Comparative study on motor imagery classification based on BP-NN and SVM

Hongru Jia\(^1,2\), Shuai Wang\(^3\), Dezhi Zheng\(^1,2\), Xiaolei Qu\(^1,2\), Shangchun Fan\(^1,2\)

\(^1\)School of Instrumentation Science and Opto-electronics Engineering, Beihang University, Beijing 100191, People’s Republic of China
\(^2\)School of Computer Science and Engineering, Beihang University, Beijing 100191, People’s Republic of China
\(^3\)Beijing Advanced Innovation Centre for Big Data-based Precision Medicine, Beihang University, Beijing 100191, People’s Republic of China

Abstract: Motor imagery (MI) classification based on electroencephalogram (EEG) with high speed and accuracy is a key issue in brain–computer interface (BCI) technology. This study compared the support vector machine (SVM) and back-propagation neural network (BP-NN) for MI classification. In this study, EEG data of four subjects provided by BCI competition 2008 was employed. For the comparison of classification accuracy (CA), there were three steps. First, EEG feature extraction for MI was implemented by using a common spatial pattern. Second, SVM and BP-NN were used to classify MI by cross-validation. Finally, the CA rate, receiver operating characteristic curve and area under the curve (AUC) were given to evaluate two classifiers. The average CA rates obtained on the four subjects using SVM and BP-NN were 75.20 and 80.73\%, respectively. Furthermore, the mean AUCs of SVM and BP-NN were 0.7860 and 0.9462, respectively. Both average CA rate and AUC indicate that BP-NN has better accuracy of classification than SVM.

1 Introduction

During the past decades, brain–computer interface (BCI) has drawn increasing attention as a technology that can directly translate the thoughts in brains to readable messages and commands. BCI emerges as a useful communication and control technology for severely paralysed people [1], advanced control systems, entertainments etc. One of the most important tasks of a BCI system is to distinguish the electroencephalogram (EEG) signal including EEG feature extraction and pattern classification. Pattern classification of the EEG signal is considered as the key technique in decoding the brain activities. Since the accuracy and the calculation speed of classifying the EEG signal are of high importance concerning the BCI system, great efforts have been made in finding appropriate pattern classification algorithms for years.

Following the development of pattern classification algorithms, many different methods have been studied to classify the EEG signal. There are four different categories for classification algorithms in traditional machine learning: linear classifiers, non-linear classifiers, nearest neighbour classifiers and neural networks (NNs) [2]. The Bayesian network is one of the non-linear classifiers and it has a high precision in solving discrete problems [3]. However, information may be lost when dealing with a continuous signal such as EEG when using the algorithm. The k-nearest neighbour (KNN) algorithm belongs to nearest neighbour classifiers. KNN can approximate any function which enables it to produce non-linear decision boundaries because it has a high value of k [4]. However, the KNN algorithms are known to be sensitive to the large dimension difference of sample data. Both non-linear classifiers and nearest neighbour classifiers are not as widespread as linear classifiers or NNs in BCI applications. To solve the classification problems of small sample data and non-linearity, support vector machine (SVM) is a good choice to be used for the research of motor imagery (MI) classification with its good robustness and high efficiency [5]. Li used SVM to classify the emotion recognition from the EEG signal which was extracted by power spectrum density [6]. Another algorithm usually used in the studies of MI classification recently is the artificial NN (ANN) [7]. Jin used Hilbert–Huang transform and back-propagation NN (BP-NN) to solve the problems of MI classification [8]. By analysing the potential rules between the input and output data provided in advance, the ANN algorithms get the characteristics of self-organisation, self-adaption and self-learning. There are three major NN models: feed-forward NN, feedback NN and self-organisation NN. The BP-NN is a kind of the feed-forward NN. It has the ability of generalisation and fault tolerance. Mostly, in the actual application situation, both the accuracy and processing speed affect the performance of the BCI system directly, so it is necessary to analyse the classification algorithms from these two angles at the same time. However, at present, specific research about the accuracy and efficiency of two commonly used classification algorithm of SVM and BP-NN is seldom considered.

In this work, the left and right hands MI classification were studied for comparison analysis of the SVM algorithm and the BP-NN algorithm. Common spatial pattern (CSP) was used for the EEG feature extraction. After that, the classification accuracy (CA) rate, receiver operating characteristic (ROC) curve and area under the curve (AUC) served as the primary performance measures of the CA. At last, the efficiency of the two algorithms was also analysed.

2 Theories

2.1 BP neural network

Fig. 1 shows a classical model of an artificial neurone, which is the basic processing unit of a NN. It can be defined as

$$\frac{d w_i}{dt} = -u_i(t) + \sum w_j x_j(t) - \theta_i$$

(1)

$$y_i(t) = f[u_i(t)]$$

(2)

where $w_i$ is the network weight; $x_j$ is the input of a neurone; and $y_i$ is the output of a neurone [9].

An ANN is a network that consists of such artificial neurones and their connections. BP-NN can theoretically approximate any non-linear continuous function with the condition of the reasonable structure and appropriate weights. Error gradient descent algorithm is utilised to minimise the mean square error between the output values of the network and the actual outputs. The structure of a BP-NN is shown in Fig. 2. A BP-NN includes three or more layers including the input layer, the hidden layer and the output layer.
When the classification \( x = \exp + 1, -1 \), \( x = + = \text{sgn} x \) where \( \omega - 1 - 1 \geq 0 \), \( + \sum J. \text{Eng.} \)

According to the sequential calculation along the path – from the input layer to the hidden layer to the output layer. (ii) Use error hyperplane described as

SVM is a pattern recognition method based on statistical learning theory. The basic idea of the algorithm is to map the input space in two categories and get the largest intervals of them. With the hyperplane in the space. The optimal hyperplane can separate the samples collection can be linearly classified as (3) Data acquisition

The learning process of the BP-NN algorithm contains two steps. (i) The output values \( y_i \) of the network are acquired according to the sequential calculation along the path – from the input layer to the hidden layer to the output layer. (ii) Use error gradient descent algorithm to minimise the mean square error between the output values of network and the actual outputs, and adjust the network weights along the backward path – from the output layer to the input layer.

2.2 Support vector machine

SVM is a pattern recognition method based on statistical learning theory. The basic idea of the algorithm is to map the input space in a high-dimensional space through the non-linear transformation defined by the inner product. Then solve the generalised optimal hyperplane through the non-linear transformation between the output values of network and the actual outputs, and adjust the network weights along the backward path – from the output layer to the input layer.

By using the Lagrange function and the Kuhn–Tucker condition, the SVM can be obtained as

\[
f(x) = \text{sgn} \left( \sum_{i=1}^{n} a_i y_i x_i + b \right)
\]

Different inner product functions will lead to different SVM algorithms. A non-linear kernel function is utilised to solve the problem of MI classification between Gaussian kernel and polynomial kernel [11]. Owing to the poor performance of the polynomial kernel, only the following Gaussian kernel function was used in the experiments:

\[
K(x, x) = \exp \left( -\frac{||x - x||^2}{\sigma^2} \right)
\]

3 Methods

3.1 Data acquisition

The dataset used in the experiment was obtained from the data sets 1 of BCI competition 2008 which is provided by the Berlin BCI group from the Machine Learning Laboratory of Berlin Institute of Technology. Four sets of data of subjects B, C, E and G were used in the experiment. The four subjects were instructed to imagine moving the left and right hands in accordance with the direction arrow displayed on a computer monitor. Fig. 3 shows the sequence diagram of a single trial, with a total length of 8 s. A fixation cross appeared at the centre of the screen for 2 s to alert the subjects. Then, cues were displayed for a period of 4 s during which the subjects were instructed to perform the MI task according to the direction arrow. After that, the blank screen appeared for 2 s. Every set of data consists of 200 trials with an equal time of left and right hands imagery movement [12]. After every 15 trials, a break of 15 s was given for relaxation. Signals were sampled at 100 Hz and were collected by 59 EEG channels.

3.2 Signal pre-processing and feature extraction

In MI-based BCI systems, the imagination of body movement results in oscillations called event-related synchronisation and event-related desynchronisation in the sensorimotor cortex in the \( \mu (8–12 \text{ Hz}) \) and \( \beta (18–25 \text{ Hz}) \) frequency bands [13]. According to this, 2–26 Hz bandpass filter was chosen to filter the data first in the experiment.

Then, the CSP algorithm was chosen to extract the features of the EEG signal because CSP is suitable for the analysis of multi-channel EEG signal. The basic principle of CSP is that it finds optimal spatial filters to maximise the ratio of average variances belonging to two different classes [14]. Noted the original EEG signal of a trial is \( E_{i, t, T} \), in which \( N \) is the number of the EEG channels and \( T \) is the sampling points of a single training. Both left and right hands imagery movement were sampled for \( n \) trials in each case. Moreover, the average covariances of the trials of left and right hands imagery movement \( C_L \), \( C_R \) can be calculated. Then, define the covariance of the mixed space and diagonalise the covariance matrix. Therefore, the whitening matrix \( P \) is defined by the eigenvector matrix \( U_c \) and the eigenvalue diagonal matrix \( A_c \)

\[
P = A_c^{-1/2} U_c^T
\]

Then the covariance matrices \( C_L \) and \( C_R \) can be transformed to

\[
S_{L, R} = P C_{L, R} P^T
\]

After the whitening, \( S_L \) and \( S_R \) have the same eigenvector and can be diagonalised as

\[
S_{L, R} = B A_{L, R} B^T
\]

Thus, the spatial filter \( W \) can be obtained as
\[ W = P^2 B \]  

Then, the characteristics \( Z_L, Z_R \) can be obtained after the application of \( W \) to filter the original EEG signal

\[ Z_{L,R} = W \times E_{L,R} \]  

As a result, \( f_L \) and \( f_R \) are chosen as the feature vectors of left and right hands MI, with the definition as

\[ f_{L,R} = \frac{\text{VAR}(Z_{L,R})}{\text{sum}(\text{diag}(\text{VAR}(Z_{L,R})))} \]  

At last, seven channels from the experimental data \((C_1, C_2, C_3, C_4, C_5, C_{10}, C_2)\) which are most relevant to the MI were selected for feature extraction using the above CSP algorithm. Moreover, the results of \( f_{L,R}^{1,100} \) and \( f_{L,R}^{7,100} \) were obtained.

### 3.3 MI classification

The aim of BCI classification is to distinguish EEG trials to the classes of the associated mental tasks. The data extracted with CSP were classified by SVM and BP-NN.

To train classifiers with the training data effectively and reasonably, cross-validation was adopted to avoid the occasional events and to get the best results. In the experiment, the ten-fold cross-validation method was used to disorganise the training data randomly and divided the data into ten parts. Each part was taken as the test data and the other nine were taken as training data. Each trial got a CA rate after the calculation. Finally, the average value of the ten results was estimated as the accuracy of the SVM classification algorithm. In addition, the selection of kernel function is very important in the SVM algorithm. It was found that using the Gaussian kernel is superior to other kernel functions in the SVM classification of left and right hands MI.

The design of the structure of BP-NN is a very important part of the BP-NN algorithm. With seven features obtained by the CSP algorithm, the number of neurones in the input layer is seven. The category in the experiment of MI is of two kinds. In general, when there are \( m \) classes, the number of neurones in the input layer is \( \log m \). As a result, there was one neurone in the output layer. When the output is ‘1’, it means the imagery movement of the left hand, and the output of ‘−1’ means the imagery movement of the right hand. As for the design of the hidden layers, it is simpler and faster to improve the accuracy of the network learning by adding the neurones in the hidden layer than increasing the numbers of hidden layers. Therefore, a single hidden layer was chosen in the experiment. The choice of the optimal hidden layer satisfies (14), and \( N \) represents the number of neurones in the hidden layer

\[ N = \sqrt{n + m + a} \]  

in which \( m \) is the number of output neurones, \( n \) is the number of input neurones and \( a \) is the constant between [1,10]. According to (14), \( m \) is 2, \( n \) is 7 and the neurones in the hidden layer are between [4, 13]. After a lot of trials, the comprehensive effect of network error and running time was the best when the number of neurones is 13.

The training and testing steps of BP-NN are as follows:

(i) **Data normalisation**: The purpose of the data normalisation is to avoid the dimensions of input data having differences in the order of magnitudes, and the data were normalised to \([-1,1]\).

(ii) **Training data**: Ten-fold cross-validation method was selected in the BP-NN classification. The training parameters were set up as follows: The maximum frequency of training was set to 100, the network learning rate was set to 0.01 and the accuracy of the training request was set to 0.001. The transfer function of neurones in the hidden layer was S-type tangent function.

### 3.4 Evaluation of classifiers

In this work, the CA rate, ROC curve and AUC were used to evaluate the classification models. The CA rate refers to the percentage of all correctly predictive categories, with the definition as

\[ \text{CA} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \]

where TP means true positive, TN means true negative, FP means false positive and FN means false negative. It conveys information more intuitive and it is easy to be understood [15]. Moreover, the ROC curve and AUC can reflect the error types of classifiers and the potential distribution of the test data, which make the evaluation more reliable. The ROC curve can sort the samples according to the predictive results of the classifier. The curve is drawn by using the value of a TP rate as ordinate and the value of an FP rate as abscissa each time. When the AUC is larger, the ROC curve is closer to the top-left corner, the classifier performance is better.

### 4 Results and discussion

SVM and BP-NN were evaluated, respectively, in two aspects: the accuracy and the speed of classification. CA rate was first adopted for the evaluation. Since the ten-fold cross-validation method was used in two classification models, the average CA of ten results and the best CA from ten results are received. Fig. 4 shows the average CA and the best CA of four sets of data of SVM and BP-NN. The green left bars are the average CA of SVM classification and the blue right bars are the average CA of BP-NN classification.

Through Fig. 4, it is very intuitive to see that both the average CA and the best CA of four sets of data of BP-NN are better than that of SVM. The mean results of average CA rates obtained on four subjects of SVM and BP-NN classifiers are 75.20 and 80.73%, respectively. The mean results of best CA rates obtained on four subjects of SVM and BP-NN classifiers are 88.80 and 91.6%.

Then, the ROC curve and AUC were adopted to evaluate two classifiers. Taking the data of subject B as an example, the ROC curves of the two classifiers are shown in Fig. 5a. It shows that the ROC curve of BP-NN is closer to the top-left corner than the curve of SVM based on the results, which represents that BP-NN has a better CA than SVM. The other three sets of data have the same trend. Fig. 5b shows the AUC results of the four subjects by two classifiers and the AUCs of BP-NN are all larger than the results of SVM. The mean AUCs obtained on four subjects of SVM and BP-NN are 0.7860 and 0.9462, respectively. The results reflect that the CA of BP-NN is more dominant than that of SVM.

As for the comparison of classification speed of two classifiers, the timer is set at the programme of training data. To ensure the accuracy, the time of ten times classifications of the two classifiers
are counted. The average of the four sets of data of ten times classification of the SVM classifier is only 0.012 s, while ten times classification of BP-NN classifier costs 3.685 s. The classification speed of SVM is much faster than that of BP-NN.

Compared to the BP-NN classifier, both advantages of training error and extension ability are reflected in the SVM classifier. SVM is more suitable for solving problems of a small sample and high-dimensional patterns. However, the SVM algorithm is difficult to be utilised to train large samples. Since it is a binary classification method, SVM needs to be converted for solving the multi-classification situation. On the other hand, the ANN algorithm simulates the human brain in terms of its structure and function, so it reflects some basic characteristics of human brain function. BP-NN has the ability of self-study. The disadvantage of BP-NN is that the gradient descent method it used will lead to slow training speed. In the process of training, the situation of overlearning or underlearning will occur.

5 Conclusion

In conclusion, both the CA and the efficiency of SVM and BP-NN were studied comparatively for application of left and right hands MI, of which the EEG signal features were extracted by CSP. It is found that the CA is improved when using BP-NN under the evaluation of average CA rate, ROC and AUC than SVM. The classification speed of the SVM algorithm is much faster than BP-NN. It is hoped that in future research, the advantages of the two classification algorithms can be combined. By applying the results of SVM to the training process of BP-NN algorithm, both the CA and the classification speed can be improved, which can make progress in the performance of practical BCI systems. The accuracy and the speed of classification can be improved for actual and meaningful situation simultaneously.

6 Acknowledgments

The work was supported by The National Youth Talent Support Program held by the Beihang University. For this research, we thank the innovation team of the ministry of education for advanced sensing technology at the Beihang University. More importantly, I would thank my supervisor Dr. Zheng and Dr. Wang for their trust and support.

7 References

[1] Yuksel, A., Olmez, T.: ‘A neural network-based optimal spatial filter design method for motor imagery classification’, PLoS One, 2015, 10, (5), p. e0125039
[2] Lotte, F., Congedo, M., Lécuyer, A., et al.: ‘A review of classification algorithms for EEG-based brain–computer interfaces’, J. Neural Eng., 2007, 4, (2), p. R1
[3] Liu, B., Luo, C., Wei, M.R., et al.: ’Motor imagery EEG signal analysis based on Bayesian network’, Comput. Eng., 2014, 40, (7), pp. 128–132
[4] Blankertz, B., Curio, G., Müller, K.R.: ’Classifying single trial EEG: towards brain computer interfacing’, Advances in Neural Information Processing Systems, Vancouver, Canada, 2002
[5] Cristianini, N., Shawe-Taylor, J.: ‘An introduction to support vector machines: and other kernel-based learning methods’ (Cambridge University Press, 2000)
[6] Youjun, L.I., Ning, Z., Jiajin, H., et al.: ’Human emotion multi-classification recognition based on the EEG time and frequency features by using a Gaussian kernel function SVM’, J. Beijing Univ. Technol., 2018, 44, (2), pp. 80–89
[7] Yu, F., Xu, X.: ‘A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network’, Appl. Energy, 2014, 134, (134), pp. 102–113
[8] Jin, H., Zhang, Z.: ‘Research of movement imagery EEG based on Hilbert–Huang transform and BP neural network’, J. Biomed. Eng., 2013, 30, (2), pp. 249–253
[9] Zhang, A.H., Zhang, X.W.: ’Recognition of mental task of pressing key about left-right hand based on BP neural networks’, China Med. Eng., 2007, 15, (3), pp. 239–241
[10] Wang, Y., Qiao, X., Li, P., et al.: ’Classification of motor imagery task based on wavelet packet entropy and support vector machines’, Chin. J. Sci. Instrum., 2010, 31, (12), pp. 2729–2735
[11] Shawe-Taylor, J., Cristianini, N.: ’Kernel method for pattern analysis’, J. Am. Stat. Assoc., 2004, 101, (467), p. 1730
[12] Tangermann, M., Müller, K.R., Aertsen, A.: ‘Review of the BCI competition IV’, Frontiers in Neuroscience, 2012, 6, (55), p. 55
[13] Pfurtscheller, G., Lopes, D.S.F.H.: ‘Event-related EEG/MEG synchronization and desynchronization: basic principles’, Clin. Neurophysiol., 1999, 110, (11), pp. 1842–1857
[14] Chong, L., Zhao, H.B., Li, C.S., et al.: ’CSP/SVM-based EEG classification of imagined hand movements’, J. Northeastern Univ., 2010, 31, (8), pp. 1098–1101
[15] Herman, P., Prasad, G., McGinnity, T.M., et al.: ‘Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification’, IEEE Trans. Neural Syst. Rehabil. Eng., 2008, 16, (4), pp. 317–326

Fig. 5 Evaluation of ROC curve and AUC
(a) ROC curve of subject B. The upper line is the result of BP-NN and the lower line is the result of SVM. (b) AUC of four subjects. The green left bars are the results of SVM classification and the blue right bars are the results of BP-NN classification