HIDDEN CONDITIONAL RANDOM FIELDS FOR CLASSIFICATION OF IMAGINARY MOTOR TASKS FROM EEG DATA

Jaime F. Delgado Saa†‡, Müjdat Çetin†

† Signal Processing and Information Systems Laboratory, Sabanci University
Orhanli, Tuzla, 34956 Istanbul, Turkey
‡ Department of Electrical and Electronics Engineering, Universidad del Norte
Barranquilla, Colombia
[email: delgado, mcetin]@sabanciuniv.edu

ABSTRACT

Brain-computer interfaces (BCIs) are systems that allow the control of external devices using information extracted from brain signals. Such systems find application in rehabilitation of patients with limited or no muscular control. One mechanism used in BCIs is the imagination of motor activity, which produces variations on the power of the electroencephalography (EEG) signals recorded over the motor cortex. In this paper, we propose a new approach for classification of imaginary motor tasks based on hidden conditional random fields (HCRFs). HCRFs are discriminative graphical models that are attractive for this problem because they involve learned statistical models matched to the classification problem; they do not suffer from some of the limitations of generative models; and they include latent variables that can be used to model different brain states in the signal. Our approach involves auto-regressive modeling of the EEG signals, followed by the computation of the power spectrum. Frequency band selection is performed on the resulting time-frequency representation through feature selection methods. These selected features constitute the data that are fed to the HCRF, parameters of which are learned from training data. Inference algorithms on the HCRFs are used for classification of motor tasks. We experimentally compare this approach to the best performing methods in BCI competition IV and the results show that our approach overperforms all methods proposed in the competition. In addition, we present a comparison with an HMM-based method, and observe that the proposed method produces better classification accuracy.

1. INTRODUCTION

A brain-computer interface (BCI) is a system that provides an alternative communication pathway for patients who have lost their ability to perform motor tasks due to a disease or an accident [1]. In addition, applications for healthy subjects in the fields of multimedia and gaming have started to incorporate these technologies in recent years as well [2]. BCIs aim to use brain signals to help subjects control external devices and interact with their environment. In the case of execution of (real or imaginary) motor tasks, it is known that electroencephalography (EEG) signals measured over the motor cortex exhibit changes in power related to the movements. These changes primarily involve increase and decrease of power in the alpha (8Hz-13Hz), sigma (14Hz-18Hz), and beta (18Hz-30Hz) frequency bands. These phenomena are known as event related synchronization and desynchronization [3]. This information can be used to classify different imaginary motor tasks by comparison of the power levels of the EEG signals recorded in a number of positions on the scalp. In particular, changes of the signal power in different frequency bands with time provide useful information. Based on this observation, methods based on time-frequency analysis of

the EEG signals have been proposed [4, 5, 6]. Furthermore, algorithms involving stochastic time series models taking into account changes of the signal power with time, such as hidden Markov models (HMMs) [7, 8, 9, 10] have been used in combination with features describing the temporal behavior of the EEG signals [11, 12]. We share the perspective with this latter body of work that changes in the power of the signals during execution of motor tasks reflect the underlying states in the brain and that the sequence of states provides useful information for discrimination of different imaginary motor tasks. In previous work based on HMMs, we have shown that this approach provides good results [10]. Nevertheless, if the EEG signal is modeled by an HMM, which is a generative model, the distribution of the data must be estimated and conditional independence assumptions of the data given the underlying states should be incorporated in order to make the inference problem tractable. A remedy for this problem, used in other application domains, is the use of the conditional random fields (CRFs) [13]. Although this is a discriminative model that does not require the estimation of the distribution of the data, there is one more issue for the case of the BCI applications, where, unlike the analysis of sleep EEG signals based on CRFs as proposed in [14], the sequence of states in the BCI problem is unknown, making it necessary to incorporate latent variables into the model. This issue can be addressed through the use of so-called hidden conditional random fields (HCRFs) [15]. In this paper, we present an HCRF-based approach for classification of imaginary motor tasks. We perform feature extraction and selection through time-frequency analysis of the signals based on auto-regressive modeling. These extracted features at specific frequency bands constitute the data to be fed to an HCRF. Intermediate brain states are defined and represented by latent variables in the HCRF model. Model parameters are learned from labeled training data, and inference algorithms on HCRFs are used for classification. To the best of our knowledge, this is the first use of HCRFs for analysis of EEG signals in general, and in the context of BCI in particular. We present experimental results demonstrating the improvements provided by our HCRF-based approach over the best performing methods in BCI competition IV as well as over the HMM-based method in [10].

2. HIDDEN CONDITIONAL RANDOM FIELDS

In the task of labeling sequence data, one of the most widely used tools is the hidden Markov model [16], a finite automaton which contains discrete-valued states Q emitting a data vector X at each time point, the distribution of the data at each time point depends on the current state. Given that models of this kind are generative, it is necessary to determine the joint probability over observations and labels, which requires all possible observation sequences to be enumerated. In order to make the inference problem tractable, assumptions about independence of the data at each time point conditioned on the states should be made. Such assumptions are violated in many practical scenarios. CRFs are discriminative models that overcome such issues [13], avoiding the need to determine the data distribution as well as the need for the independence assumptions.

In CRFs, the conditional probability of labels $\overline{\mathbf{Y}}$ given the data $\mathbf{X}$ (to be labeled) is given by [13]:

$$P_\theta(\overline{\mathbf{y}}|\mathbf{x}) \propto \exp\{\sum_{e \in E,k} \lambda_k f_k(e,\overline{\mathbf{y}}|_e,\mathbf{x}) + \sum_{v \in V,k} \mu_k g_k(v,\overline{\mathbf{y}}|_v,\mathbf{x})\} \quad (1)$$

where V and E are the vertices and edges in the graph $G = (V,E)$, and $\overline{\mathbf{y}}|_S$ refers to the set of components of $\overline{\mathbf{Y}}$ associated with the subgraph S. The *CRF-features*[1] f_k and g_k are related to the edges and vertices, respectively, and are given and fixed. One has to estimate the parameters λ_k and μ_k based on training data. More detailed description of CRFs is beyond the scope of this paper, for which we refer the reader to [13].

This approach overcomes the problems stated above for HMMs, however in the BCI problem, which is of interest in this paper, the values for the assumed states in the EEG signals during the execution of mental tasks are unknown and are not available in the training stage. Because of that, it is necessary to incorporate hidden state variables. Such a model has been proposed in [15], and is called the hidden conditional random field (HCRF). HCRFs are able to capture intermediate structures through hidden states, combined with the power of discriminative models provided by CRFs. Furthermore, unlike CRFs, they also provide a way to estimate the conditional probability of a class label for an entire sequence. An HCRF is constructed as follows. The task is to predict the class y from the data $\mathbf{x}$, where y is an element of the set Y of possible labels for the entire data and $\mathbf{x}$ is the set of vectors of temporal observations $\mathbf{x} = \{x_1, x_2, ..., x_m\}$. The subindex m represents the number temporal observations. Each local observation is represented by a feature vector $\phi(x_j) \in R^d$ where d is the dimensionality of the representation. The training set contains a set of labeled samples $(\mathbf{x_i}, y_i)$, for $i = 1...n$ where $y_i \in Y$ and $\mathbf{x_i} = \{x_{i,1}, x_{i,2}, ..., x_{i,m}\}$. For any $\mathbf{x_i}$ a vector of latent variables $\mathbf{h} = \{h_1, h_2, ..., h_m\}$ is assumed, providing the state sequence of the data. Each possible value for h_j is member of a finite set H of possible hidden states. The joint probability of the labels and states given the data is described by:

$$P(y, \mathbf{h}|\mathbf{x}, \theta) = \frac{exp(\Psi(y, \mathbf{h}, \mathbf{x}; \theta))}{\sum_{y', \mathbf{h}} exp(\Psi(y', \mathbf{h}, \mathbf{x}; \theta))} \quad (2)$$

where θ are the parameters of the models and $\Psi(y, \mathbf{h}, \mathbf{x}; \theta)$ is a potential function $\in R$. The conditional probability of the labels given the data can be found by:

$$P(y|\mathbf{x}, \theta) = \sum_{\mathbf{h}} P(y, \mathbf{h} | \mathbf{x}, \theta) = \frac{\sum_{\mathbf{h}} exp(\Psi(y, \mathbf{h}, \mathbf{x}; \theta))}{\sum_{y', \mathbf{h}} exp(\Psi(y', \mathbf{h}, \mathbf{x}; \theta))} \quad (3)$$

According to [15] the estimation of parameter values, using the training data, is performed using the following objective function:

$$L(\theta) = \sum_i logP(y_i|\mathbf{x_i}, \theta) - \frac{1}{2\sigma^2} \|\theta\|^2. \quad (4)$$

where the first term in (4) is the log-likelihood of the data. The second term is the log of a Gaussian prior with variance σ^2. Under this criterion gradient ascent can be used to search for the optimal parameter values $\theta^* = \arg\max_\theta L(\theta)$. Given a new test example $\mathbf{x}$ and parameter values θ^* induced from the training set, the label for the example is taken to be $\arg\max_{y \in Y} P(y|\mathbf{x}, \theta^*)$.

HCRFs use indirected graph structure, where the hidden variables h_m are the vertices of the graph. Based on this, the potential
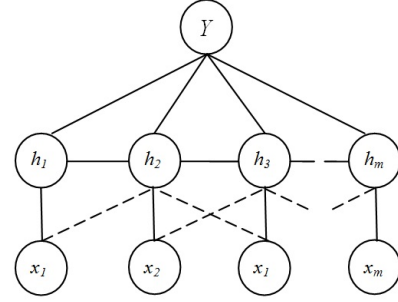
---

[1]These are simply called features in the CRF literature. However to distinguish them from features to be extracted from EEG signal, we call them CRF-features.
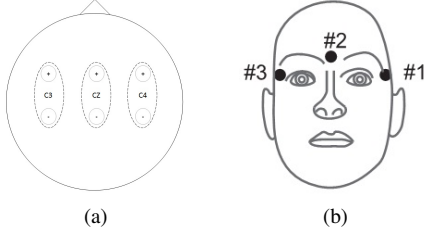

Figure 1: A simple HCRF model.

function $\Psi(y, \mathbf{h}, \mathbf{x}; \theta)$ is defined as:

$$\begin{aligned} \Psi(y, \mathbf{h}, \mathbf{x}; \theta) &= \sum_{j=1}^{m} \sum_{l \in L_1} f_{1,l}(j, y, h_j, \mathbf{x})\theta_{1,l} \\ &+ \sum_{(j,k) \in E} \sum_{l \in L_2} f_{2,l}(j, k, y, h_j, h_k, \mathbf{x})\theta_{2,l} \quad (5) \end{aligned}$$

where L_1 and L_2 are the set of node and edge *HCRF-features*, respectively, and, $f_{1,l}$ and $f_{2,l}$ are functions defining the *HCRF-features* in the model. The structure of the graph is assumed to be a tree, then exact methods for inference and parameter estimation can be used (i.e Belief Propagation). In the case of the BCI problem presented in this paper the structure proposed involves a chain of hidden states, as shown in Figure 1.

3. DESCRIPTION OF THE PROPOSED METHOD AND EXPERIMENTS

3.1 Problem and Dataset Description

In a typical BCI applications based on the imagination of motor activity, the subject is requested to execute imaginary motor tasks following a visual cue. It is known that the imagination of motor activities produces synchronization and/or desynchronization of the electrical signals recorded over the motor cortex and that this process has an asymmetrical spatial distribution during the imagination of the motor task (e.g., imagination of movement of a particular leg produces changes in the power of electrical signals in the contralateral region of the brain). Given a number of training sessions containing data from multiple trials in which the subject has been requested to imagine several motor tasks, the first task is to learn a model. Then, given some new (test) data, the task is to run an inference algorithm to perform classification of the imaginary motor task.

In this work the Dataset IIb of BCI competition IV [17], which consists of bipolar EEG recordings over scalp positions for electrodes C3, Cz and C4 (see Figure 2(a)) in 9 subjects, was used. The cue-based BCI paradigm involved two classes, represented by the imagination of the movement of left hand and right hand, respectively. The time scheme of the sessions is depicted in Figure 3. At the beginning of each trial, a fixation cross and a warning tone are presented. Three seconds later, a cue (indicating left or right movement) is presented and the subject is requested to perform the imaginary movement of the corresponding hand. The dataset contains five sessions, two for training and the remaining three for testing. Some of these sessions involved feedback, indicating to the subject how well the imagination of the motor task has been executed, and others did not. In our work we have used the sessions with feedback. Temporal behavior of the EEG signals could be modified due to the feedback influence [18].

Figure 2: (a) Montage used to extract the signal on C3, C4 and Cz, (b) EOG Channels.
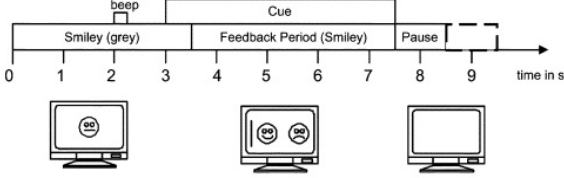

Figure 3: Time scheme for the experimental procedure.

3.2 Artifact Reduction

In order to reduce the interference of electrooculographic (EOG) signals in the EEG recordings, linear regression was employed, using the EOG data recorded at $N = 3$ channels using electrode locations shown in Figure 2(b). In this approach, the signal recorded by the EEG electrodes is modeled as the summation of the actual underlying EEG signal and the noise, represented by a linear combination of the EOG signals interfering into the EEG electrodes [19]:

$$w(n) = s(n) + u(n).b \qquad (6)$$

where n represents the discrete time index, $w(n)$ and $s(n)$ represent the noisy and the actual EEG signals at M electrodes, and $u(n)$ represents the EOG signal at N electrodes. Representing $w(n)$, $s(n)$, and $u(n)$ as row vectors, b is an *unknown* matrix of size $N \times M$ representing the set of coefficients that explain how the EOG signals have propagated by volume conduction to each of the points on the scalp where the EEG measurements are made. The problem is to recover $s(n)$ from measurements of $w(n)$ and $u(n)$. Given that the EOG signals are large in magnitude compared to the EEG signals, the interference of EEG in the EOG recordings $u(n)$ can be neglected [19]. If we knew b, the original EEG signal could be found by $s(n) = w(n) - u(n).b$. We describe a procedure to estimate b, which can then be used in this equation to estimate $s(n)$. Multiplying the signal $s(n)$ by $u(n)^T$ and taking expectation, we obtain:

$$E[u(n)^T w(n)] = E[u(n)^T s(n)] + E[u(n)^T u(n)b] \qquad (7)$$

Under assumption that there is no correlation between the EEG signal $s(n)$ and the EOG signals $u(n)$ we obtain an expression for estimating the coefficient matrix b:

$$\hat{b} = E[u(n)^T u(n)]^{-1} E[u(n)^T w(n)] \qquad (8)$$

The coefficients were calculated using a set of EOG measurements available in the dataset for each one of the subjects as described in [19]. These measurements involve the execution of different ocular movements enabling the estimation of b before the start of the motor task classification sessions.

3.3 Feature Extraction

For extraction of information from the EEG signal about the task being executed by the subject, we consider and compare two types of features in this paper: Hjorth parameters and auto-regressive power spectrum.

3.3.1 Hjorth Parameters

Hjorth parameters provide a representation of the general characteristics of the EEG signal [11]. Hjorth parameters of the EEG signal $s_i(n)$ namely Activity, Mobility, and Complexity are defined by:

$$Activity_i = var(s_i(n)) \qquad (9)$$

$$Mobility_i = \sqrt{\frac{Activity_i(\frac{ds_i(n)}{dn})}{Activity_i(s_i(n))}} \qquad (10)$$

$$Complexity_i = \frac{Mobility_i(\frac{ds_i(n)}{dn})}{Mobility_i(s_i(n))} \qquad (11)$$

where $i = 1, 2, ..., M$. The Hjorth parameters have physical interpretations. Activity corresponds to the power of the signal, Mobility represents the mean frequency, and Complexity represents the change in the frequency. This representation has been widely used in BCI applications [12, 7].

In this work, the time-varying Hjorth parameters are calculated using a sliding window of 1 second over the EEG signal with an overlapping of 80% of the windows size. The feature vector **x** to be fed to the HCRF is obtained by using the parameters calculated over the signals recorded at electrode positions C3 and C4:

$$\mathbf{x} = \begin{aligned}[t] &[Activity_{C3}(\tau), Mobility_{C3}(\tau), Complexity_{C3}(\tau), \\ &Activity_{C4}(\tau), Mobility_{C4}(\tau), Complexity_{C4}(\tau),] \end{aligned} \qquad (12)$$

where τ represents an artificial time index corresponding to the temporal window number used in computation of the Hjorth parameters.

3.3.2 Power Spectral Density Estimation using Auto-Regressive parameters

The power spectrum of the signal is computed by parametric methods involving the calculation of autoregressive (AR) models of the signal. In this paper, we use Burg's method because it provides better stability than the Yule-Walker method, by minimizing the error in backward and in forward direction [20]. The power spectrum of the EEG signal is estimated as the frequency response of the autoregressive model:

$$s_i(n) = \sum_{k=1}^{p} a_k s_i(n-k) + g(n). \qquad (13)$$

where, n represents the discrete time index, p is the model order, a_k is the $k^t h$ coefficient of the model and $g(n)$ is the system input or noise function. Then we can compute the system function in the z-domain:

$$H_i(z) = \frac{S_i(z)}{G(z)} = \left(1 - \sum_{k=1}^{p} a_k z^{-k}\right)^{-1} \qquad (14)$$

The AR spectrum can be obtained by evaluating $H(z)$ on the unit circle where $z = exp(j\omega)$ [21].

For estimating the AR parameters, we use a 1-second sliding window, as in the case of Hjorth parameters, for electrodes C3 and C4 . For each signal segment of 1 second, the model is estimated and the frequency response is obtained. The overlap of the segments was fixed to 80% of the window length. This produces a time-frequency map for each signal. From this time-frequency representation with frequency resolution of 0.25Hz providing a set of 109 components, it is necessary to select the frequency bands which provide more information about the task. We use feature selection methods for this purpose, which we describe next.

Subject	Zheng Yang et al.	Huang Gan et al.	D. Coyle et al.	AR-Power+HMM	HJORTH+HCRF.	AR-Power+HCRF.
1	0.40	0.42	0.19	**0.46**	**0.53**	**0.56**
2	0.21	0.21	0.12	**0.27**	**0.23**	**0.24**
3	**0.22**	0.14	0.12	0.19	0.14	0.18
4	**0.95**	0.94	0.77	0.93	0.94	**0.95**
5	0.86	0.71	0.57	**0.88**	**0.86**	**0.93**
6	0.61	0.62	0.49	**0.64**	**0.68**	**0.72**
7	0.56	**0.61**	0.38	0.54	0.40	0.51
8	**0.85**	0.84	**0.85**	0.71	0.84	0.80
9	0.74	**0.78**	0.61	0.75	0.64	0.71
Average	**0.60**	0.58	0.46	**0.60**	0.58	**0.62**

Table 1: Comparison with the results of the BCI competition IV (Kappa Values)

3.4 Feature Selection

When features based on auto-regressive power spectrum are used, the feature dimension is large and it is necessary to select the bands of frequencies which provide more information about the oscillatory brain activity related to the mental tasks executed by the subjects. For this task, we analyze the training dataset and calculate the degree of separability of the frequency components during the realization of imaginary motor activity, using the Fisher score. The Fisher score is a measure of the class separability for two classes i and j [22], and is defined as follows:

$$Fisher_{(f_k,\tau,i,j)} = \frac{(m_{(i,f_k,\tau)} - m_{(j,f_k,\tau)})^2}{\sigma^2_{(i,f_k,\tau)} + \sigma^2_{(j,f_k,\tau)}}. \qquad (15)$$

Where $m_{(i,f,\tau)}$ and $m_{(j,f,\tau)}$ correspond to the inter-trial average of the autoregressive power spectrum at time τ and frequency component f_k for each class, where $i \neq j$. Similarly, $\sigma^2_{(i,f_k,\tau)}$ and $\sigma^2_{(i,f_k,\tau)}$ are the inter-trial variances of the auto-regressive power spectrum for each class. For a two-class problem, this produces a matrix each entry of which contains the Fisher score for a specific value of time and frequency as shown in the top row of Figure 4. For feature selection, we compute the temporal averages of Fisher scores, as shown in the bottom row of Figure 4, and pick the top five bands for each electrode. Since we use the EEG signals from two electrodes (C3 and C4), this results in a feature vector $\mathbf{x}$ of length 10.
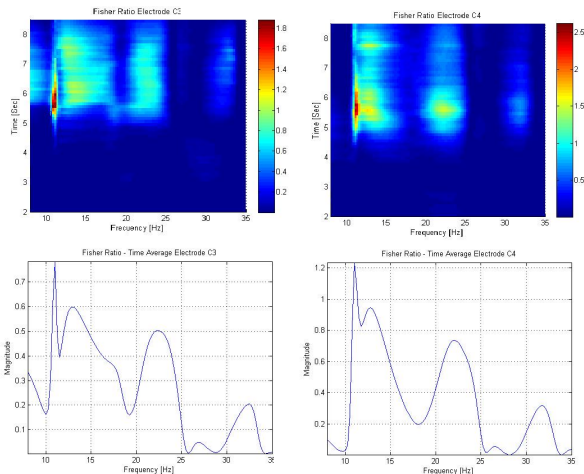

Figure 4: Fisher score computed from the training data measured by C3 (left) and C4 (right) electrodes, displayed as a function of time and frequency (top) and after temporal averaging each frequency component (bottom).

3.5 Model Selection and Classification

Feature vectors were obtained using Hjorth parameters or the autoregressive power spectrum as described previously, constitute the data $\mathbf{x}$ to be fed to the HCRF-based inference algorithm to be labeled.

Following [15], the potential function $\Psi(y, \mathbf{h}, \mathbf{x}; \theta)$ is defined as follows:

$$\Psi(y, \mathbf{h}, \mathbf{x}; \theta) = \sum_j \phi(x_j).\theta(h_j) + \sum_j \theta(y, h_j) + \sum_{(j,k) \in E} \theta(y, h_j, h_k) \qquad (16)$$

where $\theta(h_j) \in R^d$ is a parameter corresponding to the latent variable h_j. The inner product $\phi(x_j).\theta(h_j)$ can be interpreted as the compatibility between the observation x_j and the hidden state h_j, $\theta(y, h_j) \in R$ can be interpreted as a measure of the compatibility between latent variable h_j and category label y, and each parameter $\theta(y, h_j, h_k)$ measures the compatibility between an edge with labels h_k and h_j and the label y. In this work no further transformation is applied to the input data, so we take $\phi(x_j) = x_j$. Given that the definition of the potential function in (16) can be written in the same form as (5), and that the graph structure proposed for the modeling of the EEG signals is a chain, algorithms as belief propagation can be used for estimation of parameter values θ^*.

One important issue in the BCI problem treated here is that the number of different brain states encountered during the execution or imagination of motor tasks is not obvious. In order to find the number of states that explain the signal well, a three fold cross validation is performed over the training data, with possible values of 2,3,4,5 for the number of distinct states.[2] From this set of models, with different numbers of hidden states, the model which provides the best classification accuracy after the cross validation process, over the training data, is selected.

Once the model is selected, classification is done by selecting as the label y for a test sequence $\mathbf{x}$ to be:

$$\hat{y} = \arg\max_{y \in Y} P(y/\mathbf{x}; \theta^*). \qquad (17)$$

As was explained before, there are exacts methods for inference given the structure of the graph, as the objective function in (4) and its gradient can be written in terms of marginal distributions [23] which can be computed using belief propagation.

4. RESULTS

We evaluate the performance of the HCRF-based approach presented above on BCI Competition IV dataset 2b. We compare the results of our approach to the top three results in the competition for this dataset. In addition, we also present a comparison with the HMM-based method in [10].

Following the methodology used in the competition, we use the Kappa values [24] as the metric for comparing different methods:

---

[2]The value of 1 was not considered because it is physically inconsistent with phenomena involving changes (synchronization and desynchronization) in the EEG signal.

$$\kappa = \frac{C \times P_{cc} - 1}{C - 1} \qquad (18)$$

where C is the number of classes and P_{cc} is the probability of correct classification. Relatively larger kappa values indicate better performance.

The results of our experiments are shown in Table 1. Here, the results of the HMM-based method and the HCRF-based methods are compared to the top results in the BCI competition. We observe that the proposed HCRF-based method with AR spectrum features provides the best performance in five out of nine subjects compared to the BCI competition results. Furthermore, the average kappa value achieved by this approach is better than those of the the best performing methods in the BCI competition, as well as that of the HMM-based method in [10].

5. CONCLUSION

We have proposed a new method for classification of imaginary motor tasks, based on HCRFs. The autoregressive modeling of the EEG signal, followed by the computation of the power spectrum and the selection of the frequency bands according to the Fisher score, produces the feature vector that is fed to the HCRF-based classifier. The discriminative nature of this method makes it unnecessary to model the distribution of the data or make assumptions about independence. Experimental results demonstrate the improvements in the classification accuracy provided by this approach over other methods. Furthermore, this method is based on modeling of the temporal changes of the EEG signal and the analysis of the states sequences could provide insights into the physical phenomena underlying the execution of the imaginary motor tasks. This last point raises an interesting question about the physiological meaning of the states, which is the focus of our future work.

REFERENCES

[1] N. Birbaumer and C. Leonardo G, "Brain computer interfaces: communication and restoration of movement in paralysis," *The Journal of Physiology*, vol. 579, no. 3, pp. 621–636, 2007.

[2] A. Nijholt, B. Reuderink, and D. O. Bos, "Turning Shortcomings into Challenges: Brain - Computer Interfaces for Games," in *Intelligent Technologies for Interactive Entertainment* (O. Akan, P. Bellavista, and Cao, eds.), vol. 9, pp. 153–168, Springer Berlin Heidelberg, 2009.

[3] G. Pfurtscheller and L. d. S. Fernando H, "Event - related EEG / MEG synchronization and desynchronization: basic principles," *Clinical Neurophysiology*, vol. 110, no. 11, pp. 1842 – 1857, 1999.

[4] R. Magjarevic, A. Yonas, P. A S, and T. L. Mengko, "Time - Frequency Features Combination to Improve Single - Trial EEG Classification," in *World Congress on Medical Physics and Biomedical Engineering* (O. Dssel and W. C. Schlegel, eds.), vol. 25/4, pp. 805–808, Springer Berlin Heidelberg, 2010.

[5] R. Palaniappan, "Brain Computer Interface Design Using Band Powers Extracted During Mental Tasks," *Conference Proceedings. 2nd International IEEE EMBS Conference on Neural Engineering, 2005.*, pp. 321–324, 2005.

[6] Z. Mu, D. Xiao, and J. Hu, "Classification of Motor Imagery EEG Signals Based on Time Frequency Analysis," *International Journal of Digital Content Technology and its Applications*, pp. 116–119, 2009.

[7] B. Obermaier, C. Guger, C. Neuper, and G. Pfurtscheller, "Hidden Markov Models for online classification of single trial EEG data," *Pattern Recognition Letters*, vol. 22, pp. 1299–1309, Oct. 2001.

[8] S. H. Il and L. S. Whan, "Two - Layer Hidden Markov Models for Multi - class Motor Imagery Classification," *First Workshop on Brain Decoding: Pattern Recognition Challenges in Neuroimaging*, pp. 5–8, Aug. 2010.

[9] A. O. Argunsah and M. Cetin, "AR-PCA-HMM Approach for Sensorimotor Task Classification in EEG-based Brain-Computer Interfaces," *2010 20th International Conference on Pattern Recognition*, pp. 113–116, Aug. 2010.

[10] J. Delgado Saa and M. Cetin, "Modeling Differences in the Time - Frequency presentation of EEG Signals Through HMMs for classification of Imaginary Motor Tasks," *Technical Report, Sabanci University ID SU-FENS-2011/0003. Available: http://research.sabanciuniv.edu/16498*, May 2011.

[11] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalography and Clinical Neurophysiology*, vol. 29, no. 3, pp. 306 – 310, 1970.

[12] C. Vidaurre, N. Krämer, B. Blankertz, and A. Schlögl, "Time Domain Parameters as a feature for EEG-based Brain - Computer Interfaces," *Neural Networks*, vol. 22, no. 9, pp. 1313 – 1319, 2009. Brain-Machine Interface.

[13] L. John D, A. McCallum, and P. Fernando C N, "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data," in *Proceedings of the Eighteenth International Conference on Machine Learning*, ICML '01, (San Francisco, CA, USA), pp. 282–289, Morgan Kaufmann Publishers Inc., 2001.

[14] L. Gang and M. Wanli, "Subject-adaptive real-time sleep stage classification based on conditional random field.," *AMIA Annual Symposium proceedings*, pp. 488–492, 2007.

[15] A. Quattoni, S. Wang, L.-P. Morency, M. Collins, and T. Darrell, "Hidden Conditional Random Fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 1848–1852, 2007.

[16] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," in *Readings in speech recognition*.

[17] G. Pfurstcheller *et al.*, "Data set 2b," in *BCI Competition 2008*, (Graz, Austria), 2008.

[18] C. N. A. Schlögl and G. Pfurstcheller, "Enhancement of left-right sensorimotor EEG differences during feedback-regulated motor imagery," *Clin. Neurophysiol*, vol. 16, no. 4, pp. 373–382, 1999.

[19] A. Schlögl, C. Keinrath, D. Zimmermann, R. Scherer, R. Leeb, and G. Pfurtscheller, "A fully automated correction method of EOG artifacts in EEG recordings.," *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, vol. 118, pp. 98–104, Jan. 2007.

[20] P. Stoica and L. M. Randolph, *Introduction to Spectral Analysis*. Prentece Hall, 1997.

[21] B. Jansen, J. Bourn, and J. Ward, "Autorregressive Estimation of Short Segment Spectra for Computarized EEG Analysis," *IEEE Transaction on Biomedical Engineering*, vol. BME-28, no. 8, pp. 630–637, 1981.

[22] X. Pei and C. Zheng, "Classification of Left and Right Hand Motor Imagery Tasks Based on EEG Frequency Component Selection," in *Bioinformatics and Biomedical Engineering, 2008. ICBBE 2008. The 2nd International Conference on*, pp. 1888 –1891, May 2008.

[23] S. B. Wang, A. Quattoni, L.-P. Morency, D. Demirdjian, and T. Darrell, "Hidden Conditional Random Fields for Gesture Recognition," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 2, pp. 1521 – 1527, 2006.

[24] A. Schögl and J. Kronegg, *Toward Brain-Computer Interfacing*, ch. 19. Cambridge, Massachusetts: MIT Press, 2007.