CLUSTERING METHODS TO DIFFERENTIATE SEISMOLOGICAL DATA INTO EARTHQUAKE AND QUARRY BLAST

Kashyap Agravat[1], Jasmine Jha[2]

[1] M.E. Student , Post Graduation Department (Computer Engineering), L.J. Institute of Engineering and Technology, Gujarat, India

[2] Assistant Professor, Post Graduation Department (Computer Engineering), L.J. Institute of Engineering and Technology, Gujarat, India

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. In this paper , I calculated several parameters from wave form of an event. Clustering is good idea to differentiate them. Experimental result show that simple partitioning methods are not useful to differentiate seismic data into mining blast and earthquake. While FCM method gives 90-95% accurate result to identify cluster over noisy data.

Keyword : - *seismic data clustering, data clustering applications, seismic data differentiation, application of clustering algorithms.*

1. INTRODUCTION

Modern digital seismographs may record seismic waves of earthquakes and significant explosions occurring sequentially or simultaneously. These sequential or simultaneous occurring characteristics would be harmful to properly explain the recorded seismic waves and might beget some false conclusions. So it is very meaningful to separate earthquake events and explosion events which may occur sequentially or simultaneously from recorded seismic waves. The separation of simultaneously occurring earthquake and explosion events is beyond the scope of our current researches. And should be investigated in further research by some special signal processing techniques such as independent component analysis [1]. 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 compression motion and the S – wave move in 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 P- wave spectra because they have good signal-to-noise (STN) ratio over much wider bandwidth than the S-wave spectra. Magnitude of lower earthquake and quarry blast is same. Sometimes heavily loaded vehicles also cause high magnitude that can’t be discriminated by visualizing their signals.

k-means, k-medoids, fuzzy clustering are well known data partitioning methods having simplicity and dealing with large amount of data as an advantages.

In this paper I applied some clustering methods by finding the several parameters of an events. We show experimental results of clustering techniques on seismic data discrimination, and conclude that the k-means and subtractive techniques of clustering is less appropriate in such discrimination where having overlapping data problem. While on other hand FCM has its advantages against simple partitioning methods and seismic data can be clustered with better accuracy of 90-95%.

2. DATA AND METHODS

The research area is selected in Saurashtra-Kutchh that is an active mining region with many types of mines and quarries that are also Gujarat's top producer of hard coal, rock and sands. However, the region also has prevalence of natural seismicity due to active faults in Surendranagar, Bhavnagar, Kutchh, Bhuj, Chotila and Junagadh. These factors make the region an ideal test bed for clustering techniques examinated to separate data from earthquakes and mining blasts, including time - of – day analysis, spectral ratios, and complexity analysis. In order to assemble a database containing natural and probable mining-related events, before the data selection step, we first determined the site of major mining and the sites of earthquake sources.

Pandhor Mining blasts that take place in Kutchh, Kutchh-Bhuj are ripple fire blasts in open pit, sand, rock mine. The geographic coordinates of the mining area was derived by satellite image from Google Earth program (http://earth.google.com) and from the seismological map of Kutchh-Saurashtra from ISR (Institute of Seismological Research), Gandhinagar. The information about blasting times on an average of forty times monthly and the average amount of explosives varying from 20 to 20.000 kg were obtained.

The statistical distribution of an events with Md$\leq$ 5.0, with a 150-km radius of the mining area between 2008 to 2013 that was recorded by ISR, in daytime and night time was investigated and effectively outlines regions of mining activity (where the dominant percentage of daytime events occur in regions of known mine locations). Figure1 reveals that an unusually high number of daytime events are seen in the Pandhro mining area that is a likely sign of blasting activity. The number of nightly events is very low compared to the one of daytime events.

Fig -1: Location of mining site in Google map near Pandhro (Kutchh) that produce rock, lignite(Coal) and sands.

In this paper all earthquake and explosion seismic wave records are came from ISR data centre. All the events occur from 2008 to 2013 year, magnitudes are between 0.5 and 5.5. Data formats are SEED. 139 seismic waves in vertical direction are used. Each wave correlates to a different event. Of the 139 events, 80 events are earthquakes; an other 59 events are explosions. The basic parameters of some these events are listed in table 1 and table 2.

Table 1: The basic parameters of 10 earthquake events

Date	Time	Magnitude	Lat	Long	Depth
21-01-2016	2:16 PM	1.1	23.184	68.707	27
11-06-2015	3:38 PM	1.6	23.02	68.704	8.4
11-05-2015	6:11 AM	1.8	23.192	68.703	9.9
11-02-2015	2:42 PM	1.8	23.165	68.712	9.1
20-09-2015	12:59 AM	1.3	23.191	68.710	6.1
09-11-2015	2:44 PM	2.3	23.148	68.603	6.8
08-08-2015	11:02 AM	1.8	23.169	68.604	6.8
22-06-2015	10:24 PM	1.5	23.048	68.617	13.1
03-03-2015	4:04 PM	2.3	23.073	68.594	9.7
01-05-2015	7:47 AM	1.6	23.246	68.449	15

Table2: The basic parameters of 10 explosion events

Date	Time	Magnitude	Lat	Long	Depth
19-06-2012	11:15 AM	3.2	23.685	68.754	11.1
14-11-2012	3:03 AM	1.2	23.688	68.758	21.7
17-11-2012	2:01 AM	1.7	23.695	68.755	3.1
26-11-2012	6:07 AM	1.7	23.684	68.759	1.7
12-04-2012	3:16 PM	2.1	23.681	68.810	8.4
18-12-2012	6:19 PM	1.8	23.663	68.753	11.8
28-12-2012	6:30 AM	1.8	23.680	68.715	6.5
01-04-2013	2:43 AM	1.7	23.698	68.759	11.1

01-11-2013	6:23 AM	1.4	23.694	68.698	6.1
01-11-2013	11:12 PM	0.9	23.701	68.732	6.1

3. EXPERIMENTAL RESULTS

3.1 Data Preparation

I extracted numbers of events from original data file. The extraction and conversation needed to findings the parameters to apply as an input of clustering techniques. Seisan is an effective seismic analysis tool that I used to prepare the data from waveforms. I first, extracted the data base from file. Than merge the same station file into one. Finally, using the same tool, seisan, converted into txt format from ASCII. It gives amplitude of each sensor that measured independently from each other. Applied these amplitude values as input of Matlab graph to generate constant wave form.

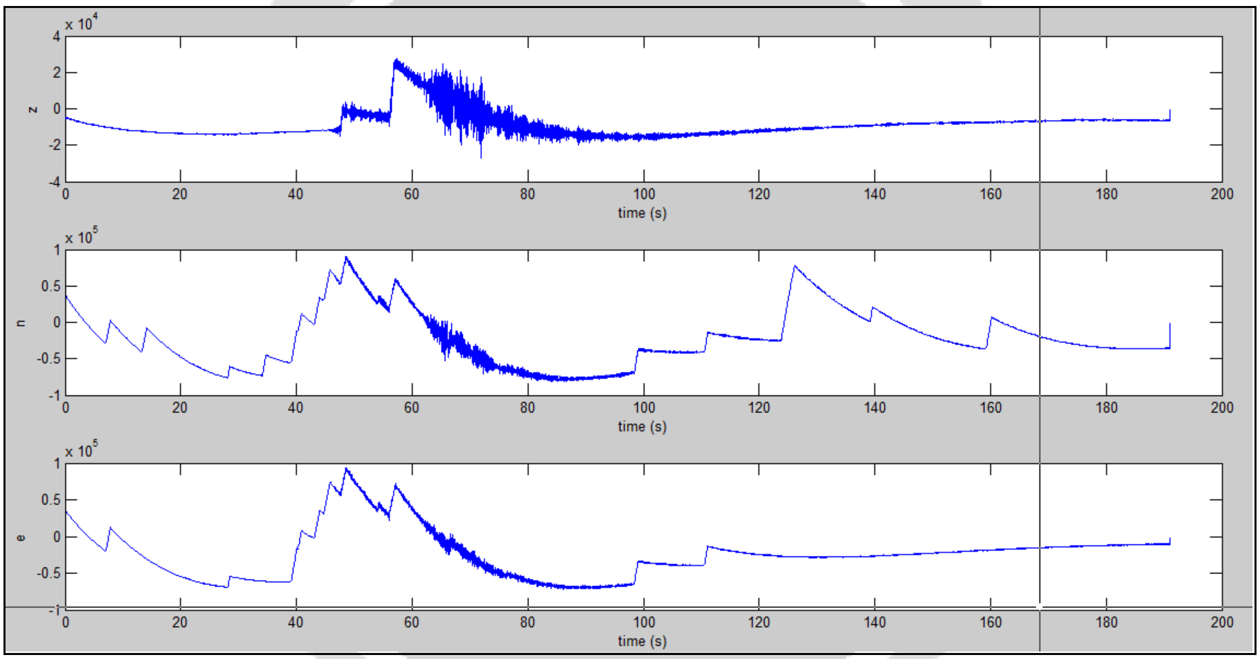

Fig -2: Generation of wave forms according to the amplitude of event

Only the variation in waveform is important, not all records. So it is advisable to capture and select portions of waveform that clearly indicates the variations in graph. Numbers formula given in Matlab built in function, I calculated several parameters such as starting point of captured waveform, end of line between starting and ending point, spectral ratio, and complexity.

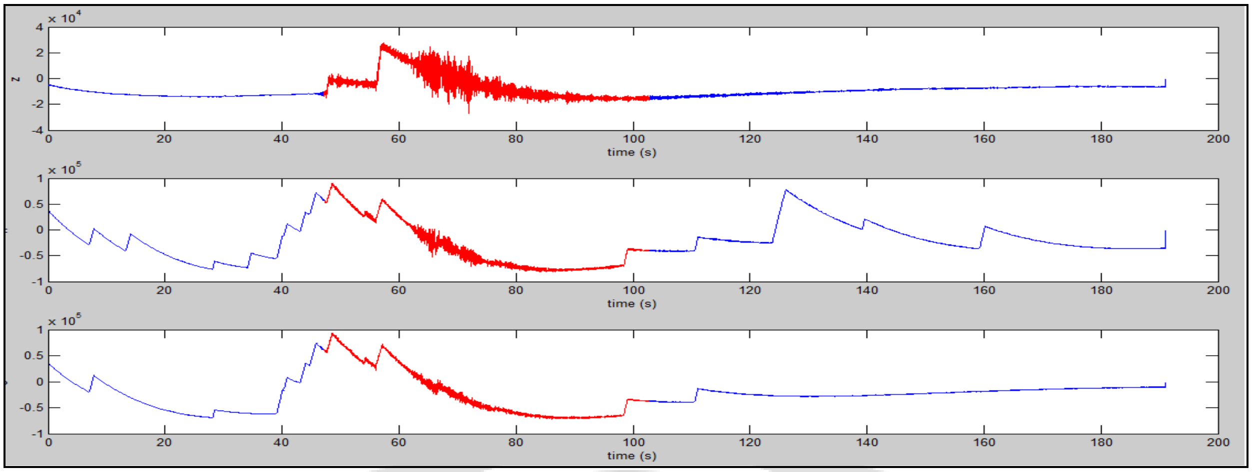

Fig -3: selecting waveform of an event having variation in it

Complexity played key role. The calculations of complexity, spectral ratio and other parameters are as follow:
X = starting of waveform
Y = end point of selection
NumLine = Number of data between X and Y
T0 = starting data point of waveform (same as X)
T2 = ending data point of waveform (same as Y)
T1 = middle point data of selected waveform

Y1 = $\int_{t1}^{t2} s2 \; dt$

Y2 = $\int_{t0}^{t1} s2 \; dt$

Complexity = y1 / y2

Spectral Ratio = $\int_{h1}^{h2} (a) \; dt$ / $\int_{h0}^{h1} (a) \; dt$

3.2 K-Means Clustering

k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results is a partitioning of the data space into Voronoi cells. In Matlab I prepared program for k-means clustering algorithm. I set the number of iterations are 5, numbers of clusters were 2, Matlab gives the functionality to use distance in some vary form like cosine; I used distance, and the input parameters for X axes I gave Complexity and for Y axes I left it empty, so by default it took time as second parameter. The result gave 2 separate clusters having centroid. By analyzing numbers of records clustered in k-means technique, 4.75 be a threshold figure to differentiate both event nicely. But k-means algorithm suffer from its accuracy and arbitrary shape of cluster as shown in figure 4 k-medoids algorithm gave the same reslut as k-means with slightly variation in reslut that will discuss in comparison.

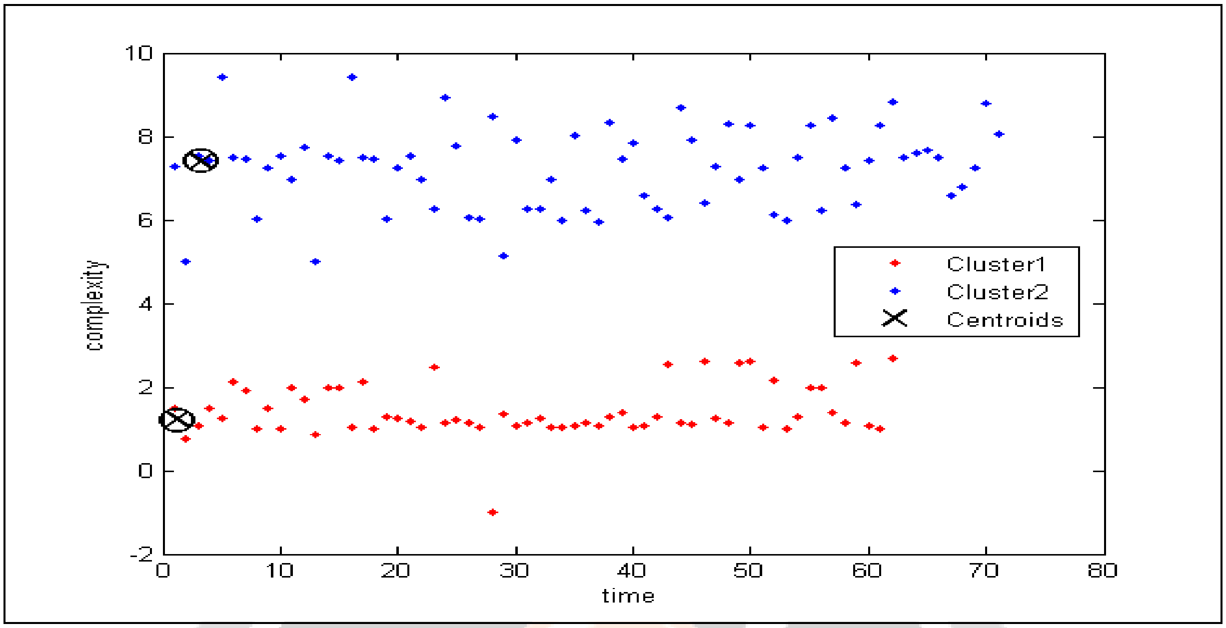

Fig -4: K-Means clustering

3.3 Subtractive Method

In above method I set only complexity as a parameter of clustering. Spectral ratio remains same for any amplitude value, complexity vary accordingly to the amplitude and spectral ratio. I used both as an input to subtractive clustering algorithm as shown in figure 5.

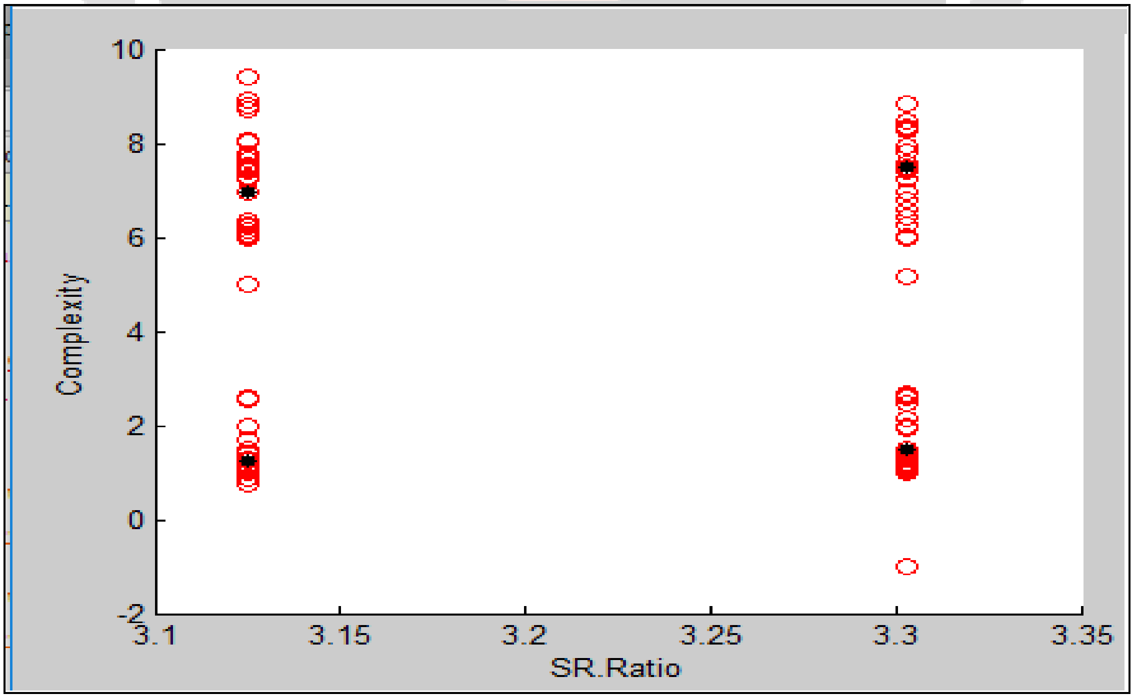

Fig- 5 : Subtractive method, spectral ratio and complexity as an input parameter

As shown in above figure, here I could not predefine number of clusters. As discussed early, spectral ratio remains same for any complexity value. I clustered data within two values of frequency domain; 0-3HZ - 3-5HZ and 0-5HZ – 5-10HZ. It gives spectral ratio 3.1250 and 3.3030 respectively. for better discrimination of frequency, I set same frequency throughout whole experimental. Here, cluster centers may vary by adding more number of data into it and reduce the accuracy from near about 75-80% to 60-75%.

3.4 FCM Method with magnitude and complexity

Early experimental results shows that Fuzzy Clustering Method is effective to locating cluster over noisy data. I had 139 files having magnitude recorded by ISR. I write the exponential value of fcm is 1.02 to 1.20. exponentail value decides the distance of object to its center while it iterate. And I took 1 to 100 iteration to cluster data. I wrote a code for fcm technique. Thus the input parameters were exponential value, numbers of iterations and obviously complexity and spectral ratio. The fcm clusters the data as shown bellow in figure 6.

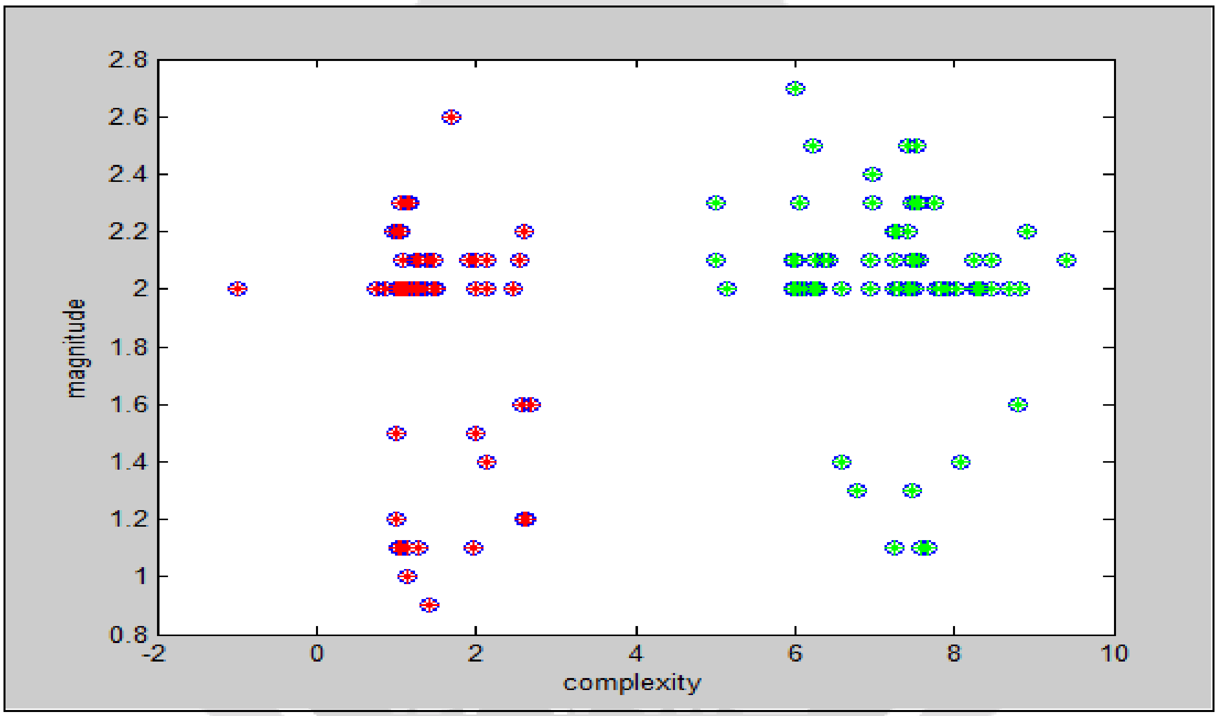

Fig- 6 : fcm method using code with complexity and magnitude as input parameters

Here, I could not identified cluster center and distance of data from its center. Also all data are far from each other and could not bound in any cluster. The result is negative. The accuracy is poor compare to early described both method.

3.4 FCM Method with magnitude and complexity(using GUI)

There were last chance to apply the same method using GUI. The best part of Interface is that I just worried about only data, I could tune up the number of iterations and exponential values. By performing same methods several time I found the optimal parameters of fcm method to differentiate seismic data. I increased exponential value 1.02 to 1.50 and gave its maximum height of 2.10. Also numbers of iterations are important for locating cluster over fuzzy data. Now, Iterations are raised to 0-200 from 0-100. Completing the set up of parameters I just applied both input parameters as I applied in coding, and mentioned the numbers of cluster equal to two, surprised results came by this experiment, now accuracy of clustering the objects were 90-95%. Which was optimal against all discussed above.

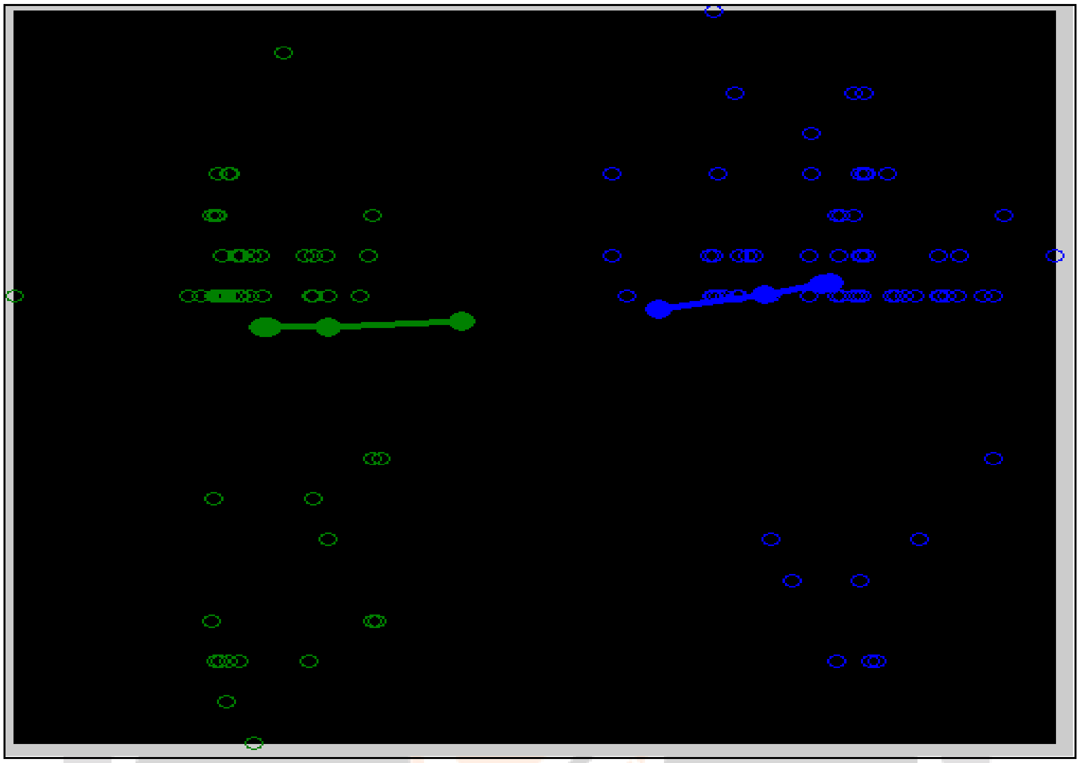

Fig- 7 : fcm method using code with complexity and magnitude as input parameters

3.5 Comparison analysis

First compare the graph before applying clustering method with the graph of same data after applying the clustering

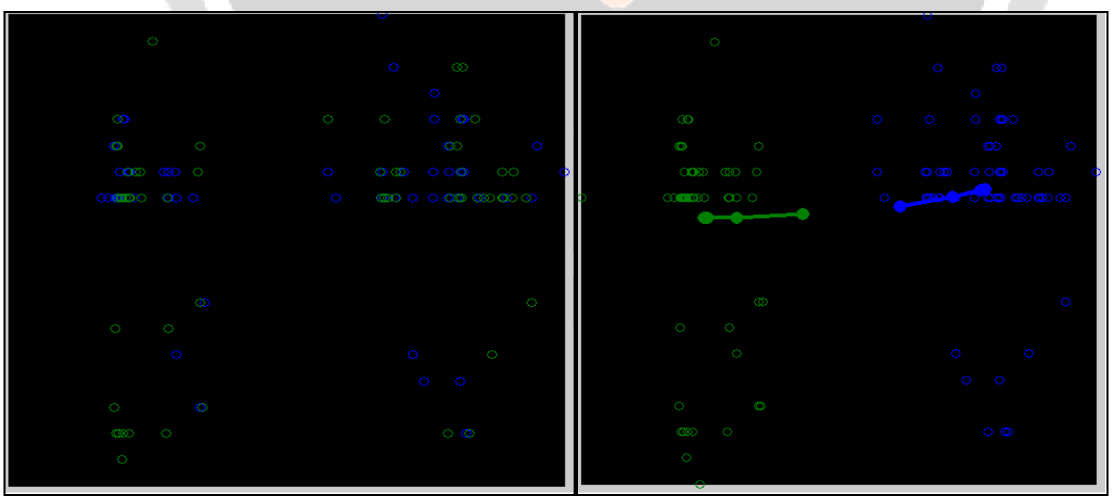

Fig- 8 : Comparison simple data graph with clustered data objects

From above figure it is clearly indicate that clustering is again proved is important role to differentiate data. Now, the question is which clustering technique is more appropriately for such type of rigid data. The best approach for clustering seismic data is fuzzy algorithm, that clearly identified in bellow table.

Table 3 : Comparison of various clustering methods that used to differentiate seismic data

Clustering Method	K-Means	K-medoids	Subtractive	Fcm using code	Fcm using gui
Number of records	133	133	133	133	133
Iterations	5	5	0-100	0-150	0-150
Parameters	SR.Ratio, Complexity	SR.Ratio, Complexity, Magnitude	Complexity, magnitude	Complexity, magnitude, sr.ratio,depth	Complexity, magnitude, sr.ratio, depth
Number of clusters	2	2	2	2	2 (may vary according to requirments
Exponential Value	1.02	1.02	1.5-2.5	2-3.5	2-3.5
Cluster Accuracy	78%	82%	86.9-88.3%	90-92%	92.5-97.5%

4. CONCLUSIONS

There is no doubt that clustering is best technique to differentiate data without consulting its class label. Experimental results show that Fuzzy clustering gives the higher accuracy compare simple partitioning methods. Also the accuracy of fuzzy clustering is higher, it may vary according to exponent, iteration and noise data. The outlier analysis can be applicable to detect outlier data as noise.

5. ACKNOWLEDGEMENT

During my whole research work I used ISR's (Institute of Seismological Research) seismic data.

6. REFERENCES

[1] HUANG Hanming, LI Rui and LU Shi Jun, BIAN Yin Ju, "Discrimina-tion of Earthquakes and Explosions Using Chirp-Z Transform Spectrum Features*", 2009 World Congress on Computer Science and Information Engineering, IEEE 2008, pp. 210-214.
[2] Han, Jiawei, and Micheline Kamber. Data Mining. Amsterdam: Elsevier, 2006. Print
[3] Garima, Hina Gulati, P.K.Singh, "Clustering Techniques in Data Mining: A Comparison," 2nd International Conference on Computing for Sustain-able Global Development (INDIA Com), IEEE 2015, pp. 410-415.
[4] T.HitendraSarma and P.Viswanath, B.Eswara Reddy, "A Fast Approx-imate Kernel k-means Clustering Method For Large Data sets," IEEE 2011, pp. 545-550.
[5] Danyang Cao, Bingru Yang, "An improved k-medoids clustering algo-rithm", IEEE 2010, Vol. 3, pp. 132-135.
[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.
[8] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," *IEEE Transl. J. Magn. Jpn.*, vol. 2, pp. 740-741, August 1987 [*Dig. 9th Annual Conf. Magn. Jpn.*, p. 301, 1982].
[9] M. Young, *The Technical Writer's Handbook*. Mill Valley, CA: University Science, 1989.