

Study Of Mining Techniques Used In The Recovery Process After The Occurrence Of Catastrophic Events

G.Kalpana, Dr.J.Chockalingam

Abstract: Catastrophic events are frequent occurrences in day to day life. Even the occurrence of a disaster is forecasted and action is taken, it is very difficult to recover the affected area after that event. Rapid recovery is the only solution used for restoring the normalcy in those areas. In order to start the recovery process, communication must be established to know the status of the areas. Crisis Management is a popular activity deployed after the occurrence of disastrous events and tries to establish a crisis response through proper communication. This paper presents a study of mining techniques that assist crisis management for establishing communication through the Internet with the aid of websites, blogs, and social networks.

Index Terms: Big Data Analytics, Catastrophic events, Classification, Clustering, Crisis management, Data Mining, Machine Learning

1 INTRODUCTION

Catastrophic events are the frequent occurrence in day to day life. There are different types of such events include earthquakes, hurricanes, floods, tsunamis, and tornadoes, etc., and disturbing the normal routines of individuals, communities. A disaster produces innumerable loss and impacts over the affected areas [1]. Even the occurrence of a disaster is forecasted and action is taken, it is very difficult to recover the affected area after that event. Rapid recovery is the only solution for restoring the normalcy in the affected areas. In order to start the recovery process, first communication must be established to know the status of those areas. Before the ages of the Internet, telecommunications, mass media devices like televisions and newspapers played a vital role to convey situations from the affected areas of a disaster and each type of communication used itself is having drawbacks. The telecommunication might be broken up during the happening of a disaster and television broadcasting involves one-way communication. The broadcasting can also be disconnected based on the environmental conditions like power failure etc., The newspaper conveyed the late news with a minimum period of one day after the disturbance of a disaster. Although the field of Science and Technology are having great development, the management of disaster was still a question mark on that stages, because there were no archived data maintained for producing affective responses [2]. After the emergence of the Internet, all the media are converted into electronic format and passing the messages to the outside world through their respective web sites. In addition to the publishing of news regarding a topic, the electronic channels also began to archive the collection online. In the initial stages of the Internet, users have to only view the published articles and came to know about a fact and communicated with others through email and chat services.

After the beginning of blogs and forums etc., the users started to discuss a topic within a website. The advent of Web.2.0 technology made a great impact in the field of the Internet and allowed the users to produce their own data on the web. As a result, Social Networks like Facebook, Twitter, etc., were appeared and enhanced the two-way communicative process. The commencement of smartphones in the field of mobile technology gave a newer dimension to the Internet including blogs, forums and Social Networks with location-based services [3]. Instant messaging services like WhatsApp were also introduced on smartphones to provide instant communication across the world. Now, the entire field of the Internet made the people connect together by creating a virtual world [4]. So, today it is very easy to communicate with any part of the world to any extent by taking the mobile anywhere with the support of Internet technologies. A person in the catastrophic can publish a photo, text, audio or video to explain the current situation. Such kinds of data are very useful for the recovery team to take needed action in time. Though the improvements on the Internet, connect the people on the world from anywhere the nature of data collected from the Internet is in an unstructured format and the size of the data is also huge in the count. The size is subjected to grow every second because people are continuously producing their own data through blogs, forums, and Social Networks, etc. Newer methods have been developed by combining the field of Data Mining, Big Data Analytics and Machine Learning for collecting and extracting the knowledge from that unstructured data. Different techniques have been adopted in each field for the purification, extraction and discovering of knowledge from the given data. The present work analyzes techniques used for extracting the knowledge by combing the Data Mining techniques, Machine Learning and Data Science techniques used in the field of the recovery process.

2 CRISIS MANAGEMENT

Crisis Management is a primary system used in the recovery process of catastrophic events. The system uses the data obtained from the Internet as the basis for a high degree of the recovery process and applies the mixture of mining, machine learning and data science techniques to get an exact knowledge about the data collected. Crisis Management involves three phases namely, preparedness, response, and recovery [5]. The preparedness stage is taking the needed actions like training,

- G.Kalpana is currently pursuing full-time Ph.D. program in computer science in Khadir Mohideen College, Adirampattinam, Affiliated to Bharathidasan University, Trichy, Tamil Nadu, India PH-9361335223. E-mail: kalpanakmccs@gmail.com
- Dr.J.Chockalingam is working as an Associate Professor in the PG & Research Department of computer science at Khadir Mohideen College, Adirampattinam, Affiliated to Bharathidasan University, Trichy, Tamil Nadu, India, PH-9786734456. E-mail: chock_abipkt@yahoo.com

advising the people, etc. The response contains activities that are reactions taken for the impact of the already occurred disaster and stabilizes the situation. Recovery is giving importance to the steps needed for reestablishing the damaged infrastructure. In all three phases, data-centric websites are getting importance to effective crisis management. Today, people from disastrous event areas are publishing their area status by using the mobiles along with any one of the interactive websites like blogs, forums or Social Networks. The published post of a user from a Social Network reaches the outside world fast than the mass media. Most of the emergency agencies, government and non-governmental organizations refer such websites for extracting the news related to a disastrous event to plan for better crisis management.

3 DATA MINING

Data mining is the process of pulling out meaningful information from the huge volume of data [6]. Data Mining analyzes the observed results and discovers the veiled knowledge [7]. Traditional mining techniques are using data from the local databases and various techniques like Association rule mining, classification, and clustering, etc. The Internet is having a vast amount of streaming and unstructured data which cannot be mined easily with the traditional techniques. Moreover, text(nominal) data appear at a high degree on the Internet and keep moving continuously through websites and social networks. The field of crisis management needs grouping of data from the various resources of the Internet and applies clustering and classification techniques with machine learning and data science techniques etc. The following sections address a foundation of clustering and classification techniques.

3.1 Classification

Classification is the most used mining technique and classifies the samples with known labels. The technique is also used in Machine Learning and is referred to as a supervised learning technique because it classifies the data with known classes. The concept of classification starts with the preparation of training and test data for the building of a model over the pre-defined classes and assign new data to a class, based on the model built. In crisis management, classification is mainly used for analyzing the occurrence of incidents during the timing of disastrous events. Different techniques are doing classifications of which the decision tree is the best one for doing the classification. The overall classification is displayed in Fig.2.

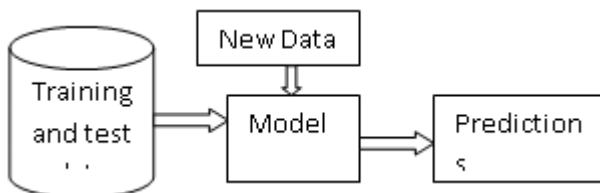


Fig.1 Classification Technique.

The outcome of the model is the prediction in Fig.1 and is to be evaluated for the performance. There are variations of classification techniques used today to mine the data in the field related to crisis management. Techniques like Linear

Discriminant Analysis (LDA), Logistic Regression and Naïve Bayes, etc., are doing linear classification and Artificial Neural Networks (ANN), K-Nearest-Neighbors (KNN), Decision Tree, etc., are the non-linear classifiers used for classifications. The decision tree is the better classifier that performs well classification than other techniques and the Support Vector Machine (SVM) is the next classifier well suited for both linear and non-linear classifications.

3.2 Clustering

Clustering the suitable technique for grouping the larger volume of unstructured data without knowing the classes(groups) in advance. In the field of Machine Learning, the technique is referred to as unsupervised learning. Clustering is used for acquiring a high amount of knowledge from the vast volume of raw data with redundancy and noise. The technique uses an intrinsic measure known as centroid for similar item detection without knowing the topic in advance. There are two types of clustering techniques namely, partitioned clustering and hierarchical clustering. Partitioned clustering is a technique that groups the data with a specified number of groups. The popular partitioned clustering technique is the K-Means method, which clusters the data with k-number of groups. The hierarchical clustering is generating a tree-like structure, instead of specifying k number of groups in advance. Agglomerative clustering the famous hierarchical method used for grouping similar points. Partitioned clustering is widely adapted for cluster the Internet data. Initially, the training and test data are selected after the preprocessing techniques. Next, the initial centroid is fixed with a number of iterations. After fixing the centroid, the K-Means groups the nearest points of assigned centroid and recalculates the centroid for the next group formation. The process is repeated until the assigning of all points to the related groups and attaining a condition that there are no more centroids. The steps involved in partitioned clustering is shown in Fig.2.

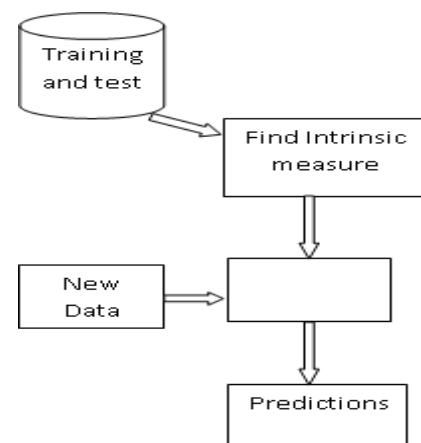


Fig.2 Steps in Partitioned Clustering

Clustering is the broadly used technique for grouping the set of users who produced the same data on the Internet. But the data produced by the users on the Internet is dynamic in nature, the basic clustering techniques used on the Internet is slightly modified as per the need of grouping. The intrinsic measure shown in Fig.2 involves any distance calculations like Euclidian distance or similarity measurements like cosine similarity or collaborative filtering. In case, the data is having a

vast number of features, dimensionality reduction is also applied before applying intrinsic measures for better clustering.

4 BIG DATA ANALYTICS

The field of Crisis Management maximum collects the data from the Internet. The data on the Internet is huge in volume, and hence the traditional Data Mining techniques need the help of Big Data to handle the vast amount of unstructured data from the Internet. Big Data Analytics is a field that discusses a larger amount of structured, semi-structured and unstructured data. Big Data Analytics involves 5Vs namely, Volume, Variety, Velocity, Veracity, and Value [8]. Big Data Analytics is mainly doing insights, used for better decision making. Big Data Analytics also analyzes the distribution and accessing of data from a huge amount of collection. Crisis management uses big data for the identification of needed resources and the status of logistics management, during emergency periods. Importantly, the field of big data is used for providing on-demand data related to the activity of the recovery process from various resources on the Internet.

5 MACHINE LEARNING

Machine Learning is the field of making the machines to learn from the existing data. According to Layman's, Machine Learning is of educating the computers on how to perform complex tasks that humans cannot know how to accomplish [9]. Machine Learning is mainly used for doing reviews of the selected data for making better decisions. Machine Learning builds a model by taking the available data or the portion from the huge volume of data by using the data mining techniques. Two types of data namely, trained and test data are used for training the model. After learning the model is subjected to predicting the newer results from the learned knowledge. The problem of recovering from the catastrophic event is becoming simple with the help of the Internet data, Data mining techniques, Machine Learning and Data Science. The combination of Data mining and Machine Learning is the most used method in crisis management.

6 DISCUSSIONS

Table.1 summarizes the methods used for the better and faster recovery process in the field of crisis management.

TABLE 1
METHODS USED FOR CRISIS MANAGEMENT

S.No	Technique	Description
1	Data Mining	Extracts the knowledge from the available data. Based on the need, the available techniques have been slightly modified. Classification and Clustering are mostly used for grouping in crisis management
2	Big Data Analytics	Handles the huge amount of dynamic data on the Internet and helps the mining for better grouping.
3	Machine Learning	Builds a model by applying the mining techniques and data from Big data sources for getting new knowledge with optimum memory and a high amount of speed.

The present paper discussed data mining, big data analytics, and machine learning are collectively used for handling the data in a better way for extracting knowledge and grouping the users of the Internet to do effective and rapid recovery process in the field of crisis management.

7 CONCLUSION

The present paper discussed a survey of mostly used Data mining techniques used along with Big Data analytics and Machine Learning for detecting the knowledge used for the effective management of Crisis Management to faster the recovery process in the disastrous area. Based on the need and the availability of data from the resources of Big Data, a better model is built with the aid of mining techniques for assisting the fast crisis management in the catastrophic event areas.

8 REFERENCES

- [1] TAO LI et.al , "Data-Driven Techniques in Disaster Information Management," ACM Computing Surveys, Vol. 50, No. 1, Article 1, Publication date: March 2017.
- [2] Chuanjie Yang et.al," Using Big Data to Enhance Crisis Response and Disaster Resilience for a Smart City", IEEE, International Conference on Big Data,2017.
- [3] K. Mohan Kumar and B. Srinivasan, "Point-of-Interest Based Classification of Similar Users by Using Support Vector Machine and Status Homophily", International Journal of Machine Learning and Computing vol. 9, no. 5, pp. 615-620, 2019
- [4] Melinda Laituri and Kris Kodrich, "People as Sensors of High Magnitude Disasters Using Internet GIS," Sensors 2008, 8, 3037-3055; DOI: 10.3390/s8053037 On Line Disaster Response Community
- [5] Daniela Pohl, " Social Media Analysis for Crisis Management: A Brief Survey ", E-Letter , STCSN, E-Letter Vol.2 No.1, 2013
- [6] Deepthy K et .al, " Survey on Data Mining Techniques to Enhance Intrusion Detection",2012 International Conference on Computer Communication and Informatics (ICCCI -2012), Jan. 10 – 12, 2012, Coimbatore, INDIA
- [7] Xiao- bin Wang et .al, "Review on the application of Artificial Intelligence in Antivirus Detection System", 2008 IEEE publications
- [8] <https://www.educba.com/big-data-vs-data-mining>
- [9] <https://www.educba.com/big-data-vs-machine-learning>
- [10] Ning Chen et.al, "THE ROLE OF DATA MINING TECHNIQUES IN EMERGENCY MANAGEMENT", In Proceedings of the 11th International Conference on Enterprise Information Systems - Artificial Intelligence and Decision Support Systems, pages 118-123 DOI: 10.5220/0001961601180123
- [11] R.Agrawal, "Mining association rules between sets of items in large databases", SIGMOD '93 Proceedings of the 1993 ACM SIGMOD international conference on Management of data
- [12] G.X.Intong et.al. "Brief survey of crowdsourcing for data mining", Elsevier, 2014.
- [13] P.Avesani, "Applying cbr to emergency management", New review of
- [14] Applied Experts Systems, 6:175-190,1999.