Review of Various Clustering Methods Used To Categorize Seismic Data into Earthquake and Mining Blast

KashyapAgravat, Madhushree B., Jasmine Jha

Abstract:- **Earthquake mainly produces P- and S- wave at the point of occurrence. Mining site that mines the geological material from earth often makes the quarry blasts on the surface of earth to accelerate the mining process instead of drilling. This mining blast also causes the same wave as the earthquake. It is very difficult to differentiate between two of them as the characteristics of waves produced by these two events are similar. Sometimes mining sites are located near to residential area and discrimination of earthquake and quarry blast data is important to analyze the geological activities to create awareness. Data clustering is one of the effective techniques that differentiate the data into a number of clusters where a cluster contains similar data.**

ISR is a Government organization, captures the real time seismic data, which also measure the mining blast data and vehicle vibration and produce the waveform. Segregation of earthquake, mining blast and vehicle vibration data is the key problem here. In this paper, we studied and analyzed the complexity and limitations of several clustering and other data discrimination methods that can be applied on seismographic database to discriminate earthquake and quarry blast data.

***Index Terms*—Clustering, seismological data clustering, similarity measure, data differentiation, clustering methods comparision**

I.INTRODUCTION

A Challenge in seismic monitoring is to uniquely discriminate between natural seismicity and anthropogenic events such as mining blasts. Two basic types of elastic waves are generated from seismic events like earthquake and mining blast namely P- and S- waves. The P-waves move in a compressional motion and the S-waves move in a shear motion perpendicular to the wave direction. These waves result in shaking of earth surface and may cause damages. It is good to focus on S- wave spectra because they have a good signal-to-noise (STN) ratio over much wider bandwidth than the S-wave spectra. Magnitude of lower earthquake and query blast may be same. Sometimes heavily loaded vehicles also cause high magnitude that can't be discriminated. In this paper, we focus on studying various clustering methods that can be used to classify seismic mixed data into earthquake and mining blast data.

Clustering is the process of grouping a set of physical or abstract objects into classes of similar objects. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters [1]. Clustering is widely used as one of the important steps in the exploratory data analysis. Clustering algorithms are used to find the useful and unidentified classes of pattern. Clustering is used to divide the data into groups of similar objects. The objects that are dissimilar are placed into separate cluster. Depending upon the metric chosen, a data object may belong to a single cluster or it may belong to more than one cluster. [2]

We studied several papers related to clustering techniques and provide comparative analysis of several clustering algorithms which can be helpful to differentiate earthquake and mining blast data that have same parameters and region of events. The paper is organized as follows: Section II is dedicated for the literature review. Section III gives a comparison of various clustering algorithms which is followed by Conclusion in Section IV.

II. LITERATURE REVIEW

In [2], Garima*et al.* discussed clustering techniques and divided them into five major categories namely: Partitioning-Based, Hierarchical-Based, Density-Based, Grid-Based and Model-Based.

In partitioning-based clustering methods, data sets are divided into a number of partitions. Later these partitions are referred as clusters. Each cluster or partition must have at least one object and no overlapping objects. In hierarchical-based algorithms, Dendogram – a tree of clusters, is constructed. This tree is based on medium of proximity. Grid based techniques are used in spatial applications where the large space is divided into number of cells. We have a large and continuous data set of earthquake, quarry blast and vibration of vehicles. So the methods like Density-Based, Grid-Based, Model-Based and traditional partitioning methods of mining are not applicable. Some improvements are required in these traditional methods.

T.Hitendra Sharma, P.Viswanath and B.Eswara Reddy introduced a fast approximation kernel k-means clustering method for large Data set in [3]. They presented their algorithm in three steps. First they used kernel based leaders clustering method to find a set of prototypes. The output of kernel based leaders clustering method are applied as an input to the k-means method of clustering which produces a partition of prototypes that are early generated. Finally, they get the partition of the entire dataset by replacing each prototype with its followers.

An improved k-medoids method for clustering large data set is more effective compared to traditional k-medoids

method, according to Danyang Cao and Bingre in [4]. They modified k-medoids clustering algorithm and constructed improved k-medoids clustering algorithm which is based on the clustering features of BIRCH algorithm. They preserve all the training data in a CF-Tree, and then they apply k-medoids to cluster the CF in leaf nodes of CF-Tree. Eventually they get k clusters from the root of the CF-Tree. The time complexity, scalability on large dataset and convex space of this algorithm is better as compare to k-medoids algorithm.

In [5], Huang Hanming, LI Rui and LU Shi Jun from Guangxi Normal University, Guilin, China, introduced a method to discriminate earthquakes and explosions using chirp-Z transform spectrum features. In their study, their overall spectrum layout is acquired by Fourier transform (FFT). Based on this overall layout seismic signals, suitable frequency range in the spectrum which contains most discriminative information is selected. Then the proposed Chirp-Z transform is applied to get finer resolution spectrum, and at last they achieved more accurate spectrum features corresponding to different types of signals. They mainly described an algorithm that employs CZT to derive dominant frequency and associated average energy. They selected 40 earthquake events and 40 explosion events which occur in a near neighborhood region in north China.

In [6], Qing *et al.* introduced a scalable clustering algorithm to discover frequently repeated trips from large scale, event-based telemation data sets collected via a satellite-based tracking system. They showed in their paper that the moving objects are challenging to analyze because of the enormous amount of data, the data quality and the approximate nature of the spatial data type. First, they indexed the trips based on a grid indexing method, and then compared only trips sharing the same grid neighborhood instead of an exhaustive pair-wise comparison of all the trips to get advantage of grid indexing are to significantly decrease the size of the data space needed to run the distance computation during the hierarchical clustering process.

In 2006, Yang Peijie*et al.* in [7], used Fuzzy Clustering approach for seismic data analysis. This Fuzzy Clustering method is useful for locating clusters embedded in background noise. Professor Jim Bezdek originally introduced this technique in 1981. [7] They presented that the algorithm attempts to partition a finite collection of elements into a collection of '*c'* fuzzy cluster with respect to some given criterion.

III. COMPARISON OF CLUSTERING ALGORITHMS

In this section we have analyzed various clustering algorithms. We find that some of clustering techniques can be useful to differentiate the seismological data like earthquake and mining blast. The comparison of various algorithms is represented in Table 1. The table also answers whether we can apply the algorithm on seismological data or not.

Table 1.Analysis On Several Clustering Techniques And Its Applications

Authors	Technique/ Clustering Method	Dataset	Advantages	Applicable to seismological data?	Reason
T.Hitendra Sharma, P.ViswanathB.Eswara Reddy	Kernal K-Means Clustering Method	Homogeneous Dataset	It is simple to implement and gives more reliable results as compared to traditional k-means algorithm	No	There may be some loss in the accuracy of clustering. For our data set, it isimpossible to define the threshold(t) as they had done.
Dalal, Harale	K-Means Clustering	Numerical Data (Crisp Data Set)	Large datasets are processed easily. It is simple to implement and results are easy to infer.	No	Sensitive to noise, Depends on initial value of k.
Danyang Cao, Bingre	Improved K-Medoids Algorithm	Homogeneous Dataset	It overcomes the drawbacks of the k-medoids algorithm, such as the time complexity, scalability on large dataset, incompetence to catch clusters of different size and the convex shapes.	No	Deriving CF-Tree from data set requires considerable effort and also data may be assorted during the implementation.
Huang Hanming, Li Rui, Lu Shi Jun	Chirp-Z Transform Spectrum Features	Homogeneous Dataset	It improves the accuracy of the solution to discrimination problem.	Yes	It is useful for limited number of readings. As the events and their readings are added, the accuracy and time complexity are reduced.
Qing Cao, BourchraBouquta, Patricia Mackenxie, Daniel Messier, JoshephJ.Salvo	Grid-Based Clustering Method	Real Time Homogeneous Dataset	Proposed algorithm significantly reduces the computational time required for clustering in Hierarchical clustering algorithm	No	Same pattern reorganization in earthquake and quarry blast pattern (data), More applicable on spatial data
Yang Peijie, Yin Xingyao, Zhang Guangzhi	Fuzzy Clustering	Training Dataset	This proposed method is simple and the algorithm is robust and has very good actual using value	Yes	By removing the constraint of limited number of parameters and cross plotting problem, it can be used to differentiate earthquake data from mining blast with some modifications. It may yield the best possible results.

IV. CONCLUSION

The field of clustering analysis, as one of the key technologies of data mining, has broad development prospects. This paper introduces several clustering methods that can be applied to differentiate large seismic data in ISR (Institute of Seismological Research) that contain same frequency magnitude of earthquake, quarry blast and vehicle vibration data.

By studying various methods of clustering we found that traditional clustering algorithm like simple partitioning, density based method, grid based and hierarchical clustering methods are not efficient enough to differentiate earthquake from vibration and explosion blast data. Fuzzy clustering algorithm may be useful by reducing its limitations of limited number of parameters.

V. ACKNOWLEDGMENTS

This work is supported by Institute of Seismological Research (ISR), Gandhinagar, India.

REFERENCES

[1] Han, Jiawei, and Micheline Kamber. Data Mining. Amsterdam: Elsevier, 2006. Print.

[2] Garima, Hina Gulati, P.K.Singh, "Clustering Techniques in Data Mining: A Comparison," 2nd International Conference on Computing for Sustainable Global Development (INDIA Com), IEEE 2015, pp. 410-415.

[3] T.HitendraSarma and P.Viswanath, B.Eswara Reddy, "A Fast Approximate Kernel k-means Clustering Method For Large Data sets," IEEE 2011, pp. 545-550.

[4] Danyang Cao, Bingru Yang, "An improved k-medoids clustering algorithm", IEEE 2010, Vol. 3, pp. 132-135.

[5] HUANG Hanming, LI Rui and LU Shi Jun, BIAN Yin Ju, "Discrimination of Earthquakes and Explosions Using Chirp-Z Transform Spectrum Features*", 2009 World Congress on Computer Science and Information Engineering, IEEE 2008, pp. 210-214.

[6] Qing Cao, BouchraBouquta, Patricia D. Mackenzie, Daniel Messier, Josheph J. Salvo, "A Grid-Based Clustering Method For Mining Frequent Trips From Large-Scale, Event-Based Telematics Datasets," Proceedings of the 2009 IEEE Internal Conference on Systems, Man, and Cybernetics San Antonio, TX, USA-October 2009, pp. 2996-3001.

[7] YANG Peijie, YIN Xingyao, ZHANG Guangzhi, "Seismic Data Analysis Based on Fuzzy Clustering," ICSP 2006 Proceedings, IEEE 2006.

First A. Author complited Diploma in Computer Engineering from Gujarat Technological Univerisity in 2011.Completed Bechlor of Engineering in Computer in 2014 from the same university.now currently studying in L.J. Institute of Engineering and Technology, Ahmedabad, under the Gujarat Technological University. And doing research work on data mining subject. Main aim to differentiate the earthquake data from noisy data like quarry blast data and vehical vibration data. got ISTE (Indian Society for Technical Education) & NIT (National Institute of Technology) National Award for Best Project during the bechlor of degree carrier, this award was given by Mr. OmPrakashKohli (Governor of Gujart) in 2015. Also got the i-SCALE Award (Innovation Student's Co-creation Award for Leadership and Excellence) in 2016, this award given by vice vice chancellor of Gujarat Technological University.

Second B. Author Jr.currently at the position of Assistant of Professor in L.J. Institue of Engineering and Technology, ahmedabad

Third C. Author.currently at the position of Assistant of Professor in L.J. Institue of Engineering and Technology, ahmedabad. Experience of industries as well as in education field. worked in patani computers, gandhinagar in past.