Score Normalization of Ensemble SVMs for Brain-Computer Interface P300 speller

Sourav Kundu
Department of Electronics and Communication Engineering
National Institute of Technology, Rourkela
Odisha-769008, India
Email: souravkundu90@gmail.com

Samit Ari
Department of Electronics and Communication Engineering
National Institute of Technology, Rourkela
Odisha-769008, India
Email: samit@nitrkl.ac.in

***Abstract*—Brain-computer interface (BCI) P300 speller can be used as a powerful aid for severely disabled people in their everyday life. The character recognition using P300 speller involves two stages for classification. First stage is to detect the P300 signal and second one is to determine the right character from the detected P300. Classification of P300 is a challenging task in character recognition process. Ensemble of classifiers is a robust method for classification as it reduces the classifier variability. In multiclassifier system the averaged score can be effected by one classifier as the score of different classifiers are not in the same level. To reduce the effect of one classifier, the score of the each classifiers are normalized. The proposed method includes different score normalization techniques for ensemble of SVMs (ESVM) for classification. Here min-max normalization, Z-score normalization and median and median absolute deviation (MAD) normalization techniques are used. The proposed algorithms have been evaluated on data set II of the BCI Competition III. It is observed that the performance of the proposed normalization technique is better compared to the earlier reported techniques for 5^{th} and 15^{th} epoch to classify different characters.**

***Index Terms*—Brain-computer interface (BCI), score normalization, ensemble support vector machine (ESVM), electroencephalogram (EEG), P300.**

I. Introduction

Brain-computer interface (BCI) might be the only medium of communication for individuals who are not able to convey through ordinary means because of severe motor disabilities like spinal cord injuries or amyotrophic lateral sclerosis (ALS) [1]. There are many alternative ways of communication for disabled people like voice or gesture based systems. However, these systems are not suitable for those individuals who suffer neuromuscular impairments. They are incapable of any muscular movement but have some cognitive abilities. The BCI system analyzes electroencephalogram (EEG) signal and sends the command to outside world. Several types of EEG signals are used for BCI system like P300, steady-state visually evoked potential (SSVEP), event-related desynchronization/synchronization (ERD/ERS) produced by motor imageries [2], etc. The BCI framework for character recognition used in this work is based on P300 which is a typical response of the brain to some predefined stimulus.

A P300 signal appears in EEG data due to the infrequent auditory or visual stimuli. It is named as P300 as a positive peak has appeared after 300ms of stimuli. When the P300 has been detected, it occurs for the stimuli appear before 300ms. From the detected P300 signal and flashing row-column information the character information can be extracted. The row-column intersection gives the character position in the speller board.

Over the last few years, several P300 classification algorithms are developed for character recognition. Ensemble support vector machine (ESVM) as a classifier and a recursive channel elimination method for channel reduction are reported in [1]. The recursive channel elimination is a time consuming task. Wavelet based feature with ensemble of fisher's linear discriminant (FLD) classifier is used in [3]. In [4] a multi-resolution approximation based feature selection is done and linear discriminant analysis (LDA) is used as a classifier.A regularied discriminative frame work is proposed in [5]. To classify the P300 signal convolutional neural networks (CNN) and temporal feature are used in [6]. A semi-supervised classifier based on least squares support vector machine (LS-SVM) is reported in [7]. Binary de-based channel selection and ensemble support vector machines (ESVMs) [8] is used for P300 detection. A novel distance coupled hidden markov models (HMM) classifier is proposed in [9]. In [10] genetic algorithm is used as a channel selection method and Bayesian linear discriminant analysis (BLDA) as a classifier.

In BCI system classifier selection, arrangement of classifiers and fusion of classifiers scores are an important task. In ensemble classifier system the score are coming from different classifier models. The scores from the different models are not in the same level, as a result one classifier's score can dominant the averaged score. Normalization is applied to the score to convert these scores into a common domain. In literature, there is no common rule for normalizing the score of different classifiers [11]. Here min-max normalization [12], [13], Z-score normalization [14] and median and median absolute deviation (MAD) [15] normalization techniques are proposed for ensemble of SVMs (ESVM).

This paper is organized as follows: Section II briefly describes the data set which is provided by the BCI competition and the description on speller paradigm. In Section III, the details about SVM is mentioned. The proposed framework is explained in Section IV. Finally, Section V represents the experimental results and comparisons with earlier reported works and conclusions of the work is given in Section VI.

II. The Data Set

The P300 speller is based on the oddball paradigm which states that when a rarely expected stimulus occurs, a positive deflection is observed in EEG signal after about 300ms. The data set is provided by the organizer of the BCI III competition.

A. BCI Paradigm

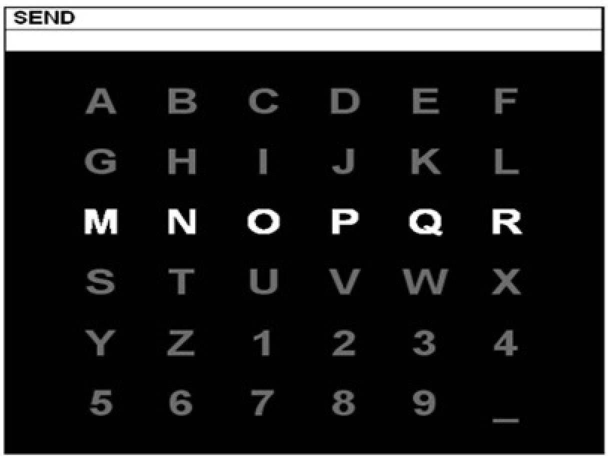

Fig. 1. P300 speller Paradigm [16]

The user interface for speller matrix contains 36 characters in 6×6 matrix. The user has to focus his attention on one character at one time [16]. The rows and columns of the matrix are intensified randomly and successively. The flashing rate is 5.7Hz. Two flashing contained the desired character out of 12 intensifications of rows or columns i.e., one row out of 6 rows and other for out of 6 columns. The responses evoked by these rare stimuli are not exactly the same those evoked by the stimuli that don't contain the desired character. These signals are called P300 signal as previously reported by Farwell and Donchin [17].

B. Database Used

The BCI competition III [16] data set II is used here. Two different subjects have participated in data collection and the data is collected in five different sessions. Every session is made out of various runs and for every run, a subject is asked to spell a character. The character matrix is intensified for 100 ms and blank for 75 ms. For one round there are 12 flashing and for one character the sets of 12 intensifications are repeated 15 times (i.e., each row/column is intensified 15 times and thus there are $12*15 = 180$ total intensifications for single character). Each repetition is called epoch. so, each character data consist of fifteen epochs. The EEG data is collected continuously from 64-channel. After bandpass-filtering from 0.1 - 60 Hz, the signal is digitized at a sample rate of 240 Hz. The database is composed of 85 training and 100 test characters of each subject.

III. Support Vector Machine (SVM)

SVM is an excellent tool for classification problems with a good generalization performance. Vapnik [18] designed this classifier for binary class problem. Let considered a training data set of N points $(x_i, y_i)_{i=1}^{N}$, where $x_i \in \mathbb{R}^m$ is ith input pattern and $y_i \in \{-1, 1\}$ is ith output pattern. To construct an optimal hyperplane which maximizes the margin boundary and minimizes the error (ξ). To solve this optimization problem quadratic programming (QP) problem is used.

$$\min_{w,\xi} \left[\frac{1}{2} \|w\|^2 + C \sum_{i=1}^{N} \xi_i \right] \tag{1}$$

where, w is weight vector and C is the regularization parameter. The regularization parameter C plays an important role in classification [19]. Smaller value of C ignores the points near to margin and increases the margin boundary, whereas the larger value of C considered all the points and to do so it is reduced the boundary. The Lagrangian representation of above function is

$$\begin{aligned} & \max_{\alpha} \left[\sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j \, y_i y_j k(x_i, x_j) \right] \\ & w = \sum_{i=1}^{N} y_i \alpha_i \Phi(x_i) \\ & \sum_{i=1}^{N} \alpha_i y_i = 0, \qquad 0 \le \alpha_i \le C, \ \forall i \end{aligned} \tag{2}$$

where α_is are Lagrange multipliers related to each training point, $k(x, x_i)$ represent the kernel function. The constructed SVM decision function is

$$f(x) = \sum_{i=1}^{N} \alpha_i y_i k(x, x_i) + b \tag{3}$$

where bias b is a real constant.

IV. Proposed framework

Three successive stages followed in the EEG based character recognition algorithm are preprocessing, feature extraction and classification. The basic steps of BCI systems are shown in Fig 2.

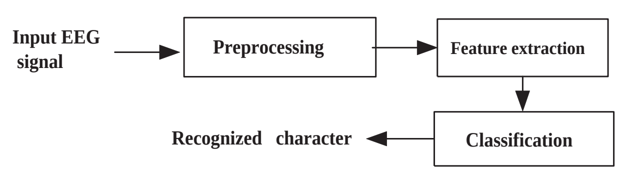

Fig. 2. Flowchart of proposed EEG based character recognition algorithm

A. Preprocessing and Feature extraction

The preprocessing stage involves the following sub-stages [1]: (i) From each channel a data of duration 0 to 667 ms is extracted after each flashing. As from the previous knowledge about P300 a positive peak will appear after 300 ms of stimulus. Therefore, it is postulated that a 667 ms window i.e. 160

samples are large enough to capture all necessary information for classification. These windows are overlapping windows. (ii) Each extracted signal has been filtered by a 8^{th} order bandpass Chebyshev filter of Type I and cut-off frequency lies within 0.1 and 20 Hz. (iii) Then these post-stimulus signal means 160 samples from each channel has been decimated according to the high cut-off frequency. After decimation, form a single channel 14 sample are taken. Then the decimated samples are transformed into a vector by concatenation of all 64 channels. Thus, for a single subject, the training set is composed of $15300 = 12 * 15 * 85$ post-stimulus vectors x_i of dimension $896 = 14 * 64$.

B. Model Selection

Here SVM is applied as a classifier. The regularization parameter C plays an important role in classifier performance. To select a proper C for SVM, a model selection procedure has been followed [1]. The training data is divided into 17 equal part which contains five characters in each part. Now each classifier is trained on one of the 17 partitions. These 17 partitions are divided into two subset as mentioned in [1]. At the time of model selection we have used one of it as a training and rest are for testing. Before classification the training data should be normalized to zero mean and unit variance. According to the normalised parameters obtain from the training dataset, the testing data is also normalised. The margin-error trade-off parameter for each SVM classifier has been selected by running the model selection procedure for different values of C. Then by select the C value that maximizes the score C_{cs}.

$$C_{cs} = \frac{t_p}{t_p + f_p + f_n} \tag{4}$$

where f_p, f_n, t_p are the number of false positive, false negative and true positive respectively for the validation set. Here true negative value is ignored because the data is unbalanced and the target is to detect the positive responses which are fewer compare to negative response. In this case, different values of C are $[0.01, 0.05, 0.1, 0.5, 1.0]$.

C. Ensemble of Support Vector Machine (ESVM)

ESVM is based on the averaging classifiers score as its reduced the classifier variability [20]. Now if there are K number of classifiers and numbers of sequences are J, then the ESVM decision function is written as

$$f_{avg}(x) = \frac{1}{K}\frac{1}{J}\sum_{k=1}^{K}\sum_{j=1}^{J} f_k(x) \tag{5}$$

$$f_{avg}(x) = \frac{1}{K}\sum_{k=1}^{K}\sum_{i=1}^{N} \alpha_i y_i k(\frac{1}{J}\sum_{j=1}^{J} x, x_i) + b \tag{6}$$

where x is the post stimuli vector and $f_{avg}(x)$ is the score of row or column. The classification score of the different classifier is normalized first, then the scores are averaged out as shown in Fig. 3. Among the six rows, which score is more, the desired character belong to that row and same is happened to the column. The intersection of row and column gives the desired character.

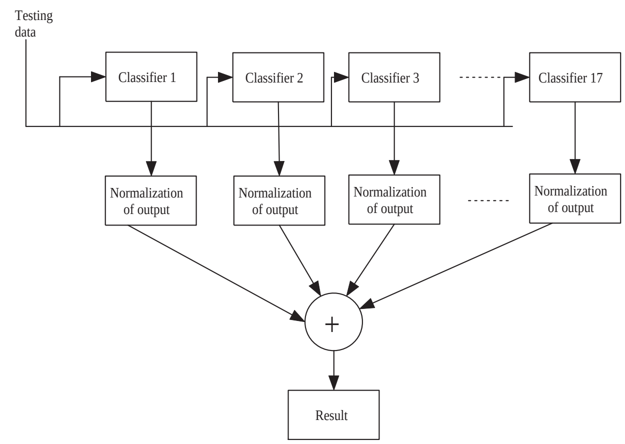

Fig. 3. Flowchart of proposed classification algorithm

D. Score normalization techniques

In the case of multiclassifier systems, there is no specific type of score normalization technique in literature. Three types of score normalization method are discussed here. Let the score of k^{th} classifier is represented by C_k.

- Min-max normalization [12]: The min-max normalized score is represented by C_{nk}.

$$C_{nk} = \frac{C_k - C_{\max}}{C_{\max} - C_{\min}} \tag{7}$$

  where $C_{\max}$ and $C_{\min}$ are the maximum and minimum of the score respectively. This type of normalization maintain the original distribution of the score, but it is sensitive to the outliers data. Only the amplitude of the score is maintain in the range of [0;1].
- Z-score normalization [14] : The normalized Z-score is represented as

$$C_{nk} = \frac{C_k - \mu}{\sigma} \tag{8}$$

  where μ is the mean and σ is the variance of the score. It is biased towards Gaussian distributions and does not guarantee a common numerical range for the normalized scores.
- Median and Median Absolute Deviation (MAD) [15]: MAD is normalized technique which insensitive to the outliers. It is defined as follows

$$C_{nk} = \frac{C_k - median}{MAD} \tag{9}$$

  where $MAD = median(|C_k - median|)$.

V. Results and Discussion

To evolute the proposed approach, the training data set is divided into seventeen equal parts as the training data consist of 85 character. Thus, it can only be divided into five equal parts consist of seventeen characters or vice versa. The second method is preferred as it gives more classifiers.

TABLE I
NUMBER OF CORRECTLY CLASSIFIED SYMBOLS FOR BCI COMPETITION III DATA SET

		Epochs														
Method	Subject	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Min-max normalization	A	18	32	57	62	70	73	83	82	84	87	89	93	96	97	97
	B	36	61	71	71	82	85	88	91	93	94	95	98	98	97	97
	Mean	27	46.5	64	66.5	76	79	85.5	86.5	88.5	90.5	92	95.5	97	97	97
Z-score normalization	A	18	33	59	63	70	72	83	83	84	87	89	93	96	97	97
	B	36	59	71	70	81	85	88	91	94	94	95	98	98	97	98
	Mean	27	46	65	66.5	75.5	78.5	85.5	87	89	90.5	92	95.5	97	97	97.5
MAD	A	17	32	59	61	71	73	82	83	83	88	89	93	96	97	97
	B	38	57	71	71	81	85	87	91	94	94	94	98	98	97	99
	Mean	27.5	44.5	65	66	76	79	84.5	87	88.5	91	91.5	95.5	97	97	98

For different normalization technique, the performance of character recognition accuracy is shown in Table I. For each subject the accuracy is calculated for different epochs and also the average accuracy is calculated. From the Table I, it is observed that as the number of epoch or number of sequences increases the percentage of character recognition accuracy increases. It is also observed that after 13^{th} epoch the increase rate of character recognition is slow. The aim of the BCI competition III is to report the classification result using all fifteen flash sequences (epochs) and additionally, only the first five flash sequences. A comparison between proposed method and other earlier reported methods is shown in Table II using the first five flash sequences and all 15 sequences. In Table II the result of BCI III competitor and some other algorithms results are shown. The BCI results have been received from the BCI cmpetition website [21]. From the result, it is observed that the performance of proposed classifier score normalization techniques are better compared to earlier reported techniques at 5^{th} and 15^{th} epoch.

TABLE II
PERFORMANCE COMPARISON OF THE PROPOSED TECHNIQUES WITH EARLIER REPORTED TECHNIQUE.

Method	Epoch	
	5	15
Yandong [21]	55.0	90.5
LDA [22]	60.5	92.0
WT-EFLD [3]	71.5	95.0
Tomika *et al.* [5]	75.0	97.0
GA-BLDA [10]	75.0	98.0
ESVM [1]	73.5	96.5
MCNN-1 [6]	69.0	95.5
MAD-ESVM	76.0	98.0

The results referred in [1] is 73.5% after 5^{th} epoch and 96.5% after 15^{th} epoch whereas Min-max normalization method achieves 76.0% and 97.0% respectively. In case of Z-score, the accuracy is 75.5% and 97.5% and for MAD normalization the accuracy is 76.0% and 98.0% after 5^{th} and 15^{th} epoch respectively.

The performance of the proposed method is better compared to other because the score of the classifiers are normalized. As the training data set is divided into several part and the different SVM models are generated from these data sets. So, the score of the different classifiers are heterogeneous in nature. For combining these classifiers, score normalization has been proposed to transform these scores into a common domain. MAD normalization gives better result compare to min-max and z-score normalization as it is insensitive to outliers. The median makes MAD normalization robust against extreme points

VI. CONCLUSION

In this paper, the effect of different score normalization techniques on the performance of character recognition accuracy in P300 speller is discussed. Feature extraction, feature selection and classification are the important steps in character recognition. In classification step, we used multiclassifier system and the score of the different classifiers are normalized according to min-max, z-score and MAD technique. The regularization parameter C is selected based on the model selection procedure. In training phase, according to the C the weight vector change itself to minimize the error. The performance of the above algorithm is evaluated on dataset II of BCI competition III, which is a benchmark data available on-line. The proposed method gives better result compared to other reported techniques when the number of epochs are 5 and 15. This model will not work well if the noise level is high as EEG signal is highly sensitive to artifacts.

ACKNOWLEDGMENT

The authors would like to thank A. Rakotomamonjy for his help towards the ESVM method. This work is supported by the Young Faculty Research Fellowship under Visvesvaraya Ph D scheme for Electronics & IT, DeitY, India.

REFERENCES

[1] A. Rakotomamonjy and V. Guigue, “BCI competition III: dataset II-ensemble of SVMs for BCI P300 speller,” *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 3, pp. 1147–1154, 2008.

[2] L. F. Nicolas-Alonso and J. Gomez-Gil, “Brain computer interfaces, a review,” *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012.

[3] M. Salvaris and F. Sepulveda, “Wavelets and ensemble of flds for p300 classification,” in *2009 4th International IEEE/EMBS Conference on Neural Engineering*. IEEE, 2009, pp. 339–342.

[4] U. Celik and S. Anca, “Classification of the visual evoked eeg using multiresolution approximation based on excitatory post-synaptic potential waveform,” in *Electrical and Electronics Engineers in Israel (IEEEI), 2010 IEEE 26th Convention of*. IEEE, 2010, pp. 929–933.

[5] R. Tomioka and K.-R. Müller, “A regularized discriminative framework for eeg analysis with application to brain–computer interface,” *NeuroImage*, vol. 49, no. 1, pp. 415–432, 2010.

[6] H. Cecotti and A. Graser, “Convolutional neural networks for p300 detection with application to brain-computer interfaces,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 433–445, 2011.

[7] Z. Gu, Z. Yu, Z. Shen, and Y. Li, “An online semi-supervised brain–computer interface,” *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 9, pp. 2614–2623, 2013.

[8] R. K. Chaurasiya, N. D. Londhe, and S. Ghosh, “Binary de-based channel selection and weighted ensemble of svm classification for novel brain–computer interface using devanagari script-based p300 speller paradigm,” *International Journal of Human–Computer Interaction*, pp. 1–17, 2016.

[9] S. Zhong and J. Ghosh, “Hmms and coupled hmms for multi-channel eeg classification,” in *Proceedings of the IEEE International Joint Conference on Neural Networks*, vol. 2, 2002, pp. 1254–1159.

[10] C.-Y. Kee, S. Ponnambalam, and C.-K. Loo, “Multi-objective genetic algorithm as channel selection method for p300 and motor imagery data set,” *Neurocomputing*, vol. 161, pp. 120–131, 2015.

[11] G. Pirlo and D. Impedovo, “Adaptive score normalization for output integration in multiclassifier systems,” *IEEE Signal Processing Letters*, vol. 19, no. 12, pp. 837–840, 2012.

[12] S. Tulyakov, S. Jaeger, V. Govindaraju, and D. Doermann, “Review of classifier combination methods,” *Machine Learning in Document Analysis and Recognition*, pp. 361–386, 2008.

[13] A. Jain, K. Nandakumar, and A. Ross, “Score normalization in multimodal biometric systems,” *Pattern recognition*, vol. 38, no. 12, pp. 2270–2285, 2005.

[14] H. J. Jung and M. Lease, “Improving consensus accuracy via z-score and weighted voting.” in *Human Computation*, 2011.

[15] J. Kittler, M. Hatef, R. P. Duin, and J. Matas, “On combining classifiers,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 20, no. 3, pp. 226–239, 1998.

[16] B. Blankertz, K.-R. Muller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlogl, G. Pfurtscheller, J. R. Millan, M. Schroder, and N. Birbaumer, “The BCI competition III: Validating alternative approaches to actual BCI problems,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2, pp. 153–159, 2006.

[17] L. A. Farwell and E. Donchin, “Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials,” *Electroencephalography and clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, 1988.

[18] V. Vapnik, *The nature of statistical learning theory*. Springer Science & Business Media, 2013.

[19] B. Zhang, “Is the maximal margin hyperplane special in a feature space,” *Hewlett-Packard Research Laboratories Palo Alto*, 2001.

[20] Y. Grandvalet, “Bagging equalizes influence,” *Machine Learning*, vol. 55, no. 3, pp. 251–270, 2004.

[21] B. Blankertz, “BCI Competition III-Final Results,” 2008. [Online]. Available: http://www.bbci.de/competition/iii/results/

[22] N. Liang and L. Bougrain, “Averaging techniques for single-trial analysis of oddball event-related potentials,” in *4th International Brain-Computer Interface Workshop*, 2008.