Weighted sparse representation for classification of motor imagery EEG signals

S. R. Sreeja, Himanshu, Debasis Samanta(✉), Monalisa Sarma

***Abstract*— Motor imagery (MI) based brain-computer interface systems (BCIs) are highly in demand for many real-time applications such as hands and touch-free text entry, prosthetic arms, virtual reality, movement of wheelchairs, etc. Traditional sparse representation based classification (SRC) is a thriving technique in recent years and has been a successful approach for classifying MI EEG signals. To further improve the capability of SRC, in this paper, a weighted SRC (WSRC) has been proposed for classifying two-class MI tasks (right-hand, right-foot). WSRC constructs a weighted dictionary according to the dissimilarity information between the test data and the training samples. Then for the given test data the sparse coefficients are computed over the weighted dictionary using l_0-minimization problem. The sparse solution obtained using WSRC gives better discriminative information than SRC and as a consequence, WSRC proves to be superior for MI EEG classification. The experimental results substantiate that WSRC is more efficient and accurate than SRC.**

***Index Terms*—Electroencephalography, motor imagery, sparsity-based classification, weighted sparse representation**

I. INTRODUCTION

Brain-computer interface systems (BCIs) aim to provide a better and quality life for the people with disabilities [1]. BCIs allow the people to control the external devices through direct brain activity. With BCI, the recent trend is to use electroencephalogram (EEG) activity induced by movement imaginations as it provides high degrees of freedom and easily detectable in both healthy and disabled individuals [1]. Thus, people with severe motor disabilities can express their intensions and communicate with the environment via motor imagery (MI) based BCIs instead of normal pathways such as peripheral nerves and muscle movements.

The processing of MI EEG data in many real-life applications, is a non-trivial task. Many initiatives, in this regard, applying machine learning algorithms have been reported in the recent literature. Although, machine learning techniques solve the problems, but accuracy and applicability is far from feasibility. Of late, sparse representation based classification (SRC) has become an emerging topic in many fields such as image classification [2], face recognition [3], speech recognition [4], etc. Sparsity approach has also been successfully implemented for MI EEG signal classification [5], [6], [7]. In sparse representation, signals can be represented as a linear combination of few columns (also called as atoms) taken from a dictionary. The dictionary can be learned from a given data or predefined bases like Fourier, wavelets, Dirac, etc., can be used for the purpose. Once the dictionary is designed or fixed, sparse coefficients are then estimated. By calculating the residual between the test data and the sparse linear combinations, SRC assigns the test data to the MI class which yields the least residual error.

SRC works fine with noisy EEG data. However, for the same MI from the same subject at different timings the sparse coefficients were scattered at different positions. It proves that SRC approach employs only linear information but fails with data locality (distance), which is necessary for classification. Also, from [8], it is evident that locality is more essential than sparsity, because locality will lead to sparsity but not necessarily vice-versa. To overcome this limitation, Lu et al. [9] and Fan et al. [3] proposed a variant of sparse representation based classifier for facial recognition problems, which is more robust than SRC. Motivated by the above-mentioned algorithms, a method has been proposed to calculate distance between the test data and the training data. This distance information proves the significance of the training data in representing the test data. Using this distance information, a weighted SRC (WSRC) model has been build to enhance the classification result of SRC.

Dictionary atoms closer to the test data tends to have greater significance in representing the same. Taking this distance or dissimilarity information into account, WSRC allots weight for each training samples (also called atoms). The weights for the atoms closer to the test data will be larger. In this way, the test data will be represented by the meaningful atoms taking both linearity and locality into account. When the weighted dictionary is employed to represent the test data, the result should be sparser than that obtained using unweighted dictionary (SRC approach). Finally, using the sparse representation, classify the test data to the MI class with the lowest reconstruction error. Thus, WSRC approach consists of the following steps:

1) Compute the weights between the atoms and the test data using Euclidean distance based Gaussian kernel.
2) Design a locality-constrained (weighted) dictionary.
3) Sparse representation over weighted dictionary.
4) Classify the test data using minimal residual (least reconstruction error).

II. MI DATA CLASSIFICATION TECHNIQUES

MI data has been classified using various techniques and they are broadly categorized into three groups: 1) Machine learning (ML) algorithms, 2) Deep learning (DL) architectures and 3) Sparse representation based classifiers. In

S. R. Sreeja and Debasis Samanta are with Department of Computer Science and Engineering, Indian Institute of Technology, Kharagpur, West Bengal, India-721302 {sreejasr, dsamanta}@iitkgp.ac.in

Himanshu is with Department of Mathematics and Monalisa Sarma is with Subir Chowdhury School of Quality and Reliability, Indian Institute of Technology Kharagpur, West Bengal, India 721302

this section, for each category, the recent works have been reviewed and reported.

A. ML algorithms based techniques

ML-based techniques help to identify the class labels by extracting useful information from the brain activities. Feature extraction (FE) is a major task and an efficient FE can achieve good accuracy with a suitable classifier. Several FE methods available for MI data, such as common spatial pattern (CSP) [10], [11], wavelet packet decomposition [12], cross-correlation [13], etc. These FE methods are combined with different types of classifiers, such as logistic regression (LR) [13], support vector machine (SVM) [10], [14], linear discriminant analysis (LDA) [15], [11], Gaussian Naïve Bayes (GNB) [16], etc., to classify the MI data.

B. DL architectures based approaches

There is also a continued interest in an advanced ML algorithms called DL algorithms. Many research works on MI EEG signal classification using DL methods have been carried out. The main advantage of DL methods is that they can decode task-related information from the raw EEG data without handcrafted features. Some of the recent architectures on MI EEG data classification include deep learning with convolutional neural networks (deep ConvNets) for EEG decoding [17], denoising autoencoder (DAE) [18], Restrict Boltzmann machine (RBM) [19] and deep neural networks (DNN) [20], [21].

C. Sparsity based classification techniques

The latest researches on sparse representation and dictionaries have shown that the dictionaries learned from the data itself outperforms the pre-defined dictionaries [22]. Earlier, a work on channel selection using sparsity was carried out by Arvaneh et al. [23]. Later, SRC for MI-based BCI [5] have been developed by Shin et al. Zhang et al. [10] found a sparse method to automatically select significant filter bands to improve classification. Sreeja et al. [6] classified MI-based EEG signals making use of band-pass filter and CSP, extracted wavelet energy as the feature to construct the dictionary and SRC for classification. Recently, Jiao et al. [7] applied sparse group representation model (SGRM) for MI EEG data classification.

III. Data and Methodology

This section describes the MI data used in this study and the proposed WSRC for classifying the different MI tasks.

A. Dataset Description

The publicly available dataset IVa from BCI competition III[1] has been used to validate the proposed approach. The dataset consists of EEG recorded data from five subjects (aa, al, av, aw, ay) who performed right-hand and right-foot MI tasks during each trial. According to the international 10-20 system, MI signals were recorded from 118 channels. For each subject, there were 140 trials for each task, and therefore 280 trials totally. The measured EEG signal was filtered using a bandpass filter between 0.05 - 200 Hz. Then the signal was digitized at 1000 Hz with 16 bit accuracy and it is downsampled to 100 Hz for further processing. The total number of trials was splitted into different number of labeled and unlabeled trials for each subject.

[1]http://www.bbci.de/competition/iii

B. Methodology

This section explains the dictionary construction and the proposed WSRC approach in detail.

1) Dictionary construction: The raw EEG data from 30 channels (FC2, FC4, FC6, CFC2, CFC4, CFC6, C2, C4, C6, CCP2, CCP4, CCP6, CP2, CP4, CP6, FC5, FC3, FC1, CFC5, CFC3, CFC1, C5, C3, C1, CCP5, CCP3, CCP1, CP5, CP3 and CP1) present over the motor cortex is considered for the experiment [6]. Each trial of MI signal of *four* seconds is divided into eight segments with half a second data in each segment, that is, $\mathbf{X} \in \mathbb{R}^{N_d \times N_{ch}}$. Here, N_{ch} denotes number of channels (30) and N_d denotes number of samples in half a second data (50). For each segment the wavelet energy $\mathbf{x_w} \in \mathbb{R}^{1\times(2\times N_{ch})}$ using *coif1* wavelet is estimated. This wavelet energy $(\mathbf{x_w})^T$ of each segment of each class $C_i = \{1, 2\}$ concatenates to form the training samples of the sub-dictionary matrices $\mathbf{D}_1$ and $\mathbf{D}_2$. Finally, the sub-dictionary matrices of right-hand and right-foot MI classes are concatenated to form an over-complete dictionary $\mathbf{D} := [\mathbf{D}_1, \mathbf{D}_2] \in \mathbb{R}^{(2N_{ch})\times n}$ where $n = \sum n_i$ and n_i is the total number of training samples extracted for the class C_i. For each subject, an individual dictionary has been constructed to make the approach work better for an individual subject in a lesser time.

2) WSRC: The framework of the proposed WSRC approach is given in Fig 1. The WSRC algorithm consists of two major steps. In the first step, it measures the dissimilarity (distance) information between the dictionary atoms and the given test data $\mathbf{y}_i$. This dissimilarity information is used for defining the weights of the dictionary atoms [3]. Each dictionary atom is then multiplied by its weight to build a locality-constrained weighted dictionary. In the second step, it performs the traditional SRC using the weighted dictionary matrix. A smaller value of distance indicates a smaller dissimilarity between the test data and the training samples. Thus, WSRC can generate more discriminative sparse coefficients that can be used to reconstruct the test data more robustly [2].

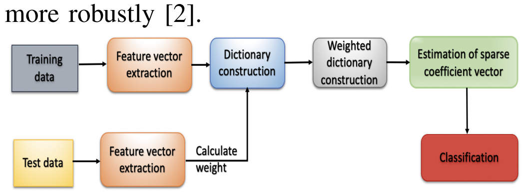

Fig. 1: Framework of the proposed WSRC approach

The dissimilarity can be measured using distances such as Mahalanobis distance (MD), absolute distance (AD), Euclidean distance (ED), χ^2 distance, cosine similarity, etc. In this work, the distance between the test data and the dictionary atom is estimated using ED and it is defined as:

$$distance = \|\mathbf{y_i} - \mathbf{d_j}\|_2^2, \quad j = 1, 2, ..., n \qquad (1)$$

Then the weights $\mathbf{w_j}$ for each dictionary atom is estimated using Gaussian kernel. This Gaussian kernel helps to capture the non-linear information within the given EEG data. The weight $\mathbf{w_j}$ is defined as

$$\mathbf{w_j} = e^{\frac{-distance}{\sigma}} \tag{2}$$

where σ is the average distance of the dictionary atoms, which helps to adjust the weights. The training atoms that are similar to the test data will have a larger weight than the non-similar ones. After obtaining the weights $\{\mathbf{w_j}\}$, the locality-constrained dictionary $\mathbf{D'}$ is constructed by performing element-wise multiplication of weight with each atom and is defined as

$$\mathbf{D'} = [w_1 d_1, w_2 d_2, ..., w_j d_j] \in \mathbb{R}^{(2N_{ch}) \times n} \tag{3}$$

With the newly designed weighted dictionary matrix $\mathbf{D'}$, for the given test data $\mathbf{y_i}$, the sparse coefficient vector is estimated by solving the l_0-minimization problem as follows.

$$\underset{\alpha_\mathbf{i}}{min} \|\alpha_\mathbf{i}\|_0 \text{ subject to } \mathbf{y_i} = \mathbf{D'}\alpha_\mathbf{i} \tag{4}$$

Orthogonal matching pursuit (OMP) algorithm [24] is used to estimate the sparse coefficients. The class of the test data $\mathbf{y}_i$ is determined by the minimal residual between $\mathbf{y}_i$ and $\mathbf{D'}\alpha_\mathbf{i}$ and it is defined as

$$Class(\mathbf{y}_i) = \arg\min_i \|\mathbf{y_i} - \mathbf{D'}\alpha_\mathbf{i}\|_2 \tag{5}$$

The traditional SRC algorithm is based only on the sparse linear representation, whereas WSRC is based on both locality and sparsity using distance information. Researches have shown that locality is more essential, as locality leads to sparsity but not necessarily vice-versa [8]. For this reason, WSRC integrates the locality structure of data into basic sparse representation. This way, locality-constrained WSRC will be more discriminative than SRC.

IV. Experimental results

In this section, the various experiments and the results obtained using WSRC method are presented. All the experiments have been executed on a machine with 3.2 GHz CPU, 4 GB RAM, and software (Python 2.7)[2]. The results have been recorded and evaluated using various quantitative measures. The results include comparison of the WSRC approach with the existing SRC approaches and ML - DL based classification approaches.

A. Performance comparison of WSRC with the existing sparsity-based techniques

To prove the efficiency of the proposed WSRC approach, the results obtained were compared with the existing sparsity-based approaches. In WSRC, for each subject, unique wavelet dictionary has been constructed as the main objective is to make the classification method work faster. A k-fold cross validation was performed on the dataset to split the entire data into k folds, from which $k-1$ folds were used to build the dictionary and *one* fold for testing the model. The k value was chosen as 10 and each fold was used for testing iteratively and the accuracies were calculated. The subject-wise performance of WSRC and the existing sparsity-based approaches are given in Table I. The values listed in Table I proves that WSRC performs better than existing sparsity-based approaches in terms of accuracy.

[2]https://github.com/BCI-HCI-IITKGP/Weighted-Sparse-classification

TABLE I: Subject-wise performance comparison of WSRC with the existing sparsity-based techniques

Methodology	Classification Accuracy (%)					
	aa	al	av	aw	ay	Mean
Sparse common spatial pattern (SCSP) [23]	80.71	97.14	57.14	85	91.42	82.28
CSP, SRC using band-power features [5]	98.83	100	95.71	97.86	91.79	96.85
CSP, SRC using wavelet energy as features [6]	98.98	97.02	96.45	98.21	93.90	96.91
Sparse group representation model (SGRM) [7]	73.90	94.50	59.50	80.70	79.90	77.70
Proposed WSRC	**99.88**	**99.90**	**98.42**	**98.54**	**93.18**	**97.98**

B. Comparison with ML and DL classifiers

In order to prove the efficiency of the proposed WSRC algorithm, the results have been compared with some widely used machine learning and deep learning algorithms. In order to make the comparison worthy, the kappa (κ) coefficient, classification error and average accuracy were calculated. Table II summarizes the results achieved by the proposed approach and the existing ML and DL methods. On an average, the proposed WSRC approach performed significantly better than the existing methods.

TABLE II: Performance comparison of WSRC with ML and DL classifiers

Methodology	Classification Accuracy (%)	κ score	Classification Error
CSP + LDA [15]	90.5 ± 7.6	-	-
SFBCSP [10]	-	-	7.95
CSP + DNN [20]	-	-	9.80
SFFS + SVM [14]	84.2 ± 10.6	-	-
DFBCSP [11]	-	0.832	-
CSP + GNB [16]	95.47	0.91	
WSRC	**97.98 ± 1.27**	**0.95**	**6.45**

C. Performance analysis

It is observed that WSRC outperforms other SRCs as a classification tool. Here below, the superiority of WSRC has been illustrated by an example. Fig. 2 (top row) shows the sparse coefficients (SC) obtained by SRC [6] for a subject who performed the same MI tasks (say, right-foot) at different trials. However, the positions of the SC in different trials are quite different in SRC approach. On the other hand, the SC obtained by WSRC for the same MI tasks at different trials are very similar to each other (see Fig. 2 (bottom row)) and have a smaller distance apart. From Fig. 2, WSRC appears to better represent the MI EEG signal, taking both sparsity and locality into account. Moreover, the performance measures obtained for the subject al using SRC and WSRC are shown in Fig. 3 and it proves that the weighted approach performs significantly better than the unweighted approach. Further, the quantitative measures calculated for SRC and WSRC approaches are listed in Table III. The values listed in Table III proves that WSRC performs better than SRC approach in terms of accuracy and time. This is because, WSRC approach can represent the data perfectly with a very few non-zero elements. For example, in SRC, the subject al achieves better performance with 60 non-zero elements whereas WSRC requires only 5 non-zero elements (see Fig. 2). This makes the WSRC algorithm run faster.

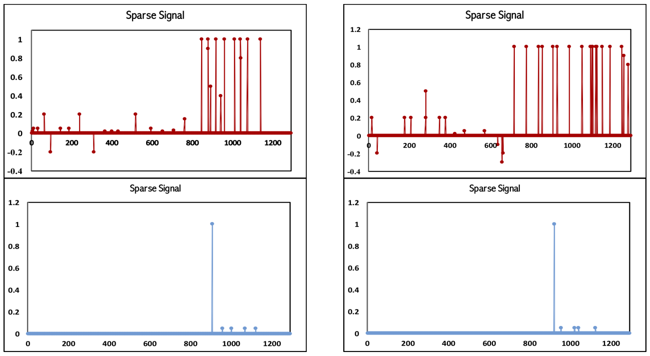

Fig. 2: SC obtained for the same MI task at different time for the same subject using SRC (top) and WSRC (bottom).

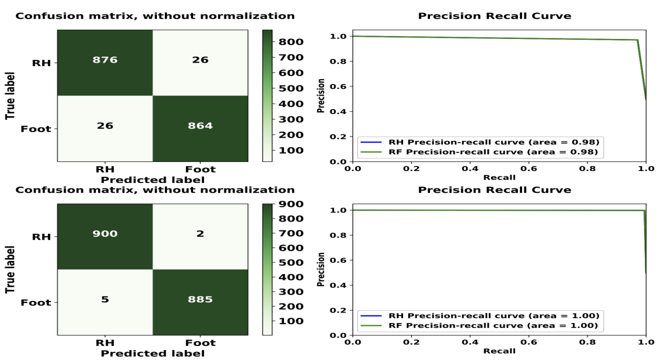

Fig. 3: Confusion matrix, precision-recall curve for the subject *al* using SRC (top row) and WSRC (bottom row).

TABLE III: Performance comparison of SRC and WSRC

Methods	Accuracy (%)	κ score	f_1 Score	Runtime (s)
SRC	96.91 $\pm$ 1.83	0.94	0.96	0.027
WSRC	**97.98 $\pm$ 1.27**	**0.95**	**0.99**	**0.016**

V. Conclusion

The existing SRC approach suffers from data locality and hence, as an alternative to SRC, in this work, a locality-constrained sparsity based approach has been investigated for classifying two-tasks MI EEG signals. The result is certainly in favour of the development of any real-time BCI applications. Here, the scalability, as far as the dictionary construction is concerned is not an issue as only wavelet energy extracted from the raw EEG signals is enough. It may be noted that there is no pre-processing task required, which enables the technique more faster. For the given test signal, distance information with the dictionary atoms are calculated, weights have been alloted for each atoms and the weighted dictionary is constructed. Later, the test signal is represented using significant dictionary atoms and the classification is carried out. The results produced by WSRC approach is good in terms of accuracy and time when compared with SRC approach. Moreover, WSRC performs well with all the subjects irrespective of the number of training signals. It may be concluded that, WSRC can be used as a classification tool for real-time MI based BCI applications.

References

[1] H. Yuan and B. He, "Brain-computer interfaces using sensorimotor rhythms: current state and future perspectives," *IEEE Trans. Biomed. Engg.*, vol. 61, pp. 1425–1435, 2014.

[2] L. Gan, J. Xia, P. Du, and Z. Xu, "Dissimilarity-weighted sparse representation for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 11, pp. 1968–1972, 2017.

[3] Z. Fan, M. Ni, Q. Zhu, and E. Liu, "Weighted sparse representation for face recognition," *Neurocomputing*, vol. 151, pp. 304–309, 2015.

[4] P. Sharma, V. Abrol, and A. K. Sao, "Deep-sparse-representation-based features for speech recognition," *IEEE/ACM Trans. Audio, Speech, and Language Process.*, vol. 25, no. 11, pp. 2162–2175, 2017.

[5] Y. Shin, S. Lee, J. Lee, and H.-N. Lee, "Sparse representation-based classification scheme for motor imagery-based brain–computer interface systems," *J. Neural Engg.*, vol. 9, no. 5, p. 056002, 2012.

[6] S. R. Sreeja, J. Rabha, D. Samanta, P. Mitra, and M. Sarma, "Classification of motor imagery based EEG signals using sparsity approach," in *Intl. Conf. IHCI*. Springer, 2017, pp. 47–59.

[7] Y. Jiao, Y. Zhang, X. Chen, E. Yin, J. Jin, X. Wang, and A. Cichocki, "Sparse group representation model for motor imagery EEG classification," *IEEE J. Biomed. Health Info.*, vol. 23, pp. 631–641, 2019.

[8] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *IEEE Conf. CVPR*, 2010, pp. 3360–3367.

[9] C.-Y. Lu, H. Min, J. Gui, L. Zhu, and Y.-K. Lei, "Face recognition via weighted sparse representation," *J. Visual Communication and Image Representation*, vol. 24, no. 2, pp. 111–116, 2013.

[10] Y. Zhang, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Optimizing spatial patterns with sparse filter bands for motor-imagery based brain–computer interface," *J. Neurosci. Meth.*, vol. 255, pp. 85–91, 2015.

[11] S. Kumar, A. Sharma, and T. Tsunoda, "An improved discriminative filter bank selection approach for motor imagery EEG signal classification using mutual information," *BMC bioinf.*, vol. 18, p. 545, 2017.

[12] J. Kevric and A. Subasi, "Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system," *Biomed. Sig. Process. and Control*, vol. 31, pp. 398–406, 2017.

[13] Y. Li, P. P. Wen, *et al.*, "Modified CC-LR algorithm with three diverse feature sets for motor imagery tasks classification in EEG based brain–computer interface," *Computer Methods and programs in Biomedicine*, vol. 113, no. 3, pp. 767–780, 2014.

[14] Z. Qiu, J. Jin, H.-K. Lam, Y. Zhang, X. Wang, and A. Cichocki, "Improved SFFS method for channel selection in motor imagery based BCI," *Neurocomputing*, vol. 207, pp. 519–527, 2016.

[15] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Muller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Sig. Proces. Mag.*, vol. 25, no. 1, pp. 41–56, 2008.

[16] S. R. Sreeja, J. Rabha, K. Nagarjuna, D. Samanta, P. Mitra, M. Sarma, *et al.*, "Motor imagery EEG signal processing and classification using machine learning approach," in *IEEE Intl. Conf. New Trends in Computing Sci. (ICTCS)*, 2017, pp. 61–66.

[17] Z. Tang, C. Li, and S. Sun, "Single-trial EEG classification of motor imagery using deep convolutional neural networks," *Optik-International J. Light and Electron Optics*, vol. 130, pp. 11–18, 2017.

[18] J. Li, Z. Struzik, L. Zhang, and A. Cichocki, "Feature learning from incomplete EEG with denoising autoencoder," *Neurocomputing*, vol. 165, pp. 23–31, 2015.

[19] N. Lu, T. Li, X. Ren, and H. Miao, "A deep learning scheme for motor imagery classification based on restricted boltzmann machines," *IEEE Trans. Neural Sys. Rehab. Engg.*, vol. 25, no. 6, pp. 566–576, 2017.

[20] S. Kumar, A. Sharma, K. Mamun, and T. Tsunoda, "A deep learning approach for motor imagery EEG signal classification," in *IEEE 3rd APWC on CSE*, 2016, pp. 34–39.

[21] I. Sturm, S. Lapuschkin, W. Samek, and K.-R. Müller, "Interpretable deep neural networks for single-trial EEG classification," *J. Neuroscience Methods*, vol. 274, pp. 141–145, 2016.

[22] S. R. Sreeja, R. R. Sahay, D. Samanta, and P. Mitra, "Removal of eye blink artifacts from EEG signals using sparsity," *IEEE J. Biomedical and Health Informatics*, vol. 22, no. 5, pp. 1362–1372, 2018.

[23] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, "Optimizing the channel selection and classification accuracy in EEG-based BCI," *IEEE Trans. Biomed Engg.*, vol. 58, no. 6, pp. 1865–1873, 2011.

[24] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Infn. theory*, vol. 50, pp. 2231–2242, 2004.