

Detection and Visualization of Events from Online News

Swapnil Khatdeo, Sanskruti Shrawane, Pranjali Kumbhare, Prof. M.S. Nimbarte
Department of Computer Engineering
Bapurao Deshmukh College of Engineering,
Sevagram, Wardha-442001

Abstract—Online News is now the easiest way to know and understand the real-world events. However, it is difficult for users to hold all details of big events. In addition, how to inform human to know about important rare events, which may trigger the sub-sequent emergency events, remains to be a challenge. To address these issues, we present Detection and Notification from Online News. Firstly, we introduce an effective approach that detect event from online data or news which combines topic modelling and Chance Discovery, and detects events more effectively by coupling multiple term-relations. Secondly, we propose a heterogeneous event-graph layout algorithm which takes the significance and visualization of events into consideration by leveraging latent co-occurrence relations to represent important rare events and thus enhance human cognition.

I. INTRODUCTION

Data mining has gained a lot of attention in recent in recent year especially with the advent of big data. With the fastest development of internet, the online data access is containing the great growth. Especially, online news has become rapidly growing due to the increasing amount of information available in electronic form such as World Wide Web.

The news event is defined as a specific thing happens of a specific time and place, which may be consecutively reported by online news or data. The rich or big information within data can access to revel some important topics or growth of certain social phenomena. Beside, it can also be exploiting for detecting some emergency event or nature disasters [10]. In short, it is easier to human to detect and visualize the real world events from online news. We can define a news event (Something happens at specific time and place) of interest to the (news) media [14].

Event detection and visualization of various real world events from an online news has attracted significant attention because it can help the decision maker to detect meaningful events and show it effectively. It provides service to the user who retrieve the information as well as send the notification to prevent and aware from various event from online news [13]. Here gives some example: the latent omens such as the abnormal behaviours of some animals may reveal that the disasters such as earthquake will occur soon; the early incubation of the diseases may trigger the subsequent cancer. In March 11th 2011, an earthquake with magnitude nine occurred in Japanese northeast coast, which was the largest earthquake since observation record begin. However, the cause of this accident were some important rare events such as nuclear leaks and tsunami [10].

In addition, how to design the proper system that detect and visualize the various events from online news. As we know the online news or data should be providing important and meaningful information as per the requirement of users. Also the motivation of user about the rare events is important consideration. Therefore, two important question raise: (1)

how to detect events, and (2) how to visualize the events meaningful for human to perceive the important rare events.

To examine that question we are developing the proper system that detect the events from various news sites also some other sources and provide that data in effective manner to the users. We are developing the application that provide the alert or notification that gives the information and motivate human about the important rear events.

Organizing of the paper in this as section II describes the literature review, Section III provides the proposed work, Section IV provide conclusion and last section provides the references.

II. RELATED WORK

Mr. Allan et. al proposed the system that extracting knowledge from any meaningful text automatically. They provide services to the users who retrieve the information by firing query on internet. They provide some extracting data while firing query on internet. They also present an emerging technique used for data extraction from web page called Top-K web pages that describe stop-k instances. [2]. Hongin Zhu et. al proposed system of data mining based on classifier in class label prediction of coal mining data. This specify the main goal toward the coal mining, it should provide the basic idea but the classifier with various technique. The system provides a powerful class label prediction tool as well as increasing knowledge of data classification models [7]. The framework used Event Detection method include Hybrid Event Detection method and Chance Discovery as well as Event Graph Layout algorithm to detect and visualize the events from online news Chen Zhang [13]. Takshi Sakaki et. al proposed model detects the earthquake shake by tweets of the people. Semantic analysis was applied to tweet to classify them into a positive and negative class. He considers twitter users as a sensor and set a problem to detect an event based on sensory observation. It can be detecting the earthquake shake by tweets of the people [10]. Allocation aware news feed system that provide a new platform for its users to get spatially related message updates from either their friends or favourite news sources. Wenjian Xu et. al proposed, a

location-aware news feed (LANF) system that generates news feeds for a mobile user based on her spatial preference (i.e. Her current location and future locations) and non-spatial preference (i.e. Her interest). Existing LANF systems simply send the most relevant geo-tagged messages to their users. A location-aware news feed framework takes the relevance and diversity of news feeds into account when scheduling news feeds for moving users [8]. Fanner et. al proposed the evaluation of microblogging and social network services opens up great opportunities for various kind of knowledge based intelligence activity which required tracking of real time events [3]. Bryan k. et. al proposed model of twitter reporter breaking news detection and visualization through the geo-tagged tweeter network [5]. To improve the real-time, situation-aware and event triggered monitoring is become important issue. Srivathsan Srinivasagopalan et. al proposed multi-agent-based CEP (Complex Event-Processing) techniques can become powerful tools for real-time, situation-aware, event triggered monitoring of highly-scalable distribute cyber-physical systems such as power-grids. Our techniques have established with formal rigor that distributed active agents can be deployed in near-real time in response to events that have the potential to scale to massive numbers [6]. S.C.Gowri et. al proposed text mining techniques are used to extract the valuable information from the raw text data and then integrate to build a structured database. Information extraction (IE) is the task of automatically extracting structured information from unstructured or semi-structured text. In this paper discussed the various methods for document representation used a full inverted index as the basis for the operation on string vectors. It is better than performing traditional approach to represent the document by minimizing document pre-processing time and feature dimensionality also it provides the potential easiness for tracing why each document is classified under the category. In this research work to suggest the encode documents into string vectors by extracting phrase than into numerical vectors [9]]. Konstantina Papanilaou et. al proposed a platform for automated data processing in the context of computational journalism. he proposes a general methodology for event extraction from different data sources [12].

III. PROPOSED WORK

However, it is increasingly difficult for user to hold the full picture of big events comprehensive perspective. In additional how to motivate user to perceive the important rare events, which may trigger subsequent emergency events, remains to be challenge in the existing system. There are some other challenges in the existing Event Detection Technique (ETD) system. They neglect the important rare event and detect incomplete information as well as co-occurrence of relation which are hidden from noisy and other useless data collection. In previous system they do not provide any reminder or alert

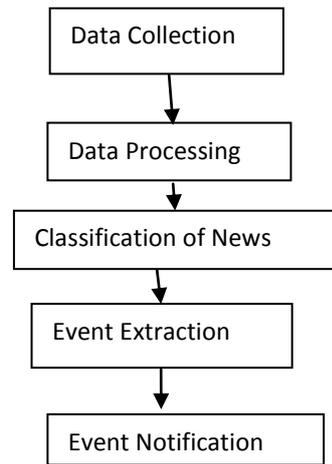


Fig. 1. Block Diagram of Proposed Methodology

to the user from various rare events and aware that events happens in the real world.

To motivate the user for identify the rare events in the world we create effective system to detect a visualize the events from online news.

The proposed method attempts to signify the arteries and veins from the patients' retinal image. In the proposed system we are developing a proper system that easier to human to browse, review and understand the real world. An events consider as something non-trivial happening of specific date, time and location. We are detecting some important rare events such as social phenomena, natural disaster, emergency events, entertainment, political, etc. as per the user requirement and sort according to category also visualize that events using web mining or mobile application that provide information as a remainder or alert type for protection and understanding the rare events.

The Fig 1. shows, how the system will collect the news from different online sources. Then categories that information as per the user choices. The categories contain the different options such as Political events, Sports events, Social events, Technological events, etc. The system will notify the user about upcoming events that will be happen in the future. We are developing the system with the following new features: The news will be arranging in category The future events will be notify to the user The system will be mobile based so it will be available to more number of user. The system will aware to user about the emergency events. The system is easy to used.

IV. IMPLEMENTATION OF PROPOSED METHODOLOGY

The Detection and Visualization of Events from Online News is provide proper system that detect the various events from different web sites or data sources and visualize as per

the user's requirement in efficient structure. The proposed work goes through a number of stages like data collection, data pre-processing, classification of news, event extraction and finally event notification. The detail explanation of these processes are given in following subsections.

A. Data Collection

The news from the different sources as well as from Different web sites is should be collect and that news or data should be store in our database. Text data such as online news and some social media bear valuable insight regarding important rare events and response to such events. Event Detection and Tracking (EDT) on the text data has attracted significant research in machine learning, information revival and social media modelling [2, 5, 10, 13]. The detection contains the two parts: New Event Detection (NED) and Future Event Detection (FED). The NED aims at detection newly occurred events from online text streams [2]. The new events may focus on the news should be happens in that particular date. The FED aims at detecting the new that happens on the future date. The FED news is detecting from micro blogging and social network services opens up the great opportunities for various kind of knowledge based intelligence activity and about the emergency events which required tracking of real time event [3].

The data collection is also called "Event Detection" that should be detect the various events from "Main Stream News" and "Blogs".

B. Data Pre-processing

The detected events as well as collected news are stored in the database. This unprocessed news from the database are improved in visibility using pre-processing. This pre-processing stage improves the quality of unprocessed rare events and some news. Pre-processing stage is used to remove the noise and eliminate irrelevant information. Firstly, the 'extraction of date references' is pre-processed i.e. it extract the date wise events or various news from different data sources. Then 'cross lingual article matching' and finally, 'detection of article duplicate' ??.

C. Classification of News

The classification of news is a method that used various algorithm to classified the various news detect from various data sources. The classification of news is a process of partitioning a various news event as per category i.e. political, sports, entertainment, Hollywood, Bollywood, country etc. For encryption purpose and provide the consistent data the "AES algorithm" is used that the AES algorithm is used to encrypt the data to increased the security of the system. As user need to register in the application, we used AES to encrypt user information to secure the user personal information. There is also "scheduling algorithm" is also used. The number of news is encounter in the online news world wide web updated and extracted after specified time provide

in the algorithm. Synchronization is done in after some period of time continuously update the news contents.

A classification of news contains the same technique i.e. Data synchronization and JSON. Data synchronization contain the process of establishing consistency among data from a source

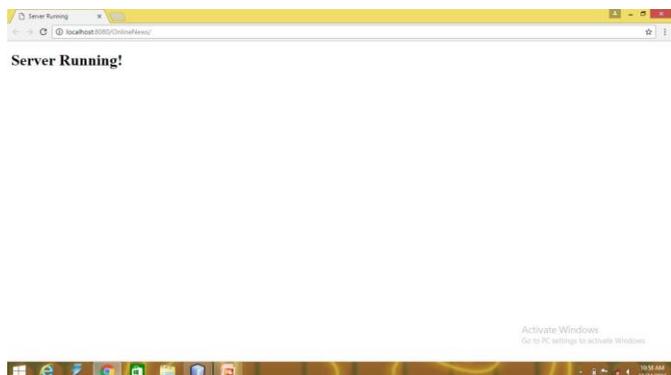


Fig. 2. Implementation of Proposed Methodology: Sample Scene of Server Output.

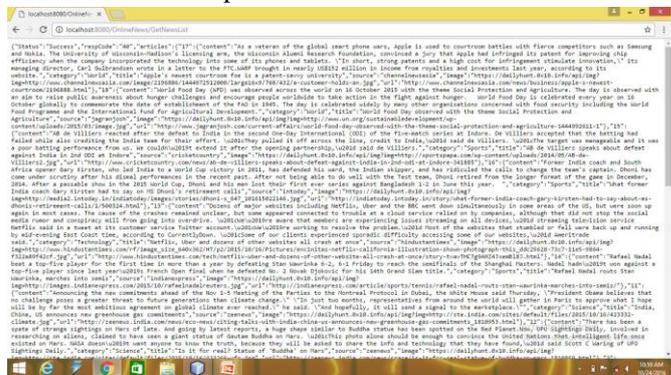


Fig. 3. Implementation of Proposed Methodology: Sample screen of Server Output in Row Data to a target data storage and vice versa and the continuous harmonization of the data over time. Sync (): This method is used to extract the news from various online news web sites. The overall data that should be extracted is in the row form that is not suitable for use that for that purpose we used "JavaScript Object Notation" technique. JSON is a minimal, readable format for structuring data. It is used primarily to transmit data between a server and web application, as an alternative to XML.

V. EXPERIMENTAL RESULTS

The output of this method is the extraction of different news from the News website and then store the extracted news on server database. After that it is converted by using Jason parser online software into the text format and finally it shown in the tabular format using MySQL.

In Figure 2, the server output is shown graphically. The GUI is created by using tools in NetBeans. In Figure 3, show

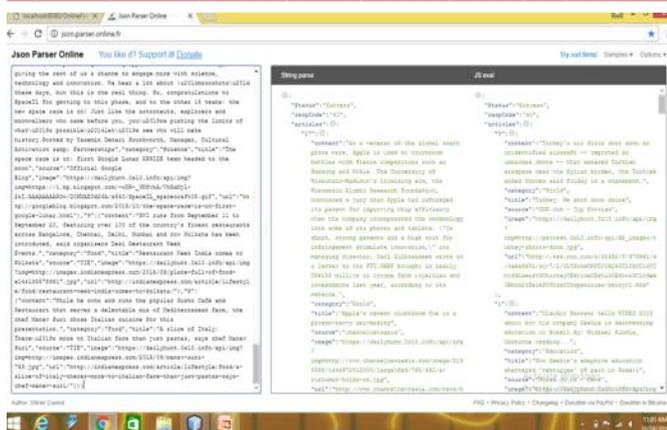


Fig. 4. Implementation of Proposed Methodology: Sample screen of JSON Parser.

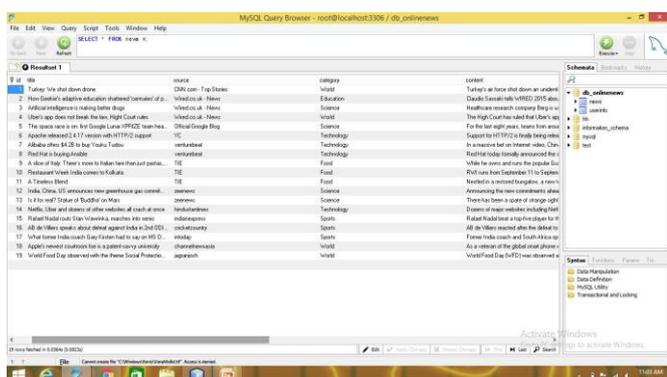


Fig. 5. Implementation of Proposed Methodology: Sample screen of Categorise Data the detected data from the news site. Here we are used the data source as a "daily hunt" site for detect the various news on particular date.

In Figure 4, it shows the data in proper form that should be suitable for convert, for that purpose we used the JSON parser that convert the detected row data in readable form. JSON is a minimal, readable format for structuring data. It is used primarily to transmit data between a server and web application, as an alternative to XML.

In Figure 5, at last it can show the category wise data. It shows which type of news, time, location and much more things related with it.

VI. CONCLUSION

In this paper we discussed the way to detect and visual-size the events from online news. A preliminary experiment demonstrates the superiority of our way by comparing with several benchmarks. However, the experiment is limited to relatively small datasets. Also, we can visualize the news related to our interest. For easier way of accessing, we proposed mobile application which gives the notification about rare events and future events. Our system is able to detect the natural disasters like earthquake, Tsunami, etc. This system aware and motivate the people to perceive the important rare events.

In addition, for further case studies on validating the visualization result and its function of supporting decision making should be conducted. Besides the event graph interaction needs to be improved for better visual exploration. Lastly, further work can focus on the extension on using the incremental clustering for new event detection task to fit the needs in the era of Big Data.

REFERENCES

- [1] Rozenshtein, P., Anagnostopoulos A., Gionis A. And Tatti N. Event Detection in activity networks. SIGKDD, pp. 1176-1185, 2014.
- [2] Allan J., Papka, R. And Lvrenko V. Online new event detection and tracking. SIGIR, pp. 37-45,1998.
- [3] F. Wanner, A. Stoffel, D. Jackle, B. C. Kwon, A. Weiler and D. A. Keim. State of the art report of visual analysis for event detection in text data stream, 2014.
- [4] Ommera Jan , Heena Khana. An Analysis of Miss-classification Error Detection Using Data Mining Techniques, 2015.
- [5] Bryan K., Santosh Y., Kim B. Twitter reporter: Breaking news de-tection and visualization through the geo-tagged twitter network 26th international conference on Computer and its application, 2011.
- [6] Srivathsan Srinivasagopalan, Supratik Mukhopadhyay a Complex-Event-Processing Framework for Smart-Grid Management, 2010.
- [7] Haixa Xi, Dan Cyo and Hongin Zhu, Application of data mining based on classifier in class label prediction of coal mining data. International journal of Security and its application, 2015.
- [8] Wenjian Xu, Chi-Yin Chow, Member, IEEE, A Location- and Diversity-aware News Feed system for Mobile Users, 2009.
- [9] S.C.Gowri, Dr. K.Meenakshi Sundaram, A Study on Information Re-trieval and Extraction for Text Data Words using Data Mining Classifier, 2010.
- [10] Takeshi Sakaki, Makoto Okazaki, Earthquake shakes Twitter users: Real-time Event Detection by Social Sensors, 2016.
- [11] Giridhar Kumaran and James Allan, Text Classification and Named Entities for new event detection, 2014.
- [12] Konstantiana Papanikolaou, Haris Papageorgiou, Nikos Pappasantopou-los, Detecting protest Events in media outlets and Twitter, 2008.
- [13] Chen Zhang , Hao Wang, Cuixia Ma. Event Panorama: A Framework for Event Detection abd Visualization from Online News, IEEE 49th Hawaii International Conference on System Science, 2016.