



An Intelligence EEG Signal Recognition Method via Noise Insensitive TSK Fuzzy System Based on Interclass Competitive Learning

Tongguang Ni, Xiaoqing Gu* and Cong Zhang

School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou, China

OPEN ACCESS

Edited by:

Mohammad Khosravi,
Persian Gulf University, Iran

Reviewed by:

Shan Zhong,
Changshu Institute of
Technology, China
Juan Yang,
Suzhou University, China

*Correspondence:

Xiaoqing Gu
guxq@cczu.edu.cn

Specialty section:

This article was submitted to
Neuroprosthetics,
a section of the journal
Frontiers in Neuroscience

Received: 21 June 2020

Accepted: 20 July 2020

Published: 04 September 2020

Citation:

Ni T, Gu X and Zhang C (2020) An
Intelligence EEG Signal Recognition
Method via Noise Insensitive TSK
Fuzzy System Based on Interclass
Competitive Learning.
Front. Neurosci. 14:837.
doi: 10.3389/fnins.2020.00837

Epilepsy is an abnormal function disease of movement, consciousness, and nerve caused by abnormal discharge of brain neurons in the brain. EEG is currently a very important tool in the process of epilepsy research. In this paper, a novel noise-insensitive Takagi–Sugeno–Kang (TSK) fuzzy system based on interclass competitive learning is proposed for EEG signal recognition. First, a possibilistic clustering in Bayesian framework with interclass competitive learning called PCB-ICL is presented to determine antecedent parameters of fuzzy rules. Inherited by the possibilistic c-means clustering, PCB-ICL is noise insensitive. PCB-ICL learns cluster centers of different classes in a competitive relationship. The obtained clustering centers are attracted by the samples of the same class and also excluded by the samples of other classes and pushed away from the heterogeneous data. PCB-ICL uses the Metropolis–Hastings method to obtain the optimal clustering results in an alternating iterative strategy. Thus, the learned antecedent parameters have high interpretability. To further promote the noise insensitivity of rules, the asymmetric expectile term and Ho–Kashyap procedure are adopted to learn the consequent parameters of rules. Based on the above ideas, a TSK fuzzy system is proposed and is called PCB-ICL-TSK. Comprehensive experiments on real-world EEG data reveal that the proposed fuzzy system achieves the robust and effective performance for EEG signal recognition.

Keywords: noise insensitive, TSK fuzzy system, Bayesian framework, possibilistic clustering, Ho–Kashyap procedure, asymmetric expectile term

INTRODUCTION

Epilepsy occurs randomly and may occur multiple times in a day. In the case of epileptic seizures, the patients have a sudden physical convulsions and loss of consciousness, which bring great physical and psychological pain to patients (Ahmadlou and Adeli, 2011; Gummadavelli et al., 2018; Cury et al., 2019). Seizures will lead to brain cell death, affect brain function, and even threaten patients' lives in serious cases. The incidence of epilepsy is high, and the age range is very wide, including children, adolescents, and the elderly, but the incidence of children and adolescents is the highest. Both men and women are likely to have the disease, and men are more likely to have this disease than women. As an important clinical means of monitoring and diagnosing epilepsy, EEG provides a more rapid and stable low-cost and non-invasive technology in monitoring the brain activity of the cerebral cortex. It provides information that other physiological methods

cannot provide. The specific waveforms such as spike, sharp, and complex wave can be reflected by EEG. Therefore, the prevention and treatment of epilepsy research for epilepsy patients is of great significance. In the process of diagnosis and treatment of epilepsy, EEG plays an irreplaceable role. Doctors usually judge the condition of patients by observing their EEG.

The traditional way to judge the EEG signal is not only inefficient, but also because of the difference of experts' subjective experience, the automatic detection of EEG signal is still one of the hot issues in biomedical research (Jiang et al., 2017a; Martinez-Vargas et al., 2017; Li et al., 2019). An automatic epilepsy detection method can help doctors improve the accuracy of epilepsy diagnosis and also greatly save time. The research of automatic epilepsy detection is of great value to the prevention, diagnosis, and treatment of epilepsy. At present, epilepsy can be detected by machine learning and data mining. Firstly, the effective feature information is extracted from EEG and preprocessed for data analysis; secondly, the preprocessed EEG data are sent to the classifier for analysis and detection of epileptic and non-epileptic EEG data. In the above implementation process, the key research is to design an effective prediction and discrimination method that can be applied to normal EEG signal and epileptic EEG signal. Many effective methods have been successfully applied to automatic epilepsy detection system, including extreme learning machine (ELM), artificial neural network, Bayesian linear discriminant analysis, support vector machine (SVM), and fuzzy system (Kabir and Zhang, 2016; Qi et al., 2017; Akhavan and Moradi, 2018; Truong et al., 2018; Hossain et al., 2019; Liu et al., 2019; Sreej and Samanta, 2019; Xia et al., 2020). The fuzzy system is a model constructed to deal with the thinking, analysis, reasoning, and decision-making processes in production and practice. It can directly translate natural language into computer language. Due to its ability to process uncertain and ambiguous information, it has a high degree of interpretability and strong learning ability (Juang et al., 2007; Gu et al., 2017a; Jiang et al., 2017b,c; Gu and Wang, 2018). However, the traditional fuzzy system has poor robustness and anti-interference ability, and the classification accuracy is not high in data noise scenarios. But in real life, the classification of noise data is widely used. For example, in actual application scenarios, due to differences in machine advices or scanning technology, such as different rotation angles and noise, the quality of medical images may vary greatly (Siuly and Li, 2015; Hussein et al., 2019; Razzak et al., 2019).

Based on key technology of fuzzy system modeling, this paper proposes a novel noise-insensitive Takagi–Sugeno–Kang (TSK) fuzzy system. How to determine the antecedent and consequent parameters is the key to modeling the noise-insensitive fuzzy system (Takagi and Sugeno, 1985; Jiang et al., 2015). For the antecedent part of fuzzy rules, clustering is one kind of a commonly used strategy, such as fuzzy c -means (FCM) clustering (Bezdek et al., 1984), fuzzy $(c + p)$ clustering (Leski, 2015), Bayesian fuzzy clustering (BFC) (Glenn et al., 2015), and possibilistic c -means (PCM) clustering (Krishnapuram and Keller, 1993). However, FCM, fuzzy $(c + p)$, and BFC are sensitive to noise and will lead to unsatisfactory partition in noisy scenarios. PCM inherits the practicability and flexibility of fuzzy

clustering and greatly enhances the clustering performance of data with noise or outliers. However, the unsupervised nature of PCM makes it unable to use the class label information of samples, which easily causes the insufficient fuzzy space partition, thus further affecting the learning of antecedent parameters of fuzzy rules. The principle of antecedent parameter learning using PCM clustering is shown in **Figure 1A**. PCM clustering is directly used on whole datasets or on samples in each class, and then the antecedent parameters are learned using the obtained clustering results. Then the data samples are simply divided into several clusters, without fully taking advantage of the geometry of data and the label information of samples. In this case, in the data overlapping regions, the distance between clustering centers may be too small or the centers may overlap.

In this paper, we first propose a noise-insensitive possibilistic clustering in Bayesian framework with interclass competitive learning called PCB-ICL. Inherited by PCB, PCB-ICL is noise insensitive; meanwhile, different classes of cluster centers will produce a competitive relationship during the learning process. That is, in the sample overlapping area, the clustering centers are attracted by the samples of the same class and also excluded by the samples of other classes and pushed away from the heterogeneous data. The principle of antecedent parameter learning using PCB-ICL clustering is shown in **Figure 1B**. PCB-ICL integrates the competitive learning mechanism of clustering centers among different classes in the Bayesian framework. PCB-ICL considers the structure information of samples in the clustering procedure and realizes the competition between clustering centers among different classes. We obtain the antecedent part of fuzzy rules by performing PCB-ICL alternatively on each class samples. Then, a Ho–Kashyap procedure (Leski, 2003) with an asymmetric expectile term (Huang et al., 2014a,b) is adopted to estimate the consequent parameters of fuzzy rules. Due to the statistical characteristics of the asymmetric expectile term, it is insensitive to noise; so the asymmetric expectile term is used to measure the misclassification error. Based on the above idea, the TSK fuzzy system called PCB-ICL-TSK is developed, which learns antecedent parameters by PCB-ICL clustering and consequent parameters by the Ho–Kashyap procedure with an asymmetric expectile term. We apply the proposed algorithm on the Bonn EEG dataset, and the experimental results on several noisy classification tasks demonstrate that PCB-ICL-TSK can achieve satisfactory performance in EEG signal classification. The novelty of our study is as follows. (1) Both the PCB-ICL and Ho–Kashyap procedure with an asymmetric expectile term are insensitive to noise; thus, the obtained antecedent and consequent parameters are noise insensitive. (2) With the Bayesian framework, the clustering results of PCB-ICL are globally optimal. In addition, the competitive relationship strategy between cluster centers enhances the interpretability of the antecedents of fuzzy rules. (3) The experiments on real-word EEG datasets confirm the effectiveness of PCB-ICL-TSK.

The detailed chapters are arranged as follows. Section Backgrounds introduces the TSK fuzzy system and PCM clustering. Section Possibilistic Clustering in Bayesian With Interclass Competitive Learning explores PCB-ICL clustering. Section Noise-Insensitive TSK Fuzzy System via Interclass

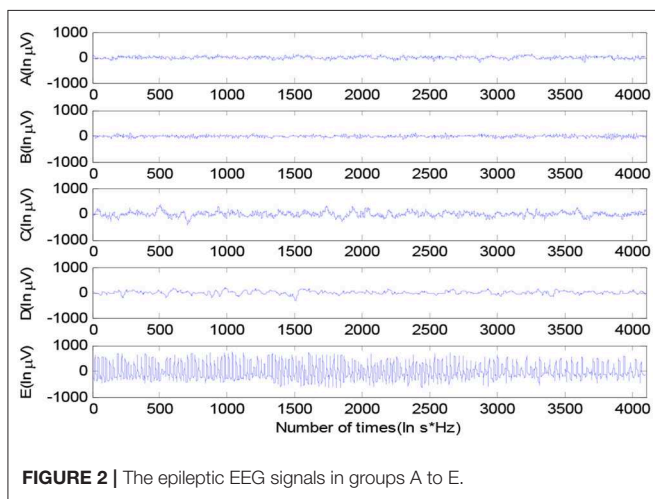
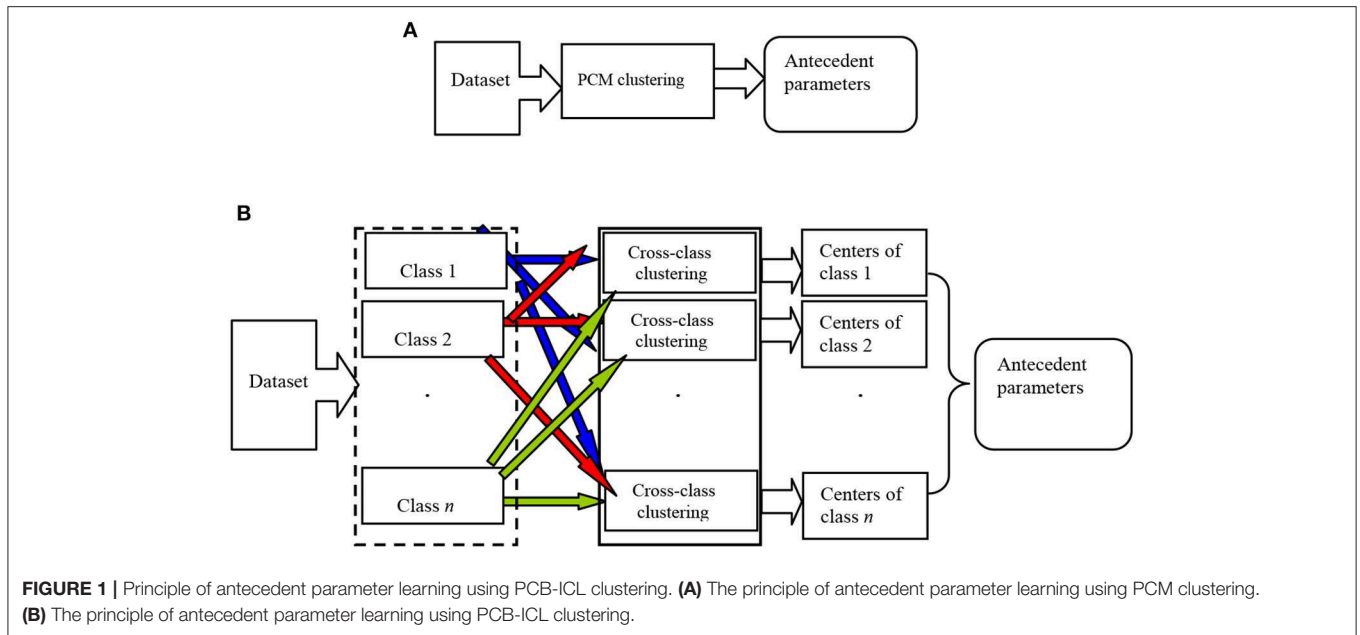


TABLE 1 | The basic information of EEG data groups of A–E.

	Group	Description
Healthy volunteers	A	EEG signals of healthy volunteers in an awakened state with eyes open
	B	EEG signals of healthy volunteers in an awakened state with eyes closed
Patients	C	EEG signals of patients in hippocampal formation of the opposite hemisphere of the brain
	D	EEG signals of patients in the epileptogenic zone during periodic lulls
	E	EEG signals of patients during seizure activity

Competitive Learning explores the noise-insensitive TSK fuzzy system PCB-ICL-TSK. Section Experiment is experiments on noisy EEG data. Section Conclusion is the conclusion.

BACKGROUNDS

Dataset

The epileptic EEG in the experiment is the Bonn dataset from Bonn University, Germany (Tzallas et al., 2009). The Bonn EEG dataset consists of five groups of data, namely, A to E, shown in **Figure 2**. Each group of data contains 100 EEG signal segments of 23.6 s, which were selected from continuous single-channel EEG recordings. The EEG signals were recorded under different conditions with five patients and five healthy volunteers. The basic information of groups A–E is shown in **Table 1**.

TSK Fuzzy System

The most commonly used rule in the zero-order TSK fuzzy system can be represented by

Rule R_k : IF x_1 is $A_{k,1}$ and x_2 is $A_{k,2}$ and ... and x_d is $A_{k,d}$,

$$\text{then } f_k(\mathbf{x}) = P_{k,0}, (k = 1, 2, \dots, K) \tag{1}$$

where x_1, x_2, \dots, x_d are input variables, $A_{k,i}$ is a fuzzy subset, and K is the number of fuzzy rules. For an input vector \mathbf{x} , the output of the corresponding TSK fuzzy system is represented by

$$y_{\text{output}} = \frac{\sum_{k=1}^K \mu_k(\mathbf{x})p_{k,0}}{\sum_{k=1}^K \mu_k(\mathbf{x})} = \sum_{k=1}^K \tilde{\mu}_k(\mathbf{x})p_{k,0}, \tag{2}$$

where the fuzzy membership $\mu_k(\mathbf{x})$ and the normalized fuzzy membership $\tilde{\mu}_k(\mathbf{x})$ is

$$\mu_k(\mathbf{x}) = \prod_{i=1}^d \mu_{A_{k,i}}(x_i), \tag{3}$$

$$\tilde{\mu}_k(\mathbf{x}) = \frac{\mu_k(\mathbf{x})}{\sum_{k'=1}^K \mu_{k'}(\mathbf{x})}. \tag{4}$$

For the sample \mathbf{x}_i , we can rewrite it by

$$d(\mathbf{x}_i) = [\tilde{\mu}_1(\mathbf{x}_i), \tilde{\mu}_2(\mathbf{x}_i), \dots, \tilde{\mu}_K(\mathbf{x}_i)]^T, \tag{5}$$

Generally, antecedent and consequent parameters of rules are determined separately. A popular way to estimate antecedent parameters is to use a certain fuzzy clustering method (Takagi and Sugeno, 1985; Gu et al., 2017b; Salgado et al., 2017). Then $\mu_{A_{k,i}}(x_i)$ can be computed by

$$\mu_{A_{k,i}}(x_i) = \exp\left(-\frac{(x_i - y_{k,i})^2}{2\delta_{k,i}}\right), \tag{6}$$

where the width parameter $\delta_{k,i}$ can be obtained by

$$\delta_{k,i} = \frac{h \cdot \sum_{j=1}^N u_{k,j}(x_{ji} - y_{k,i})^2}{\sum_{j=1}^N u_{k,j}}, \tag{7}$$

where h is the scale parameter and $u_{k,j}$ is the fuzzy membership of the j th input sample \mathbf{x}_j belonging to the k th cluster.

Then the learning of consequent parameters can be represented by

$$\min_{\mathbf{p}} \sum_{i=1}^N |l_i d(\mathbf{x}_i)^T \mathbf{p} - 1| \tag{8}$$

Using the least square solution to minimize the squared loss, Equation (8) can be written by

$$\min_{\mathbf{p}} J(\mathbf{p}) = (\mathbf{D}\mathbf{p} - \mathbf{1}_{N \times 1})^T \mathbf{H}(\mathbf{D}\mathbf{p} - \mathbf{1}_{N \times 1}) + \tau \mathbf{p}^T \mathbf{p}, \tag{9}$$

where $\mathbf{D} = [l_1 d(\mathbf{x}_1)^T, \dots, l_N d(\mathbf{x}_N)^T]^T$, the matrix $\mathbf{H} = \text{diag}(h_1, h_2, \dots, h_N)$, $h_i = 1/|l_i d(\mathbf{x}_i)^T \mathbf{p} - 1|$ for $l_i d(\mathbf{x}_i)^T \mathbf{p} - 1 < 0$, and $h_i = 0$ otherwise. τ is the regularization parameter. Using the Ho-Kashyap iterative method (Leski, 2003), \mathbf{p} can be computed by

$$\mathbf{p} = (\mathbf{D}^T \mathbf{H} \mathbf{D} + \tau \mathbf{I})^{-1} \mathbf{D}^T \mathbf{H} \mathbf{1}, \tag{10}$$

where \mathbf{I} is the identify matrix.

PCM Clustering

PCM clustering is a probability clustering based on FCM. Based on the framework of possibility theory, PCM not only takes into account the general criteria of clustering with the minimum distance within one class and the maximum distance between classes but also emphasizes the principle of the maximum membership value to avoid ordinary solution problems. The objective function of PCM is

$$\min_{\mathbf{U}, \mathbf{Y}} \sum_{n=1}^N \sum_{c=1}^C u_{nc}^m (\mathbf{x}_n - \mathbf{y}_c)^2 + \sum_{n=1}^N \sum_{c=1}^C \eta_c (1 - u_{nc})^m, \tag{11}$$

s.t. $u_{nc} \in [0, 1], \forall n, c$

The closed solution of \mathbf{U} and \mathbf{Y} can be obtained by minimizing the objective function with respect to u_{nc} and \mathbf{y}_c by.

$$\mathbf{y}_c = \frac{\sum_{n=1}^N u_{nc}^m \mathbf{x}_n}{\sum_{n=1}^N u_{nc}^m} \tag{12}$$

$$u_{nc} = \frac{1}{1 + \left(\frac{(\mathbf{x}_n - \mathbf{y}_c)^2}{\eta_c}\right)^{\frac{1}{m-1}}} \tag{13}$$

POSSIBILISTIC CLUSTERING IN BAYESIAN WITH INTERCLASS COMPETITIVE LEARNING

Objective Function

A clustering method implements data partition with some certain degree of similarity. In the clustering process, the samples of one class will have a repulsive effect on the clustering center of other classes, especially in the overlapping regions of different classes of samples; the greater the overlap density, the greater the repulsive force. In these sample overlapping regions, clustering centers of different classes form the competitive learning relationship. On the one hand, the clustering centers are attracted by samples of this class; on the other hand, the clustering centers are excluded by different classes of samples and far away from the overlapping region. In this paper, this idea is embedded into PCM clustering. Based on the Bayesian framework, we propose the possibilistic clustering in Bayesian with interclass competitive learning.

Suppose a given binary classification dataset $\mathbf{X} = \{\mathbf{x}_n, l_n\}_{n=1}^N$, in which $\mathbf{X}_1 = \{\mathbf{x}_n, l_n\}_{n=1}^{N_1}$ and $\mathbf{X}_2 = \{\mathbf{x}_n, l_n\}_{n=N_1+1}^N$ represent two class samples and $l_n \in \{+1, -1\}$ is the class label of the n th sample. Let the cluster number of one class samples be C_1 and the cluster centers of the other class \mathbf{Z} be priorly known $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_{C_2}]^T$, where the cluster number is C_2 . We suppose data \mathbf{X} follows the normal distribution, and each sample \mathbf{x}_i has an independent probability distribution. The maximum

posterior estimation of data and parameters in \mathbf{X}_1 is expressed by

$$p(\mathbf{X}_1, \mathbf{U}, \mathbf{Y}) = p(\mathbf{X}_1 | \mathbf{U}, \mathbf{Y}) p(\mathbf{U} | \mathbf{Y}) p(\mathbf{Y}) \propto \exp \left\{ -\frac{1}{2} \left(\sum_{n=1}^{N_1} \sum_{c=1}^{C_1} u_{nc}^m \|\mathbf{x}_n - \mathbf{y}_c\|^2 + \sum_{n=1}^{N_1} \sum_{c=C_1+1}^{C_1+C_2} u_{nc}^m \|\mathbf{x}_n - \mathbf{z}_c\|^2 \right) \right\} \times \left[\prod_{n=1}^{N_1} \prod_{c=1}^{C_1+C_2} \exp \left(-\frac{1}{2} \eta_c (1 - u_{nc}^m) \right) \times \exp \left\{ -\frac{1}{2} \sum_{c=1}^{C_1} (\mathbf{y}_c - \boldsymbol{\mu}_y)^T \sum_{y=1}^{-1} (\mathbf{y}_c - \boldsymbol{\mu}_y) \right\} \right], \quad (14)$$

where $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{C_1}]^T$ is the unknown cluster center matrix of one class sample. By taking the logarithm of Equation (14), the objective function of PCB-ICL method can be obtained as

$$J(\mathbf{X}_1, \mathbf{U}, \mathbf{Y}) = \sum_{n=1}^{N_1} \sum_{c=1}^{C_1} u_{nc}^m \|\mathbf{x}_n - \mathbf{y}_c\|^2 + \sum_{n=1}^{N_1} \sum_{c=C_1+1}^{C_1+C_2} u_{nc}^m \|\mathbf{x}_n - \mathbf{z}_c\|^2 + \sum_{n=1}^{N_1} \sum_{c=1}^{C_1+C_2} \eta_c (1 - u_{nc}^m)^m + \sum_{c=1}^{C_1} (\mathbf{y}_c - \boldsymbol{\mu}_y)^T \sum_{y=1}^{-1} (\mathbf{y}_c - \boldsymbol{\mu}_y). \quad (15)$$

From Equations (14) and (15), we can see that (1) the PCB-ICL method shows the competition relationship between clustering centers of different classes. Different from the traditional PCM clustering method, PCB-ICL not only considers the label information of samples but also considers the competition relationship between clustering centers, as shown in the first two items. On the premise that the clustering centers of the other class are priori known, the clustering centers of the current class will inevitably have a competition relationship with these known clustering centers in the overlapping region. (2) Due to simultaneously utilizing the global distribution structure and the discrimination information of the samples, the obtained antecedent part of fuzzy rules by PCB-ICL can realize the clarity of fuzzy space partition and enhance the interpretability of the fuzzy rules.

Parameter Learning

To obtain the optimal fuzzy partition matrix \mathbf{U} , the PCB-ICL method uses the Metropolis–Hastings method (Chib and Greenberg, 1995; Elvira et al., 2017) to construct a Markov chain to make $p(\mathbf{U} | \mathbf{X}_1, \mathbf{Y})$ stable. The conditional distribution $p(\mathbf{U} | \mathbf{X}_1, \mathbf{Y})$ is proportional to the joint distribution $p(\mathbf{X}_1, \mathbf{U}, \mathbf{Y})$ when the sample and clustering center are known and also is proportional to the conditional distribution $p(\mathbf{U} | \mathbf{X}_1, \mathbf{Y})$. Therefore, we only need compute $p(\mathbf{x}_n, \mathbf{u}_n | \mathbf{Y})$ of the sample \mathbf{x}_n :

$$p(\mathbf{x}_n, \mathbf{u}_n | \mathbf{Y}) = p(\mathbf{x}_n | \mathbf{u}_n, \mathbf{Y}) p(\mathbf{u}_n | \mathbf{Y}) \propto \exp \left\{ -\frac{1}{2} \left(\sum_{c=1}^{C_1} u_{nc}^m \|\mathbf{x}_n - \mathbf{y}_c\|^2 + \sum_{c=C_1+1}^{C_1+C_2} u_{nc}^m \|\mathbf{x}_n - \mathbf{z}_c\|^2 \right) \right\} \times \prod_{c=1}^{C_1+C_2} \exp \left(-\frac{1}{2} \eta_c (1 - u_{nc}^m) \right). \quad (16)$$

Thus, the process of the i th iteration of the Markov chain is

1) Generate a new state \mathbf{u}_n^+ of \mathbf{u}_n with a uniform distribution as

$$\mathbf{u}_n^+ \sim \text{Uniform}(0, 1), \forall n \quad (17)$$

2) The newly generated membership \mathbf{u}_n^+ is accepted by the probability a_u as

$$a_u = \min \left\{ 1, \frac{p(\mathbf{x}_n, \mathbf{u}_n^+ | \mathbf{Y})}{p(\mathbf{x}_n, \mathbf{u}_n | \mathbf{Y})} \right\} \quad (18)$$

Then accepting a_u as the current state with probability \mathbf{u}_n ,

$$\mathbf{u}_n = \begin{cases} \mathbf{u}_n^+, & \mu \leq \alpha_u \\ \mathbf{u}_n, & \mu > \alpha_u \end{cases} \quad (19)$$

where μ is a random number in $[0, 1]$. The distribution of the new state \mathbf{u}_n^+ obtained by sampling is independent of the current sample, and the state $\mathbf{u}_n^+ / \mathbf{u}_n$ is independent, so a_u does not need Hasting correction.

3) Compare $p(\mathbf{x}_n, \mathbf{u}_n^+ | \mathbf{Y}^*)$ and $p(\mathbf{x}_n, \mathbf{u}_n^* | \mathbf{Y}^*)$, where \mathbf{Y}^* and \mathbf{u}_n^* are the optimal values of \mathbf{Y} and \mathbf{u}_n . If, $p(\mathbf{x}_n, \mathbf{u}_n^+ | \mathbf{Y}^*) > p(\mathbf{x}_n, \mathbf{u}_n^* | \mathbf{Y}^*)$ \mathbf{u}_n^+ is replaced by \mathbf{u}_n^* .

When the matrix \mathbf{U} is fixed, we use Metropolis–Hastings to sample the conditional distribution $p(\mathbf{Y} | \mathbf{X}, \mathbf{U})$. In this case, $p(\mathbf{Y} | \mathbf{X}, \mathbf{U})$ is proportional to the joint distribution $p(\mathbf{X}, \mathbf{U}, \mathbf{Y})$. We estimate \mathbf{y}_c by using the Gaussian distribution as

$$\mathbf{y}_c^+ \sim \mathcal{N} \left(\mathbf{y}_c, \frac{1}{\sigma} \sum \mathbf{y} \right) \quad (20)$$

where \mathbf{y}_c^+ centers on the current value \mathbf{y}_c . σ is a positive number and is used to control the compactness of cluster centers. In the experiment, we empirically set σ to 10.

For the newly generated \mathbf{y}_c^+ , it is independent of other clustering centers. Then the conditional distribution $p(\mathbf{X}, \mathbf{y}_c | \mathbf{U})$ is represented by

$$p(\mathbf{X}, \mathbf{y}_c | \mathbf{U}) = p(\mathbf{X} | \mathbf{U}, \mathbf{y}_c) p(\mathbf{y}_c) \propto \exp \left\{ -\frac{1}{2} \sum_{n=1}^{N_1} u_{nc}^m \|\mathbf{x}_n - \mathbf{y}_c\|^2 \right\} \times \exp \left\{ -\frac{1}{2} (\mathbf{y}_c - \boldsymbol{\mu}_y)^T \sum_{y=1}^{-1} (\mathbf{y}_c - \boldsymbol{\mu}_y) \right\}. \quad (21)$$

Similarly, the newly generated membership \mathbf{y}_c^+ is accepted by the probability a_y as

$$a_y = \min \left\{ 1, \frac{p(\mathbf{X}, \mathbf{y}_c^+ | \mathbf{U})}{p(\mathbf{X}, \mathbf{y}_c | \mathbf{U})} \right\} \quad (22)$$

Since the Gaussian distribution is symmetric, a_y does not need Hasting correction.

Finally, we compute $p(\mathbf{X}, \mathbf{U}^*, \mathbf{Y}^*)$ using Equation (15) and compare it with the current $p(\mathbf{X}, \mathbf{U}, \mathbf{Y})$. If $p(\mathbf{X}, \mathbf{U}, \mathbf{Y}) > p(\mathbf{X}, \mathbf{U}^*, \mathbf{Y}^*)$, the $\{\mathbf{U}, \mathbf{Y}\}$ is replaced by $\{\mathbf{U}^*, \mathbf{Y}^*\}$.

Based on the above analysis, we give the procedure of the PCB-ICL method in **Algorithm 1**.

Algorithm 1 | PCB-ICL method.

Input: Dataset X_1 of one class, the number of clustering C , priori known clustering center matrix Z of the other class;

Output: Fuzzy partition matrix U^* and clustering center matrix Y^* .

Step 1 Initiate $u_n^+ \sim Uniform(0, 1), \forall n$;
 Step 2 Initiate $y_c^+ \sim \mathcal{N}(y_c, \frac{1}{\sigma} \sum y)$, $\forall c$;
 Step 3 Set $u_n^* = u_n^+, y_c^* = y_c^+$;
 For $iter = 1, 2, \dots, N_{iter}$
 For $n = 1, 2, \dots, N$
 Step 4 Sample u_n^+ using Equation (17) and accept it as u_n using Equations (18) and (19);
 Step 5 If $\rho(x_n, u_n^+ | Y^*) > \rho(x_n, u_n^* | Y^*)$, then $u_n^* = u_n^+$;
 Endfor
 For $c = 1, 2, \dots, C$
 Step 6 Sample y_c^+ using Equation (20) and accept it as y_c^* using Equations (21) and (22);
 Step 7 If $\rho(X, y_c^+ | U^*) > \rho(X, y_c^* | U^*)$, then $y_c^* \leftarrow y_c^+$;
 Endfor
 Step 8 If $\rho(X, U^*, Y^*) > \rho(X, U, Y)$, then $U^* \leftarrow U, Y^* \leftarrow Y$;
 Endfor.

Algorithm 2 | Fuzzy partition on the whole data.

Input: Two class samples X_1 and X_2 , the numbers of clustering C_1 and C_2 in two classes;

Output: Fuzzy partition matrix $U^{(1)*}, U^{(2)*}$ and clustering center matrix $Y^{(1)*}, Y^{(2)*}$.

Step 1 Initiate $u_n^{(1)}, u_n^{(2)} \sim Uniform(0, 1)$ in two classes;
 Step 2 Initiate in two classes;
 Step 3 Set $u_n^{(1)*} = u_n^{(1)}, u_n^{(2)*} = u_n^{(2)}, y_{c_1}^{(1)*} = y_{c_1}^{(1)}, y_{c_2}^{(2)*} = y_{c_2}^{(2)}$,
 $iter = 0$;
 Do
 Step 4 Perform **Algorithm 1** on X_1 ;
 Step 5 Perform **Algorithm 1** on X_2 ;
 $iter = iter + 1$;
 Until $Y^{(1)*}$ is $|Y^{(1)*}(v) - Y^{(1)*}(v - 1)| \leq \epsilon$ or $iter > N_{iter}$

NOISE-INSENSITIVE TSK FUZZY SYSTEM VIA INTERCLASS COMPETITIVE LEARNING

Antecedent Parameter Learning in PCB-ICL-TSK

In this section, we compute the antecedent parameters in PCB-ICL-TSK. The premise of PCB-ICL clustering in **Algorithm 1** is that the clustering centers of other class are priori known, which is obviously not feasible in practical application. To perform the fuzzy partition on the whole data set, we take the strategy of an alternating cycle to perform **Algorithm 1** on different classes. In this case, the clustering results of one class influence the ones of the other class. Taking binary classification as an example, we perform **Algorithm 1** on positive class X_1 and negative class X_2 alternately. The detailed fuzzy partition of the whole data is shown in **Algorithm 2**.

The numbers of clustering in two classes are C_1 and C_2 , and the cluster centers in two classes are Y_1 and Y_2 , respectively. After applying **Algorithm 2** on the whole data, the center matrix Y can be described by $Y^* = [Y^{(1)*}; Y^{(2)*}]$.

Consequent Parameter Learning in PCB-ICL-TSK

In this section, we compute the noise-insensitive consequent parameters in PCB-ICL-TSK. As discussed before, using the obtained the antecedent parameters, the dataset $X = \{x_i, l_i\}_{i=1}^N$ is represented as $S = \{(\tilde{\mu}(x_i), l_i)\}_{i=1}^N$, where $\tilde{\mu}(x_i) = [\tilde{\mu}_1(x_i)^T, \tilde{\mu}_2(x_i)^T, \dots, \tilde{\mu}_{(C_1+C_2)}(x_i)^T]^T$. Defining the vector $d(x_i) = [\tilde{\mu}_1(x_i)^T, \tilde{\mu}_2(x_i)^T, \dots, \tilde{\mu}_{(C_1+C_2)}(x_i)^T, 1]^T$, the consequent vector $p^* = [p_0^1, p_0^2, \dots, p_0^{(C_1+C_2)}, w]^T$ can be

computed by

$$f(x_i) = (p^*)^T d(x_i) = p_0^T \tilde{\mu}(x_i) + w \begin{cases} \geq 0, & x_i \in X_1 \\ < 0, & x_i \in X_2 \end{cases} \quad (23)$$

where the vector $p_0 = [p_0^1, p_0^2, \dots, p_0^{(C_1+C_2)}]^T$ and w is the decision threshold. If we multiply Equation (23) by the class label, Equation (23) is represented as $l_i(p^*)^T d(x_i) \geq 0$ ($i = 1, \dots, N$). Then, the vector p^* can be computed by

$$l_i(p^*)^T d(x_i) \geq \epsilon_0 \quad (24)$$

In particular, $\epsilon_0 = 1$ leads to the classical SVM. For simplicity, we set $\epsilon_0 = 1$, and Equation (24) can be written as $l_i(p^*)^T d(x_i) \geq 1$. Thus, Equation (24) can be written as

$$J(p^*) = \sum_{i=1}^N (l_i(p^*)^T d(x_i) - 1)^2 \quad (25)$$

Denote the matrix $D = [l_1 d(x_1)^T, l_2 d(x_2)^T, \dots, l_N d(x_N)^T]^T$ and the error vector $e = D^* p^* - \mathbf{1}$. Equation (25) can be rewritten as

$$\min_{p^*} J(p^*) = \frac{1}{2} (D p^* - \mathbf{1})^T H (D p^* - \mathbf{1}) \quad (26)$$

where the matrix $H = (\lambda/N) \text{diag}(h_1, h_2, \dots, h_N)$, with $h_i = 0$ for error $e_i \geq 0$ and 1 otherwise.

However, the misclassification error in Equation (24) is noise sensitive. To further improve the robustness of the TSK fuzzy system, we use the asymmetric expectile term, which is noise insensitive, especially to noise around the decision boundary. The weight h_i of the i th sample can be expressed by

$$h_i = \begin{cases} q, & e_i \geq 0 \\ (1 - q), & e_i < 0 \end{cases} \quad (27)$$

where h_i is the q (lower) expectile parameter. Obviously, when $q = 0$, the loss term obtained in Equation (27) is equal to the

Algorithm 3 | Learning algorithm for consequent parameters.

Input: The dataset \mathbf{X} ; the number of clusters ($C_1 + C_2$); the cluster centers $\mathbf{Y}_1, \mathbf{Y}_2$ and the membership matrix $\mathbf{U}_1, \mathbf{U}_2$; the expectile parameter q ; and the regularization parameter τ ;

Output: Consequent parameters \mathbf{p}_0 .

Step 1 Run **Algorithm 2** to obtain the antecedent parameters;

Step 2 Compute the membership function $d(\mathbf{x}_i)=[\tilde{\mu}_1(\mathbf{x}_i)^T, \tilde{\mu}_2(\mathbf{x}_i)^T, \dots, \tilde{\mu}_{(C_1+C_2)}(\mathbf{x}_i)^T, 1]^T$ by using Equations (5)–(7);

$k = 0$;

Do

Step 3 Obtain the parameters $(\mathbf{p}^*)^{(k)}$ using Equation (29);

Step 4 Compute the parameter $\mathbf{e}^{(k)}$ using $\mathbf{e}^{(k)} = \mathbf{D}(\mathbf{p}^*)^{(k)} - \mathbf{1}$;

Step 5 Compute the parameter $\mathbf{H}^{(k+1)}$ using Equation (27);

$k = k + 1$;

Until \mathbf{p}^* is convergence or $k > k_{\max}$.

hinge loss, and when $q = 0.5$, the loss term is equal to the l_2 loss in Huang et al. (2014a,b).

At the same time, considering the regularization term, Equation (26) can be rewritten as

$$\min_{\mathbf{p}^*} J(\mathbf{p}^*)^{(k)} = \frac{1}{2}(\mathbf{D}^*(\mathbf{p}^*)^{(k)} - \mathbf{1})^T \mathbf{H}^{(k)}(\mathbf{D}^*(\mathbf{p}^*)^{(k)} - \mathbf{1}) + \frac{\tau}{2}(\mathbf{p}_0^{(k)})^T \mathbf{p}_0^{(k)} \quad (28)$$

where τ is the regularization parameter. $\mathbf{p}^*(k)$, $\mathbf{H}^{(k)}$, and $\mathbf{e}^{(k)}$ are the k th iteration of \mathbf{p}^* , \mathbf{H} , and \mathbf{e} , respectively.

The condition for optimality of Equation (28) in the k th iteration is obtained by setting $dJ/d\mathbf{p}^* = 0$:

$$(\mathbf{p}^*)^{(k)} = ((\mathbf{D}^*)^T \mathbf{H}^{(k)} \mathbf{D}^* + \tau \tilde{\mathbf{I}})^{-1} (\mathbf{D}^*)^T \mathbf{H}^{(k)} \mathbf{1} \quad (29)$$

where $\tilde{\mathbf{I}}$ is the identity matrix with the last element on the main diagonal set to 0.

The consequent parameter learning in IB-TSK-FC on dataset \mathbf{X} is shown in **Algorithm 3**.

EXPERIMENT

Experimental Settings

The real-world EEG signals have characters of high dimensionality and instability. Feature extraction is a necessary stage before classification for EEG signal recognition. In general, time domain and frequency domain feature extractions are two types of feature extraction methods (Wen and Zhang, 2017). In our experiments, we extract EEG features using kernel principal component analysis (KPCA) and short-time Fourier transform (STFT) (Blanco et al., 1997). The former is the time domain feature extraction, and the latter is the frequency domain feature extraction. In the experiment, we design eight classification tasks, namely, four binary classification and four three-class classification tasks, as shown in **Table 2**. We corrupt the original

TABLE 2 | EEG classification tasks in the experiment.

Tasks	Number of classes	Datasets
T1	Two classes	A and C
T2		A and E
T3		B and D
T4		B and E
T5	Three classes	A, C and E
T6		A, D, and E
T7		B, C, and E
T8		B, D, and E

TABLE 3 | Parameter settings for all methods in the experiment.

Methods	Parameter settings
ϵ -margin-TSK-FS FS-FCSVM	Number of rules $\in \{1, 2, \dots, 12\}$, regularization parameter $\in \{10^{-3}, 10^{-2}, \dots, 10^3\}$, scale parameter $\in \{0.4^2, 0.6^2, \dots, 3^2\}$, fuzzy index = 2
IB-TSK-FC	Model sparsity parameter $\in \{1, 2, \dots, 6\}$, fuzzy index = 2, number of particles = 10, convergence thresholds = 10^{-3} , convergence threshold = 10^{-3} , strength parameter = 3
CS-SVM	Gaussian kernel parameter $\in \{10^{-2}, 10^{-1}, \dots, 10^2\}$, regularization parameter $\in \{10^{-3}, 10^{-2}, \dots, 10^3\}$
FRSVM-ANCH	Gaussian kernel parameter $\in \{10^{-2}, 10^{-1}, \dots, 10^2\}$, regularization parameter $\in \{10^{-3}, 10^{-2}, \dots, 10^3\}$, pinball loss parameter = 0.05
PCB-ICL-TSK	Fuzzy index = 2, number of rules $\in \{1, 2, \dots, 12\}$, convergence threshold = 10^{-3} , strength parameter = 3, expectile parameter = 0.05, maximum number of iterations = 1,000 and 200 in Algorithms 1 and 2 , respectively

datasets with different amounts of random noises at 5, 10, and 15% noise levels.

The experimental environment in this study is a computer with Intel Core i3-3317U 3.40-GHz CPU and 8-GB RAM. To validate the performance of MST-TSK, we compare three fuzzy systems (FS-FCSVM; et al., 2007, ϵ -margin-TSK-FS; Leski, 2005, and IB-TSK-FC; Gu et al., 2017b) and two robust classification methods (CS-SVM; Iranmehr et al., 2019 and FRSVM-ANCH; Gu et al., 2019). The Gaussian kernel is used for two SVM methods. The parameter settings for all methods are listed in **Table 3**. All parameters are obtained by a 5-fold cross-validation strategy.

Classification Performance Comparison

In this section, eight EEG classification tasks are used to verify the classification performance of PCB-ICL-TSK. **Tables 4, 5** show the experimental results of six classification algorithms using STFT and KPCA feature extraction methods at the 5% noise level. **Tables 6, 7** show the experimental results of six

TABLE 4 | The classification accuracy for the 5% noise level using STFT features.

Tasks	FS-FCSVM	ϵ -margin-TSK-FS	IB-TSK-FC	CS-SVM	FRSVM-ANCH	PCB-ICL-TSK
Task 1	94.56	94.69	95.22	96.27	96.73	96.74
Task 2	94.88	94.81	95.34	96.19	96.51	96.62
Task 3	94.64	94.57	95.39	96.45	96.70	96.69
Task 4	94.21	94.34	95.57	96.58	96.54	96.67
Task 5	93.46	93.25	93.98	96.27	96.32	96.29
Task 6	93.37	93.42	93.96	95.89	96.14	96.12
Task 7	93.25	93.34	93.86	95.78	95.97	96.04
Task 8	93.38	93.29	93.75	95.86	96.09	96.10

The bold values indicate the best classification performance in the tasks.

TABLE 5 | The classification accuracy for the 5% noise level using KPCA features.

Tasks	FS-FCSVM	ϵ -margin-TSK-FS	IB-TSK-FC	CS-SVM	FRSVM-ANCH	PCB-ICL-TSK
Task 1	94.53	94.70	95.23	96.28	96.63	96.75
Task 2	94.92	94.78	95.39	96.23	96.50	96.59
Task 3	94.62	94.55	95.40	96.42	96.72	96.72
Task 4	94.23	94.33	95.56	96.62	96.51	96.65
Task 5	93.47	93.25	93.95	96.26	96.33	96.31
Task 6	93.38	93.44	93.97	95.94	96.11	96.19
Task 7	93.23	93.36	93.85	95.79	95.95	96.01
Task 8	93.37	93.34	93.76	95.89	96.10	96.11

The bold values indicate the best classification performance in the tasks.

TABLE 6 | The classification accuracy for the 10% noise level using STFT features.

Tasks	FS-FCSVM	ϵ -margin-TSK-FS	IB-TSK-FC	CS-SVM	FRSVM-ANCH	PCB-ICL-TSK
Task 1	92.79	92.84	93.44	94.64	95.09	95.29
Task 2	93.12	93.01	93.58	94.57	95.21	95.28
Task 3	92.76	92.76	93.59	94.86	95.39	95.30
Task 4	92.48	92.52	93.76	95.03	94.99	95.21
Task 5	91.60	91.45	92.16	94.63	94.66	94.93
Task 6	91.61	91.65	92.23	94.23	94.71	94.70
Task 7	91.44	91.58	92.09	94.18	94.61	94.68
Task 8	91.58	91.55	91.94	94.19	94.56	94.67

The bold values indicate the best classification performance in the tasks.

TABLE 7 | The classification accuracy for the 10% noise level using KPCA features.

Tasks	FS-FCSVM	ϵ -margin-TSK-FS	IB-TSK-FC	CS-SVM	FRSVM-ANCH	PCB-ICL-TSK
Task 1	92.75	92.80	93.39	94.61	95.24	95.27
Task 2	93.12	92.98	93.53	94.56	95.21	95.21
Task 3	92.77	92.72	93.63	94.85	95.14	95.26
Task 4	92.49	92.58	93.84	95.08	94.95	95.20
Task 5	91.63	91.42	92.20	94.66	94.77	94.96
Task 6	91.53	91.66	92.22	94.24	94.58	94.69
Task 7	91.45	91.52	92.10	94.24	94.51	94.68
Task 8	91.61	91.58	91.87	94.19	94.54	94.69

The bold values indicate the best classification performance in the tasks.

TABLE 8 | The classification accuracy for the 15% noise level using STFT features.

Tasks	FS-FCSVM	ϵ -margin-TSK-FS	IB-TSK-FC	CS-SVM	FRSVM-ANCH	PCB-ICL-TSK
Task 1	90.73	90.77	91.39	92.98	93.48	93.88
Task 2	91.13	90.94	91.49	92.84	93.85	93.90
Task 3	90.70	90.76	91.53	93.18	93.70	93.88
Task 4	90.48	90.57	91.84	93.33	93.85	93.84
Task 5	89.68	89.39	90.19	92.91	93.07	93.57
Task 6	89.55	89.74	90.20	92.59	92.92	93.33
Task 7	89.49	89.48	90.05	92.48	93.04	93.28
Task 8	89.62	89.67	89.85	92.50	93.28	93.32

The bold values indicate the best classification performance in the tasks.

TABLE 9 | The classification accuracy for the 15% noise level using KPCA features.

Tasks	FS-FCSVM	ϵ -margin-TSK-FS	IB-TSK-FC	CS-SVM	FRSVM-ANCH	PCB-ICL-TSK
Task 1	90.73	90.77	91.39	92.96	93.47	93.87
Task 2	91.13	90.94	91.49	92.86	93.34	93.82
Task 3	90.70	90.76	91.53	93.20	93.73	93.90
Task 4	90.48	90.57	91.84	93.33	93.82	93.82
Task 5	89.68	89.39	90.19	92.93	93.12	93.49
Task 6	89.55	89.74	90.20	92.52	92.91	93.32
Task 7	89.49	89.48	90.05	92.47	92.99	93.24
Task 8	89.62	89.67	89.85	92.49	93.20	93.22

The bold values indicate the best classification performance in the tasks.

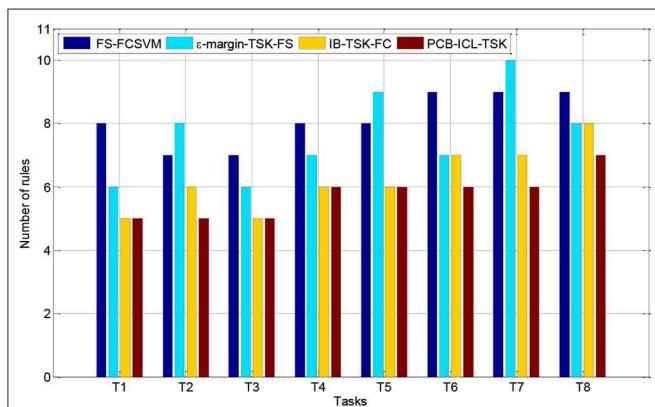


FIGURE 3 | The rules obtained by four fuzzy systems on the 5% noise level using KPCA features.

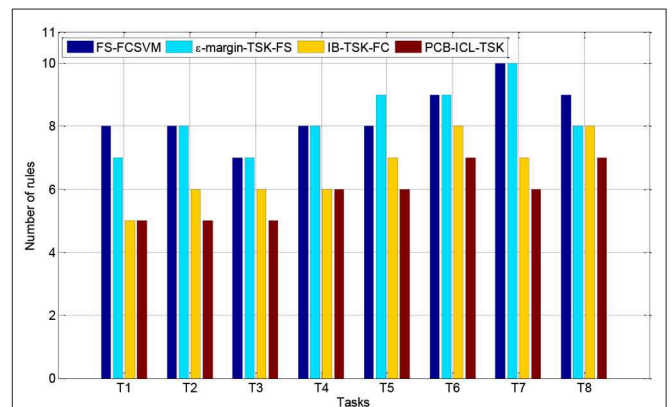


FIGURE 5 | The rules obtained by four fuzzy systems on the 5% noise level using STFT features.

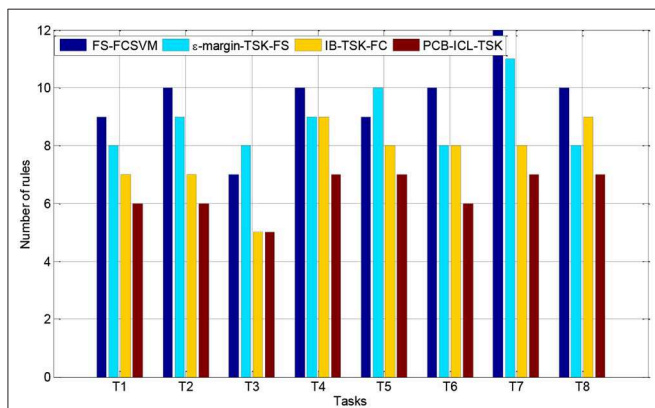


FIGURE 4 | The rules obtained by four fuzzy systems on the 15% noise level using KPCA features.

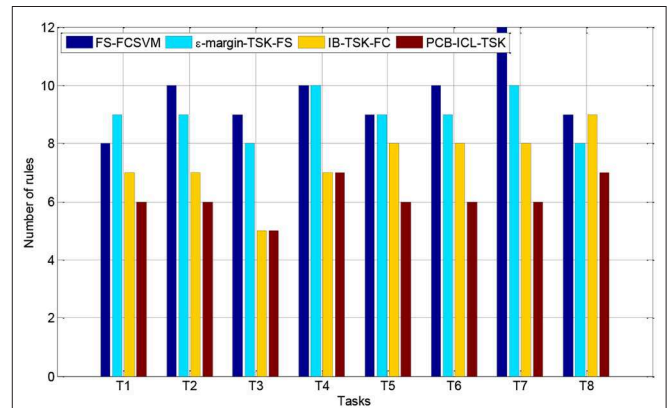
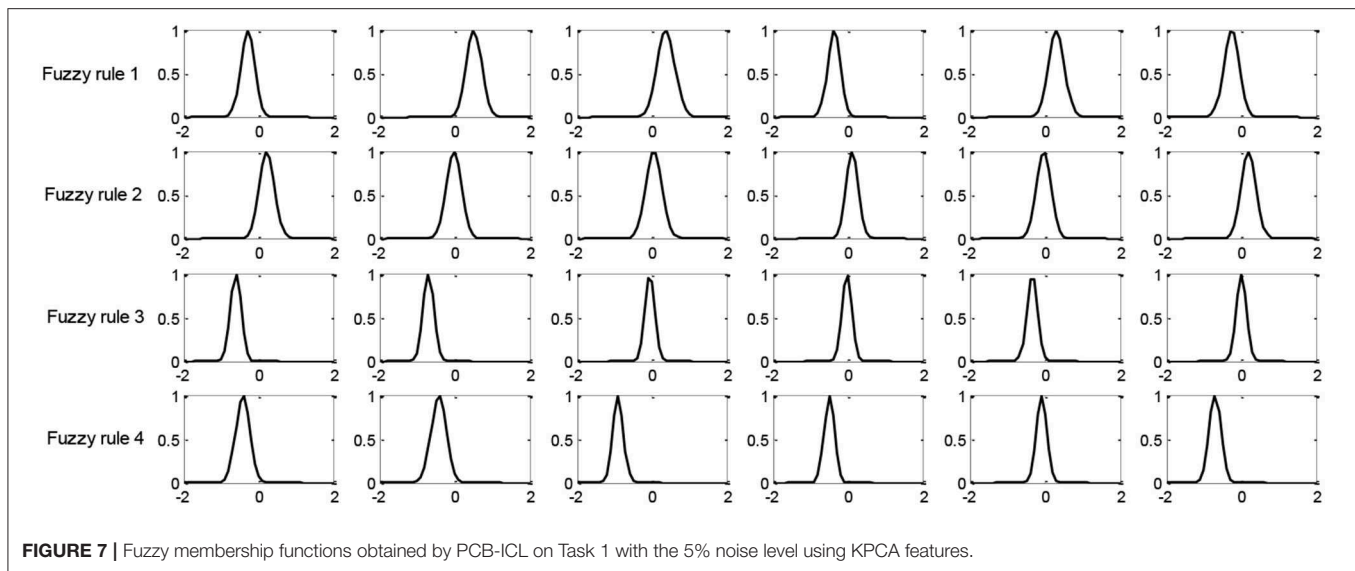


FIGURE 6 | The rules obtained by four fuzzy systems on the 15% noise level using STFT features.



classification methods using STFT and KPCA feature extraction methods at the 10% noise level. **Tables 8, 9** show the experimental results of six classification methods using STFT and KPCA feature extraction methods at the 15% noise level. From the experimental results, it can be seen that the noise data seriously affect the classification performance of the method. During the learning process, considering the noise of the data is helpful to promote the classification performance. Therefore, the performances of FS-FCSVM, ε -margin-TSK-FS, and IB-TSK-FC are poor. CS-SVM, FRSVM-ANCH, and PCB-ICL-TSK are not sensitive to noise, and they can achieve good classification results. In particular, PCB-ICL-TSK shows excellent classification performance in different levels of noise occasions, and it reflects strong robustness. Since PCB-ICL-TSK uses the PCB-ICL and Ho-Kashyap procedure with an asymmetric expectile term to compute antecedent and consequent parameters of fuzzy rules, it is noise insensitive. In addition, in the Bayesian framework, PCB-ICL obtains global optimal clustering results, and the strategy of competitive relationship of clustering centers can enhance the interpretability of the antecedents of fuzzy rules.

Interpretability Comparison

In this section, we compare the number of fuzzy rules of four fuzzy systems in Task 8. **Figures 3, 4** show the number of fuzzy rules on the 5 and 15% noise levels for four fuzzy systems using KPCA features. **Figures 5, 6** show the number of fuzzy rules on the 5 and 15% noise levels for four fuzzy systems using STFT features. From the results in **Figures 3–6**, compared with the three fuzzy systems, the number of fuzzy rules obtained by PCB-ICL-TSK is the least in all EEG classification tasks. It is known that for fuzzy systems, the interpretability of fuzzy rules is related to the number of fuzzy rules and the definition of fuzzy subsets. The fuzzy membership function obtained by PCB-ICL on Task 1 at the 5% noise level using KPCA features is shown in **Figure 7**.

Because PCB-ICL clustering considers the influence of clustering centers of different classes in the process of clustering, that is, the competition relationship between different classes of clustering centers, PCB-ICL clustering can obtain clustering centers with a large interval, which guarantees the partition clarity of feature space and the classification accuracy of the obtained fuzzy system and the interpretation of rules.

CONCLUSION

The noise-insensitive PCB-ICL-TSK fuzzy system is proposed in this paper. In the learning of rule antecedent parameters, the proposed noise-insensitive PCB-ICL clustering based on the Bayesian probability model is used. PCB-ICL clustering considers the repulsion between different clustering centers, which can ensure the interpretability of the rule antecedent. PCB-ICL can learn the global optimal solution of clustering results by using the Markov model. PCB-ICL-TSK learns consequent parameters using the Ho-Kashyap procedure with an asymmetric expectile term. Thus, it not only has strong noise resistance but also has high classification performance. The experimental results of a real EEG dataset show that PCB-ICL-TSK has achieved satisfactory results in classification performance and high interpretability. Our future work is to further improve its practicability when the sample dimension is large.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: The dataset analyzed for this study can be found in the Department of Epileptology University of Bonn [<http://epileptologie-bonn.de/cms/upload/workgroup/lehnertz/eegdata.html>].

AUTHOR CONTRIBUTIONS

TN and XG conceived and developed the theoretical framework of the manuscript. All authors carried out the experiment and data process and drafted the manuscript.

REFERENCES

- Ahmadlou, M., and Adeli, H. (2011). Functional community analysis of brain: a new approach for EEG-based investigation of the brain pathology. *Neuroimage* 58, 401–408. doi: 10.1016/j.neuroimage.2011.04.070
- Akhavan, A., and Moradi, M. H. (2018). Detection of concealed information using multichannel discriminative dictionary and spatial filter learning. *IEEE Trans. Inform. Foren. Secur.* 13, 2616–2627. doi: 10.1109/TIFS.2018.2825940
- Bezdek, J., Ehrlich, R., and Full W. (1984). FCM: the fuzzy c-means clustering algorithm. *Comp. Geosci.* 10, 191–203. doi: 10.1016/0098-3004(84)90020-7
- Blanco, S., Kochen, S., Rosso, O. A., and Salgado, P. (1997). Applying time frequency analysis to seizure EEG activity. *IEEE Eng. Med. Biol. Mag.* 16, 64–71. doi: 10.1109/51.566156
- Chib, S., and Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *Am. Stat.* 49, 327–335. doi: 10.1080/00031305.1995.10476177
- Cury, C., Maurel, P., Gribonval, R., and Barillot, C. (2019). A sparse EEG-informed fMRI model for hybrid EEG-fMRI neurofeedback prediction. *Front. Neurosci.* 13:1451. doi: 10.3389/fnins.2019.01451
- Elvira, V., Míguez, J., and Djurić, P. M. (2017). Adapting the number of particles in sequential Monte Carlo methods through an online scheme for convergence assessment. *IEEE Trans. Signal Process.* 65, 1781–1794. doi: 10.1109/TSP.2016.2637324
- Glenn, T. C., Zare, A., and Gader, P. D. (2015). Bayesian fuzzy clustering. *IEEE Trans. Fuzzy Syst.* 23, 1545–1561. doi: 10.1109/TFUZZ.2014.2370676
- Gu, X., Chung, F., and Wang, S. (2017a). Bayesian Takagi-Sugeno-Kang fuzzy classifier. *IEEE Trans. Fuzzy Syst.* 25, 1655–1671. doi: 10.1109/TFUZZ.2016.2617377
- Gu, X., Chung, F. L., Ishibuchi, H., and Wang, S. (2017b). Imbalanced TSK fuzzy classifier by cross-class Bayesian fuzzy clustering and imbalance learning. *IEEE Trans. Syst. Man Cybernet. Syst.* 47, 2005–2020. doi: 10.1109/TSMC.2016.2598270
- Gu, X., Ni, T., and Fan, Y. (2019). A fast and robust support vector machine with anti-noise convex hull and its application in large-scale ncRNA data classification. *IEEE Access.* 7, 134730–134741. doi: 10.1109/ACCESS.2019.2941986
- Gu, X., and Wang, S. (2018). Bayesian Takagi-Sugeno-Kang Fuzzy model and its joint learning of structure identification and parameter estimation. *IEEE Trans. Indust. Inform.* 14, 5327–5337. doi: 10.1109/TII.2018.2813977
- Gummadavelli, A., Zaveri, H. P., Spencer, D. D., and Gerrard, J. L. (2018). Expanding brain-computer interfaces for controlling epilepsy networks: novel thalamic responsive neurostimulation in refractory epilepsy. *Front. Neurosci.* 12:474. doi: 10.3389/fnins.2018.00474
- Hossain, M. S., Amin, S. U., Alsulaiman, M., and Muhammad, G. (2019). Applying deep learning for epilepsy seizure detection and brain mapping visualization, ACM Trans. *Multimed. Comput. Commun. Appl.* 15, 1–17. doi: 10.1145/3241056
- Huang, X. L., Shi, L., Pelckmansb, K., and Suykens, J. A. K. (2014a). Asymmetric v-tube support vector regression. *Comput. Stat. Data Anal.* 77, 371–382. doi: 10.1016/j.csda.2014.03.016
- Huang, X. L., Shi, L., and Suykens, J. A. K. (2014b). Support vector machine classifier with pinball loss. *IEEE Trans. Pattern Anal. Mach. Intell.* 36, 984–997. doi: 10.1109/TPAMI.2013.178
- Hussein, R., Palangi, H., Ward, R. K., and Wang, Z. J. (2019). Optimized deep neural network architecture for robust detection of epileptic seizures using EEG signals. *Clin. Neurophysiol.* 130, 25–37. doi: 10.1016/j.clinph.2018.10.010
- Iranmehr, A., Shirazi, H. M., and Vasconcelos, N. (2019). Cost-sensitive support vector machines. *Neurocomputing* 343, 50–64. doi: 10.1016/j.neucom.2018.11.099
- Jiang, Y., Deng, Z., Chung, F., Wang, G., Qian, P., Choi, K. S., et al. (2017c). Recognition of epileptic EEG signals using a novel multiview TSK fuzzy system. *IEEE Trans. Fuzzy Syst.* 25, 3–20. doi: 10.1109/TFUZZ.2016.2637405
- Jiang, Y., Deng, Z., Chung, F., and Wang, S. (2015). Multi-task TSK fuzzy system modeling using inter-task correlation information. *Inform. Sci.* 298, 512–533. doi: 10.1016/j.ins.2014.12.007
- Jiang, Y., Deng, Z., Chung, F., and Wang, S. (2017b). Realizing two-view TSK fuzzy classification system by using collaborative learning. *IEEE Trans. Syst. Man Cybernet. Syst.* 47, 145–160. doi: 10.1109/TSMC.2016.2577558
- Jiang, Y., Wu, D., Deng, Z., Qian, P., Wang, J., Wang, G., et al. (2017a). Seizure classification from EEG signals using transfer learning, semi-supervised learning and TSK fuzzy system. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 2270–2284. doi: 10.1109/TNSRE.2017.2748388
- Juang, C. F., Chiu, S. H., and Shiu, S. J. (2007). Fuzzy system learned through fuzzy clustering and support vector machine for human skin color segmentation. *IEEE Trans. Syst. Man Cybernet. Part A Syst. Hum.* 37, 1077–1087. doi: 10.1109/TSMCA.2007.904579
- Kabir, E., and Zhang, Y. (2016). Epileptic seizure detection from EEG signals using logistic model trees. *Brain Inform.* 3, 93–100. doi: 10.1007/s40708-015-0030-2
- Krishnapuram, R., and Keller, J. M. (1993). A possibilistic approach to clustering. *IEEE Trans. Fuzzy Syst.* 1, 98–110. doi: 10.1109/91.227387
- Leski, J. M. (2003). Ho-Kashyap classifier with generalization control. *Pattern Recogn. Lett.* 24, 2281–2290. doi: 10.1016/S0167-8655(03)00054-0
- Leski, J. M. (2005). TSK-fuzzy modeling based on ϵ -insensitive learning. *IEEE Trans. Fuzzy Syst.* 13, 181–193. doi: 10.1109/TFUZZ.2004.840094
- Leski, J. M. (2015). Fuzzy (c+p)-means clustering and its application to a fuzzy rule-based classifier: towards good generalization and good interpretability. *IEEE Trans. Fuzzy Syst.* 23, 802–812. doi: 10.1109/TFUZZ.2014.2327995
- Li, X., Yang, H., Yan, J., Wang, X., Li, X., and Yuan, Y. (2019). Low-intensity pulsed ultrasound stimulation modulates the nonlinear dynamics of local field potentials in temporal lobe epilepsy. *Front. Neurosci.* 13:287. doi: 10.3389/fnins.2019.00287
- Liu, C. L., Xiao, B., Hsiao, W. H., and Tseng, V. S. (2019). Epileptic seizure prediction with multi-view convolutional neural networks. *IEEE Access.* 7, 170352–170361. doi: 10.1109/ACCESS.2019.2955285
- Martinez-Vargas, J. D., Strobbe, G., Vonck, K., Van Mierlo, P., and Castellanos-Dominguez, G. (2017). Improved localization of seizure onset zones using spatiotemporal constraints and time-varying source connectivity. *Front. Neurosci.* 11:156. doi: 10.3389/fnins.2017.00156
- Qi, F., Li, Y., and Wu, W. (2017). RSTFC: a novel algorithm for spatio-temporal filtering and classification of single-trial EEG. *IEEE Trans. Neural Netw. Learn. Syst.* 26, 3070–3082. doi: 10.1109/TNNLS.2015.2402694
- Razzak, I., Hameed, I. A., and Xu, G. D. (2019). Robust sparse representation and multiclass support matrix machines for the classification of motor imagery EEG signals. *IEEE J. Transl. Eng. Health Med.* 7, 2168–2372. doi: 10.1109/JTEHM.2019.2942017
- Salgado, C. M., Viegas, J. L., Azevedo, C. S., Ferreira, M. C., Vieira, S. M., and Sousa, J. M. C. (2017). Takagi-Sugeno fuzzy modeling using mixed fuzzy clustering. *IEEE Trans. Fuzzy Syst.* 25, 1417–1429. doi: 10.1109/TFUZZ.2016.2639565
- Siuly, S., and Li, Y. (2015). Designing a robust feature extraction method based on optimum allocation and principal component analysis for epileptic EEG signal classification. *Comput. Methods Prog. Biomed.* 119, 29–42. doi: 10.1016/j.cmpb.2015.01.002

FUNDING

This work was supported in part by the National Natural Science Foundation of China under Grant 61806026 and by the Natural Science Foundation of Jiangsu Province under Grant BK20180956.

- Sreej, S. R., and Samanta, D. (2019). Classification of multiclass motor imagery EEG signal using sparsity approach. *Neurocomputing* 368, 133–145. doi: 10.1016/j.neucom.2019.08.037
- Takagi, T., and Sugeno, M. (1985). Fuzzy identification of systems and its application to modeling and control. *Trans. Syst. Man Cybernet.* 15, 116–132. doi: 10.1109/TSMC.1985.6313399
- Truong, N. D., Nguyen, A., Kuhlmann, D. L., Bonyadi, M. R., Yang, J. W., Ippolito, S., et al. (2018). Integer convolutional neural network for seizure detection. *IEEE J. Emerg. Select. Top. Circuits Syst.* 8, 849–857. doi: 10.1109/JETCAS.2018.2842761
- Tzallas, A. T., Tsipouras, M. G., and Fotiadis, I. D. (2009). Epileptic seizure detection in EEGs using time-frequency analysis. *IEEE Trans. Inform. Technol. Biomed.* 13, 703–710. doi: 10.1109/TITB.2009.2017939
- Wen, T., and Zhang, Z. (2017). Effective and extensible feature extraction method using genetic algorithm-based frequency-domain feature search for epileptic EEG multiclassification. *Medicine* 96:e6879. doi: 10.1097/MD.0000000000006879
- Xia, K., Ni, T., Yin, H., and Chen, B. (2020). Cross-domain classification model with knowledge utilization maximization for recognition of epileptic EEG signals. *IEEE/ACM Trans. Comput. Biol. Bioinformatics.* doi: 10.1109/TCBB.2020.2973978. [Epub ahead of print].

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Ni, Gu and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.