CLUSTERING OF BISPECTRAL INDEX MEASUREMENTS DATA BY USING THE FUZZY NEIGHBORHOOD RELATIONS

Efendi Nasibov

Dokuz Eylul University, Izmir, Turkey, *efendi.nasibov@deu.edu.tr*

Cluster analysis has an important role in analysis of the ElectroEnsepholoGraphy (EEG) signals of the brain activities [Escalona-Moran et al., 2007; Jin S-H. Et al., 2005; Van Hese et al., 2008]. The primary objective of clustering is to simplify statistical analysis by grouping similar objects in a cluster. Clustering methods can be divided into five main groups such as hierarchical, prototype-based, density (or neighborhood)-based, model-based, and grid-based [Han&Kamber, 2001].

From another point of view, the clustering methods can be investigated whether they are crisp or fuzzy clustering methods. Most of the proposed fuzzy clustering methods are based on the Fuzzy c-means (FCM) algorithm [MacQueen J., 1967; Han&Kamber, 2001]. These methods conceive the fuzziness of clustering as being assigned to some clusters with certain degrees of membership. Nevertheless, FJP algorithm handles the fuzziness from a different point of view [Nasibov, 2006; Nasibov&Ulutagay, 2006; Nasibov&Ulutagay, 2007]. The basic property of the FJP method is that it handles the fuzziness with respect to levels. Thus, it investigates in how much detail the elements are handled in the formation of clusters.

DBSCAN is another algorithm which has a low complexity, *i.e.* it runs fast and it uses two parameters which determine neighborhood radius and neighborhood threshold respectively. Unfortunately, these parameters should be set properly according to the scale of the dataset and the density of clusters. So, it means that, the speed of this algorithm is meaningful with an effective setting of these parameters [Ester et al., 1996].

Another method FJP runs slower than DBSCAN algorithm principally. However, it is easier to determine correct structure of clusters in a wide change interval of these parameters regardless of the scale of dataset and the density within clusters. So, it is plausible to say that FJP algorithm is more robust than DBSCAN with respect to the parameters. Another difference of FJP algorithm from DBSCAN is that it uses fuzzy neighborhood relations instead of classical neighborhood analysis as DBSCAN does.

In the paper Nasibov (2007) the Fuzzy Neighborhood-DBSCAN (FN-DBSCAN) algorithm which combines the speed of DBSCAN and robustness of FJP algorithms is proposed.

In this sudy, FN-DBSCAN algorithm is compared with classical FCM method by using BIS data which were recorded in sleep time by using EEG.

1. Dataset. The Bispectral Index (BIS) is a continuous processed EEG parameter that correlates to the patient's level of hypnosis, where 100=awake and 0=flat line EEG. BIS was designed to correlate with "hypnotic" clinical endpoints (sedation, lack of awareness, and memory) and to track changes in the effects of anesthetics on the brain.

The main purpose of this study is not to show the effectiveness of any method numerically, but to show that studies on neighborhood-based cluster analysis could provide more effective results. 22 datasets each of which are registered in every five seconds during sleep for a 25-minute periods are used to compare FCM and FN-DBSCAN clustering algorithms. Thus, each dataset contains 306 BIS measurements. In order to use in learning process, experts determined the BIS stage values corresponding to each measurement moment. Another purpose of the study is to predict the stage intervals and levels to the utmost. Because of this, we present not all of the datasets, but just one of them in order to point out the difference between two methods.

2. FN-DBSCAN algorithm. The main objective of the neighborhood-based clustering algorithms is to grow the concerned cluster until its density is greater than a specified threshold. Namely, each point in the concerned cluster should contain at least minimum number points within a certain threshold. Such a method could be used in order to eliminate outlier points and to determine clusters with arbitrary shapes.

As mentioned previously, FN-DBSCAN algorithm, integrates the advantages of DBSCAN and FJP algorithms in such a way that it combines the speed of DBSCAN algorithm and robustness and fuzziness of FJP algorithm.

The parameters ε and MinPts are used in classical DBSCAN algorithm. However, since ε represents the direct value of neighborhood radius, it gets values in different intervals corresponding to the scale of data. For instance, if data are between [0,1], the value of ε will be in this range. But if data are between [0,100], then ε will take values from this interval. Such a case causes some problems in trimming the values of ε. In order to eliminate this problem, we can normalize data and get an ε value between [0,1] by using the following transformation

$$x_{ij} = \frac{x_{ij} - x_j^{\min}}{(x_j^{\max} - x_j^{\min})\sqrt{m}}, \quad j = \overline{1,m}\,. \tag{5}$$

where $x_j^{\min} = \min\limits_{i=\overline{1,n}} x_{ij}$ and $x_j^{\max} = \max\limits_{i=\overline{1,n}} x_{ij}$, $j = \overline{1,m}$.

On the other hand, we can use the formula given below in order to invert the value of MinPts to the interval [0,1] and indicate it by ε_2 parameter:

$$MinPts = \varepsilon_2 \cdot w^{\max} \tag{6}$$

where $w^{\max} = \max\limits_{i=\overline{1,n}} w_i$, and w_i, is the cardinality of the point x_i with a certain ε_1 parameter. In general words, regarding fuzzy situation, w_i is the sum of the membership degrees of points with ε_1 parameter to the neighborhood set. Thus,

$$w_i = \sum_{k=1}^{n} N_{x_i}(x_k) \tag{7}$$

where $N_{x_i}(x_k)$ is the neighborhood degree of point x_k to the point x_i. In classical case, for obtaining $N_{x_i}(x_k)$ the following formula is used:

$$N_{x_i}(x_k) = \begin{cases} 1, & if \;\; d(x_i, x_k) \le \varepsilon_1 \\ 0, & otherwise \end{cases} \tag{8}$$

where $d(x_i, x_k)$ is the distance between x_i and x_k. If we enlarge it to the fuzzy neighborhood case, N_{x_i} function can be formed as any neighborhood membership function.

One of the fundamental advantages of using fuzzy neighborhood function is that the neighborhood membership degrees of the points with different distances from core point also differ. Such a neighborhood membership function used in NRFJP algorithm is as follows:

$$N_x(y) = \left\{ \begin{matrix} 1 - \dfrac{d(x,y)}{d^{\max}}, & d(x,y) \le \varepsilon_1 \;\; ise, \\ 0 \qquad\qquad , & diger. \end{matrix} \right\} \tag{9}$$

In order to explain FN-DBSCAN Algorithm based on fuzzy neighborhood analysis, we give some definitions. Let's determine neighborhood not with radius, but with level sets.

Definition 1: (fuzzy neighborhood). The neighborhood set of point $x \in X$ with ε_1 parameter is as follows:

$$FN(x;\varepsilon_1) = \{ y \in X \mid N_x(y) \ge \varepsilon_1 \}\,. \tag{10}$$

$N_x : X \to [0,1]$ is any membership function that determines neighborhood relation between points. ε_1 parameter used in Eq. (10) determines the minimal threshold of neighborhood membership function. Thus, $FN(x;\varepsilon_1)$, in fact, determines the ε_1- level set of the fuzzy neighborhood set of the point x.

Definition 2: (fuzzy core point). x is called a fuzzy core point with parameters ε_1 and ε_2 if

$$card\ FN(x;\varepsilon_1) \equiv \sum_{y \in N(x;\varepsilon_1)} N_x(y) \geq \varepsilon_2$$

holds for any point $x \in X$.

<u>*FN-DBSCAN algorithm:*</u>

Step 1. Specify parameters ε_1 and ε_2.
Step 2. Mark all the points in the data set as unclassified.
Step 3. Find an unclassified fuzzy core-point with parameters ε_1 and ε_2.
Step 4. Mark p to be classified. Start a new cluster to be the current cluster and assign p to the current cluster.
Step 5. Find all the unclassified points in the set $FN(p;\varepsilon_1)$. Create a set of seeds and put all these points into the set.
Step 6. Get a point q in the seeds, mark q to be classified, assign q to the current cluster, and remove q from the seeds.
Step 7. Check if q is a fuzzy core-point with parameters ε_1 and ε_2; if so, add all the unclassified points in the set $FN(q;\varepsilon_1)$ to the set of seeds.
Step 8. Repeat step 6 through Step 7 until the set of seeds is empty.
Step 9. Find a new fuzzy core point p with parameters ε_1 and ε_2 and repeat Step 4 through Step 7.
Step 10. Mark all the points, which do not belong to any cluster, as noise.
End.

3. Experimental results. In this section, FCM and FN-DBSCAN algorithms are compared by using BIS data recorded by EEG during sleep (Fig. 1). Prototype-based FCM algorithm was not successful in determining correct structure of BIS data set. Especially, the long term unchanged intervals were partitioned into some clusters artificially. On the other hand, neighborhood-based FN-DBSCAN algorithm determined only jumping situations, and did not respond to permanent changes. Thus, it is obvious that FN-DBSCAN algorithm is advantageous in detecting the sudden reactions for series in BIS data.

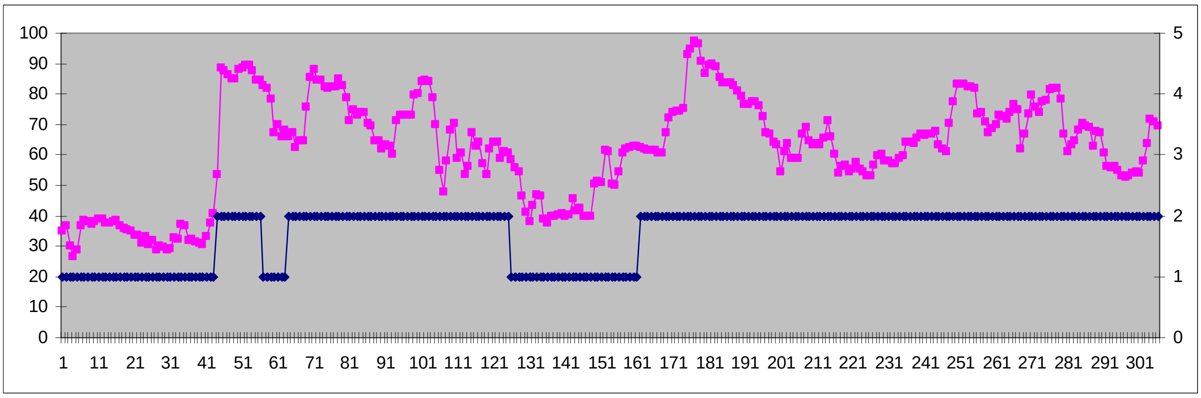

Figure 1. BIS data and stages.

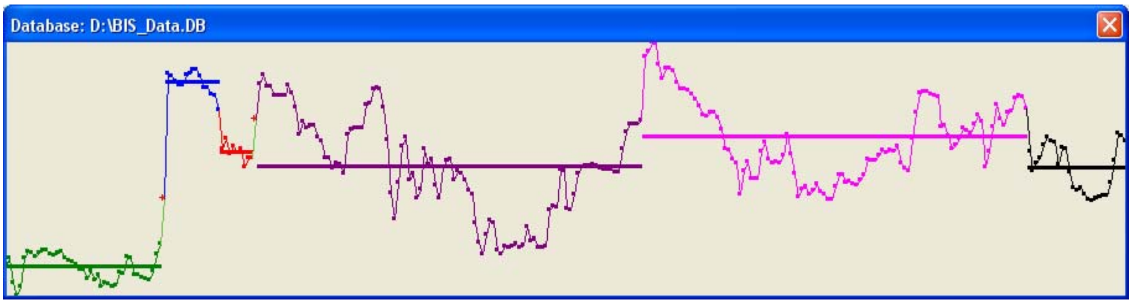

Figure 2. Clustering results of FN-DBSCAN algorithm with parameters $\varepsilon_1 = 0.76$ and $\varepsilon_2 = 0.1$

4. Conclusion. In this study, we mentioned a novel density-based clustering algorithm, Fuzzy Neighborhood DBSCAN (FN-DBSCAN). This algorithm integrates the speed of well-known DBSCAN algorithm and the robustness of fuzzy relation-based FJP algorithm.

In the experimental part of the study, instant reactions of the brain during sleep to the certain factors were analyzed by using two clustering algorithms, FCM and FN-DBSCAN. The experimental results showed that, since FCM algorithm partitions the data set into approximately equal-size clusters, it is not successful in the analysis of BIS data. However, density-based clustering algorithms are more successful in detecting clusters with various sizes. In this manner, FN-DBSCAN algorithm gives a wider perspective in that kind of data analysis.

Acknowledgment

This study is supported by the Scientific and Technological Research Council of Turkey (TUBITAK, Project No. 106T312).

References

[1] Escalona-Moran M., Cosenza M.G., Guillen P., Coutin P., 2007. Synchronization and clustering in electroencephalographic signals. Chaos, Solitons and Fractals 31: 820–825.

[2] Ester, M., Kriegel, H.P., Sander J., Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining, 226–231.

[3] Han, J., Kamber, M., 2001. Data Mining Concepts and Techniques. Morgan Kaufmann Publishers, 550 pp.

[4] Jin S-H, Ham B-J, Kim S. Y., 2005. Functional clustering in EEG photic and auditory driving in schizophrenia. Int. J. of Psychophysiology 56: 249– 259.

[5] MacQueen J., 1967. Some Methods for Classification and Analysis of Multivariate Observations. 5th Berkeley Symp. Math. Statist. Prob., 1: 281-297.

[6] Nasibov E.N., 2006. An Alternative Fuzzy-Hierarchical Approach to Cluster Analysis. In proc. 7th Int. Conf. On Application of Fuzzy Systems and Soft Computing, Siegen, Germany, pp. 113-123.

[7] Nasibov E.N., 2007. The Analysis of Fuzzy Neighborhood Relation in Density-Based Clustering. In proc. of the 1st Int. Conf on Soft Computing Technologies in Economy, Baku, Azerbajian, pp. 124-136.

[8] Nasibov E.N., Ulutagay G., 2007. A new unsupervised approach for fuzzy clustering. Fuzzy Sets and Systems 158:2118-2133.

[9] Nasibov E.N., Ulutagay G., 2006. On the fuzzy joint points method for fuzzy clustering problem. Automatic Control and Computer Sciences 40(5):33-44.

[10] Van Hese P., Vanrumste B., Hallez H., Carroll G.J., Vonck K., Jones R.D., Bones P.J., D'Asseler Y., Lemahieu I., 2008. Detection of focal epileptiform events in the EEG by spatio-temporal dipole clustering. Clin Neurophysiol, doi:10.1016/ j.clinph. 2008.04.009.