2017 International Conference on Innovations in information Embedded and Communication Systems (ICIIECS)

Motor imagery signal classification using Semi Supervised and Unsupervised Extreme Learning Machines

M.Dharani
ME(Applied Electronics), Dept of EEE
Kongu Engineering College
Perundurai, Erode
Dharanimurugesan93@gmail.com

M.Sivachitra
Assistant Professor (SL.G.) Dept of EEE
Kongu Engineering College
Perundurai, Erode
sivachitra@kongu.ac.in

***Abstract*— Brain computer interface (BCI) aims at providing a brand new communication approach without brain's traditional output through nerve and muscle. "Electroencephalography" has been widely used for BCI system as it is a non-invasive approach. Recently, various classifiers have been used for the analysis of EEG signals measured under the planning and relaxed state. The major work addressed in the paper is the classification of EEG signals (motor imagery) measured under planning and relaxed state using advanced learning classifiers. The dataset of planning and relaxed state is a benchmark data and it is taken from UCI (University of California, Irvine) repository. Semi supervised ELM (SS-ELM) and unsupervised ELM (US-ELM) are recently developed networks and used for the EEG signal classification task. Both of these algorithms can be fit in to a unified framework and handle multi-class classification or multi-cluster clustering. SS-ELM algorithm performs better than US-ELM and other real valued algorithms in classifying planning and relaxed states. The improvement is due to the use of spectral techniques in embedding and clustering.**

Index Terms: Clustering, embedding, planning and relaxed dataset, Extreme learning machine, Manifold regularization, semi-supervised learning and unsupervised learning.

I. Introduction

In recent times, artificial neural networks are applied in numerous domains like classification [1]-[7], prediction [8]-[10], management systems, workstations, physics, engineering and human activity recognition [11]. The immediate goal of BCI analysis is to produce communications capabilities to severely disabled people who are totally paralyzed. A BCI is a man-made intelligence system that acknowledges a particular set of patterns in brain signals with five stages namely, signal acquisition, preprocessing, feature extraction, classification, and also the management interface for the practical implementation of the system. Generally EEG based recording of electrical potentials could be a harmless and efficient for the observation of brain activity. A good classifier need to deliver high classification accuracy with less machine complexity. In this paper, classification part of the BCI system is focused.

Extreme Learning Machine (ELM) is a fast learning algorithm introduced for training single-layer feed-forward networks (SLFNs). The major benefits of ELM are that the networks input weights are assigned randomly and output weights are found analytically [12-15]. Most of the present day algorithms is meant for training SLFNs, such as the illustrious back-propagation algorithm [16] and also the Levenberg-Marquardt algorithm [17], adopt gradient strategies for optimizing the weights within the network.

ELMs are primarily used for supervised learning tasks such as classification and regression, which greatly limits their pertinence. Sometimes, in a few applications like information retrieval text classification, and fault identification, obtaining labels for fully supervised learning are time consuming and expensive.

It is appreciated that the supervised, unsupervised and semi-supervised ELMs can be put into a unified framework and they possess two major steps 1) random input feature mapping and 2) finding output weights. The first step is to construct the hidden layer by randomly generating hidden neurons. The second stage is to find the weights between the hidden layers and the output layer, and this can be the difference of supervised, unsupervised semi-supervised and ELMs lies [18].

In this paper, we have applied semi-supervised ELM (SS-ELM) and unsupervised ELM (US-ELM) [18] to classify planning and relaxed state data. The results show that the SS-ELM and US-ELM algorithms are competitive with ELM algorithms in terms of efficiency.

The organization of the paper is as follows, Section II and III describes about the learning algorithms namely, SS-ELM and US-ELM respectively. Section IV explains EEG data and the comparison results of the same. The conclusion and future work is discussed in the section V.

II. Semi supervised Extreme Learning Machine

The SS-ELM algorithm recently developed by gaohuang et.al [18] has been elucidated briefly. In semi-supervised learning, few labeled and many unlabeled data are used. The

978-1-5090-3294-5/17/$31.00 ©2017 IEEE

SS-ELM includes the manifold regularization to control unlabeled data to increase the classification accuracy. This process is carried out when the labeled data are limited. Skewed data causes poor generalization fit for testing dataset. Skewed data is one in which some classes have more data than the other. Therefore, the penalty factor is introduced to eliminate the effects and hence improves the generalization performance of the testing set. Suppose that x_i belongs to class t_i that has N_{t_i} training patterns, then we can associate e_i with a penalty of

$$C_i = \frac{C_0}{N_{t_i}} \quad (1)$$

Where C_0 could be a user defined parameter as given in earliest ELMs. As a result of this approach, the patterns from the dominant classes will not be over fitted and patterns from the class with fewer samples will not get neglected.

The formulation of SS-ELM is given by

$$\min_{\gamma \varepsilon R^{n_h X n_0}} \quad \frac{1}{2}||\gamma||^2 + \frac{1}{2}||C^{\frac{1}{2}}(\tilde{Y} - H\gamma)||^2 + \frac{\lambda}{2} T_r(\gamma^T H^T L H\gamma) \quad (2)$$

Where $Y \in R^{(la+un)Xn_0)}$ is that the training target with the initial la rows equal to Y_{la} and the remaining equal to zero.

Further, tthe gradient of the objective function is computed with respect to γ

$$\nabla L_{SS-ELM} = \gamma + H^T C(\tilde{Y} - H\gamma) + \lambda H^T L H \gamma \quad (3)$$

SS-ELM Algorithm is given by,

<u>Algorithm 1</u>

SS-ELM algorithm

Input used:

Labelled patterns, $\{X_{la}, Y_{la}\} = \{x_i, y_i\}_{i=1}^{la}$.

Unlabelled patterns, $X_{un} = \{x_i\}_{i=1}^{un}$.

Output:

The mapping function of SS-ELM: $f: R^{n_i} \to R^{n_0}$.

Learning Algorithm steps:

(i) Graph Laplacian L is constructed each X_{la}, and X_{un}.

(ii) Construct an ELM network of n_h hidden neurons, random input weights and biases, and calculate the output matrix of the hidden neurons $H \in R^{(la+un)Xn_h}$

(iii) Select the trade-off parameter C_0 and λ.

(iv) If n_h *is less than* N

Calculate the output weights using,

$$\gamma^* = (I_{n_h} + H^T CH + \lambda H^T L H)^{-1} H^T C\tilde{Y}$$

Else

Calculate the output weights using,

$$\gamma^* = H^T (I_{l+u} + CHH^T + \lambda LHH^T)^{-1} C\tilde{Y}.$$

return the mapping function $f(x) = h(x)\,\gamma$.

III. Unsupervised Extreme Learning Machine

In this section, the US-ELM algorithm recently developed by gaohuang et.al [18] has been briefly explained. This algorithm is introduced for learning data without targets. Additionally, it has also been stated that US-ELM gives favorable performance compared to state-of-the-art clustering algorithms.

In this algorithm, the complete training data $X\{x_i\}_{i=1}^{N}$ are unlabeled (N is the training patterns with different classes) and our aim is to search the optimized structure for the initial data. The origination of US-ELM follows from the formulation of SS-ELM. For the unlabeled data,Eq (2) is reduced to

$$\min_{\gamma \varepsilon R^{n_h X n_0}} \quad ||\gamma||^2 + \lambda T_r(\gamma^T H^T L H\gamma) \quad (4)$$

This formulation reaches its minimum at γ=0. After adding constraints, the formulation is given by

$$\min_{\gamma \varepsilon R^{n_h X n_0}} \quad ||\gamma||^2 + \lambda T_r(\gamma^T H^T L H\gamma) \quad (5)$$

$$s.t. \quad (H\gamma)^T H\gamma = I_{n_0}$$

US-ELM Algorithm is given by,

<u>Algorithm 2</u>

The USELM algorithm

Input:

The training data: $X \in R^{N\times n_i}$;

Output:

For embedding task: N_+ The embedding in a no-dimensional space: $E \in R^{N\times n_0}$

For clustering task:

The label vector of cluster index: $y \in N_+^{N\times 1}$

(i) Graph Laplacian L is constructed from X.

(ii) Construct an ELM structure of n_h hidden neurons with random input weights, and calculate the output matrix of the hidden neurons. $H \in R^{N\times n_h}$

(iii)

If $n_h \le N$

Find the generalized Eigen vectors $v_2, v_3, \ldots, v_{n_0+1}$ of $(I_{n_h} + \lambda H^T L H)v = \Upsilon H^T H v$

Corresponding to the second through the n_{0+1} smallest Eigen values.

Let $\gamma = [\widetilde{v_2}, \widetilde{v_3}, \ldots \ldots \ldots \tilde{v}n_{0+1}]$, where

$\widetilde{v_i} = v_i / \| Hv_i \|$,i= 2,….., n_{0+1}

Else

The generalized eigenvectors of

$(I_u + \lambda LHH^T)u = \Upsilon HHH^T u$.

$u_2, u_3, \ldots, u_{n_0+1}$ are found corresponding to the second from the $n_0 + 1$smallest eigenvalues.

Let $\gamma = [\widetilde{u_2}, \widetilde{u_3}, \ldots \ldots \ldots \tilde{u}n_{0+1}]$,

Where $\widetilde{u_i} = u_i / \|HH^T u_i\|$,*i*= 2,….., n_{0+1}

(iv) Compute the embedding matrix: E = $H\gamma$

(v) (For clustering): Consider each row of *E as a* point*,* and N points are clustered into K clusters with the help of k-means clustering algorithm. For all the points, let y be the label vector of cluster index.

return *E* (for the embedding task) or *y* (for the clustering task).

The eigenvectors obtained from US-ELM are not directly used for data representations, nevertheless the same are used as the factors of the network, i.e., the output weights. After the US-ELM model is trained, it can be used for any presented data in the original input space. Hence, USELM provide an open way for handling new patterns without recomputing eigenvectors

IV. Dataset Description And Performance Comparison Of US-ELM and SS-ELM For EEG Data

The dataset is obtained from UCI (University of California Irvine) repository [20]. The total numbers of samples present in the dataset are 182, out of which 91 samples are used for training and 91 samples are used for testing. Subsequently, the dataset used in [19] has 50% training and 50% testing data and the same numbers are used here. Wavelet transform is used for obtaining the features, with 12 being input feature vectors and one target vector. More details about the dataset can be seen from [19]. The results generated for semi-supervised ELM (SS-ELM) and unsupervised ELM (US-ELM) are shown in the Table I. All the networks presented in the table are real valued networks. NHN represents the number of hidden neurons. The results for ELM, SRAN (GA), McNN(GA), Mc-FIS, Fuzzy classifier has been obtained from [21]. From Table I, it can be inferred that SSELM performs better than US-ELM and ELM. The performance efficiency is better due to the use of spectral techniques for embedding and clustering.

TABLE I. CLASSIFICATION PERFORMANCE OF EEG SIGNAL

Classifier	**Overall Testing Efficiency**	**NHN**
ELM	71	10
US-ELM	**71.978**	**41**
SS-ELM	**73**	**80**
SRAN(GA)	70.8	20
McNN(GA)	70.33	20
Mc-FIS	71.429	20
Fuzzy classifier	71.42	20

V. Conclusion

To overcome the drawbacks of supervised learning algorithms, unsupervised and semi-supervised algorithms are developed [18]. These algorithms are used for EEG signal classification task. A quick response has been obtained with all the networks. In a few cases, similar to text classification, information retrieval and fault identification, getting labels for completely supervised learning are time consuming, whereas a large number of unlabeled data is open and cost effective to gather. The two additions of ELMs namely, semi-supervised and unsupervised learning uses spectral techniques for embedding and clustering and are probable to greatly enlarge the applicability of ELMs, and offer new insights for the extreme learning models. The overall testing efficiency of SSELM is 73% and USELM is 71.978% respectively. Moreover, the results clearly indicate that the performance of SSELM and USELM classifier are better than other real valued classifier considered for comparison. Hence, it depicts that these two networks provide good classification efficiency and it can be used in real world applications. As a future work, hybrid techniques can be incorporated in these networks for better classification.

References

[1] S. Suresh, N. Sundararajan, and P. Saratchandran, "Risk sensitive loss function for sparse multicategory classification problems," Information Sciences, vol.178, no.12, pp.2621-2638, 2008.

[2] S. Suresh, R. V. Babu, and H. J. Kim, "No reference image quality assessment using modified extreme learning machine classifier," *Applied Soft Computing*, vol. 9, no. 2, pp. 541–552, 2009.

[3] R. Savitha, S. Suresh, N. Sundararajan, and H. J. Kim "Fully Complex-valued ELM Classifiers for Real-valued Classification Problems", Neurocomputing, Vol: 78, No:1, pp. 104-110, 2012.

[4] K. Subramanian, S. Suresh, N. Sundararajan, "A meta-cognitive neuro-fuzzy inference system (McFIS) for sequential classification problems, IEEE Trans. on Fuzzy Systems, Vol.21,no.6,pp 1080-1095,2013.

[5] S. Suresh, R. Savitha, and N. Sundararajan, "A sequential learning algorithm for complex valued self-regulating resource allocation network-CSRAN," IEEE Transactions on Neural Networks, vol.22, no.7, pp.1061-1072,2011.

[6] S. Suresh, N. Sundararajan, and R. Savitha, Supervised Learning Algorithm in Complex Domain, Springer-Verlag, 2012.

[7] Sivachitra, M.,Savitha, R.,Suresh, S.,Vijayachitra,.2014,S . "A Fully Complex-valued Fast Learning Classifier (FC-FLC) for Real-valued Classification Problems" ,Neurocomputing, Vol.149,pp-198-206.

[8] E. Sathish, M. Sivachitra, Ramasamy Savitha, and S.Vijayachitra, "Wind profile prediction using a metacognitive fully complex-valued neural network," IEEE Internation conference on Advanced computing (ICoAC),pp1-6, 2012.

[9] S. Gokul, M. Sivachitra, and S. Vijayachitra, "Parkinson's disease prediction using machine learning approaches," IEEE, Fifth International Conference on Advanced Computing (ICoAC), 2013.

[10] M.Sivachitra, and S.Vijayachitra, A Meta-cognitive Fully Complex Valued Functional Link Predictor Network for solving real valued prediction problems. International Conference on CCIP (2015),IEEE, pp.1-6.

[11] K..R. Muller, C. W. Anderson, and G. E. Birch, "Linear and nonlinear methods for brain computer interfaces," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol.11, pp.165–169, 2003.

[12]G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine a new learning scheme of feedforward neural networks," in Int. Joint Conf. Neural Networks, vol. 2. IEEE, 2004, pp. 985–990.

[13] E. D. Karnin, "A simple procedure for pruning back-propagation trained neural networks," *IEEE Trans. Neural Netw.*, vol. 1, no. 2, pp. 239–242, 1990.

[14]G.-B. Huang,Q.-Y.Zhu,C.-K.Siew, Extreme learning machine :theoryand applications, Neurocomputing70(1)(2006)489–501.

[15]G.-B. Huang, X.Ding ,H.Zhou, Optimization method based extreme learning machine for classification,Neurocomputing74(1)(2010)155–163.

[16]R. Savitha, S. Suresh, and N. Sundararajan, "Fast learning complex valued classifiers for real valued classification problems," International Journal of Machine Learning and Cybernetics, vol. 4, no.5, pp.469-476, 2013.

[17] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp.533–536, 1986.

[18] Gao Huang, Shiji Song, Jatinder N. D. Gupta, and Cheng Wu "Semi-supervised and unsupervised extreme learning Machines ",IEEE Transactions on cybernetics, Vol. 44,no.12,pp.2405-2417,2014.

[19]Shweta Sahu, and Rajen B. Bhatt , "Automatic classification of Electroencephalography signals Using packet analysis and fuzzy decision trees", Proc. of 28th National Systems Conference (NSC-2004), Dec. 16-18, Vellore, INDIA, 2004.

[20] C. Blake and C. Merz, "UCI repository of machine learning databases,"Department of Information and Computer Sciences, University of California, Irvine, [URL: http://archive.ics.uci.edu/ml/], 1998.

[21] M.Sivachitra, and S.Vijayachitra, "Planning and relaxed state EEG signal classification using Complex Valued Neural Classifier for Brain Computer Interface