Common Bayesian Network for Classification of EEG-Based Multiclass Motor Imagery BCI

Lianghua He, *Member, IEEE*, Die Hu, Meng Wan, *Member, IEEE*, Ying Wen, Karen M. von Deneen, and MengChu Zhou, *Fellow, IEEE*

***Abstract*—Modeling and learning of brain activity patterns represent a huge challenge to the brain-computer interface (BCI) based on electroencephalography (EEG). Many existing methods estimate the uncorrelated instantaneous demixing of EEG signals to classify multiclass motor imagery (MI). However, the condition of uncorrelation does not hold true in practice, because the brain regions work with partial or complete collaboration. This work proposes a novel method, termed as a common Bayesian network (CBN), to discriminate multiclass MI EEG signals. First, with the constraints of a Gaussian mixture model on every channel, only related channels are selected to construct a normal Bayesian network. Second, the nodes that have both common and varying edges are selected to construct a CBN. Third, the probabilities on common edges are used to learn about the support vector machine for classification. To validate the proposed method, we conduct experiments on two well-known BCI datasets and perform a numerical analysis of the propose algorithm for EEG classification in a multiclass MI BCI. Experimental results show that the proposed CBN method not only has excellent classification performance, but also is highly efficient. Hence, it is suitable for the cases where a system is required to respond within a second.**

***Index Terms*—Bayesian network (BN), brain-computer interface (BCI), electroencephalography (EEG), learning algorithm, support vector machine (SVM).**

Manuscript received November 30, 2014; revised April 23, 2015; accepted June 13, 2015. Date of publication July 27, 2015; date of current version May 13, 2016. This work was supported in part by the National Natural Sciences Foundation of China under Grant 61272267, Grant 61170220, Grant 51075306, and Grant 61273261, in part by the International S&T Cooperation Program of China under Grant 2013DFM10100, in part by the Program for New Century Excellent Talents in University under Grant NCET-11-0381, and in part by the Fundamental Research Funds for the Central Universities, State Key Laboratory of Software Engineering. This paper was recommended by Associate Editor Y. Hata. *(Corresponding authors: Lianghua He and MengChu Zhou.)*

L. He is with the Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai 201804, China (e-mail: helianghua@tongji.edu.cn).

D. Hu is with the Department of Communication Science and Engineering, Fudan University, Shanghai 200433, China (e-mail: hudie@fudan.edu.cn).

M. Wan is with the Center for Science and Technology Development, Ministry of Education, Beijing 100080, China (e-mail: wanmeng@cutech.edu.cn).

Y. Wen is with the Department of Computer Science and Technology, East China Normal University, Shanghai 200241, China (e-mail: ywen@cs.ecnu.edu.cn).

K. M. von Deneen is with the School of Life Sciences and Technology, Xidian University, Xi'an 710071, China (e-mail: vondenk@ufl.edu).

M. C. Zhou is with the Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai 201804, China and also the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102 USA (e-mail: zhou@njit.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMC.2015.2450680

I. Introduction

NON-INVASIVE brain-computer interfaces (BCIs) enable people to communicate with an electronic device by measuring electric or magnetic signals generated by the brain's nervous system via electroencephalography (EEG) and use these signals to analyze the intention of a user. The most popular EEG-based BCI systems are based on EEG signal changes in association with imaging of different types of movements [1]–[3], such as thinking about moving their arms, hands, tongue, legs, or rotating an object. Because motor imagery (MI) can trigger neuronal activities in most areas related to sensorimotor movement in a very similar way as the real executed movement, if it is properly decoded, these neuronal signals can be used as control signals to reflect a subjective movement-related mental state of a user directly without inducing the factors located outside. The framework of an EEG-based BCI system is shown in Fig. 1.

In order to analyze EEG signals, especially MI EEG signals, numerous analysis methods are proposed for classification of two or more MI classes, e.g., imaginary movements of the left hand, right hand, tongue, and foot. The most famous one is the common spatial pattern (CSP) method [4]. Its various extensions [5], [6] are widely used for extracting discriminative spatial (or joint spatio-spectral) patterns that contrast the power features of spatial patterns in different MI classes. An information theoretic feature extraction method [7] and other CSP extensions [8], [9] have been proposed to tackle multiclass problems.

Many useful discriminant methods have been proposed, e.g., locality preserving projection method [10] based on a self-regression model [11], algorithms based on a wavelet transform [12], and neural networks [13]. Recently, the probabilistic methods, such as Gaussian processes [14] and Bayesian learning [15], [16], are proposed to improve the robustness and accuracy of BCIs. Furthermore, the obtained probabilities may provide the confidence level of outputs, which is meaningful for their post-processing [17], [18]. In addition, a support vector machine (SVM) [19], [20] is also certified to give high quality results in BCI [21], but its standard version has not been applied to solve a multiclass problem in BCI due to its high complexity.

The core of an MI BCI system is its discrimination algorithms. They can be divided into three categories. The first one is based on the frequency domain and extracts frequency factors during MI to make a decision. Neural signals under

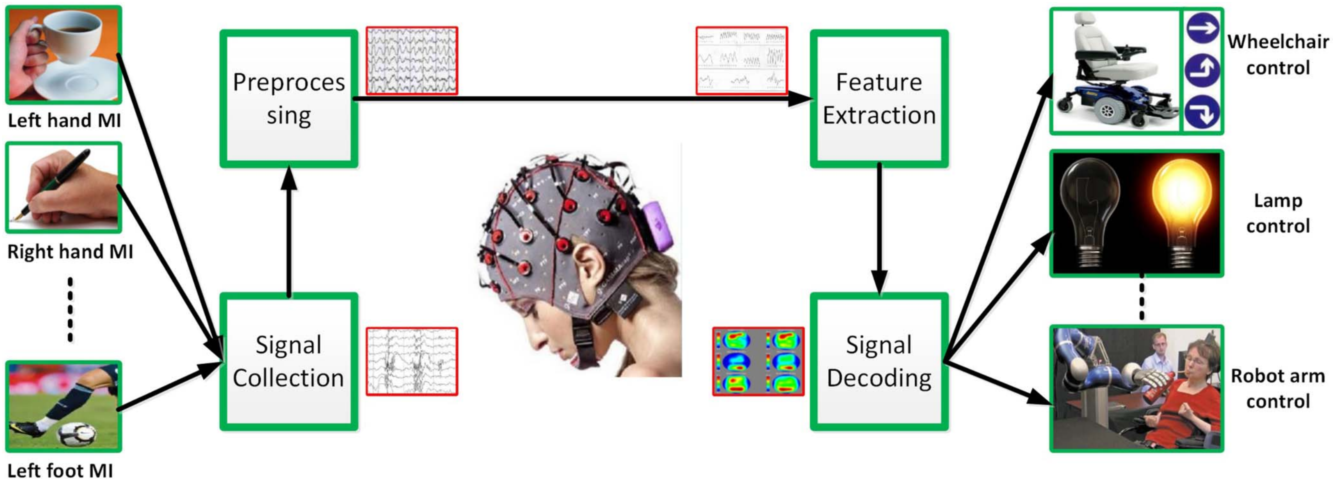


Fig. 1. Simple framework of EEG-based BCI.

different MI tasks have different frequency properties and the useful EEG signal frequency distribution can be found at a range from 0.1 to 40 Hz. Thus, it is possible to extract the frequency feature of this band to discriminate different MI tasks. The commonly used parameters are band power estimation, power spectral density, brain rhythm frequency, and wavelets.

The second method is based on temporal analysis. In general, the signal to noise ratio (SNR) of EEG signals is too low to be detected at one time. Although the effective frequency band is limited, the feature value is mixed with too much noise, which affects the final analysis results. Therefore, researchers introduce many machine learning methods for MI discrimination, e.g., Gaussian mixture network (GMM) [22], independent component analysis [23], [24], autoregressive [25], SVM, and CSP [7], [26]. With rigorous mathematical theory, these methods obtain optimal results on specific statistical standards, such as statistically uncorrelated, statistically independent, or minimum reconstruction error. Among these methods, the most widely used one is CSP, which computes spatial filters that maximize the variance ratio of x to y, where x and y are two data-conditioned MI classes. In this way, spatial filters are designed to project raw signals to a new space where mental tasks can be optimally classified by the projected variances [26]. Excellent classification results have been reported by using CSP for preprocessing in noninvasive BCIs based on MI [27]. Many of CSP extensions, such as multiclass CSP [7], are proposed to improve it on multiclass MI under complicated conditions.

The third category is based on spatial analysis. Different MI tasks produce different EEG signals whose spatial distributions differ. CSP, synchronization likelihood network [28–29], artificial neural network and Bayesian inference are classic. According to the CSP theory, the covariance matrices for different MI classes are the key points during the pursuit for the optimal projection matrix, which is formed by the EEG signal changes with spatial factors. For the other three, only these changes can make their link weight and statistical distribution meaningful. However, the spatial information used in such methods is expressed as the scattering of observed data.

As known to all, EEG is the continuous measurement of differences in electrical potential between points on the scalp whose amplitude is at the level of uV. The useful brain signals are so weak that their SNR is smaller than 0 dB. Although many effective methods are proposed to remove ocular artifact from EEG and improve SNR, it is still difficult to decrease the effects of the random noise. The most widely used feature of EEG is event-related potential (ERP) by repeating the same cognition task and taking the average. However, most MI tasks are finished with single trial and cannot be analyzed by ERP. Therefore, researchers want to search the temporal, frequency or spatial relationship between MI tasks and EEG data through models summarized above. We can see that CSP and its extensions are the most widely used method in BCI competition. The reason is that they can describe the linear spatial relationship in data. Different from it, principal component analysis (PCA) extracts linear features from data and shows much worse results since it is very hard to learn good features to analyze brain activation precisely in temporal, frequency or spacial domain alone owing to such low SNR. Therefore, we propose a novel method, named common Bayesian network (CBN), to learn about the relationship between the activation areas and MI tasks. We focus on the statistical relationship by using Bayes rules, but not features, during our analysis.

The framework of the proposed CBN is shown in Fig. 2. There are three major steps in our method: 1) Bayesian network (BN) construction; 2) CBN construction; and 3) feature discrimination. In the first step, two aspects are different from a traditional BN structure learning process. One is the novel use of GMM to learn EEG information of every BN node, instead of a single Gaussian model or discretization. EEG signals are so complicated that it is difficult to model them precisely with a single Gaussian model [22], [30], [31]. According to the knowledge of information theory, discretization causes the loss of information, including useful one. Compared to discretization or a single Gaussian model, GMM has better performance to represent EEG signals at expense of more computation. The other aspect is that several constraints are used during structure learning according to the properties of BCI data, such as physical constraints. In the second step,

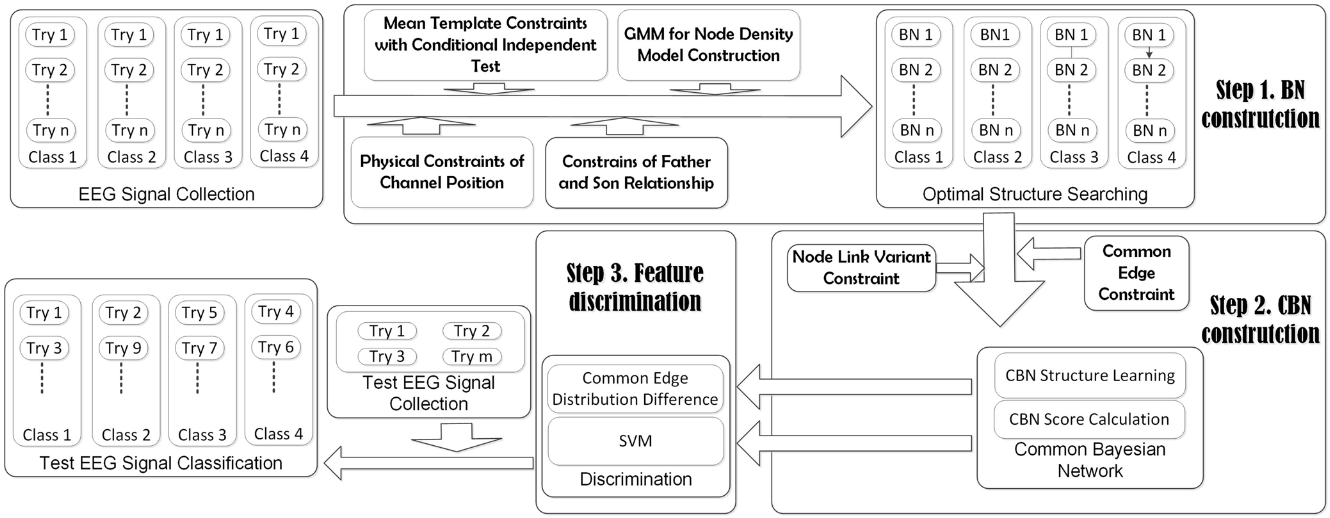


Fig. 2. Framework of the proposed CBN method.

space-time variants are extracted and organized at the same time. The spatial information is shown in the structure and time variant is represented in the edge probability. In the last step, the features are selected based on common edges on the position with structure variation in different MI tasks. These features can have more discrimination information than the edges which have no change among different MI tasks.

The primary novelty of the proposed CBN is a spatio-temporal statistical information extraction method. First, we use GMM to calculate temporal statistical information of MI intrachannel EEG data. Second, the spatial statistical information of MI interchannel EEG data is extracted through the inference probability of constructed common edges in BN. Therefore, according to statistic theory, if we have enough data, we can analyze MI with high performance that was not possible before.

The organization of this paper is as follows. Section II introduces some basic BN theory and then describes the proposed CBN; experimental results are reported in Section III and conclusions are drawn in Section IV.

II. Common Bayesian Network

A. Preliminary of Bayesian Network

A BN is represented by a directed acyclic graph. Let $X = (x_1, x_2, \ldots, x_n)$ be a collection of random variables. A BN over X is specified by a pair $(G; P)$, where $G = (V; E)$ is graph structure with node set $V = (v_1, v_2, \ldots)$ and edge set $E = (e_1, e_2, \ldots)$. If there is an arc from v_i to v_j, i.e., $(v_i, v_j) = e_k \in E$, we say that v_i is a parent of v_j and v_j is a child of v_i. P is the joint distribution. G satisfies the following two conditions: 1) node v_i corresponds to variable x_i and 2) every variable is conditionally independent of its nondescendants given its parents, i.e., $p(x_i|x_j, x_k) = p(x_i|x_j)$ if node x_k is node x_j's parent.

The learning task in BN can be separated into two subtasks: structure learning and parameter estimation. The former aims to identify the most suitable topology and implies a set of conditional independence relations among the variables involved as long as they are valid. In particular, the graph structure constrains the conditional independencies among those variables. Given a certain causal structure, only some patterns of conditional independence are expected to occur generically among the variables. Besides independencies, the BN structure can also be used in certain domains to represent cause and effect relationships through the edges and their directions. In these cases, the parent of a node is taken to be the directed cause of the quantity represented by that node. The parameters define the conditional probability distributions for a given network topology. They need to be estimated as accurately as possible.

Existing structure learning approaches follow two basic strategies: they either look over structures that maximize the likelihood of the observed data (score-based methods), or they test for conditional independencies and use these to constrain the space of possible structures. The former starts by defining a statistically motivated score that describes the fitness of each possible structure to the observed data. Such scores include Bayesian scores [32], [33] and minimum description length scores [34]. The latter tries to estimate properties of conditional independence among the attributes in the data. This is usually done via a statistical hypothesis test, e.g., χ^2 test.

In general, finding a structure that maximizes the score is known as an NP-hard problem [35]–[37]. Although the constraint satisfaction approach is efficient, it is sensitive to independence tests. It is commonly believed that an optimization approach is a better tool for structure learning from data. Most existing learning tools apply such standard heuristic search techniques as greedy hill-climbing and simulated annealing to find high-scoring structures [33], [38], [39]. Recently, new methods are proposed by formulating a structure learning problem as an integer linear program and solving it via a branch-and-cut method [40], [41].

As EEG signals are collected from several channels in different brain areas under a given cognition task, studies on brain cognition have shown that every cognition task is completed via the collaboration of many if not all brain areas under some definite cognition law [42]. It is also shown that different cognition tasks have different cognition functions which lead to different activated brain areas. Therefore, it is reasonable to deduce that different cognition tasks create different brain networks. This research aims to investigate this difference and classify it.

B. Structure Learning Based on GMM

1) GMM Construction for Channel: For an EEG channel signals with a set of N points in D dimensions, $x_1, \ldots, x_N \in R^D$ and a family P, find the probability density $p(x) \in P$ that is most likely to generate the given points. One way to define the family P is to give each of its members the same mathematical form, and to distinguish different members with different values of parameters θ. The functions in P can be mixtures of Gaussian functions

$$p(x:\theta) = \sum_{k=1}^{K} a_k g(x; m_k, \delta_k) \tag{1}$$

$$g(x; m_k, \delta_k) = \frac{1}{\left(\sqrt{2\pi}\delta_k\right)^D} e^{-\frac{1}{2}\left(\frac{\|x-m_k\|}{\delta_k}\right)^2} \tag{2}$$

where x is a D-dimensional continuous-valued data vector, $\theta = (\theta_1, \ldots, \theta_k) = ((p_1, m_1, d_1), \ldots, (p_k, m_k, d_k))$ is a $K(D+2)$-dimensional vector. $a_k, k \in \{1, 2, \ldots, K\}$ is are the k-th mixture weight satisfying $\sum_{i=1}^{K} a_i = 1$. $g(x; m_k, \delta_k)$ is a D-dimensional isotropic Gaussian function with mean vector m_k and the standard deviation δ_k.

In order to calculate θ that generates the highest probability, the likelihood function can be defined as

$$L(X;\theta) = \prod_{n=1}^{N} f(x_n;\theta) = \prod_{n=1}^{N}\sum_{k=1}^{K} a_k g(x_n; m_k, \delta_k). \tag{3}$$

The parametric density estimation problem can be defined precisely as follows:

$$\tilde{\theta} = \arg\max_{\theta} L(X;\theta). \tag{4}$$

Many maximization algorithms may be used to find $\tilde{\theta}$. The famous method that well solves this problem is Estimation–Maximization [43]. Its two steps are E and M steps that are used for computing likelihood bound and new estimates of a_k^{i+1}, m_k^{i+1}, and δ_k^{i+1}.

EM starts with initial values a_k^0, m_k^0, and δ_k^0. Based on the expectation of a logarithm shown in (4) and Jensen's inequality, the equation of the bound is constructed, called an "E step." The maximization of Jensen's inequality that yields new estimates a_k^{i+1}, m_k^{i+1}, and δ_k^{i+1} is called an M step. The detailed formula are as follows.

E step

$$a^{(i)}(k|n) = \frac{a_k^{(i)} g\left(x_n; m_{ks}^{(i)}, \delta_k^{(i)}\right)}{\sum_{m=1}^{K} a_k^{(i)} g\left(x_n; m_{ks}^{(i)}, \delta_k^{(i)}\right)}. \tag{5}$$

M step

$$m_k^{(i+1)} = \frac{\sum_{n=1}^{N} a^{(i)}(k|n) x_n}{\sum_{n=1}^{N} a^{(i)}(k|n)} \tag{6}$$

$$\delta_k^{(i+1)} = \sqrt{\frac{1}{D}\frac{\sum_{n=1}^{N} a^{(i)}(k|n)\left\|x_n - m_k^{(i+1)}\right\|^2}{\sum_{n=1}^{N} a^{(i)}(k|n)}} \tag{7}$$

$$a_k^{(i+1)} = \frac{1}{N}\sum_{n=1}^{N} a^{(i)}(k|n) \tag{8}$$

After θ is calculated, the GMM model for one channel signals is constructed.

2) Conditional Densities for Two GMMs: Given two nodes u and v in a BN, their conditional probability is the key for structure score evaluation when structure learning is performed. In order to construct it reasonably and easily, this paper makes the following premises.

a) For every node v_i, a GMM is learned with the constraint of $K = 2$.
b) For every GMM on every EEG channel, the covariance matrix δ is diagonal.
c) The relationship among different Gaussian components is independent.

Therefore, according to the Bayes' rule, the conditional density between two signals can be calculated as

$$p(y|x) = \frac{p(y,x)}{p(x)} \tag{9}$$

where $p(y, x)$ is the joint distribution among x and y. Because of the Gaussian property of x and y, we have

$$p(y,x) = \sum_{i=1}^{n} a_i N\left(y, \mu_{y,i}, \delta_{y,i}\right) N\left(y, \mu_{x,i}, \delta_{x,i}\right). \tag{10}$$

According to (9) and (10), we have

$$p(y|x) = \sum_{i=1}^{n} W_i(x) N\left(y, \mu_{y,i}, \delta_{y,i}\right) \tag{11}$$

where

$$W_i(x) = a_i N\left(y, \mu_{x,i}, \delta_{x,i}\right) \Big/ \sum_{j=1}^{n} a_j N(y, \mu_{x,i}, \delta_{x,i}). \tag{12}$$

3) Bayesian Information Criterion Score Based on GMM Nodes: There are many models to evaluate a new Bayesian structure, such as the Cheeseman–Stutz, Bayesian information criterion (BIC), and Laplace approximation scores. Compared with other score models, BIC not only measures the efficiency of a parameterized model in terms of predicting the data, but also penalizes the complexity of the model. Note that the complexity refers to the number of parameters in the model. Therefore, BIC is chosen to evaluate a Bayesian structure. For the nodes with continuous variations, the BIC score is calculated according to the likelihood function

$$\mathrm{BIC} = -2\ln p(x|M) + L \times (\ln(n) - \ln(2\pi)) \tag{13}$$

where L is the number of free parameters to be estimated. Because there are three parameters for each Gaussian component, i.e., w, δ, and μ, we have $L = 3 \cdot k_g$ where k_g is the number of Gaussian components. n is the number of data points in x, and $p(x|M)$ is the marginal likelihood of the

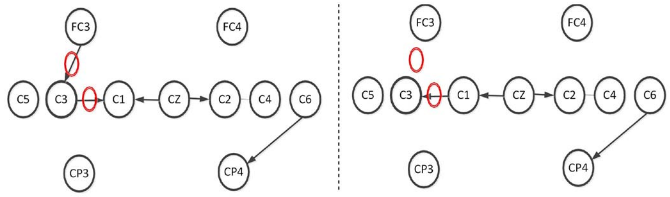


Fig. 3. Constructed BN structures from two runs for a same MI task.

observed data given model M. It is calculated through the conditional probability between parents and sons.

Finally, the structure with the minimum BIC score constructed based on the greedy search algorithm or other methods [38], [44] is chosen as the final result of learning.

C. Common Bayesian Network

After normal BN construction is finished, the next step is to construct CBN. Its major function is to classify multiclasses of MI. Thus, the features extracted from BN should have two properties: one is enough discriminated information for different kinds of MI to guarantee the performance; The other is a stable position to make the calculation possible. In fact, because of extremely low SNR of EEG signal, the structure of learned BN changes from time to time, even for the same MI task. In order to obtain the unique structure and performance-robust CBN from various learned BNs, all edges and nodes are evaluated by using the following concepts to help select its appropriate nodes and edges.

1) Edge Common Rate and Common Edge: First, we define the common property of edges. It is known that the features extracted in most recognition tasks must show the common property for one MI class. Ideally, different MI classes features should have totally different common properties. Because of the noise in EEG signals, the BN structures learned from the same MI task is sometime different. Fig. 3 shows the BN structure of two runs from the same right hand MI task with 11 channels (i.e., 11 nodes). For example, there is an edge from node $FC3$ to $C3$ in run 1, while it disappears in run 2, There is an edge direction change between node $C3$ and $C5$.

Fig. 4(a) shows the edge accumulation results of 20 runs from left hand MI, while Fig. 4(b) shows the results of 20 runs from right hand MI. Fig. 4(c) shows the edge difference accumulation results of 20 runs between left and right hand MI. The axes of X and Y are the EEG channels, which are the nodes when BN constructed. The axis of Z is the accumulated edge number in the total 20 runs. If there are 5 runs that there is an edge from node $CP3$ to node $C5$ among 20 runs after BN structure construction, the value of $(CP3, C5)$ is 5. We can see that except for a few edges, such as $(FC3, C3)$, $(C3, C1)$, and $(FC4, C4)$, most edges show common characteristics between two MI tasks. From the view point of statistics, BN structures are stable and that is the reason that the BN method can be used to classify multiclass MI tasks.

In conclusion, considering that an edge itself in a BN is a structural feature, its common property is very important for MI classification. Such a property also describes the generality of that edge. In this paper, the number of times that edge $E_k(i, j)$ appears in the same MI task is considered in evaluating the edge common character. We define

$$E_k(i,j) = \begin{cases} 0 & \text{if } (i,j) \notin E_k \\ 1 & \text{if } (i,j) \in E_k. \end{cases} \tag{14}$$

Considering the fact that only the conditional probability from node i to j is used for discrimination in this paper, we are only interested in whether there is an edge from node i to j. In order to describe the common characteristic, a common rate is defined as

$$Cr(i,j) = \sum_{k=1}^{N} (E_k(i,j) + \mathrm{E}_k(j,i)) \Big/ 2N \tag{15}$$

where N is the total number of MI runs and E_k is the edge set of BN G_k. We can see from (15) that $Cr(i, j)$ can be viewed as the probability of an edge from i to j. In order to evaluate the stable property of the selected edges, we define an edge as a common one if $Cr(i, j) = 1$.

2) Node Variation Rate and Key Node: In every MI task, EEG signals are collected through EEG channels from different brain areas. According to the BN definition, its node characteristic shows the information of the corresponding brain area. For a BN G_i, in order to calculate node characteristic easily, we first give every node a value. Given the ith run and jth node, let

$$W_i = (w_{i1}, w_{i2}, \ldots, w_{iN})$$
$$\text{with} \quad w_{ij} = \begin{cases} 0, & \text{if no edge on node } j \text{ is in } G_i \\ 1, & \text{if node } j \text{ is a parent in } G_i \\ -1, & \text{if node } j \text{ is a son in } G_i. \end{cases} \tag{16}$$

Every node has a value w_i on the i_{th} run. From the view point of machine learning, the excellent features should have a maximal interclass distance and the minimal intraclass distance, which is the basic theory of the famous Fisher rule. Thus it is used to evaluate the node distribution across different MI tasks. After the node value matrix W is calculated, the intraclass distance matrix α and interclass distance matrix β for every node are calculated as follows:

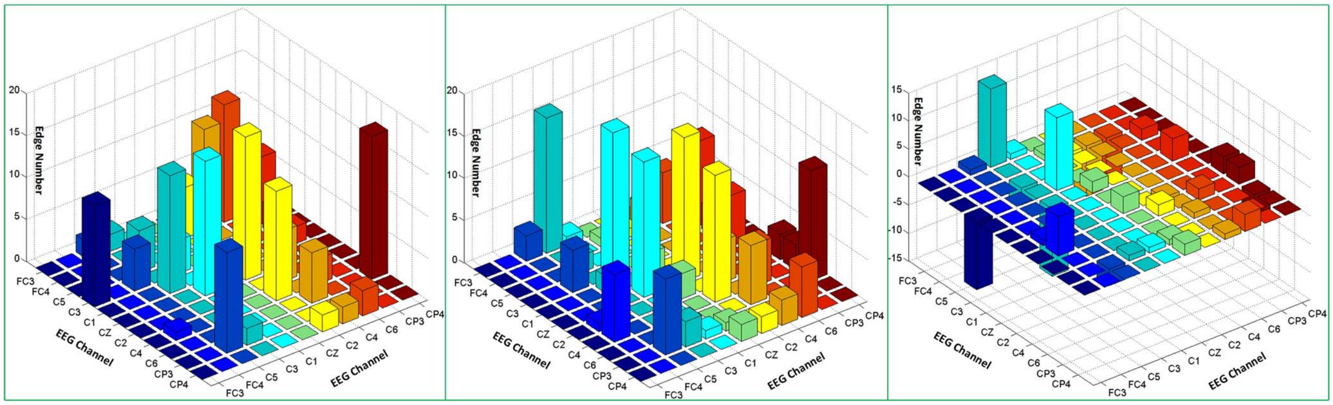
$$\mathbf{m}_i = \frac{1}{N_i} \sum_{\mathbf{w} \in W_i} w, \; i = 1, 2, \ldots, Q$$
$$\alpha_i = \sum_{\mathbf{x} \in X_i} (w - \mathbf{m}_i)(w - \mathbf{m}_i)^{\mathrm{T}}, \; i = 1, 2, \ldots, Q \tag{17}$$

where N_i is the run count of the ith MI task and Q is the number of MI tasks. For two MI tasks i and j, β is defined as

$$\beta_{ij} = \big(\mathbf{m}_i - \mathbf{m}_j\big)\big(\mathbf{m}_i - \mathbf{m}_j\big)^{\mathrm{T}}. \tag{18}$$

Thereafter, every node has P intraclass distances and $P \times P$ interclass distances. We define the node variation rate by using the ratio of α and β as follows:

$$f = \sum_{i=1}^{N_i} \alpha_i \Big/ \sum_{i=1}^{N_i} \sum_{j=1}^{P} \beta_{ij}. \tag{19}$$



(a) (b) (c)

Fig. 4. BN edges accumulation results of 20 runs. Accumulation result from (a) left hand MI and (b) right hand MI. (c) Difference between (a) and (b).

Algorithm 1 CBN Construction

1 Step: BN construction

Input: $X = (x_1, x_2, \ldots x_{N_c})$ and N_c is the EEG channel count;
1. Structural learning using physical constraints;
2. Node probability density calculation using *GMM*;

Output: (G, P) where $G = (V, E)$ and P is the reference probability;

2 Step: CBN construction

Input: G, δ, and f_0;
1. Calculate: C_r, for every $e \in E$ and f for every $v \in V$;
2. Select: s_e where $C_r \geq \delta$ and s_v where $f \geq f_0$;

Output: $S\_G = (S\_V, S\_E)$;

3 Step: Feature extraction and discrimination

Input: S_G
1. Calculate: S_P in S_G
2. Classify: $SVM(S\_P)$

Output: y, y is the label of MI tasks

It can be seen that the value of f indicates its varying property during MI. The larger it is, the more sensitive it is to MI. In other words, if one node has a large f, it is reasonable to deduce that this node has a large difference between interclass and intraclass of MI tasks and this node satisfies both properties mentioned above, that is, being stable and discriminative. In this paper, such a node is called a key node.

3) Common Bayesian Network: A CBN can be constructed in the following four steps.

a) A BN is constructed via the classic method described in [33] and [39].
b) All edges are analyzed via (14) and (15) and common edges are extracted via (15).
c) All nodes are analyzed via (16)–(19) and key nodes are extracted via (19).
d) A CBN is constructed based on key nodes and common edges.

CBN is clearly a subgraph of the original BN based on the discrimination features which are extracted to classify different MI. The pseudocode to construct it is shown in Algorithm 1.

D. Structure Constraints Based on EEG

Generally, there are no additional constraints during structure learning of a BN except for the conditional probability. However, in this paper, the BN is studied for real-time BCI task classification. Thus, the computational complexity has to be reduced as much as possible. According to the theory of BN structure learning, the complexity is only related to the number of nodes and edges. Therefore, the key question of using BN to classify MI tasks is how to select nodes properly and calculate edges quickly.

The research results of neural and brain science [45] show that: 1) the activated areas are different given different MI tasks and 2) only motor areas are activated during MI tasks. A BN node is an EEG channel and the relationship among nodes is not only calculated by signal data, but also deduced according to brain activation. Therefore, two constraints are adopted during BN structure learning: one is that only those channels around motor areas are used as nodes of a CBN as shown in Figs. 5(a)–(b). Another is that only those adjacent channels from eight neighborhoods have the father and son relationship, for example, in Fig. 5(a), only nodes 17–19, 27, 29, and 37–39 can be the father or son nodes for node $C3$.

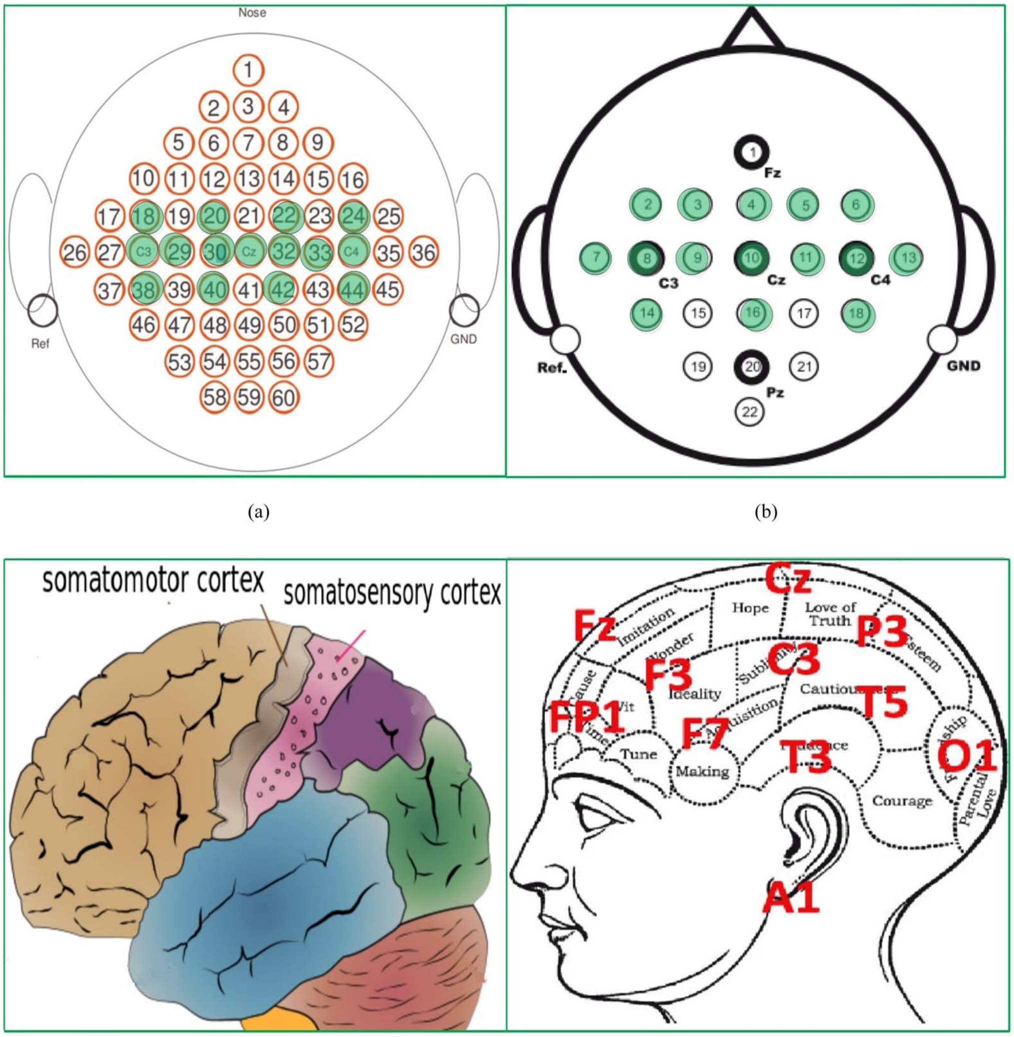
E. Feature Extraction Based on CBN

According to the definition of BN $(G; P)$, conditional density P is an important feature for CBN. It indicates the statistical relationship among nodes. Because the working mode of our brain is the same when motor imaging, it is true that all MI tasks share more than one activated brain area and the major difference among them is the activation amplitude. Therefore, the final feature vector used is $P = (p_1, p_2, \ldots, p_M)$ with p_i being the conditional probability of the common edge in CBN. An SVM method [46] is used to classify MI classes based on P.

III. Experiments and Results

A. Datasets

The dataset III2a in BCI competition III and dataset 2a in BCI competition IV, called DS1 and DS2 for short, are used



(a) (b)

(c) (d)

Fig. 5. Motor area and the selected channels in this paper on DS1 and DS2. (a) Electrode montage of DS1. (b) Electrode montage of DS2. (c) Motor area of our brain. (d) Electrode positions on our brain.

to test the performance of the proposed CBN. They can be downloaded from http://www.bbci.de/competition/.

In DS1, the data from three subjects are collected with a 64-channel EEG amplifier from the neuroscan as shown in Fig. 5(a) by using the left mastoid as the reference and the right mastoid as ground. EEG is sampled with 250 Hz and filtered between 1 and 50 Hz with the notch filter on. All three subjects are asked to image four MI classes with the left hand, right hand, tongue, or foot movement. The experiment consists of 6 runs with 40 trials for each MI class, i.e., totally there are 240 runs used for every MI class in the experiment.

In DS2, nine subjects are collected with four different MI tasks (movement of the left hand, right hand, and both feet and tongue). Twenty-two channels are used to record the EEG as shown in Fig. 5(b). The signals are sampled with 250 Hz and bandpass-filtered between 0.5 and 100 Hz. The sensitivity of the amplifier is set to 100 V. An additional 50 Hz notch filter is enabled to suppress line noise. Two sessions on different days are recorded for each subject. Each session is comprised of 6 runs with 48 trials, i.e., there are 288 trials in total, with 72 trials per MI task used in the experiment.

B. Channel Selection

In both datasets, the channel count is at least 22, which is too many to construct a BN in a reasonable time. Considering the fact that the collected EEG signals are all about MI, we locate the useful brain areas at the motor areas, such as the primary motor cortex, premotor cortex, and supplementary motor cortex. Fig. 5(c) shows the area of the motor functions in our brain from an anatomical point of view. Fig. 5(d) shows the partition of EEG under 10–20 international systems. We can see that the motor areas are located at frontal area (F), centralis area (C), and parietal area (P). Therefore, according to the prior analysis and channel distribution of the two BCI datasets, only 15 channels are selected to construct a BN shown as the shaded points in Figs. 5(a) and (b).

C. Kappa Coefficient Evaluation

Kappa coefficient [47] measures inter-rater agreement for categorical items. This measure indicates the proportion of agreement that remains after correction for agreement expected by chance. It is generally considered to be a more robust

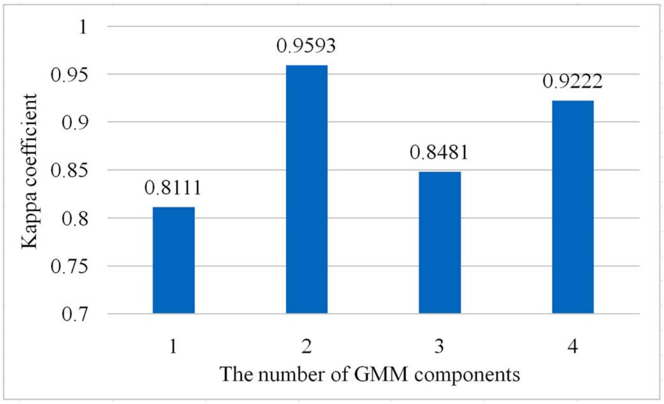


Fig. 6. Discrimination performance with different GMM component count.

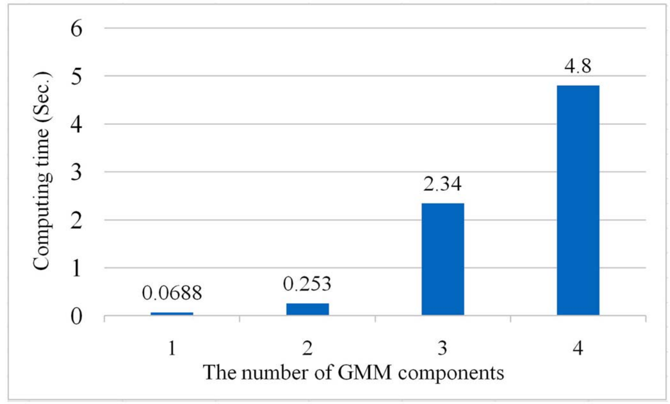


Fig. 7. Computing time with different GMM component count.

measure than simple percent agreement calculation since it takes into account the agreement occurring by chance. It is interpreted with five levels.

1) Poor agreement if $k < 0.20$.
2) Fair agreement if $0.20 \le k < 0.40$.
3) Moderate agreement if $0.40 \le k < 0.60$.
4) Good agreement if $0.60 \le k < 0.80$.
5) Very good agreement if $0.80 \le k \le 1.00$.

It is adopted to evaluate the proposed method's performance in DS1 and DS2. In order to make comparison with the other teams' results, we use it in this paper.

Mathematically, kappa coefficient is calculated as

$$k = \frac{\omega - \tau}{1 - \tau}. \tag{20}$$

In this paper, ω is the observed level of agreement that the k MI BCI classes agree, and τ is the proportion that the k MI BCI classes are expected to agree by chance alone.

D. Impact of GMM Components

From the description of CBN in Section II, we know that its structure is learned based on GMM. The number of GMM components not only affects the computational load, but also CBN performance. Therefore, the first experiment is to test how many GMM components are appropriate for MI BCI tasks.

The data used is $K3B$ from DS1 and the component count varies from 1 to 4. The computer is DELL Precision M4800 with i7-4900MQ CPU, 16 GB RAM. The discrimination performance and computational time are shown in Figs. 6 and 7.

It can be seen from Fig. 6 that the best performance is reached when the component count is 2. Considering that the data size is 100, GMM with three and more components meets some difficulty to calculate parameters accurately. If the data size increases, it may yield better performance. From Fig. 7, we can see that the relationship between computational time and component count is nonlinear. When the latter is 2, the computing time is only 0.25 s. But when it increases to 3, several seconds are needed to give discrimination results.

Therefore, considering computing time and discrimination performance together, we use two components in later experiments.

E. Experiment for Normal Four MI Class Classification

The first experiment is designed to test the comprehensive performance of the proposed method CBN compared to the competition results reported in [27] and [48].

As in the competition, only those trials with known labels are used as training data and the data with unknown labels are used as the test data. The parameter of the kappa coefficient is calculated as the comparison index. In order to reduce the amount of calculation, the number of Gaussian functions in GMM is constrained to 2. All the results are shown in Tables I and II.

We can see that on three subjects from DS1, CBN outperforms all other methods significantly with performance of 98%, 88%, and 82%, respectively. Especially for the subject of K6, the performances of all listed competition methods are less than 80%, while the proposed method reaches 88%, which means that CBN can analyze four classes of MI tasks with very good agreement. As for the mean kappa coefficients, compared to the highest result of 79% in the competition, the proposed CBN obtains a result of 90%, 11% higher, which represents great improvement.

On dataset of DS2, there are seven subjects on which our CBN method receives the highest performance with 70%, 53%, 48%, 68%, 82%, and 76%. On subjects 7 and 9, the first and second prize team have better performance than CBN with 77% and 69%, respectively. On subjects 4–6, the kappa coefficient is improved by the proposed CBN method at least one level, which is very huge. For example, we have 38% improvement on subject 5. As for the subjects 7 and 9, there is one MI class classified with abnormally bad performance, which yields the final low kappa coefficient. For example, for subject 7, the left hand MI has 22% recognition rate only. Totally, compared with the other teams' methods, our method has overwhelming advantages. As for the mean kappa coefficient, the proposed CBN is at 66%.

F. Experiment for Different Length of Imaging Time

Generally, in order to acquire enough EEG data, every MI task lasts several seconds. For example, in DS1 and DS2, the duration for one MI is at least 3 s. If we finish an action according to the EEG signals based on such designed MI task, we would look like a robot with discontinuous actions.

TABLE I
DETAILED RESULTS ON DATASET OF BCI III III2A

Methods	kappa coefficient	Subjects K3	K6	L1	Contributor
CSP + SVM	0.79	0.82	0.76	0.8	[27]
CSP+SVM+kNN+LDA	0.69	0.9	0.43	0.71	[27]
PCA + ICA + SVM	0.63	0.95	0.41	0.52	[27]
CBN+SVM	**0.9**	**0.98**	**0.88**	**0.82**	**The proposed**

TABLE II
KAPPA COEFFICIENTS OF ALL SUBJECTS ON DATASET OF BCI IV 2A

Methods	kappa coefficient	Subjects 1	2	3	4	5	6	7	8	9	Contributor
Filter Bank CSP	0.57	0.68	0.42	0.75	0.48	0.40	0.27	**0.77**	0.75	0.61	[48]
LDA and Bayesian Classifer	0.52	0.69	0.34	0.71	0.44	0.16	0.21	0.66	0.73	**0.69**	[48]
CSP+SVM	0.31	0.38	0.18	0.48	0.33	0.07	0.14	0.29	0.49	0.44	[48]
LDA + SVM	0.30	0.46	0.25	0.65	0.31	0.12	0.07	0.00	0.46	0.42	[48]
CSP + SVM	0.29	0.41	0.17	0.39	0.25	0.06	0.16	0.34	0.45	0.37	[48]
CBN + SVM	**0.66**	**0.69**	**0.51**	**0.87**	**0.85**	**0.78**	**0.42**	0.54	**0.97**	0.45	**The proposed**

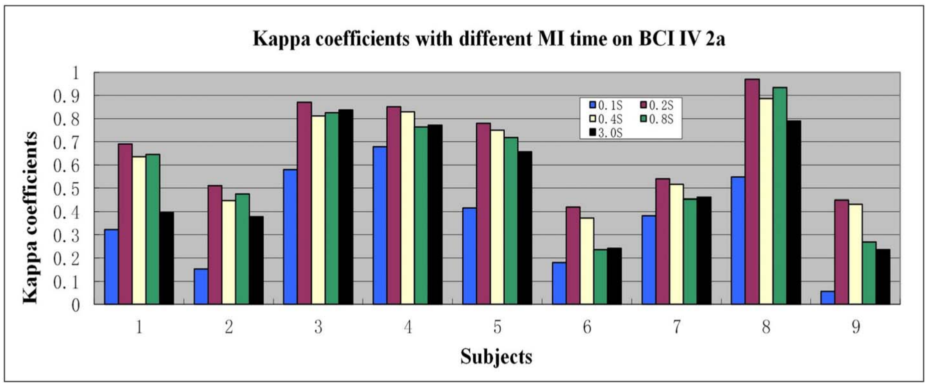


Fig. 8. Kappa coefficients under different MI time duration on BCI IV 2a.

In fact, one subject can image many times during the given time. The imaging time of one MI could only last half a second, or even shorter. Furthermore, 3 s is too long for a real-time task. Therefore, the target of this experiment is to test the effect of the length of imaging time. Such results could be very useful for real-time BCI application.

DS1 and DS2 are both used to test the effect. The results of five kinds of time length are calculated as: 0.1, 0.2, 0.4, 0.8, and 3.0 s. During calculating, we split every run into several time slices with given time length without overlap and make statistics after calculating the performance of these time slices. The results are shown in Figs. 8 and 9.

Although the above analysis shows that the performance may not be proportional to the time duration, the results are still above our expectation. When the time duration is less than 0.2 s, the performances are very bad. Similarly, when the time duration is 3 s long, the performances also decrease much. The best performances for all 12 subjects on two datasets have the duration of 0.2 s only. When the imaging time is longer than 0.2 s, the performances are inversely proportional to the imaging time.

The possible explanation is as follows. When the time duration is too short, there is not enough MI information to make discrimination. When it is too long, the density of MI information is low, which drops the discrimination performance. The data with 0.2 s is the best choice for four classes of MI BCI tasks for our proposed CBN method.

If one subject can successfully stimulate his/her MI signals within 0.8 s, considering the computing time, it is possible that the total response time is less than 1 s. Therefore, our method can be applied to the cases where the system is required to respond in 1 s, such as electric wheelchair control.

G. Experiment for Different Imaging Moment

In DS1 and DS2, the time duration for one MI is different, which is 1 and 3 s, respectively. In fact, the time duration of one MI is flexible in many multiclass MI tasks. Theoretically we do not know how long is enough. The only thing that can be concluded is that 3 s is too long for our human being to finish one MI task and it cannot meet the real-time requirement. Currently, more and more researches are paid on real-time BCI and the latest work is done by Huang *et al.* [49]. They construct a real-time multichannel EEG system through a parallel very large-scale integration architecture of a singular value decomposition processor.

Therefore, the third experiment is designed to test the performance of EEG data at different MI moments for investigating the possibility of real-time multichannel BCI based on EEG. We calculate the kappa coefficients on such moments

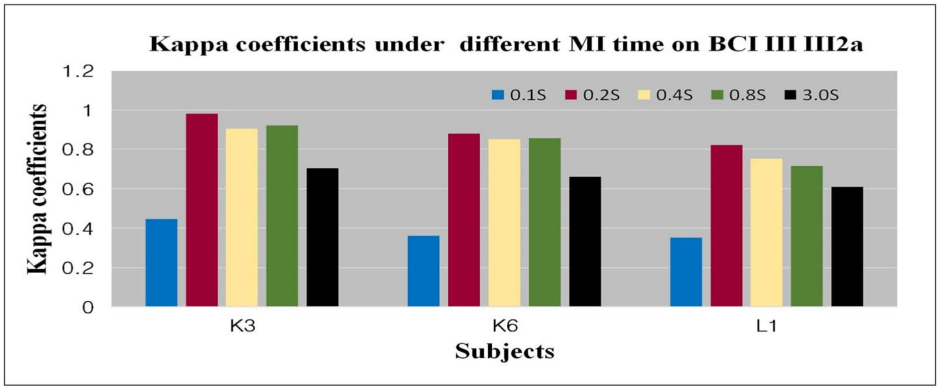


Fig. 9. Kappa coefficients under different MI time duration on BCI III III2a.

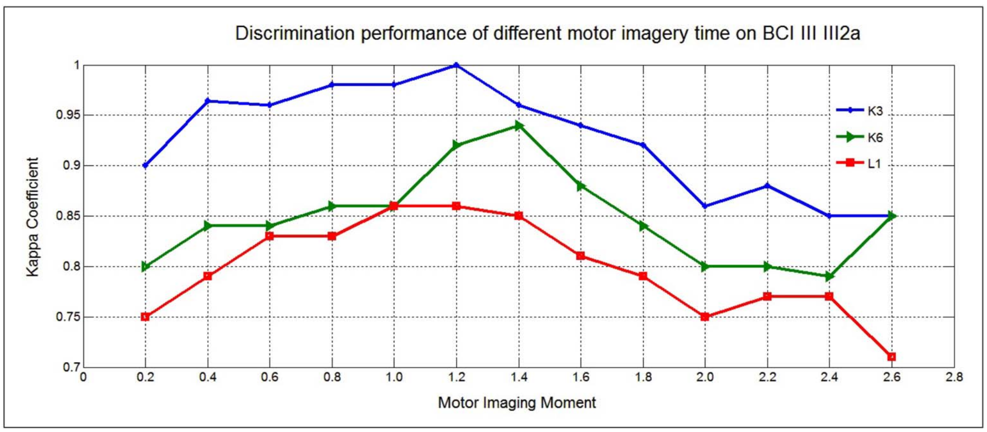


Fig. 10. Discrimination performance of different MI time on BCI III III2a.

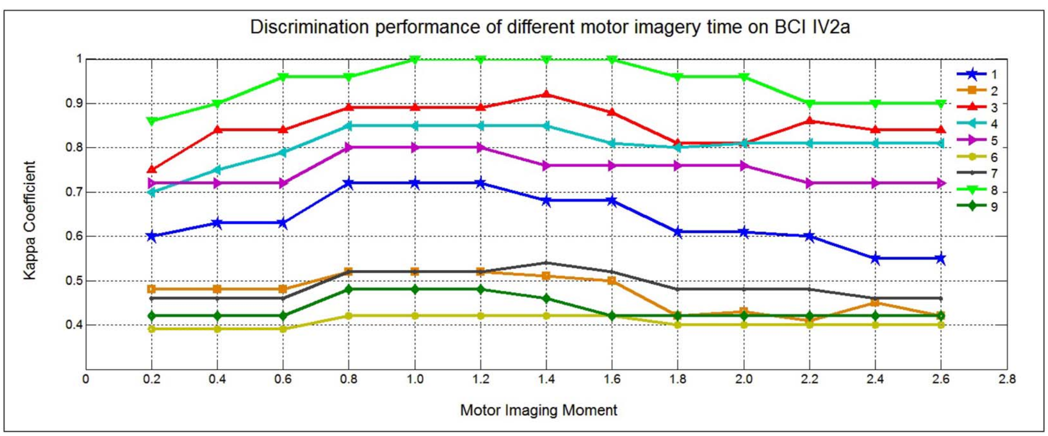


Fig. 11. Discrimination performance of different MI time on BCI IV2a.

on both datasets. According to above experimental results, only the data with 0.2 s length is analyzed through the whole run. Considering the fact that most subjects want to stop imaging when an MI task is nearing its completion, only the first 13 slices are used. Then the performance on the corresponding 13 slices for all MI runs is calculated. Figs. 10 and 11 show the kappa coefficient discrimination performance of 12 subjects from two datasets. We have the following conclusions.

First, the performance increases as the imaging task continues. Second, the performance reaches the highest score after about 1.2 s. Third, the performance decreases after the imaging time is longer than 1.4 s. Fourth, the performance difference between subjects with relatively good performance and those with relatively bad one is not much.

The results are explained as follows: 3 s is too long to finish the MI task once; and usually subjects need to image several times in one run. At the beginning, it is easy for them to focus on motor imaging. All data are useful for discrimination. High performance can be acquired. However, as the time goes by, they may feel fatigued or be distracted.

The data thus contain much noise which lowers discrimination performance.

In fact, different subjects have different imaging efficiency with many reasons, such as imagery content and imaging way. But after long training, most subjects can stimulate their MI signals faster than before training. From Figs. 10 and 11, we can see that most subjects have the highest performance after 0.8 s imaging. Thus, it is possible for one subject to stimulate enough effective MI BCI information within 0.8 s. However, it also shows the limitations of MI BCI that it is hard to be used in the application with fast time reaction requirement, such as a car race game.

IV. Conclusion

A novel method called CBN is proposed to analyze multiclass MI BCI EEG data. In order to make BN more suitable for classifying multiclass BCI tasks, two constraints are used during channel selection and BN structure learning. Furthermore, a two component GMM model is used to calculate the node's probability density that makes it more accurate comparing to a single Gaussian model. Considering the fact that the constructed BNs have variations because of the noise in EEG signal and the instability of EEG signals, we have proposed the concept of common edges which are extracted after evaluating all nodes and edges from these variants. The inference probability of common edges are used to make discrimination. The experimental results on two datasets from BCI III and IV show that the proposed CBN has excellent performance on multiclass MI BCI classification. The MI moment and MI time duration experiments results show that the best performance is acquired after 1.5 s motor imaging with only 0.2 s time duration. Therefore, after being well trained, an MI BCI system based on the proposed CBN method can respond within a second, which will extend its application greatly.

Because the proposed CBN is based on BN, it needs enough data to learn accurate conditional probability during structure learning. Thus, in practical applications, it is necessary to spend time to collect training data, which would make subjects tired. Therefore, our further research is to decide the shortest time duration needed for this purpose and to seek alternative methods, e.g., extracting accurate neuron signals to achieve the same objective.

References

[1] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proc. IEEE*, vol. 89, no. 7, pp. 1123–1134, Jul. 2001.

[2] P. F. Wei, W. He, Y. Zhou, and L. P. Wang, "Performance of motor imagery brain-computer interface based on anodal transcranial direct current stimulation modulation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 3, pp. 404–415, May 2013.

[3] V. Gandhi, G. Prasad, D. Coyle, L. Behera, and T. M. McGinnity, "EEG-based mobile robot control through an adaptive brain-robot interface," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 44, no. 9, pp. 1278–1285, Sep. 2014.

[4] G. Pfurtscheller, C. Neuper, D. Flotzinger, and M. Pregenzer, "EEG based discrimination between imagination of right and left hand movement," *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, no. 6, pp. 642–651, Dec. 1997.

[5] S. Lemm, B. Blankertz, G. Curio, and K. Muller, "Spatio-spectral filters for robust classification of single trial EEG," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 9, pp. 1541–1548, Sep. 2005.

[6] B. Blankertz *et al.*, "The Berlin brain-computer interface: Accurate performance from first-session in BCI-naive subjects," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 10, pp. 2452–2462, Oct. 2008.

[7] M. Grosse-Wentrup and M. Buss, "Multiclass common spatial patterns and information theoretic feature extraction," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 8, pp. 1991–2000, Aug. 2008.

[8] G. Dornhege, B. Blankertz, G. Curio, and K. R. Müller, "Increase information transfer rates in BCI by CSP extension to multi-class," in *Proc. 17th Annu. Conf. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, Dec. 2003, pp. 733–740.

[9] G. Dornhege, B. Blankertz, G. Curio, and K. R. Müller, "Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 993–1002, Jun. 2004.

[10] R. Xu *et al.*, "Enhanced low-latency detection of motor intention from EEG for closed-loop brain-computer interface applications," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 2, pp. 288–296, Feb. 2013.

[11] D. J. Krusienski, D. J. McFarland, and J. R. Wolpaw, "An evaluation of autoregressive spectral estimation model order for brain-computer interface applications," in *Proc. IEEE Int. Conf. Eng. Med. Biol. Soc.*, New York, NY, USA, Aug./Sep. 2006, pp. 1323–1326.

[12] T. Hinterberger, A. Kubler, and J. Kaiser, "A brain-computer interface (BCI) for the locked-in: Comparison of different EEG classifications for the thought translation device," *Clin. Neurophysiol.*, vol. 114, no. 3, pp. 416–425, Mar. 2003.

[13] G. Pfurtscheller, J. Kalcher, and C. Neuper, "On-line EEG classification during externally-paced hand movements using a neural network-based classifier," *Electroencephalogr. Clin. Neurophysiol.*, vol. 99, no. 5, pp. 416–425, Nov. 1996.

[14] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces," *J. Neural Eng.*, vol. 4, no. 2, pp. 1–13, Jun. 2007.

[15] H. Suk and S.-W. Lee, "A novel Bayesian framework for discriminative feature extraction in brain-computer interfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 286–299, Feb. 2013.

[16] M. J. Zhong, F. Lotte, M. Girolami, and A. Lécuyer, "Classifying EEG for brain computer interfaces using Gaussian processes," *Pattern Recognit. Lett.*, vol. 29, no. 3, pp. 354–359, Feb. 2008.

[17] S. J. Roberts and W. D. Penny, "Real-time brain computer interfacing: A preliminary study using Bayesian learning," *Med. Biol. Eng. Comput.*, vol. 38, no. 1, pp. 56–61, 2000.

[18] Q. Tao, G. Wu, F. Y. Wang, and J. Wang, "Posterior probability support vector machines for unbalanced data," *IEEE Trans. Neural Netw.*, vol. 16, no. 6, pp. 1561–1573, Nov. 2005.

[19] U. Hoffmann, "Bayesian machine learning applied in a brain-computer interface for disabled users," Ph.D. dissertation, Signal Processing Institute, School of Engineering, EPFL, Lausanne, Switzerland, 2007.

[20] V. N. Vapnik, *Statistical Learning Theory*. New York, NY, USA: Wiley, 1998.

[21] J. C. Sun, T. Y. Zhang, and H. Y. Liu, "Multi-class analogue modulation recognition algorithms based on support vector machines," *J. Univ. Electron. Sci. Technol. China*, vol. 35, no. 2, pp. 149–152, 2006.

[22] T. Tsuji, O. Fukuda, H. Ichinobe, and M. Kaneko, "A log-linearized Gaussian mixture network and its application to EEG pattern classification," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 29, no. 1, pp. 60–72, Feb. 1999.

[23] W. Zhou and J. Gotman, "Automatic removal of eye movement artifacts from the EEG using ICA and the dipole model," *Progr. Nat. Sci.*, vol. 19, no. 9, pp. 1165–1170, Sep. 2009.

[24] M. Fatourechi, A. Bashashati, R. K. Ward, and G. E. Birch, "EMG and EOG artifacts in brain computer interface systems: A survey," *Clin. Neurophysiol.*, vol. 118, no. 3, pp. 480–494, Mar. 2007.

[25] J. del R. Millán and J. Mouriño, "Asynchronous BCI and local neural classifiers: An overview of the adaptive brain interface project," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 159–161, Jun. 2003.

[26] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, Dec. 2000.

[27] B. Blankertz *et al.*, "The BCI competition. III: Validating alternative approaches to actual BCI problems," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 153–159, Jun. 2006.

[28] S. Kar and A. Routray, "Effect of sleep deprivation on functional connectivity of EEG channels," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 43, no. 3, pp. 666–672, May 2013.

[29] S. Hu, M. Stead, Q. Dai, and G. A. Worrell, "On the recording reference contribution to EEG correlation, phase synchorony, and coherence," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 5, pp. 1294–1304, Oct. 2010.

[30] L. Xu, P. Xu, A. T. Chen, and D. Z. Yao, "Gaussian source model based iterative algorithm for EEG source imaging," *Comput. Biol. Med.*, vol. 39, no. 11, pp. 978–988, Nov. 2009.

[31] S. Faul *et al.*, "Gaussian process modeling of EEG for the detection of neonatal seizures," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 12, pp. 2151–2162, Dec. 2007.

[32] G. F. Cooper and E. Herskovits, "A Bayesian method for the induction of probabilistic networks form data," *Mach. Learn.*, vol. 9, no. 4, pp. 309–347, Oct. 1992.

[33] D. Heckerman, D. Geiger, and D. M. Chickering, "Learning Bayesian networks: The combination of knowledge and statistical data," *Mach. Learn.*, vol. 20, no. 3, pp. 197–243, Sep. 1995.

[34] W. Lam and F. Bacchus, "Learning Bayesian belief networks: An approach based on the MDL principle," *Comput. Intell.*, vol. 10, no. 3, pp. 269–293, Aug. 1994.

[35] D. Chickering, "Learning Bayeisan networks is NP-complete," in *Learning From Data: Artificial Intelligence and Statistics V*, D. Fisher and H. J. Lenz, Eds. New York, NY, USA: Springer, 1996, pp. 121–130.

[36] D. M. Chickering, D. Heckerma, and C. Meek, "Larger sample learning of Bayesian networks is NP-hard," *J. Mach. Learn. Res.*, vol. 5, pp. 1287–1330, Dec. 2004.

[37] S. Dasgupta, "Learning polytrees," in *Proc. 15th Conf. Mach. Learn.*, Stockholm, Sweden, Jul./Aug. 1999, pp. 134–141.

[38] D. M. Chickering, "Learning equivalence classes of Bayesian network structures," *J. Mach. Learn. Res.*, vol. 2, pp. 445–498, Mar. 2002.

[39] D. Heckerman, "A tutorial on learning with Bayesian networks," *Mach. Learn.*, vol. 2, no. 3, pp. 445–498, 2002.

[40] J. Cussens, "Bayesian network learning with cutting planes," in *Proc. 27th Conf. Uncertainty Artif. Intell. (UAI)*, Barcelona, Spain, Jul. 2011, pp. 153–160.

[41] T. Jaakkola, D. Sontag, A. Globerson, and M. Meila, "Learning Bayesian network structure using LP relaxations," *J. Mach. Learn. Res.*, vol. 9, pp. 358–365, Mar. 2010.

[42] R. Cabeza and A. Kingstone, *Handbook of Functional Neuroimaging of Cognition*, 2nd ed. Cambridge, MA, USA: MIT Press, 2006.

[43] Z. Guo, M. C. Zhou, and G. Jiang, "Adaptive optimal sensor placement and boundary estimation for monitoring mass objects," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 38, no. 1, pp. 222–232, Jan. 2008.

[44] N. Friedman, I. Nachmana, and D. Peér, "Learning Bayesian network structure from massive datasets: The sparse candidate algorithm," in *Proc. 15th Conf. Uncertainty Artif. Intell.*, Stockholm, Sweden, 1999, pp. 206–215.

[45] S. J. Richard *et al.*, *Human Brain Function*, 2nd ed. Amsterdam, The Netherlands: Elsevier, 2004.

[46] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.

[47] M. Banerjee, M. Capozzoli, L. McSweeney, and D. Sinha, "Beyond kappa: A review of interrater agreement measures," *Can. J. Stat.*, vol. 27, no. 1, pp. 3–23, Mar. 1999.

[48] M. Tangermann *et al.*, "Review of the BCI competition IV," *Front. Neurosci.*, vol. 6, no. 55, pp. 1–31, Jul. 2012.

[49] K.-J. Huang, J.-C. Chang, C.-W. Feng, and W.-C. Fang, "A parallel VLSI architecture of singular value decomposition processor for real-time multi-channel EEG system," in *Proc. IEEE 17th Int. Symp. Consum. Electron. (ISCE)*, Hsinchu, Taiwan, Jun. 2013, pp. 21–22.

Lianghua He (M'14) received the B.S. degree in surveying and mapping from the Wuhan Technology University of Surveying and Mapping, Wuhan, China, in 1999, the M.S. degree in surveying and mapping from Wuhan University, Wuhan, in 2002, and the Ph.D. degree in electronic engineering from Southeast University, Nanjing, China, in 2005.

From 2005 to 2007, he was a Post-Doctoral Researcher with Tongji University, Shanghai, China. Since 2007, he has been with the Department of Computer Science and Technology, Tongji University, where he is currently an Associate Professor. His current research interests include pattern analysis, machine learning, and cognitive computing.

Die Hu received the B.S. and Ph.D. degrees in electronic engineering from Southeast University, Nanjing, China, in 2001 and 2006, respectively.

Since 2006, she has been with the Department of Communication Science and Engineering, Fudan University, Shanghai, China, where she is currently an Associate Professor. Her current research interests include wireless communications, signal processing, and 5G.

Meng Wan (M'14) received the B.S. degree in inorganic nonmetal materials from Tianjin University, Tianjin, China, in 1996, the M.S. degree in national economic plan and management from the Central University of Finance and Economics, Beijing, China, in 2000, and the Ph.D. degree in enterprise management from Wuhan University, Wuhan, China, in 2008.

He is an Associate Professor with the Science and Technology Development Center, Ministry of Education, Beijing. His current research interests include computer network architecture, network and systems management, science and technology management, and system engineering.

Prof. Wan is a member of ACM.

Ying Wen received the B.Sc. degree in industry automation from Hefei University, Hefei, China, in 1997, the M.Sc. degree in image processing and pattern recognition from Shanghai University, Shanghai, China, in 2002, and the Ph.D. degree in image processing and pattern recognition from Shanghai Jiao Tong University, Shanghai, in 2009.

From 2009–2012, she was a Postdoctoral Research Fellow with Columbia University, New York, NY, USA. She is currently an Associate Professor with the Department of Computer Science and Technology, East China Normal University, Shanghai. Her current research interests include documentary processing, pattern recognition, machine learning, and medical image processing.

Karen M. von Deneen received the A.A.S. and B.U.S. degrees in education and veterinary technology from Morehead State University, Morehead, KY, USA, in 1997 and 1998, respectively, the M.S. degree in animal science from Oregon State University, Corvallis, OR, USA, in 2002, the D.V.M. degree in veterinary medicine from the College of Veterinary Medicine and Biomedical Sciences, Colorado State University, Fort Collins, CO, USA, and the Ph.D. degree in pathobiology and large animal clinical sciences from the University of Florida, Gainesville, FL, USA, in 2009.

She is currently an Associate Professor with the School of Life Science and Technology, Xidian University, Xi'an, China. Her current research interests include functional magnetic resonance imaging and neuroimaging applications to acupuncture research in obesity and food/drug addiction.

MengChu Zhou (S'88–M'90–SM'93–F'03) received the B.S. degree in control engineering from the Nanjing University of Science and Technology, Nanjing, China, in 1983, the M.S. degree in automatic control from the Beijing Institute of Technology, Beijing, China, in 1986, and the Ph.D. degree in computer and systems engineering from Rensselaer Polytechnic Institute, Troy, NY, USA, in 1990.

He joined the New Jersey Institute of Technology, Newark, NJ, USA, in 1990, where he is currently a Distinguished Professor of Electrical and Computer Engineering. His current research interests include Petri nets, Internet of Things, semiconductor manufacturing, transportation, and energy systems. He has over 600 publications including 12 books, 300+ journal papers (majority in IEEE TRANSACTIONS), and 28 book-chapters.

Prof. Zhou is the Founding Editor of the IEEE Press Book Series on Systems Science and Engineering. He is a Fellow of the International Federation of Automatic Control and the American Association for the Advancement of Science.