The 7th Annual IEEE International Conference on
Cyber Technology in Automation, Control and Intelligent Systems
July 31-August 4, 2017, Hawaii, USA

An Approach to Minimize Calibration Time for Brain Computer Interface Based on Motion Imagery

Zou Yijun, Zhao Xingang, Xu Weiliang, Han Jianda
Shenyang Institute of Automation (SIA) ,Chinese Academy of Sciences ,Shenyang, China

***Abstract*—The long training time is a big problem that block the application of brain computer interface. This paper solve this problem using the existing dataset of many subjects. By analyzing the differences in EEG signal among different users, removing these differences as much as possible and exploits the common points of different individual, the proposed method can correct the data in existing dataset to a new dataset close to the target data. Then using the ensemble method, we combine the model of the existing dataset to a final model. The result shows that the proposed method can decrease the training time greatly while the recognize accuracy can also meet the need.**

Keywords—brain computer interface; common spatial pattern; training time; multi-subject model

I. INTRODUCTION

Brain computer interface (BCI) is a technology that allows users to interact with the computer through the brain. When brain activities occur, the brain undergoes various physiological changes: blood oxygen, temperature, electromagnetic fields, etc. These physiological changes can be observed by external devices. By analyzing these observed physiological changes, it is reasonable to predict the brain activity. This is the basic principle of brain computer interface.

In the different types of BCI, the BCI based on electrical signal has become the most widely used one in recent years because of its better time resolution and utility. However, most of today's BCI applications are still in the prototype stage, few BCI systems can be applied outside the laboratory. There are many reasons for this phenomenon, of which the most critical factors are these three: 1, the recognition accuracy is still unable to meet the needs of the actual systems. 2, the information transformation rate is limited. 3, the need for long training time.

For the problem of long training time, there are two types of approach to reduce the training time: 1, get better classification accuracy using only the small training data set through the improved algorithms. 2, use the other people's data to get better classification accuracy.

A very important approach of the first type is semi-supervised learning. Semi-supervised learning use the unlabeled data set to improve the final result of learning algorithm [1]-[3]. Jiannan Meng [3] learned an initial model from the tagged data set and uses it to classify the unlabeled data set. Then they learned a new model based on the classification results. In addition to the semi-supervised approach, some other methods have also been proposed to improve training accuracy in the case of using small training set. Fabien Lotte [4] designed a method of generating artificial signals to construct a large EEG signal set. The artificial EEG signals were created based on the existing training data and then extended the training set. Lotte's method get a much better result than the semi-supervised study, but this result is for the two types classification problem too.

Limited by the size of the training data set, the current method to reduce training time using the user's own training data cannot get a good result. Training model just using a small number of samples (10 samples per type of action) is impossible to achieve good training results. Therefore, using multi-subjects' data to improve the training effect is a necessary way.

How to use multi-subjects' data have appeal many researcher's interest. At present, BCI based on other people's data has two main directions: knowledge transfer and data fusion. Knowledge transfer refers to the transfer of knowledge from existing data to the data of new individuals [5]-[8]. In addition to knowledge transfer, data fusion is another direction of using multi-subjects' data. the fusion methods are mainly two types: data pooled methods and ensemble methods[9]-[13].

The current method either simply use the other person's data to provide a reference or constraint for the establishment of a new user's model, or establish a user-independent model directly through multi-subjects' data set. The former method has big shortage in information mining for different subjects' EEG signal, while the latter method ignores the essential differences among different individuals. To deal with these problems, this paper focuses on the analysis of the differences in EEG signal among different users, removes these differences as much as possible and exploits the common points of different individual EEG signal. In this paper, the difference between different users' EEG is divided into three aspects: 1, the mismatch of the EEG channels among different subjects, 2, the degree of ERD / ERS phenomenon 3, the frequency band that ERD / ERS phenomenon occurs. This paper focuses on eliminating the first two aspects, for the third aspect, the often used bandwidth filter can effectively reduce its impact [14]. On the other hand, as for the multi-subjects data fusion, this paper no longer use the training data set isolates, but combine the multi-subjects dataset and the target user's own data together. So the final model has both the target user's and the multi-subjects' characteristis. Secondly, this paper designs a confidence score for different users' model. The score is based on LDA's linear projection and projection variance. Then we get a final model by combining the multi-subjects' model using this confidence score.

II. METHOD

This paper presents a method to reduce the training time using the multi-subjects' data set. The key is to modify the data

in the other subjects' data set according to the target dataset. By comparing the differences between the two dataset, minimizing the differences and maximizing the commonality, we can get a better model than just using the target data. Then, when the one-to-one model is designed, the multi models is designed separately. Then we combine the multi models together. The method is shown in Figure 1.

The key points of the method are two: the data preprocessing based on the target data and the multi model confusion. The data preprocessing period can be divided into two steps: spatial filter and data correction.

A. Spatial fileter

One of the biggest differences between two person's EEG signals is that the data channels in different person are not matched. The purpose of spatial filter is adjust the channels of the other person's EEG data to the small dataset of the target person. Then the same channels in the two dataset stand for the same messages. The EEG data is represent by $\{D^{i,\ j} \in R^{C\times S}\}$, in which i=1stansfor tehe existing training set, i=2 stands for the target set, j stands for the motion type, C is the number of channels, s is sample numbers each epoch. $T \in R^{C\times C}$ is the designed spatial filters, each row of T represents a spatial filter, and each column represents a channel. After transform $D' = T \times D$ on the training set, we search for classifier based on the new dataset D'. Here, the key to find the spatial filter T is common spatial pattern (CSP) algorithm.

CSP is one of the most popular algorithms in BCI. The advantage of CSP is that CSP can maximize the differences between two types of motion and obtain the most representative channel combination (spatial filters) W for each motion. In order to match the channels of different person as much as possible, we use CSP algorithm for both dataset, and obtain CSP filters for the same motion. Then we make spatial transform $D' = T \times D$ on the training data. The goal of the transform is to minimizing the differences of the CSP filters between the new data and the target data. Then the core problem is how to get the spatial transform T.

CSP is an algorithm for two types of motion, for four types of motion, we can get six groups of spatial filters. The process of spatial transformation on the training data tis equivalent to re-sampling EEG signals on the training set, while the sample location matches the target data in greatest degree. In this process, we should guarantee than neither new information be added, nor existed information lost. Therefore, the spatial filter

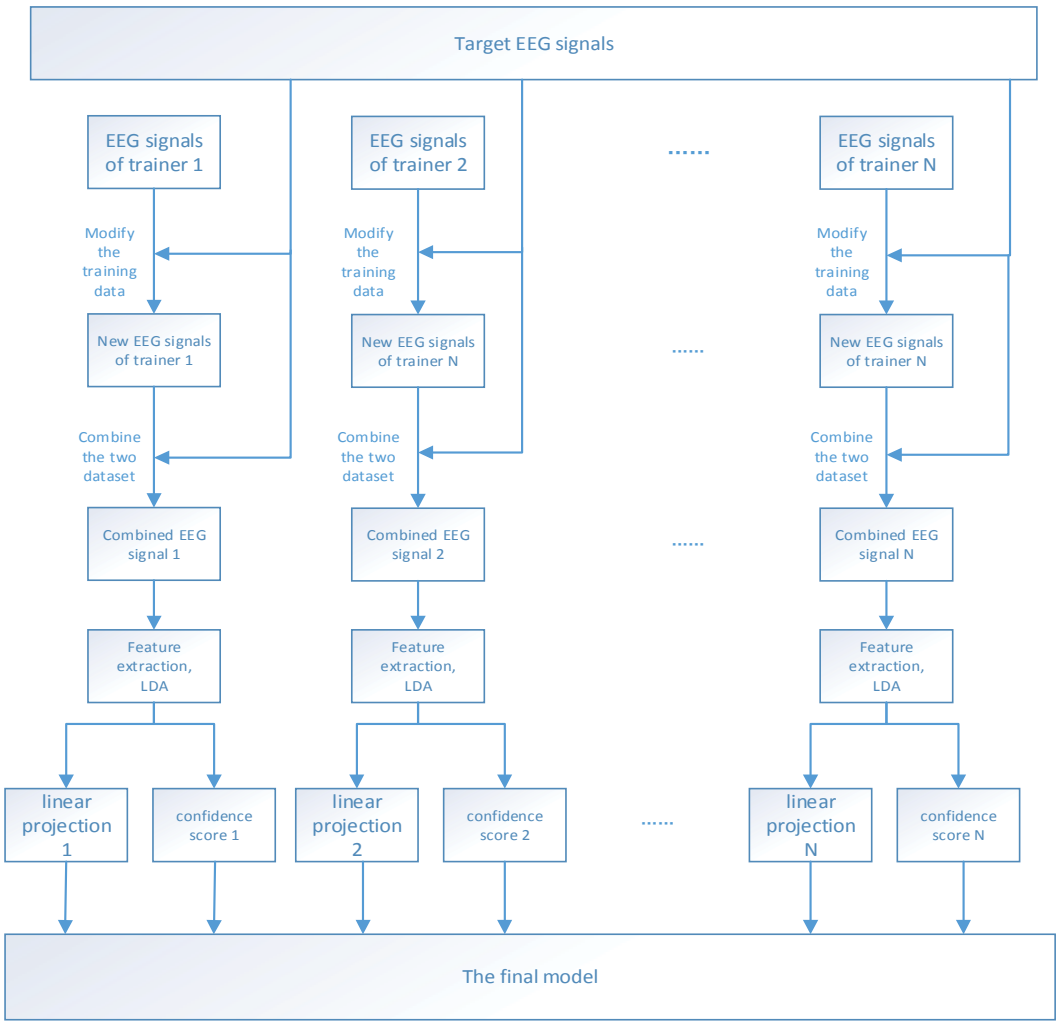

Fig.1 general view of the method

is set to an orthogonal matrix, which means that the modulo of each filter are 1, and all of them are orthogonal. The following is the detail of the calculations.

First, we calculate CSP filters for the two datasets, total twelve CSP transform matrix $\{W^{i,p} \in R^{C\times C}\}$ is calculated, in which, p = 1 ~ 6. For each dataset, we get filters group $\{W^{i} \in R^{6C\times C}\}$. To make the training group data match the target group, we need make the target group's CSP filters can be applied to the new training group data. The filtered data is $D_1' = \mathrm{T}D_1$. Then we find the CSP filters W_1' for D'. The goal is minimizing the differences between W_1' and W_2, that is, $\mathrm{T} = \mathrm{argmin}(\|W_2 - W_1'\|_F^2)$. Then we calculate W_1' for the new dataset T.

(1) Find the covariance of the two types of motion $C_i' = \sum_{k=1}^{n} \frac{D_k' \times D_k'^T}{tr(D_k' \times D_k'^T)} = \sum_{k=1}^{n} \frac{TD_k \times (TD_k)^T}{tr(D_k \times D_k^T)} = TC_iT^T$.

(2) Find the whitening matrix Q', first, $C' = C_1' + C_2' = \mathrm{TC}T^T = \mathrm{T} \times V_C \Sigma V_C^T \times T^T$ so $Q' = \Sigma^{-\frac{1}{2}} V_C^T \times T^T = \mathrm{Q}T^T$.

(3) Calculate the whitened matrix $S_i' = Q' \times C_1' \times Q'^T = QT^T \times TC_iT^T \times TQ^T = QC_iQ^T = S_i$, so S and S 'have the same eigenvector matrix U.

(4) Calculate W', $W' = U^T \times Q' = U^T \times Q \times T = W \times T^T$.

Therefore $\mathrm{T} = \mathrm{argmin}(\|W_2 - W_1 \times T^T\|_F^2), \mathrm{st}.\, T^TT = I$. However, in order to reduce the over-fitting, we need to increase the regular term. That is adding the constraint term for T according to the given information, we set the constraint as matrix P. So the final T is:

$$\mathrm{T} = \mathrm{argmin}(\|W_2 - W_1 \times T^T\|_F^2 + \alpha\|T - P\|_F^2), st.\ T^TT = I$$

The weight of the old channel i in the new lead j is the element $\mathrm{T}(j,i)$ of T, and when the training set is not processed, T is equal to the identity matrix I, which means the new channels and the original channels are the same. For the transformed channel, the closer the new channel is to the original channel, the greater the effect on the new channel. So the constraint matrix P should satisfy: The closer the channel j is to the channel i, the smaller P(i, j) should be. Here, we use exponential decay as a function of this property. Thus, $\mathrm{P(i,j)} = e^{-d(i,j)}$, where d (i, j) is the distance between channel i and channel j.

After giving the target function and the constraint matrix, the next step is to calculate T.

For the ease of calculation, firstly we transform the optimization problem with regular term to the optimization problem without regular term. Here we use the Lagrangian multiplier method.

1)First transform the target function，$\|W_2 - W_1 \times T^T\|_F^2 = \mathrm{tr}(W_1^TW_1 + W_2^TW_2) - 2 \times \mathrm{tr}(TW_1^TW_2)$，$\|T - P\|_F^2 = \mathrm{tr}\big((T-P)^T(T-P)\big) = tr(T^TT - T^TP - P^TT + P^TP) = tr(-T^TP - P^TT + I + P^TP)$， we use trace to describe the orthogonality constraints. The L function is: $\mathrm{L} = -\mathrm{tr(TA)} + \alpha tr(-T^TP - P^TT + I + P^TP) + \lambda tr\big((Q^TQ - I)^T(Q^TQ - I)\big)$, where, $\mathrm{A} = W_1^TW_2$.

2) The minimization condition is $\begin{cases} \frac{\partial L}{\partial T} = O \\ \frac{\partial L}{\partial \lambda} = 0 \end{cases}$, $\frac{\partial L}{\partial T} = -\frac{\partial}{\partial T}\big(\mathrm{tr(TA)}\big) + \alpha\frac{\partial}{\partial T}\big(tr(-T^TP - P^TT)\big) + \lambda\frac{\partial}{\partial T}\Big(tr\big((Q^TQ - I)^T(Q^TQ - I)\big)\Big) = -A^T - \alpha P - \alpha P + \lambda\frac{\partial}{\partial T}\Big(tr\big((Q^TQ - I)^T(Q^TQ - I)\big)\Big) = B^T + \lambda\frac{\partial}{\partial T}\Big(tr\big((Q^TQ - I)^T(Q^TQ - I)\big)\Big)$, where, $\mathrm{B} = (A^T + 2\alpha P)^T = A + 2\alpha P^T$

3) For problem without regular term, $\mathrm{T} = \mathrm{argmin}(\|W_2 - W_1 \times T^T\|_F^2), st.\ T^TT = I$, we have the similar situation with former, $L_1 = -\mathrm{tr(TA)} + \alpha tr(-T^TP - P^TT + I + P^TP) + \lambda tr\big((Q^TQ - I)^T(Q^TQ - I)\big)$, the minimization condition is $\begin{cases} \frac{\partial L_1}{\partial T} = O \\ \frac{\partial L_1}{\partial \lambda} = 0 \end{cases}$, $\frac{\partial L}{\partial T} = A^T + \lambda\frac{\partial}{\partial T}\Big(tr\big((Q^TQ - I)^T(Q^TQ - I)\big)\Big)$.

4) By comparing the formulas in (2) and (3), the optimization problem of the regular term can be transformed into the optimization problem without the regular term, $\mathrm{T} = \arg\max\big(tr(T\mathrm{B})\big), st.\, T^TT = I$, where, $\mathrm{B} = A + 2\alpha P^T$.

Then we solve the new problem:

1) SVD decompose matrix B：$\mathrm{B} = U\Sigma V^T$, $\mathrm{tr}(TB) = tr(TU\Sigma V^T) = tr(V^TTU\Sigma)$. Set $\mathrm{Z} = V^TTU$, then Z is a orthogonal matrix, so $|z_{ij}| \le 1, \forall i,j$.

2) Then$tr(V^TTU\Sigma) = tr(Z\Sigma) = z_{11}\sigma_1 + z_{22}\sigma_2 + \cdots + z_{nn}\sigma_n \le \sigma_1 + \sigma_2 + \cdots + \sigma_n$, the equal sign when and when$\mathrm{Z} = \mathrm{I}$, which means$V^TTU = I$, so $\mathrm{T} = \mathrm{V}U^T$.

3) So, the final spatial filter is: $\mathrm{T} = \mathrm{V}U^T$, where V、U is the result of SVD decomposition on B, that is, $V^TBU = W_1^TW_2 + 2\alpha P$.

Another problem is that the different row vector of W have different confidence. As a spatial filter, its role is to maximize the distinction between the two types of motion, the first few rows and the last few rows have the largest degree of distinction, while the middle part have the smallest degree of distinction. Therefore, when calculating the transformation T, the filter bank W must be weighted. The weighting factor is designed using an exponential structure $\mathrm{L} = \begin{cases} l^{-c} \\ l^{c-C/2} \end{cases}$ (C is the number of rows, l is the weight of the base).

B. Data correction

The differences in the ERD phenomenon of different people are mainly in three aspects: 1, the difference of the strongest position that ERD/ERS occurs. 2, the frequency band difference of ERD/ERS. 3, the difference of the degree of ERD.

For the third difference, we use the following method to decrease its effect. The specific method is as follows:

(1) Extract the EEG signals with the same motion from the training set and the target set and use CSP to calculate the spatial filter W and matrix eigenvalue vector v. W can map the signal D to a space where the data set with the eigenvalue close to 1 have the max variance (the energy of 8 ~ 30Hz is the largest), and the data with the eigenvalue close to 0 have the min variance.

(2) Calculate the correction coefficient $e = (v - 0.5)^p$ based on the eigenvalue and calculate the correction matrix $E = I + diag(e)$.

(3) Map the training set to the space $W \times D^1$. In this space, we can correct the data through the correction matrix directly and then map the data to the original space through the inverse transformation of W to return to the original space, that is, $D_0^1 = W^{-1} \times E \times W \times D^1$.

Then we correct all the four motion's data and get a new data set for training.

C. *The ensemble method*

The method described above is all data preprocess method, after applying these method to each subjects' data in the train dataset (there are many subjects' data in the train set). After preprocessing the raw data, we can use other people's data to create a target-based classification model, but the model is one-to-one, that is, using one person's model to identify another person's motion. The following problem is how to use these data to get a good final model. Here, we design of a new ensemble method to ensemble the different subjects models will be more than one model of integration. The feature extraction method is CSP, and the classifier is LDA.

When using LDA to classify the four types of motion, we need to find a projection direction that all the feature points in the projection direction have the largest intra-class dispersion and the smallest inter-class dispersion. If the linear projection of the feature point in the projection direction is positive, the output of the classifier is category 1. If the linear projection of the feature point in the projection direction is negative, the output of the classifier is class 2. And the absolute value of the linear

	No method	Using another day's data	Data pooled method	Semi-supervised	DLRCSP	WTRCSP	The proposed method
01 E	0.2734	0.7143	0.5000	0.3320	0.2969	0.2994	0.6602
01 T	0.2109	0.7455	0.4514	0.2344	0.3080	0.3770	0.6992
02 E	0.2500	0.4375	0.2813	0.2188	0.1953	0.2683	0.4648
02 T	0.2063	0.4241	0.2569	0.2070	0.2617	0.3728	0.4336
03 E	0.3555	0.7768	0.5833	0.3242	0.3750	0.4973	0.6797
03 T	0.2422	0.7679	0.5833	0.1914	0.1406	0.2035	0.7422
04 E	0.2578	0.5402	0.4132	0.2813	0.2500	0.3395	0.5000
04 T	0.2383	0.5000	0.3993	0.2695	0.2266	0.2885	0.4219
05 E	0.2695	0,3750	0.2604	0.1992	0.2070	0.2750	0.3438
05 T	0.2578	0,4018	0.2569	0.2188	0.2617	0.3245	0.3164
06 E	0.2344	0.3705	0.2813	0.2070	0.2969	0.4037	0.3516
06 T	0.2852	0,3929	0.2465	0.3008	0.2500	0.3021	0.3242
07 E	0.1133	0.7455	0.4568	0.3438	0.1836	0.2173	0.7305
07 T	0.2070	0.7054	0.5674	0.2422	0.1211	0.1412	0.6680
08 E	0.2422	0.7277	0.6181	0.2617	0.1914	0.1899	0.7617
08 T	0.1953	0.7500	0.5694	0.2969	0.1523	0.2268	0.7422
09 E	0.2188	0.7321	0.5451	0.2930	0.1875	0.2503	0.6680
09 T	0,2734	0.6652	0.5313	0.2148	0.2734	0.3404	0.5898
average	0.2406	0.6401	0.4334	0.2576	0.2322	0.2960	0.5609

Tabel.1 Different methods use 32 data points when the identification results

projection is actually related to the confidence level of the classification output.

In the muti-subjects model, each person has a LDA. We add the linear projection of different person's LDA. But different LDA classifier have different confidence for the final result. So we design a confidence score for each LDA model to weight the different LDA models when fusing them. The confidence score is inversely proportional to the intra-class dispersion after LDA projection, and is proportional to the degree of discrimination of the model itself. In order to assess each LDA, we get every LDA model's classification accuracy on its train data using cross validation.

The confidence score for the i-th classifier is: $m_i = p_i\left(w_i^T S_{W,i} w_i\right)$, Where p_i is the kappa coefficient for cross validation results for the i-th person's own data, w_i is the linear projection vector of the i-th LDA classifier, $S_{W,i}$ is the intra-class dispersion of the i-th LDA classifier.

Then we can get the weighted sum of the linear projection of every model: $L = \sum_{i=1}^{N} m_i l_i$, where l_i is the linear projection of the i-th LDA classifier. Finally, when L is greater than zero, the output of the classifier is class 1, and when L is less than zero, the output of the classifier is class 2.

III. RESULT

A. EEG data set

The data set we use is the data set 2a in the BCI Contest IV. This data set is provided by the Graz task group. The data set contains nine users' data. Each user performs two experiments in two days. Each experiment carries out the motion imagery of the four types of motion (the left hand, the right hand, the tongue and the foot). The user perform each motion 72 times, a total of 288 motion. At the beginning of each test, the motion hint appear on the screen, the user starts to imagine the motion, and the hint stops after 4s. The collected EEG data contained 22 channels.

B. Experiments based on a small amount of data

In this section, the data for each dataset is divided into two parts: training data and validation data. Unlike conventional training data, here, only a small amount of data is used as training data. Specifically, the training data set has a total of 32 data points (8 data points each motion). The rest 256 data points is the test data set. In each point we select the data from 3s to 5s. It means that only less than 5 minutes of training data is used.

For comparison, we choose the following methods: 1, data pool method, as a representative of the data fusion method. 2, semi-supervised learning, the semi-supervised learning method used here refers to the method of Jianjun Meng in the paper. 3, the regularization of the CSP method. We select two kinds of RCSP. One is DL_RCSP mentioned in Fabien Lotte's paper, which is a method that does not rely on the other person data. It just choose the identity matrix as a regular term. The other is WTRCSP mentioned in the paper, according to the conclusions of the paper, WTRCSP get the best results among many RCSP

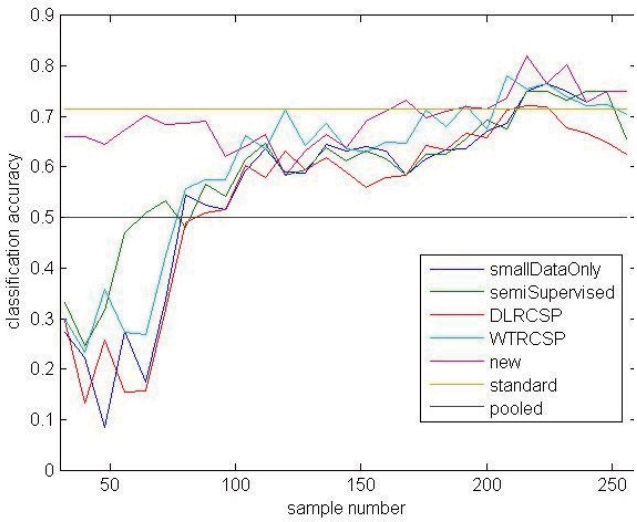

Fig. 2 Different methods to identify the accuracy of the number of sampling points with the use of changes

algorithm. Also we compare the result of the proposed method with no method. The other comparison method is to classify the test set using the subjects' another day's data.

It can be seen from Table 1 that in the case of using only a small amount of data, that is, the training samples composed of 32 data are used to classify the test set of 256 data remaining, and the average classification accuracy of 9 individuals is only 24.06%, even lower than random. That means only 32 only the data can not get any classification effect. The last column of Table 1 is the result of using the proposed method, the average recognition accuracy can reach 56.09%, as a comparison, the second column gives the results using the same subjects' another day's data (for example, using 01T data to recognize 01T). The average recognition accuracy is 64.01%. It can be seen that the average result of using this method is only 8% lower than the comparison result, while the data used is only one ninth of the latter. In some situation, the recognition result is even better than the comparative method.

To further illustrate the effectiveness of the proposed method in reducing training time, Figure 1 shows the results of classifying the remaining data points using different numbers of training data points. The number is from 32 to 256. It can be seen that, using only its own data, the recognition accuracy is gradually increased from very low until 0.7. With the use of the proposed method, the initial accuracy can reach 66.02%, and the final accuracy is more than the result of using another day's data. In comparison, the other method seems to be better than using no method, but the improvement is very limited.

In addition, in order to demonstrate the effectiveness of the proposed method, the changes of the CSP filter of the corrected data are given in Fig. 3, and there are six groups of CSP filters , The first column is the original data, the third column is the corrected target, and the second column is the data after preprocessing. It can be seen that the train set's CSP filter is made closer to the target set's CSP filter.

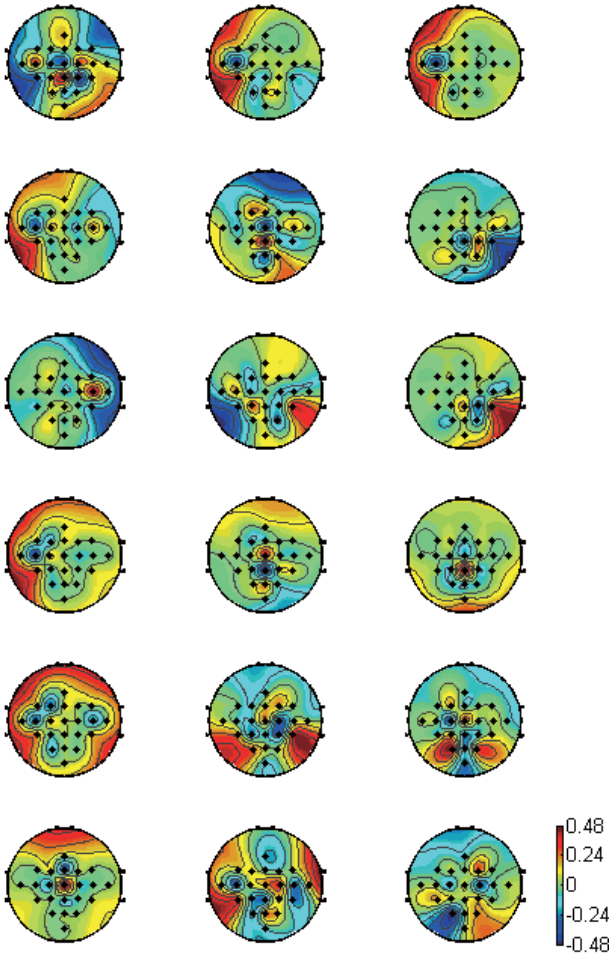

Fig.3 the CSP filters of data after the data preprocessing. There are total six lines, respectively, stands for the six combination of all four types' motion. There are total three column. The first column stands for the train set data, the third column stands for the target data, the second column stands for the data of train set after preprocessing.

IV. CONCLUSION

In this paper, we propose a method to improve the action recognition effect by using multiplayer data in the brain-computer interface based on motion imagination. There are two core points of the method: data preprocessing based on target data and fusion of multiplayer data The From the above experimental results can be seen, this method can greatly reduce the training time. The idea of this article is to introduce multiplayer data, after a lot of people have studied the method based on multiplayer data. The difference between this article is: 1, the other person data and my data combined use, so that the training model has its own characteristics, but also combines the characteristics of many people, rather than blindly pursue the establishment of user-independent model. 2, focusing on different users between the EEG differences and common ground, and try to remove the difference through the algorithm, directly extract the common points of different users, in order to improve the action recognition effect.

REFERENCES

[1] Jayaram, Vinay, et al. "Transfer learning in brain-computer interfaces." IEEE Computational Intelligence Magazine 11.1 (2016): 20-31.

[2] Tu, Wenting, and Shiliang Sun. "Semi-supervised feature extraction for EEG classification." Pattern Analysis and Applications 16.2 (2013): 213-222.

[3] Meng, Jianjun, et al. "Improved Semisupervised Adaptation for a Small Training Dataset in the Brain–Computer Interface." IEEE journal of biomedical and health informatics 18.4 (2014): 1461-1472.

[4] Lotte, Fabien. "Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain–computer interfaces." Proceedings of the IEEE 103.6 (2015): 871-890.

[5] Lotte, Fabien, and Cuntai Guan. "Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms." IEEE Transactions on biomedical Engineering 58.2 (2011): 355-362.

[6] Arvaneh, Mahnaz, Ian Robertson, and Tomas E. Ward. "Subject-to-subject adaptation to reduce calibration time in motor imagery-based brain-computer interface." Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE. IEEE, 2014.

[7] Lotte, Fabien, and Cuntai Guan. "Learning from other subjects helps reducing brain-computer interface calibration time." Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on. IEEE, 2010.

[8] Heger, Dominic, et al. "Subject-to-subject transfer for CSP based BCIs: Feature space transformation and decision-level fusion." Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE. IEEE, 2013.

[9] Krauledat, Matthias, et al. "Towards zero training for brain-computer interfacing." PloS one 3.8 (2008): e2967.

[10] Dalhoumi, Sami, Gérard Dray, and Jacky Montmain. "Knowledge transfer for reducing calibration time in brain-computer interfacing." Tools with Artificial Intelligence (ICTAI), 2014 IEEE 26th International Conference on. IEEE, 2014.

[11] Fazli, Siamac, et al. "Subject-independent mental state classification in single trials." Neural networks 22.9 (2009): 1305-1312.(ensemble)

[12] Tu, Wenting, and Shiliang Sun. "A subject transfer framework for EEG classification." Neurocomputing 82 (2012): 109-116.

[13] Reuderink, Boris, et al. "A subject-independent brain-computer interface based on smoothed, second-order baselining." Engineering in medicine and biology society, EMBC, 2011 annual international conference of the IEEE. IEEE, 2011.

[14] Pfurtscheller, Gert, and FH Lopes Da Silva. "Event-related EEG/MEG synchronization and desynchronization: basic principles." Clinical neurophysiology 110.11 (1999): 1842-1857.