International Journal of Scientific & Engineering Research, Volume 6, Issue 10, October-2015 198
ISSN 2229-5518

Estimation of workload using EEG data and classification using linear classifiers

Sushil Chandra Institute of Nuclear Medicine and Allied Science, DRDO, India.
Greeshma Sharma
Akansha Gomes
DevendraJha Scientific Analysis Group, DRDO, Delhi, India. sagdjha@gmail.com
Alok Prakash Mittal Netaji Subhas Institute of Technology, New Delhi, India. mittalap@gmail.com

Abstract—Cognitive workload is a subjective term operantly defined as a worker's perception of a work performance and work difficulty. To estimate workload through Electroencephalogram (EEG) requires good algorithm with best features. Objective of this research study was to estimate workload using linear classifiers on non linear data. Workload was presented by varying levels of Multi Attribute Task Battery II (MATB-II task). Two non linear features of Hurst exponent and Higuchi Fractal Dimension have been extracted from the data which was acquired from 28 subjects who were all male in the age group 25-40. The classification has been performed using three prominent classifiers i.e. K-Means, K Nearest Neighbor (KNN) and Support Vector Machine (SVM) to test their efficiency in the case of workload data. We have hypothesized SVM classifier to give best results out of the three classifiers. Comparing the performance accuracy of the selected classifiers, we propose a classifier that will give best results for workload classification

Index Terms—EEG, Workload Classification, Classifier, Support Vector Machine (SVM), K-Means, KNN

1 INTRODUCTION

The Electroencephalogram (EEG) signal is a voltage signal arising from synchronized neural activity. EEG is produced due to coordinated activity of millions of neurons in the brain. EEG is measured in two ways. One way is to place an electrode on or near the scalp, and other is by implanting an electrode in the skull. Synchronized neural activity varies according to development, mental state, and cognitive activity, and hence causing variations in EEG signal [1].

The field of Brain-Computer Interfaces (BCI) is a driving force for utilizing EEG technology. BCI is a field of study which interests many neurologists. It enables a direct communications pathway between the brain and the object to be controlled. The very aim of BCI is to translate brain activity into a command for a computer. To achieve this goal either regression [2] or classification [3] algorithms can be used. These algorithms are used to identify patterns of brain activity [4].

1.1 Workload classification and its relation to BCI

Cognitive workload describes the level of mental resources that a person utilizes at a given time and for a given task which affects their ability to process information, to respond to their surroundings and hence to make decisions. It can be defined as a scale with two opposite ends:

- Overload – when too much information is being processed
- Underload – when information being processed is less than the threshold level.

Monitoring the workload level of an operator tells about the present workload state of the operator and helps in controlling it through various techniques. One such technique which has been used in this study is that of meditation (SudarshanKriya). BCI system can be considered as a brain wave pattern recognition system [5],[6]. BCI development is no longer constrained to just patients or for treatment, gamers are becoming a target group that would likely to be adaptive to use EEG as a new modality [7]. In order to get general information about the user's brain wave pattern, a series of mental task scenarios are executed by the user. This information is then used to train a classification system so that it can learn to recognize and thus map different brain patterns to actions. The user can then start a game, and the classification system will continuously analyze the incoming brain waves and map them to the appropriate actions and thus control some feature(s) of the running game. Workload classification is an effective BCI approach. The brainwave pattern of the operator generated during high and low workload states are different and thus can be used in BCI to perform different actions in the two cases. The performance of pattern recognition depends on both the features and the classification algorithm employed.

1.2 RELATED WORK

The analysis of workload from EEG data has a tradition in the psychological community. A lot of Research has been done previously on workload data.
In one of the studies by Alexander J. Casson, PSD feature was taken and the data was classified using neural networks [8]. An average performance accuracy of 86% was obtained for 8 subjects. This accuracy could have been higher. The limitation introduced was due to Artificial Neural Network as it requires very large amount of data for training the network so as to

ISSN 2229-5518

give very accurate results and if there is a gap in recording, then there is a significant effect on the performance accuracy.

In other work, PSD and mean were taken as features and SVM and Artificial Neural Network (ANN) were used for comparison of classification performance to find the best classifier of the two. Neural Network being a non-linear classifier was expected to outperform SVM, but accuracy of SVM came out to be better than ANN. The reason for SVM giving high performance accuracy is that it is a linear, supervised classifier. And it has been shown to give very good results when applied with two input features. In one of the experiments, all the nonlinear features i.e. Correlation Dimension (C.D), Hurst Exponent (H.E) and Largest Lyapunov Exponent (LLE) were used for classifying non linear workload data. And the results showed significant changes in the features as the cognitive state of the operator changed [9],[10].

From the previous work it can be said that although workload EEG is a non linear signal, it is not necessary that it will be classified best with non linear classifiers like ANN. Hence, we hypothesize SVM classifier to outperform KNN and K-Means classifiers when Hurst Exponent and Fractal Dimension are taken as input features.

1.3 FEATURES

The Hurst exponent (H) is a statistical measure used to classify time series. H=0.5 indicates a random series while H>0.5 indicates a trend reinforcing series. The larger the H value is, the stronger trend. Series with The values of the Hurst exponent range between 0 and 1. Based on the Hurst exponent value H, a time series can be classified into three categories. (1) H=0.5 indicates a random series. (2) 0<H<0.5 indicates an anti-persistent series. (3) 0.5<H <1 indicates a persistent series [11].

Fractal dimension gives a measure of signal complexity and so the measure of complexity of the processes that generate the signal under consideration. Higuchi's fractal dimension (HFD) is an appropriate method for analyzing the FD of biomedical signals.[12],[13]. Hence, HFD is used here for analyzing the complexity of non-linear workload data.

1.4 CLASSIFIER

One of the classifiers we used for classification is support vector machine (SVM) . SVM is a supervised classifier which uses a discriminant hyper-plane to identify classes [14]. The selected hyper-plane is the one that maximizes the margins, i.e., the distance from the nearest training points. SVM uses a regularization parameter C that enables accommodation to outliers and allows errors on the training set. Such an SVM enables classification using linear decision boundaries, and is known as linear SVM. This classier has been applied, always with success, to a relatively large number of synchronous BCI problems [15],[16]. However, it is possible to create nonlinear decision boundaries, with only a low increase of the classifier's complexity, by using the "kernel trick".

KNN is an unsupervised algorithm. Unsupervised technique means that it does not make any assumptions on the underlying data distribution. KNN assumes that the data is in a feature space. More exactly, the data points are in a metric space. Since the points are in feature space, they have a notion of distance. This need not necessarily be Euclidean distance although it is the one commonly used. Each of the training data consists of a set of vectors and class label associated with each vector. KNN can work equally well with arbitrary number of classes. [17].

K-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The main idea is to define k centroids, one for each cluster [18]. The next step is to take each point belonging to a given data set and associate it to the nearest centroid [19]. When no point is pending, the first step is completed and an early group age is done. At this point we need to recalculate k new centroids as barycenter of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop we may notice that the k centroids change their location step by step until no more changes are done. In other words centroids do not move any more. Finally, this algorithm aims at minimizing an *objective function*, in this case a squared error function. The objective function

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \qquad 0$$

Where $\left\| x_i^{(j)} - c_j \right\|^2$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster center c_j, is an indicator of the distance of the *n* data points from their respective cluster centers [20],[21].

2. METHODS

A. Data Acquisition

We studied eight male participants, between ages of 25-40. The data was recorded as part of the Cognitive State Assessment task using a 14 channel EEG in standard 10–20 position

- *Author name is currently pursuing masters degree program in electric power engineering in University, Country, PH-01123456789. E-mail: author_name@mail.com*
- *Co-Author name is currently pursuing masters degree program in electric power engineering in University, Country, PH-01123456789. E-mail: author_name@mail.com*
  (*This information is optional; change it according to your need.)*

ISSN 2229-5518

(AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4). All channels used a mastoid reference and ground and a 128Hz sampling rate.

A total of 80 EEG recordings were performed in a noise free environment using eight subjects (10 tests on each subject) on a single day. Each recording consisted of three 15 min EEG sessions, allowing the temporal performance of the workload monitor to be evaluated on a number of scales. Within each 15 min session the EEG data can be divided into training and testing periods which are separated in time by seconds or minutes. In initial 15min EEG was taken without any training in which all subjects were asked to do MATBII task in which varying levels of difficulty was introduced. Afterwards, subjects were asked to do meditation (SudarshanKriya) for 25min and then again their EEG was recorded in which they again performed MATBII task. The task was set up so as to vary difficulty levels dynamically. A total of 5 min was spent in each state, with at least a 1 min transition present between task segments classed as high and low workload. Here, only the high and low workload monitoring data segments are analyzed and the transition segments have been strictly discarded.

B. Pre Processing and Classification

The EEG data acquired during various training and testing sessions was sampled at 128Hz sampling rate. Artifacts have always been a premium source for signal contamination. Usually eye artifacts and muscular artifacts can be found in EEG signals when recorded under noisy conditions. Hence, prior to processing the signal and extracting features, we removed the artifacts by performing Independent Component Analysis (ICA) and then data was filtered using Butterworth filter in order to have signal from 0.5 to 45 Hz. After filtering the data and removing the artifacts, nonlinear features of Hurst exponent and Higuchi fractal dimension were extracted. The extracted features were used as inputs for the specified classifiers i.e. SVM, K-Means, KNN and the performance accuracy was compared for different classifiers.

3 RESULT AND DISCUSSION

The experiment was performed successfully and the extracted features were classified. The results obtained after classification have been compared for accuracy. Across all subjects classified using all the classifiers ,the per subject performances are given in the bar graph(fig.1) and the results of classification of pre and post training data have been given in figures below:

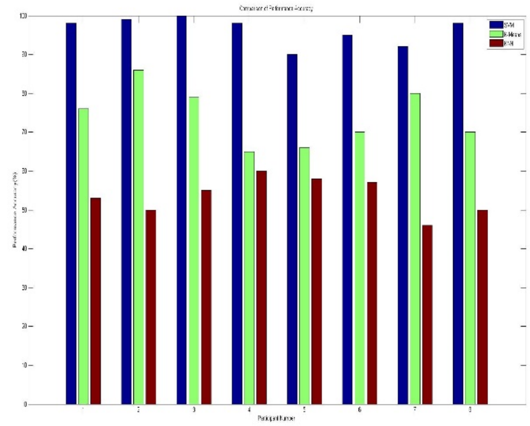


Figure 1: classification accuracy of all the classifiers used

The graph in fig.1 clearly shows that performance accuracy of SVM classifier was the highest while that of KNN was the lowest. The highest accuracy found in SVM across all the subjects was 99% while the average of performance across all the subjects was 92%. The highest accuracy in case of K-Means and KNN came out to be 86% and 58% respectively.

The results of classification of one subject for pre and post training data have been given in figures below:

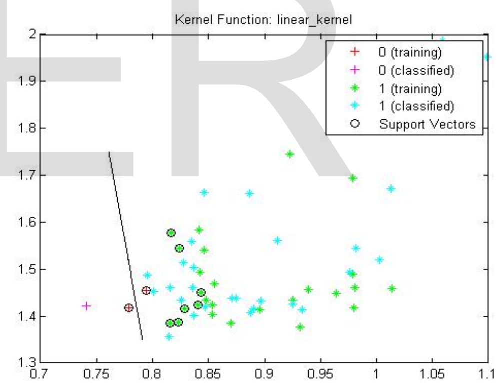
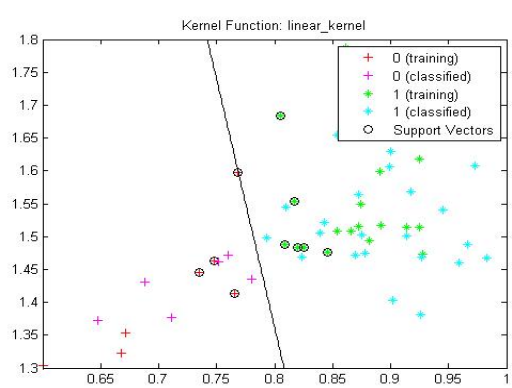


Figure 2: classification result of pre training data using SVM classifier



ISSN 2229-5518

Figure 3: classification result of post training data using SVM classifier

From the above figures (2, 3), 92% of the pre training data was classified as high load whereas only 8% was low load. Contrary to this, 60% of the data was classified as high load and 40% as low load in post training data.

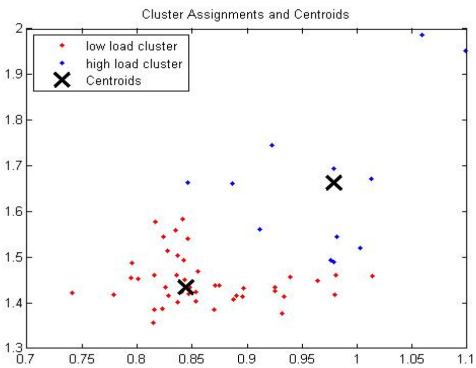


Figure 4: classification result of pre training data using K-Means classifier

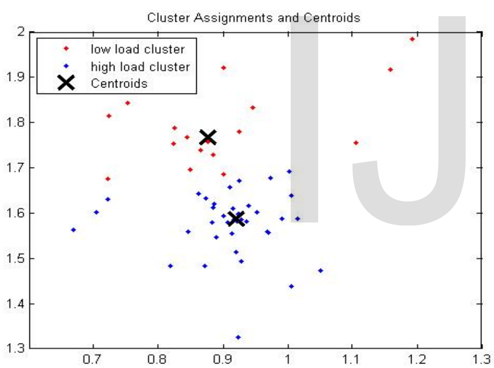


Figure 5: classification result of post training data using K-Means classifier

Comparing above figures (4, 5), 70% of the data was classified as low load and 30% as high load from pre training data. Whereas 50% of the data was classified as low load and 50% as high load in post training data (this gave an error of approximately 10%).

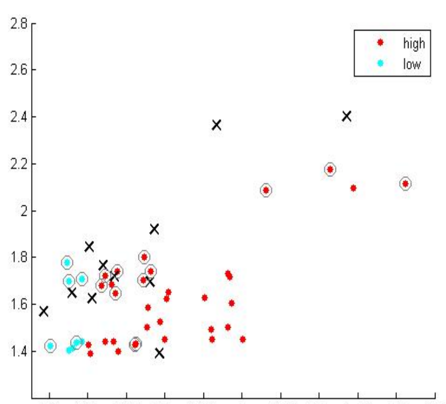


Figure 6: classification result of pre training data using KNN classifier

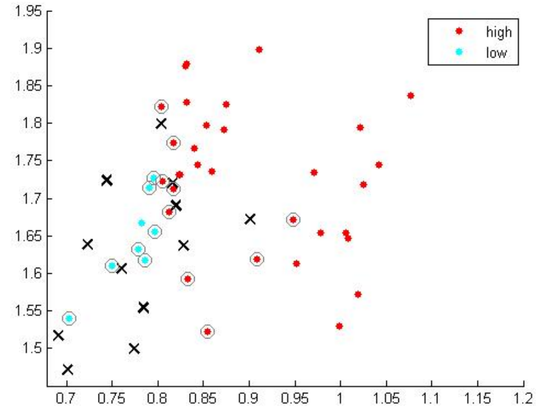


Figure 7: classification result of post training data using KNN classifier

Comparing the above figures (6, 7), 16% of the data was classified as low load and 84% as high load from pre training data. Whereas there was 38% low workload and rest 62% high workload in post training data. (This gave an error of 26%).

A satisfactory level has been achieved using all three classifiers showing that they can be used for determining the operator state. Based on comparison of classification performance accuracy, we can say that in case of workload data, SVM classifier proves to be the best among the three chosen classifiers. Besides comparing the performance of classifiers we also observed that operator was in comparative low workload state after receiving a training session in which the operator was asked to meditate and relax. This shows that meditation and relaxing have a significant impact on the workload state of the operator.

This study has presented a BCI based workload classification system investigating the performance of some of the prominent classifiers in classifying the workload data into high and low workload respectively.

The results of classification obtained by different classifiers can be seen above in figures (2-7). In this we can clearly observe two things primarily SVM classifier outperformed KNN and K-Means classifier and secondarily, Post training data proves that operator was in low workload state compared to pre training. Hence, our initial hypothesis proved to be correct as SVM outperformed other two classifiers. The reason for such high performance accuracy of SVM classifier is due to the fact that it is a linear classifier and is a supervised one. It doesn't require large data for training and hence gives very accurate results with limited data and with two input features.

ISSN 2229-5518

4 CONCLUSION

This paper investigated the performance of three prominent classifiers namely Support Vector Machine(SVM), K-Nearest Neighbours and K-Means . As per our initial hypothesis, the performance accuracy of support vector machine proved to be the best amongst the three chosen classifiers . One reason why SVM outperformed other classifiers is that it is a supervised classifier. It was trained with sufficient ammount of data before it was tested for its performance. And previously also SVM has been proven to work really well with two input features. A key result of this study is the significant change in workload state of the operator after the training. Prior to training, the operator was in a high workload state and after the training(here meditation in our experiment) the operator was found to be in low workload state. This system can be applied to detect workload in real world scenarios, such as computer work and gaming, but more detailed analyses and systematic evaluations are needed to get more insights of the capabilities and limitations of the system. For a real application in Human-System Interaction a major drawback of the proposed system is the fact that the resistance that users offer to wear an EEG cap outside experimental scenarios. Therefore need for less invasive wearable device is high and only then can EEG be used in real world applications.

REFERENCES

[1]Nitin NarhariHegde, et al., EEG signal classification using K-Means and Fuzzy C Means clustering methods, IJSTE, 2349-784X

[2] D. J. McFarland and J. R. Wolpaw. Sensorimotor rhythm-based brain-computer interface (bci): feature selection by regression improves performance. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 13(3):372{379, 2005.

[3] W. D. Penny, S. J. Roberts, E. A. Curran, and M. J. Stokes. Eeg-based communication: a pattern Recognition approach. IEEE Transactions on Rehabilitation Engineering, 8(2):214{215, 2000.

[4] D. J. McFarland, C. W. Anderson, K.-R. Muller, A. Schlogl, and D. J. Krusienski. Bci meeting 2005-workshop on bci signal processing: feature extraction and translation. IEEE Transactions On Neural Systems and Rehabilitation Engineering, 14(2):135,138, 2006.

[5] R. O. Duda, P. E. Hart, and D. G. Stork. Pattern Recognition, second edition. WILEYINTERSCIENCE, 2001.

[6] A.K. Jain, R.P.W. Duin, and J. Mao. Statistical pattern recognition: A review. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(1):4, 37, 2000.

[7] Erik Andreas Larsen, classification of EEG signals in brain computer interface systems, Norwegian University of Science and Technology, Department of Computer and Information Science

[8]Alexander J. Casson, artificial neural network classification of operator workload with an assessment of time variation and noise-enhancement to increase performance,10.3389/fnins.2014.00372

[9] K. Natarajan, R. Acharya, F. Alias, and et al., "Nonlinear analysis of EEG signals at different mental states," BioMedical Eng. Online, vol. 3, pp. 1-11, 2004.

[10] N. Kannathal, U. R. Acharya, C. M. Lim, and et al., "Characterization of EEG—A comparative study," Computer Methods and Programs in Biomedicine, vol. 80, pp. 17-23, 2005.

[11] Bo Qian, Khaled Rasheed, "Hurst exponent and financial market predictability", Department of Computer Science ,University of Georgia Athens, GA 30601 USA [Qian, Khaled]@cs.uga.edu

[12] T.Higuchi, "Approach to an irregular time series on the basis of the fractal theory", Physical D, vol. 31, pp. 277-283, 1988,

[13] A.Accardo, M.Affinito, M.Carrozzi, and F.Bouquet, "Use of the fractal dimension for the analysis of electroencephalographic time series", Biol.Cybern. vol. 77, pp. 339-350, 1997

[14] Analysis of EEG signals using fractal dimension", PhD Thesis, Supervisor W.Klonowski, Institute of Bio cybernetics and Biomedical Engineering, Polish Academy of Sciences, Warsaw, 2003.

[15] K. P. Bennett and C. Campbell 2000" Support vector machines: hype or hallelujah" ACM SIGKDD Explore. Newsletter. 2 1–13

[16] A. Rakotomamonjy, V. Guigue, G. Mallet and V. Alvarado 2005"Ensemble of SVMs for improving brain computer interface" p300speller performances Int. Conf. on Artificial Neural Networks.

[17] G. N. Garcia, T. Ebrahimi and J-M Vesin 2003 "Support vector EEG classification in the Fourier and time-frequency correlation domains "Conference Proc. 1st Int. IEEE EMBS Conf. on Neural Engineering

[18] SaravananThirumuruganathan, "A Detailed Introduction to K-Nearest Neighbor (KNN) Algorithm", May 17, 2010, *Biweekly Links – 05-17-2010*

[19] Andrew Moore: "K-means and Hierarchical Clustering - Tutorial Slides" http://www-2.cs.cmu.edu/~awm/tutorials/kmeans.html

[20] Brian T. Luke: "K-Means Clustering"

[21] Tariq Rashid: "Clustering"