VOL. 13, NO. 20, OCTOBER 2018 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

CLUSTERING WITH FUZZY C-MEANS AND LINEAR DISCRIMINANT ANALYSIS FOR EPILEPSY CLASSIFICATION

Harikumar Rajaguru and Sunil Kumar Prabhakar
Department of Electrical and Computer Engineering, Bannari Amman Institute of Technology, Sathyamangalam, India
Email: harikumarrajaguru@gmail.com

ABSTRACT

A serious neurological disorder characterized by unpredictable seizures affecting the nervous system and brain is epilepsy. A seizure is nothing but a very short disruption in the normal activity of the brain that interferes severely with the brain function. The brain is comprised of billions of cells termed neurons which communicate by means of sending and receiving electrical messages. The activity of the brain is quite a rhythmic process where all the groups of neurons communicate with other similar group of neurons. When a seizure occurs, large groups of brain cells send messages simultaneously thereby the normal brain function gets temporarily disrupted where the seizure is occurring. An electroencephalography (EEG) is a test that helps to record and manage the electrical signals of the brain. For the diagnosis of epilepsy and sleep disorders, the physicians use it widely. As the EEG recordings are generally very lengthy, processing it is difficult and hence in this paper Fuzzy C-Means (FCM) technique is used for clustering initially. Later the clustered values are then classified with the help of linear discriminant analysis (LDA) classifier. Results show that an average classification accuracy of 93.54% along with an average performance index of 84.71% is obtained.

Keywords: epilepsy, EEG, FCM, LDA.

1. INTRODUCTION

Epilepsy is a severe neurological condition affecting the brain and it is otherwise known as seizure disorder [1]. It is characterized by unpredictable and uncontrollable seizures which vary in its severity and type. Though the seizures begin in the brain, it easily affects the other parts of the human body too. A seizure is therefore a serious indication of some issue happening in the brain. The primary reasons of epilepsy are unknown and it can affect any gender or age. To determine whether a person is having seizures or not, EEG test is highly useful [2]. The main intention in treating epilepsy lies in seizure control and to minimize its side effects. Various treatment options for seizure disorder include dietary therapies, medication, surgical option etc. Causing a general awareness to the people around the epileptic patient can be of great utility during the times of emergency. One's lifestyle is severely affected by epilepsy as the patient cannot engage in his routine activities like driving, swimming etc. Depending on the severity of the epilepsy and the respective treatment to which the patient responds, the impact of epilepsy on a person's lifestyle can be assessed. Therefore, epilepsy detection and classification play an important role in the fields of clinical diagnosis, biomedical signal processing, machine learning and pattern recognition.

A few important works in epilepsy classification from EEG signals is discussed as follows. Softmax Discriminant Classifier (SDC) was utilized by Rajaguru and Prabhakar for epilepsy classification from EEG signals [3]. Based on delay permutation entropy, the epileptogenic focus detection in intracranial EEG was done by Zhu *et al* [4]. The epileptic EEG classification was based on Extreme Learning Machine (ELM) and non linear features were done by Yuan *et al* [5]. The implementation of Adaboost Classifier for epilepsy classification with dimensionality reduction techniques was carried out by Prabhakar and Rajaguru [6]. Using Discrete Wavelet Transform (DWT) and Artificial Neural Networks (ANN), the EEG signal classification into seizure and non-seizure class was done by Patted *et al* [7]. Based on the statistical pattern recognition and wavelets, the EEG signals were classified for the detection of epileptic seizures by Gajic *et al* [8]. A hybrid classification model combining Artificial Bee Colony and Particle Swarm Optimization was employed with dimensionality reduction techniques for epilepsy classification from EEG signals by Rajaguru and Prabhakar [9]. For the classification of focal and non-focal epileptic EEG signals, an Empirical Mode Decomposition technique was utilized by Sharma *et al* [10]. The organization of the paper is as follows. In section 2, the materials and methods are discussed followed by the implementation of clustering using Fuzzy C-Means in section 3. Section 4 gives the usage of Linear Discriminant Analysis (LDA) as a post classifier followed by results and discussion in section 5 and ended with conclusion in section 6.

2. MATERIALS AND METHODS

In this study, the EEG database utilized is obtained from the University of Bonn, Germany and it is available online for the usage of students and researchers. There are totally 5 different sets of data available to the public and research community. For 100 patients, the study was carried out and the state of the patients was analyzed for ictal seizure activity. The electrodes were placed within the epileptogenetic zone and the kind of electrodes used is intracranial in nature. The total number of epochs present was 100 and the duration of every epoch was 23.6 seconds. With the help of 128 channel amplifier module, the recordings of the EEG signals were carried out. The sampling rate considered here is 176.31 Hz and the Analog to Digital resolution for digitization purposes had 12 bits. The block diagram of the work is given in Figure-1.

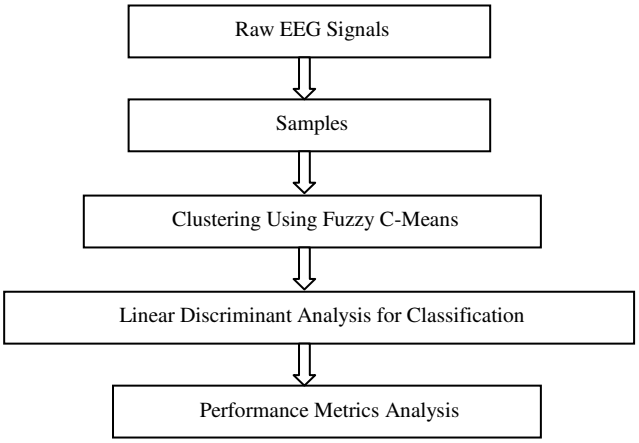

Figure-1. Block diagram of the work.

3. CLUSTERING USING FUZZY C-MEANS

One of the famous unsupervised learning techniques to group the same data points based on similarity measures tending to minimize the intracluster similarity and maximize the inter cluster similarity is clustering [11]. A typical clustering algorithm of the fuzzy portioning methodology is performed on a set of Q data points $Q=\{z_1, z_2, ..., z_Q\}$, where each $z_i \in \Re^g$ is a feature vector which consists of $'g'$ real valued measurements which explains in detail about the features of the data point z_i. The fuzzy clusters f of the data points are represented by a fuzzy membership matrix named as a fuzzy partition $V=[v_{ib}]_{f\times q}$ where v_{ib} denotes the fuzzy membership of the ith data point to the bth fuzzy cluster. The main aim of FCM is to locally minimize the objective function as it is an iterative procedure.

$$K_m = \sum_{k=1}^{f}\sum_{i=1}^{q} v_{ib}^p \left\| z_i - w_b \right\|^2$$

where $\{w_b\}_{b=1}^{f}$ is the cluster point of the clusters f and the fuzzy membership matrix is represented by the array v_{ib} as is expressed as:

$$V \in M_{gfq}$$

The Euclidean distance is nothing but the inner-product norm and is denoted as $\|\bullet\|$.

The parameter $m \in [1,\infty)$ is a simple weighting exponent on every fuzzy membership that helps to determine the fuzziness amount of the resulting classification:

$$P_{gfq} = \left\{ V \in \Re^{f\times q} \middle| \sum_{b=1}^{f} v_{ib} = 1, 0 < \sum_{i=1}^{q} v_{ib} < q \right\}$$

$$v_{ib} \in [0,1]; 1 \le b \le f; 1 \le i \le q$$

The summary of the FCM algorithm is as follows:

a) The number of fuzzy clusters $'f'$ are selected

b) The initial cluster centres $w_1, w_2, ..., w_f$ is selected

c) The elements of the fuzzy partition matrix is computed using

$$v_{ib} = \frac{1}{\sum_{d=1}^{f} \left(\left\| z_i - w_b \right\| / \left\| z_i - w_d \right\| \right)^{2/(p-1)}}$$

d) The cluster centres are computed using the following:

$$w_b = \frac{\sum_{i=1}^{q} v_{ib}^p . z_i}{\sum_{i=1}^{q} v_{ib}^p}$$

The steps (c) and (d) are repeated unless the total number of iterations $'t'$ exceeds a certain limit $\left\| W_{new} - W_{old} \right\| < \xi$, where $\xi = 0.04$. Ultimately 128 clusters are formed and then it is fed inside LDA for classification.

4. LINEAR DISCRIMINANT ANALYSIS (LDA) AS A POST CLASSIFIER

The fuzzy clustered values are then classified with the help of LDA classifier [12]. The classification is generally performed through a learning process where observations with known labeling of their respective classes are provided by the training set. LDA is a simple and yet robust classification technique based on the selection of a specific class which has the highest posteriori probability. The data was a normal distribution generally in LDA. A linear combination of predictors that can effectively split the various classes in a best manner can be easily searched by LDA. The LDA usually maximizes the function that denotes the difference between the means for a two-class problem. The difference is normalized by a specific measure of the within-class scatter represented as follows:

$$K(\beta) = \frac{\left| \tilde{\mu}_1 - \tilde{\mu}_2 \right|^2}{\tilde{S}_1^2 + \tilde{S}_2^2}$$

where $\beta = \left[\beta_1, \beta_2, ..., \beta_p\right]^T$ is an $p \times 1$ discriminant vector, $\tilde{\mu}_i = \beta^T \mu_i$ and

$$\mu_i = \frac{1}{P_i} \sum_{y \in C_i} y$$

where μ_i denotes an $p \times 1$ mean vector of class C_i (i=1 or 2) which has P_i features samples $\{y\}$, each having a size of $p \times 1$. Now as an equivalent of variance, the scatter for every class is defined as

$$\tilde{S}_i^2 = \sum_{z \in C_i} \left(z - \tilde{\mu}_i\right)^2$$

where $z = \beta^T y = \beta_1 y_1 + \beta_2 y_2 + ... + \beta_p y_p$ is the Fischer Linear Discriminant that tends to maximize the criterion functions.
Therefore,

$$S_i = \sum_{y \in C_i} \left(y - \mu_i\right)\left(y - \mu_i\right)^T$$

Therefore, for a linear model, the unknown weighting coefficient β is represented as:

$$\beta = \left(S_1 + S_2\right)^{-1} \left(\mu_1 - \mu_2\right)$$

5. RESULTS AND DISCUSSIONS

The Fuzzy C-Means Clustered values are classified with the help of Linear Discriminant Analysis (LDA) Classifier and based on the specific parameters like Classification Accuracy, Specificity, Sensitivity and Performance Index, the average results are computed and depicted in Table-1. The mathematical formulae for the Performance Index (PI), Sensitivity, Specificity and Accuracy are expressed as follows:

$$PI = \left(\frac{PC - MC - FA}{PC}\right) \times 100$$

where PC = Perfect Classification, Missed Classification = MC and the FA = False Alarm. The Sensitivity, Specificity and Accuracy measures are mathematically expressed as follows:

$$Sensitivity = \frac{PC}{PC + FA} \times 100$$

$$Specificity = \frac{PC}{PC + MC} \times 100$$

$$Accuracy = \frac{Sensitivity + Specificity}{2}$$

Table-1. Consolidated table for FCM clustered values with LDA classifier.

Name	Average
PC (%)	87.08
MC (%)	0.04
FA (%)	12.87
PI (%)	84.71
Specificity (%)	99.95
Sensitivity (%)	87.13
Accuracy (%)	93.54

6. CONCLUSIONS

Thus, classification of epilepsy is a pretty difficult task which requires a keen observation of the patient, clinical information along with the EEG recordings. The classifiers serve as a sincere diagnostic decision support system for neurologists dealing with epilepsy. In this work, Fuzzy C-Means technique was used as a clustering technique and then the clustered values are then classified with the help of LDA classifier. Results show that an average classification accuracy of 93.54% along with an average performance index of 84.71% is found out. Future works is to work with various versions of LDA and analyze its performance with the Fuzzy C-Means Clustered values for obtaining a better epilepsy classification rate.

REFERENCES

[1] H. Rajaguru, S.K. Prabhakar. 2017. Power Spectral Density and KNN Based Adaboost Classifier for Epilepsy Classification. IEEE Proceedings of the International Conference on Electronics, Communication and Aerospace Technology (ICECA 2017), Coimbatore, India, pp.441-445

[2] H. Rajaguru, S.K. Prabhakar. 2017. Metric Multidimensional Scaling and Aggregation Operators for Classifying Epilepsy from EEG signals. IEEE Proceedings of the International Conference on Electronics, Communication and Aerospace Technology (ICECA 2017), Coimbatore, India, pp.567-570

[3] H. Rajaguru, S.K. Prabhakar. 2017. Analyzing Dimensionality Reduction with Softmax Discriminant Classifier for Epilepsy Classification. 2nd IEEE International Conference on Communication and

Electronics Systems, Coimbatore, India, October 19-20.

[4] G. Zhu, Y. Li, P.P. Wen, S. Wang and M. Xi. 2013. Epileptogenic focus detection in intracranial EEG based on delay permutation entropy, Proc Int Symp Comput Models Life Sci. 31-36.

[5] Q. Yuan, W. Zhou, S. Li and D. Cai. 2011. Epileptic EEG classification based on extreme learning machine and nonlinear features, Epilepsy Res. 96: 29-38.

[6] S.K. Prabhakar, H. Rajaguru. 2017. Adaboost Classifier with dimensionality reduction techniques for Epilepsy Classification from EEG. International Conference on Biomedical and Health Informatics, Thessaloniki, Greece. 18-21 November.

[7] A. Patted, S. Bekal and V. Desai. 2016. EEG signal classification into seizure and non-seizure class using discrete wavelet transform and artificial neural network, Int. J. Innovat. Res Comput Commun Eng. 4: 1-7.

[8] D. Gajic, Z. Djurovic, S. Di Gennaro and F. Gustafsson. 2014. Classification of EEG signals for detection of epileptic seizures based on wavelets and statistical pattern recognition, Biomed Eng Appl Basis Commun. 26: 1-45.

[9] S.K. Prabhakar, H. Rajaguru. 2017. Performance Analysis of Factor Analysis and Isomap with Hybrid ABC-PSO Classifier for Epilepsy Classification. International Conference on Biomedical and Health Informatics, Thessaloniki, Greece.

[10] R. Sharma, R.B. Pachori and U.R. Acharya. 2015. Application of entropy measures on intrinsic mode functions for the automated identification of focal electroencephalogram signals. Entropy. 17: 669-691.