect events; these techniques assume all documents are relevant and consider each document to have events of specified interest. The noisy and unstructured nature of social media, scale and processing restrictions in speed make document-pivot techniques irrelevant method for detection event in social media.

3.2 Feature-Pivot Techniques:

Feature-Pivot techniques models an event as a bursty activity, as most features raises sharply in frequency as the event emerges. Similar to document-pivot techniques, feature-pivot techniques are not dealing well with noisy data, as a result event may not be properly detected. In addition, not all bursts are events of interest, some ones may not be noted as they may occur without explicit burst activity.

4. EVENT DETECTION IN ONLINE SOCIAL MEDIA DATA NETWORKS

Based on the availability of the data, Event detection methods can be generally categorized into unspecified or specified techniques. Specified event detection relies on prior data or features that are well about the event like venue, type, time and short description. We can adapt traditional information retrieval and extraction techniques for detecting specified event. But, when we do not know anything about the type of event that we are interested to detect, majority of traditional methods are useless. Hence unspecified event detection algorithms depend on temporal signals of online data streams. The methods that are used to detect unspecified event include listening to trends or bursts in data streams, clustering features with similar trends, and categorizing events into different categories. The following subsections describe the techniques for unspecified and specified event detection.

4.1 Specified Event detection

Specified event detection includes known or planned social events. The information about these events could be given partially or fully with the related content or metadata details like time, venue, location and performers. This section describes various approaches that used for detecting specified events.

Sakaki et al.[1] studied a realtime detection of occurrence of earthquake by proposing a classifier applying semantic analysis on the tweets and modeling a spatiotemporal model for detecting events. Tweets are classified automatically based on keyword, statistical and word context features of a tweet using SVM as a classifier. In this paper, each social media user is treated as a sensor and the tweets are considered as the sensor information that has timestamp and the geographical location of the user. Event was detected by applying the probability density function of the exponential distribution which handles time-series data. Bayesian filters like Kalman and particle filters are applied for detecting location of the tweets.

Lee and Sumiya [2] have introduced a geo-social local event detection system, which is used to identify local festivals by monitoring crowd behaviors via Twitter. They have collected geo-tagged tweets for the specified region and applied k-means algorithm for dividing them into several regions-of-interest (ROI) by analyzing the geographical coordinates of the collected information. The authors concluded that an increased user activity like moving inside or coming to an ROI, when combined with raising number of tweets gives strong indication of local festivals

Benson et al. [3] proposed a structured graphical model where individual messages are analyzed, clustered and induced a canonical value for each event property. The model applies a conditional random field (CRF) component at the message level for retrieving field values such as artist name and location of the event. The authors biased the decisions outputted by the CRF to be consistent with canonical record values, thereby having consistency within an event cluster. The authors applied their technique using Twitter data stream to describe entertainment event record of city calendar division of NYC.com.

Popescu and Pennacchiotti [4] presented a method for detecting controversial events in twitter stream. The authors introduced a perception of a Twitter snapshot which contains target entity, a given period, and a collection of tweets about the entity for the target period. The authors first give a controversy score for each snapshot and then gives rank to snapshots based on the controversy score. The authors conclude that Hashtags are most relevant semantic features to detect the topic of a tweet and they found that controversy detection is influenced a lot by linguistic, structural, and sentiment features.

Pohl et al. [5] presented a method using social media platforms Flickr and Youtube for detecting crisis-related event. The authors consider the Geo referenced data as a vital information source for crisis management and hence they used a two phase clustering approach for event detection. In the first phase, term-based centroids are calculated using geo referenced data with a Self Organizing Map (SOM). In the second phase, by applying reassignment and the cosine distance measure, the assignment of best fitting data points to the calculated centroids are done.

Becker et al. [6] used query building strategies to detect planned events from Social media platform Twitter. They have filtered tweets related to an event by applying a variety of simple query constructing strategies where, each one uses different combinations of the event's context features, namely, title, description, venue, time, and location. They decided the final set of high-precision query strategy by asking an annotator to give label to the results returned by each strategy. Co-location techniques and term-frequency analysis are applied by authors for improving recall for locating descriptive event terms and phrases, which are then used to define new queries iteratively.

4.2 Unspecified Event detection

Unknown events include breaking news, emerging events, and general topics that create sensation among a large number of online social media users. This section gives an overview of different approaches that are applied for detecting unspecified event.

Phuvipadawat and Murata [7] developed an algorithm for detecting breaking news in Social media platformTwitter. It has two modules - story finding and story development. In the module story finding, messages are retrieved from the Twitter API by using pre-defined search queries and a new story is formed by grouping messages that are similar. Similarity between messages is found based on tf-idf with an increased weight for proper noun terms, hashtags, and usernames. In module story development, each news story is assigned with appropriate ranking through a period of time. Each news story is ranked by applying weighted combination of count of followers (reliability) and the count of retweeted messages (popularity) with a time adjustment for keeping the freshness of the message to rank every cluster.

Chen and Roy [8] proposed a method for performing event detection using Flickr photos by using the tags in the user's annotations. As every photo does not represents an event, authors use feature-pivot approaches for detecting event-related tags before actually detecting events of photos. The method is composed of three modules: In the module Event Tag Detection, the temporal and locational attribute of tag is processed to identify event related tags by applying Rattenbury approach of Scale-structure Identification (SI). A wavelet transform is used to remove noise. In the module Event Generation, authors distinguished aperiodic-event related

and periodic-event-related tags by analyzing the distribution patterns. Event-related tags are grouped into clusters with each cluster, representing an event, contains tags with similar locational and temporal distribution patterns in addition to similarly associated photos. A density based clustering method was used applied; In Event Photo Identification, for each tag cluster, photos corresponding to the represented event is extracted.

There are some research work that uses signal processing methods for detecting events from social networks. Cordeiro [9] presented a light weighted algorithm using Twitter data stream for event detection that uses wavelet signal analysis of hashtag occurrences. Hashtags are used for building wavelet signals, instead of individual words. The author has found that a sudden raise in the number of a given hashtag is a good indicator of an event that is occurring at a given time. Hence from the tweets all hashtags were extracted and then clustered in intervals of 5 minutes. Hashtag signals are constructed over time by finding the count of the hashtag mentions in each interval and then by grouping them into separated time series, one for each hashtag, and finally all tweets that mention the hashtag during each time series are concatenated. Adaptive filters are applied for removing noise, before using the continuous wavelet transformation for providing a time-frequency representation of each signal. The author used Peak and local maxima detection was applied for detecting an event within a given time interval.

Weng et al. [10] proposed a method based on clustering of discrete wavelet signals built from individual words generated by Twitter stream. When compared with Discrete Frequency Transforms (DFT) that is used for event detection in traditional media which is localized in frequency only, Wavelet transformations are localized in both time and frequency domain. Hence they are able to find the time and the period of a bursty event within the signal. Time-dependent variant of document frequency–inverse document frequency (DF-IDF) is used for signal construction, where DF records the count of messages that includes a specific word, while IDF uses word frequency up to the current time step. A sliding window is then applied to capture the change over time using the H-measure (normalized wavelet entropy). By applying signals cross-correlation, trivial words are extracted out which measure similarity between two signals as a function of a time lag. The left over words are then clustered using a modularity based graph partitioning technique, which splits the graph into subgraphs where each graph represents an event. Finally, significant events are identified by considering the number of words and the cross-correlation among the words related to an event.

Fang et al. [11] have presented a method that uses clustering algorithm for grouping tweets by considering multi-relations that exists between tweets which are calculated using different deatils such as hashtags, textual data and timestamp.

5. CONCLUSION

Online social media has become one of major research platform in recent years. Monitoring and processing this rich and dynamic flow of data generated by social media user can result in valuable information in a timely manner which would not have been possible in traditional data

This paper provides an outline of the problems that event detections techniques face when processing online social media compared to traditional media. It also presents a survey of algorithms proposed for detecting event in online social media platforms. Classification of these techniques is based on the target event type which may be unspecified or specified event detection.

Although extensive research is done towards event detection in social media, still there is a need of efficient and reliable event detection system that continuously monitors and analyzes the activities from different social media sources and provides better support for multiple languages, and enhanced summarization and visualization methods.

REFERENCES

[1] Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes Twitter users: real-time event detection by social sensors. In: Proceedings of the 19th international conference on World Wide Web. pp. 851–860(2010),

[2] Lee, R., Sumiya, K.: Measuring geographical regularities of crowd behaviors for Twitter-based geo-social event detection. Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks pp. 1–10 (2010)

[3] Benson, E., Haghighi, A., Barzilay, R.: Event Discovery in Social Media Feeds. Artificial Intelligence 3(2-3), 389–398 (2011)

[4] Popescu, A.m., Pennacchiotti, M.: Detecting controversial events from twitter. In: Proceedings of the 19th ACM international conference on Information and knowledge management - CIKM '10. p. 1873. CIKM '10, ACM Press, New York, New York, USA (2010),

[5] Pohl, D., Bouchachia, A., Hellwagner, H.: Automatic identification of crisis-related sub-events using clustering. In: Proceedings - 2012 11th International Conference on Machine Learning and Applications, ICMLA 2012. vol. 2, pp. 333–338 (2012)

[6] Becker, H., Chen, F., Iter, D.: Automatic identification and presentation of Twitter content for planned events. Proceedings of the Fifth . . . pp. 655–656 (2011),

[7] Phuvipadawat, S., Murata, T.: Breaking News Detection and Tracking in Twitter. In: 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. WI-IAT '10, vol. 3, pp. 120–123. (2010)

[8] Chen, L., Roy, a.: Event detection from flickr data through wavelet-based spatial analysis. Proceedings of the 18th ACM conference on Information and knowledge management pp. 523–532 (2009),

[9] Cordeiro, M.: Twitter event detection: combining wavelet analysis and topic inference summarization. In: the Doctoral Symposium on Informatics Engineering - DSIE'12 (2012)

[10] Weng, J., Yao, Y., Leonardi, E., Lee, F., Lee, B.s.: Event Detection in Twitter. Development (98), 401–408 (2011)

[11] Fang, Y., Zhang, H., Ye, Y., Li, X. Detecting hot topics from Twitter: A multiview approach, Journal of Information Science, vol. 40, no. 5, pp. 578-593, 2014.