A data partitioning method for increasing ensemble diversity of an eSVM-based P300 speller

Yu-Ri Lee, Hyoung-Nam Kim

Department of Electronics Engineering, Pusan National University, Busan, Republic of Korea

ABSTRACT

A P300 speller is a device for typing words by analysing the electroencephalogram (EEG) caused by visual stimuli. Among classifying methods used for the P300 speller, the ensemble of support vector machines (eSVM) is well known for achieving considerable classification accuracy. The eSVM is composed of linear support vector machines trained by each small part of the divided training data. To obtain an ensemble model with good accuracy, it is generally important that each classifier be as accurate and diverse as possible; diverse classifiers have different errors on a dataset. However, the conventional eSVM considers only an accuracy viewpoint of an individual classifier by clustering the homogeneous training data with similar noisy components. With such a viewpoint of diversity, we propose a dataset manipulation method to divide a training dataset into several groups with different characteristics for training each classifier. We reveal that the distance between a letter on which a subject is concentrating, and an intensified line on a visual keyboard, can generate EEG signals with different characteristics in a P300 speller. Based on this property, we partition the training data into groups with the same distance. If each individual SVM is trained using each of these groups, the trained classifiers have the increased diversity. The experimental results of a P300 speller show that the proposed eSVM with higher diversity improves the letter typing speed of the P300 speller. Specifically, the proposed method shows an average of 70% accuracy (verbal communication with the Language Support Program is possible at that level) by repeating the dataset for a single letter only four times.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

A P300 speller is a popular brain computer interface (BCI) system, which enables people to spell a text on a computer through visual stimulus [1]. A P300 speller uses a P300 component, which has a positive peak at nearly 300 ms after stimulus, because the P300 component reflects a higher response by stimulation than other components. When the P300 wave is detected for a letter, the repeatedly obtained electroencephalogram (EEG) signals are averaged to increase the signal-to-noise ratio (SNR) [2]. However, the measured EEG signals are mixed with other bio-signals such as electrooculography and electromyography. Because these unexpected signals cannot be approximated by a zero-mean Gaussian random process, it is difficult to detect P300 signals by simply averaging EEG signals. To detect P300 signals efficiently, several classification methods, such as an artificial neural network, linear discriminant analysis, and support vector machine (SVM), have been introduced in a P300 speller [3–5].

To encourage further development of signal processing and classification methods for the BCI, BCI Competition III [6] was held and P300 speller data were included in the competition. The competitors proposed various algorithms for a P300 speller to estimate correct letters using training and test data, without true letter information. Among the proposed algorithms, the algorithm using an ensemble of SVMs (eSVM) exhibited the best performance in terms of estimation accuracy [7]. It was shown that the algorithm using eSVM had better performance than that using a single SVM where there were less iterations of training data [5]. Because this algorithm is a bagging method that is one of the ensemble of classifiers, it could reduce the influence of signal variability by averaging classifier outputs [5]. For individual SVM accuracy improvement, the conventional eSVM considered that the dataset, an input to a single classifier, has homogeneity, which signifies that the dataset has similar noisy components.

Since then, there have been several efforts for P300 speller performance improvement using an ensemble method, in terms of training time and accuracy. An ensemble method takes significant
training time owing to several classifiers. In order to overcome this, an attempt was made to employ wavelets and an ensemble of Fisher’s linear discriminant (FLD) [8]. Even though the algorithm could reduce the training time by using FLD with a shorter computation time than SVM, the classification accuracy decreased. Following that attempt, Perseh and Kiamini [9] achieved the reduction of training time by utilising event related potential (ERP) signals and FLD; however, this method also resulted in reduced classification accuracy. The majority of the efforts to decrease training time have suffered from lowered classification accuracy. El Dabbagh et al., in a study using an ensemble method to improve classification accuracy, proposed an ensemble of weighted SVMs and eSVM employing a new clustering of training datasets [10]. However, these methods achieved minor improvements.

Another indicator of the performance of a P300 speller is the typing speed, which is the speed at which one letter is entered. The typing speed is used as an important measure to assess the feasibility of a real BCI system because it affects user convenience. The speed is affected by the number of repetitive data, which are repetitively obtained signals increasing the SR of the signal in the P300 speller. Decreasing the number of repetitions increases the typing speed, but may lower the accuracy. However, because errors within 30% of all input letters of a P300 speller can be corrected through the Language Support Program, keeping the classification accuracy above 70% and reducing the number of repetitions can increase the typing speed and ensure the accuracy of the P300 speller. In terms of this typing speed, the subsequent studies on the eSVM [8–10] mentioned above have approximately the same typing speed as the conventional eSVM.

This paper aims to enhance this typing speed by modifying the algorithm of the conventional eSVM. Reducing the number of repeatedly obtained data enhances the typing speed; however, it may also increase signal variability. Nevertheless, the problem of increased signal variability can be reduced by improving the ensemble method. The performance of an ensemble method is known to be affected by the accuracy of an individual classifier and diversity within the ensemble [11]. Diversity means that when the same data is entered into classifiers, the classifiers produce different outputs. If the outputs of the classifiers are uncorrelated, then, high ensemble diversity exists, and the ensemble will usually achieve error reduction over the individual classifiers. In the ensemble method used in the present P300 spellers, there is no performance improvement research that considers diversity.

In this respect, this paper proposes an improved P300 speller based on eSVM using the simple data manipulation of grouping the input dataset on each classifier. Because the obtained EEG signals can have different characteristics depending on the distance between the target letter and given stimulus, we group the input data according to the distance. The individual classifiers are trained by different groups of datasets. Because the difference in the groups can reduce the correlation between the separating hyperplanes of the trained classifiers, the diversity of the ensemble can be increased. Therefore, the proposed method may increase the classification accuracy of the ensemble method owing to the increased diversity. We verified the accuracy of the proposed method by using the open source of the dataset II of BCI competition III. The competition has provided one training and one test datasets for each of the two different subjects. The EEG signal was recorded from 64 electrode channels and more details about the experimental setup are in [6].

The rest of this paper is organised as follows: In Section 2, a P300 speller is described and the eSVM is described in Section 3. A new P300 speller using dataset manipulation for ensemble diversity is proposed in Section 4. Through the experimental results, in Section 5, we compare the accuracy of the proposed P300 speller with that of the conventional P300 speller. This paper is finally concluded in Section 6.

2. A P300 speller

2.1. Visual stimulus for a P300 speller

ERPs are a measured brain response caused by specific stimulus. The recorded ERPs are very small voltages caused by the background brain activity together with other bio-signals. This signifies that ERPs are not easily detected in the EEG recording of a single trial. Thus, the EEG signals obtained from many trials are averaged to confirm a distinct ERP response to a stimulus. The P300 wave is an ERP component and shows a positive peak at nearby 300 ms after a stimulus. A P300 speller uses a P300 component, which reflects a higher stimulus response than other components.

A P300 speller presents target letters by analysing P300 waves from the EEG signals obtained when visual stimuli are provided for a subject using a display. To secure the objectivity of tested BCI data, we used dataset II from BCI Competition III [6]. The dataset was obtained using a paradigm described by Farwell and Donchin [1], in which a matrix of 6 × 6 cells is used to represent 36 letters, as shown in Fig. 1. For a single-target letter, each of the six rows and six columns is intensified and the intensifications are presented in a random sequence. A subject focuses attention on one of the 36 cells of the matrix and then, a P300 wave is evoked in response to the intensification of a row or column containing the target letter. In order to enhance the reliability of the speller, a set of 12 intensifications is repeated 15 times for each letter.

2.2. EEG data analysis for a P300 speller

A P300 speller displays one of the 36 letters by analysing the EEG caused by the visual stimuli. However, in terms of signal analysis, it deals with a problem of separating two EEG patterns according to the existence of the P300 component. As mentioned above, it provides 12 stimuli for each letter. Only two stimuli out of the 12 produce EEGs related to the desired target letter, and have a P300 component. The EEGs obtained from the remaining 10 stimuli are not related to the desired target letter and do not have a P300 component. This one set is repeated 15 times, and thus, 15 × 12 data are used to find one target letter by combining individual results after pattern recognition. A P300 speller uses machine learning to recognise and separate the two types of EEG patterns.

Machine learning is a method of automatically creating a system model to predict new data by recognising a pattern of several
datasets given in advance. It can be divided into three applications: clustering, classification, and regression, according to the desired output. A P300 speller belongs to the classification type, which divides inputs into two or more classes with known desired outputs. Fig. 2 illustrates a block diagram for classification. In a P300 speller with a sample rate of 240 Hz, a preprocess performs time windowing between 0 and 667 ms (160 samples) and frequency filtering between 0.1 and 10 Hz, resulting in the P300 signal with the bandwidth of 9.9 Hz. At this point, each filtered signal is decimated according to the high cut-off frequency, and thus, each signal is composed of 14 time samples. This feature extraction causes dimension reduction of the preprocessed dataset because EEG signals for a P300 speller have a high dimension owing to the time analysis on several electrodes for spatial resolution. Using these feature vectors and desired outputs, i.e., labels of dataset, a classification model is trained by an algorithm. The trained model is used for classification of a new dataset with unknown labels.

There are several algorithms to train a classification model such as FLD analysis [12], decision tree [13], neural network [14], and SVM [15]. Among them, the SVM has been successfully applied for a P300 speller. A data point is viewed as a u-dimensional vector, and a (u - 1)-dimensional hyperplane separates such points. There are many hyperplanes that might classify the data. In SVM, one of the possible hyperplanes can be obtained by choosing one that provides the largest margin between classes. The SVM is motivated to find the maximum margin hyperplane such that the distance from the plane to the support vectors, which are the nearest data points on each side, can be maximised. The hyperplane is given by $w \cdot x + b = 0$ with $w = (w_1, w_2, \ldots, w_u)^T$. The Euclidean distance from point $x_i$ to the hyperplane is given by
\[
dist = \frac{|w \cdot x_i + b|}{||w||},
\]
where $||w||$ denotes a two-norm of $w$, and $x_i$ is training data. For a canonical representation of the hyperplane, rescaling $w$ and $b$, such that the point closest to the hyperplane satisfies $|w \cdot x_i + b| = 1$, yields the margin of $2/||w||$. Maximising the margin is equivalent to minimising $||w||^2/2$. Thus, the SVM can be formulated as the following optimisation problem:
\[
(w_{SVM}, b_{SVM}) = \arg \min_{(w,b)} \frac{||w||^2}{2}
\text{ s.t. } y_i (w \cdot x_i + b) \geq 1 \text{ for all } i,
\]
where $y_i$ denotes a known class label (+1 or -1) corresponding to $x_i$. The problem of (2) can be solved using Lagrange multipliers, $\alpha_i$, as follows:
\[
L(w, b, \alpha) = \frac{1}{2}||w||^2 - \sum_i \alpha_i [y_i (w \cdot x_i + b) - 1].
\]
In solving this primal problem, we need to consider three variables, $w$, $b$, and $\alpha$. To simplify the optimisation problem, (3) can be changed by dual form, which has only one variable, $\alpha_i$. The $w_{SVM}$ is obtained as
\[
\nabla_w L(w, b, \alpha) = w - \sum_i \alpha_i y_i x_i = 0
\]
\[
w = \sum_i \alpha_i y_i x_i.
\]
The offset, $b_{SVM}$, can be recovered by finding $x_i$ on the margin’s boundary [16]. After the training phase above, SVM classifies test data based on the hyperplane of $(w_{SVM}, b_{SVM})$. It is known that an SVM achieves a stable classification performance for unknown test data because it selects a hyperplane maximising margin among several hyperplanes, which make errorless decisions for training data [17].

In contrast, to improve classification performance, ensemble methods, which construct a set of classifiers from multiple training data, and then, combine each result of the multiple classifiers, became a major learning paradigm [18]. This ensemble-based approach can easily solve difficult problems through a divide and conquer strategy using several classifiers. There are famous ensemble methods such as boosting and bootstrap aggregating (bagging) [11]. Boosting is to build an ensemble of sequentially trained classifiers. Assume that training an ith classifier is completed. Then, if the classifier misclassifies some of the trained data, the misclassified data are used for training the next $(i+1)$th classifier with larger weights than the correctly classified data. There are several weighting methods for boosting [19]. In general, boosting can provide high accuracy; however, it has a larger potential risk of over-fitting [20,21].

Bagging builds an ensemble model by averaging the output of classifiers, or voting. When classifying high dimensional features, a bagging method has a better classification performance [20,21]. The aggregated classifier generated by bagging is given by
\[
H(x_i) = E(h(x_i)),
\]
where $h(x_i)$ denotes a classification result for input data $x_i$ in a classifier trained by any algorithm, and $E$ averages the results of individual classifiers. Using simple algebra and inequality $E[X]^2 \leq E[X^2]$, (5) can be expressed as
\[
(f(x_i) - H(x_i))^2 \leq E[(f(x_i) - h(x_i))^2],
\]
where $f(x_i)$ denotes the ground-truth data of input data $x_i$, Eq. (6) denotes that the squared error of $H(x_i)$ is smaller than the average-squared error of $h(x_i)$. This is the reason why bagging is effective for unstable classifiers, and for reducing the variance of classification errors.

An example applying the bagging method is the ensemble of SVMs(eSVM). This is a bagging method using SVMs as classifiers and was proposed to improve the accuracy of a P300 speller. Also this algorithm solved the classification problem of a P300 speller with remarkable classification accuracy [7].
3. Ensemble of SVMs for a P300 speller

The performances of eSVM were evaluated on dataset II of BCI Competition III [6]. The dataset was recorded from a 64-channel EEG cap at 240 Hz sampling rate. The signal was collected from two subjects, A and B, in five sessions each. Each session consisted of a number of runs. In each run, the subjects focused attention on a series of letters (total training 85 letters and test 100 letters for each of the two subjects). These letters were from words; however, they had been scrambled such that spelling timeline had been lost. Fig. 3 is the training dataset organisation for dataset II of the BCI Competition III. The dataset for one single letter is obtained by 15 repeated trials and the dataset of a single trial is organised with the data obtained from 12 intensifications. For each subject, the total number of training data $x$ is 12 (intensification set) $\times$ 15 (the number of repetitions of intensification set) $\times$ 85 (the number of letters) $= 15,300$.

**Fig. 3.** The training dataset organisation in dataset II of BCI Competition III.

---

**Fig. 4.** Block diagrams of eSVM (a) in training phase and (b) test phase.

**Fig. 4** shows a block diagram of the conventional eSVM proposed in [5]. Fig. 4(a) denotes a block diagram of the training phase. Each linear SVM in the eSVM learns with some parts of the training data. Because each SVM is trained by different training data, which are partitioned from the training set, the same SVM algorithm may produce different hyperplanes. The training signals are partitioned into homogeneous groups, which are considered to have similar noisy components owing to a short acquisition time, to train a classifier in each of these groups. In [5], 180 EEG signals associated with a single letter, which came from 12 (intensification sets) $\times$ 15 (number of intensification set repetitions), presented similar noisy features, brain activities, and acquisition conditions, because they had been acquired in a relatively short period of time. Then, through extensive simulations in [5], the number of letters contained in one partition was determined to be five. Hence, the pth partition ($p = 1, 2, \ldots , P$) dataset $\{S_p\}$ is composed of 900 ($=5 \times 12 \times 15$) EEG signals, leading to 17
different partitions ($P = 17$). Training sets of each partition are used to train a single SVM.

Fig. 4(b) shows a block diagram of the test phase after the training phase. At this phase, the conventional eSVM is set by using the SVM models in the training phase. The $p$th partition of dataset $\{S_p\}$ is used to train the $p$th SVM. Before determining the class, the score function of an SVM is calculated by substituting $(\mathbf{w}, b)$ of the hyperplane, $\mathbf{w}^T \mathbf{x} + b = 0$, with $\mathbf{w}_{\text{SVM}}$ of (4) and $b_{\text{SVM}}$ of (2). Therefore, the score function of the $p$th SVM, $s_p(\mathbf{x}^{\text{test}})$, is obtained by [5]:

$$s_p(\mathbf{x}^{\text{test}}) = \sum_{i \in S_p} y_i \alpha_i^{(p)} \mathbf{x}_i^{\text{test}} \mathbf{x}_i^{\text{test}} + b_{\text{SVM}}^{(p)};$$

(7)

where $\alpha_i^{(p)}$ and $b_{\text{SVM}}^{(p)}$ are an optimal Lagrange multiplier and parameter for maximizing the margin between classes obtained after training the $p$th SVM using $i$th training data, $\mathbf{x}_i$, and label $y_i$, belonging to subset $\{S_p\}$; $\mathbf{x}^{\text{test}}$ is a new test data set; and $(\cdot, \cdot)$ denotes an operator of inner product. To combine the results regarding the test set denoted by $\mathbf{x}_i^{\text{test}}$, where $j$ is a repetition index, $r$ is a row number, and $c$ is a column number in the $6 \times 6$ user display, using 17 different SVM outputs, $s_p(\mathbf{x}_i^{\text{test}})$, from (7), eSVM uses double averaging as follows [5]:

$$A_{J, r, c} = \frac{1}{P} \sum_{j=1}^{P} \sum_{p=1}^{P} s_p(\mathbf{x}_i^{\text{test}}),$$

(8)

Note that $A_{J, r, c}$ is the averaged output depending on a test set of eSVM. The final estimated letter is obtained by selecting the letter of the row and column corresponding to $\max A_{J, r, c}$ for the given six rows and six columns.

To enhance the classification accuracy of each SVM, unnecessary channels need to be excluded. The channel selection of eSVM is basically identical to feature selection in machine learning. Feature selection methods can be sorted into a filter, wrapper, and embedded method, according to the building of a feature selection model [22]. The eSVM uses a wrapper method, which recursively searches combinations of features by evaluating the results of the trained SVM. In view of the feature search, recursive feature searching in a classifier requires significant computational time in the training phase. To reduce the computational burden, the eSVM adopts a sequential backward search. This search method starts with all the features and removes one at each iteration with the goal of improving the evaluation result.

Fig. 5 is a detailed flow chart of the feature search and selection method used in the eSVM. The methods start with 64 channels ($64 \times 14 = 896$ features) as shown in Fig. 3. In step 1, each channel is evaluated one at a time. The data corresponding to the evaluated channel are removed from the input data obtained from all the channels. Using the remaining input data, the eSVM is trained. The accuracy of the trained eSVM is a measure of the importance of the removed channel. The accuracy is calculated by

$$E_c = \frac{tp}{tp + fp + fn},$$

(9)

where $tp$, $fp$, and $fn$ are the numbers of true positive, false positive, and false negative, achieved by classifying the validation dataset, respectively. The large value of (9) denotes that the classification accuracy is high. In this step, we can achieve the same number
of $E_c$ as the evaluated channels. In step 2, the evaluated channels are ranked in ascending order by the evaluation results. Because a large $E_c$ denotes that the evaluated channel does not contribute to classification, the channel is ranked low. In step 3, the $R$ channels with low rank are removed and the evaluation result of SVM trained by input data with the remaining channels is stored. To find the channel combination with enhanced classification accuracy, those steps are continued until the remaining channels are less than or equal to $R$. We arbitrarily set $R$ to four at each iteration to reduce the training time. The final selected channels are the combination with the highest value of the stored evaluation results. The validation set $\{V_n\}$ for the evaluation of each single classifier is composed of the other dataset, which is not the input dataset in each classifier (see [5] for more details).

4. The proposed data partitioning method for diversity generation in a P300 speller

4.1. Ensemble diversity

The conventional eSVM partitions training datasets into several groups, considering only homogeneity within the group. This method can improve individual classifiers accuracy. However, the accuracy of the ensemble method also is affected by the combination of the classifier’s decisions. Because ensemble diversity makes different decisions according to the individual classifiers, this has been a fundamental issue in the ensemble method [23].

Given a single data $x$, Krogh et al. [24] defined the ambiguity of the individual learner $h$ as

$$\text{Amb}(x) = E[(h(x) - H(x))^2]. \quad (10)$$

Because (10) measures the amount of variability among individual results of individual learners, it can be considered as having a diversity. In order to confirm the relationship of the error and the diversity of the ensemble method, the average error of individual learners can be stated as follows:

$$E[(h - f)^2] = E[(h - H + H - f)^2]$$
$$= E[(h - H)^2 + (H - f)^2 + 2(h - H)(H - f)] \quad \text{(11)}$$
$$= E[(h - H)^2] + (H - f)^2.$$  

Since $H$ is the average value for classifiers $h$ as in (5) and $f$ is the ground-truth data in (11), $H$ and $f$ are not affected by $E[-]$ of averaging over the classifiers and thus $(H - f)^2$ can come out of $E[-]$. From (11), we can confirm that the quadratic error of the ensemble method is decomposed as

$$(H - f)^2 = E[(h - f)^2] - E[(h - H)^2], \quad \text{(12)}$$

where for notational convenience, all $x$’s are omitted at $h(x)$, $f(x)$, and $H(x)$. In (12), the first term, $E[(h - f)^2]$, is the average error of the individuals, and the second term, $E[(H - H)^2]$, is from (10). Because each term in (12) has positive values, to reduce the quadratic error of the ensemble method, $(H - f)^2$, the first term of (12) should be decreased and second term of (12) should be increased. This signifies that with both higher accuracy of the individuals and diversity of the learners, the error of the ensemble method will be lower.

There are heuristic, but effective, mechanisms for increasing diversity in an ensemble method. The common basic idea is to inject some randomness into the learning process. Popular mechanisms include manipulating the data samples, input features, learning parameters, and output representations. Among these, data manipulation is known to be the most popular mechanism [11], where a given dataset is resampled and then, the individual classifiers are trained from each resampled dataset.

Fig. 6 shows the proposed dataset partitioning method. Fig. 6(a) depicts the assignment of index $l$, which represents the different
six row and six column intensifications. The distance \( d \) can be expressed as

\[
d = \begin{cases} 
  l - l_{c}^{(m)} & \text{for } l \leq 6 \\
  l - l_{r}^{(m)} & \text{for } l > 6
\end{cases}
\]

(13)

where \( l_{c}^{(m)} \) and \( l_{r}^{(m)} \) is the column and row index of the \( m \)th target letter. The index \( l \) satisfying \( d = 0 \) is the column or row index of the \( m \)th target letter. In the case of the display given in Fig. 6(a), the largest \( d \) is 5. For example, one target letter ‘K’, \( l_{c}^{(m)} \) is 5, and \( l_{r}^{(m)} \) is 8. The indices \( l \) satisfying \( d = 1 \) are 4, 6, 7, and 9 under (13), as shown in Fig. 6(b). Fig. 6(c)–(e) shows the indices \( l \) that cause \( d = 2–4 \), respectively.

When using an eSVM, each single classifier has its own partitioned training data. If there are several training datasets for each classifier, the error of the ensemble method converges to a steady error rate as the number of classifiers in the ensemble method increases [26]. However, regarding practical cases, such as dataset II of BCI Competition III, because a finite number of data is given, we need to increase the number of classifiers and provide a sufficient number of datasets for each classifier.

In order to increase both the number of classifiers and ensemble diversity, we divide the training dataset into \( G \) groups as follows:

\[
S_{g} = \left\{ \mathbf{x}^{(m)} \right\} \left\{ \begin{array}{l} 
M \left\lfloor \frac{m - 1}{G} \right\rfloor + 1 \leq m \leq \frac{M}{G} \cdot G \\text{if } g = G, \\
M \left\lfloor \frac{m - 1}{G} \right\rfloor + 1 < m \leq \frac{M}{G} \cdot G 
\end{array} \right\}
\]

(14)

where \( g \) is the index of groups \((g = 1, 2, \ldots, G)\) and \( \mathbf{x}^{(m)} \) is the dataset according to the \( m \)th letter \((m = 1, 2, \ldots, M)\). Then, the data of \( \{ S_{g} \} \) is divided into six subsets \( \{ S_{g,d} \} \) according to (14):

\[
S_{g,d} = \left\{ \mathbf{x}_{j,l}^{(m)} \in S_{g} \mid l - l_{c}^{(m)} = d, \text{ for } l \leq 6, \right. \\
\left. l - l_{r}^{(m)} = d, \text{ for } l > 6 \right\},
\]

(15)

where \( \mathbf{x}_{j,l}^{(m)} \) is a dataset with the \( l \)th intensification index \((l = 1, 2, \ldots, 12)\) of the \( j \)th repetition \((j = 1, 2, \ldots, 15)\) of \( m \)th letter and \( d \) is the distance between a target letter and other intensifications \((d = 0, 1, \ldots, 5)\).

In the conventional eSVM [5], the study suggested separating the training signals into different subsets, each having homogeneity in that the training signals included in the same subset have similar noisy components. This approach had considered that the signals related to a single letter presented similar noisy features because they were acquired in a relatively short period of time. Indeed, during this short time, they can suppose that brain activities and acquisition conditions have not changed considerably. We also separate the dataset related to a single letter into the same group by (14), which is considered to be homogeneous [5].

In the case of dataset II of BCI Competition III, we divide the training dataset \( \mathbf{x} \) (the total number is 15,300) according to the 85 letters into three groups \((G = 3)\) as in (14), and thus, each one of the divided datasets becomes a dataset with 28 or 29 letters. Then, the dataset of each group is partitioned into six subsets according to the partitioning rule described in (15). Among those datasets in a group, dataset \( \{ S_{g,d} \} \) with \( d = 0 \), represents a target dataset, and five datasets, \( \{ S_{g,d} \} \) with \( d \neq 0 \), represent non-target datasets. The five non-target datasets are divided according to the value of \( d \) and each non-target data with its own \( d \) value is used as a training dataset for the five corresponding SVMs. On the other hand, the target dataset with \( d = 0 \) is used as a training dataset for all five SVMs. Because the number of groups is three, the total number of SVMs becomes 15, which is shown in the proposed eSVM of Fig. 7.
In addition, as the conventional eSVM uses channel selection to improve the accuracy of each SVM, the proposed eSVM uses the same channel selection method, as shown in Fig. 5. However, as the datasets used in the training phase of the proposed eSVM are different from those of the conventional eSVM, the validation dataset for evaluating each proposed SVM is also differently defined. For channel selection in the proposed eSVM, each SVM is evaluated by validation datasets not used for training SVMs. Thus, when the first to fifth SVMs are trained on group 1, then, group 2, 3, or both, can be their respective validation set. For training the sixth to tenth SVMs on group 2, group 3, 1, or both, can be a validation set. That is, the validation dataset \( V_{g,d} \) of the SVM trained by the dataset \( S_{g,d} \) is composed of the non-target datasets with a distance of \( d \) and the target-data sets in a group \( g' \neq g \). In the proposed method, we choose one group out of two possible ones as follows:

\[
V_{g,d} = \{ S_{g',d'} \mid g = (g \mod G) + 1, d = 0 \text{ or } d' \} .
\] (16)

After the channel selection, each SVM is trained with the training data associated with the selected channels. This is the final trained model of the proposed eSVM.

4.3. Diversity measurement

To show that EEG signals are affected differently depending on the distance between a target letter and other intensifications, we used a measure of ensemble diversity, named by difficulty [25]. This measure was originally proposed by Hansen and Salamon [18], and explicitly formulated by Kuncheva and Whitaker [25]. A random variable \( Z \) has values in \( \{ 0, \frac{1}{2}, \ldots, 1 \} \) and denotes the proportion of the number of classifiers that correctly classify a randomly drawn input \( x \) to the total number of ensemble classifiers, \( P \). The probability mass function of \( Z \) can be estimated by running the \( P \) classifiers on the training dataset.

Considering the distribution shape of \( Z \), if the same input is predicted into the wrong class for all classifiers and the other inputs are predicted into the correct class for all classifiers, the distribution shape is with two separated peaks. On the other hand, if the inputs predicted into the wrong class for some classifiers, are predicted into the correct class for other classifiers, the distribution shape is one off-centred peak. The goal is a measure of diversity based on the distribution of disagreement. Because the variance of \( Z \) is able to capture the distribution shape, the difficulty \( \theta \) can be defined as [25]

\[
\theta = \text{var}(Z) .
\] (17)

This signifies that the lower the value of \( \theta \), the larger the diversity. Therefore, we use this value as a measure of diversity for comparison between the conventional and proposed eSVM.

5. Experimental results

We used dataset II of BCI Competition III [6] and arranged the dataset by extracting features using the same method as [5]. Because the P300 wave appears near 300 ms after an intensification stimulus, 160 time-samples for 0–667 ms were extracted at a sampling rate of 240 Hz. The datasets were filtered by an 8th
order bandpass Chebyshev Type I filter with the pass band range of 0.1–10 Hz, resulting in the P300 signal with the bandwidth of 9.9 Hz. Then, to reduce the data size, we decimated the band-pass filtered data with a sampling rate of 20 Hz.

First, we proved that the proposed eSVM using data partitioning, according to the distance from the target to the other stimulus, increases the diversity. The difficulty of (17) was used as a measure to compare degrees of diversity. We then verified how this affects the accuracy of the P300 speller by increasing the diversity.

Figs. 8 and 9 show histograms of Z for the conventional eSVM ($P=17, N=12 \times 15 \times 85 = 15,300$ data points) and proposed eSVM ($P=15, N=15,300$ data points) using the datasets of subjects A and B in dataset II of BCI Competition III. For convenience, we scaled linearly onto $[0, 1]$. In Fig. 8, the conventional and proposed methods have a distribution with one off-centred peak on the right side. The difficulty of Fig. 8(a) is 0.0621 and the value of Fig. 8(b) is 0.0605. Although the difference is imperceptible, the difficulty resulting from the proposed method is smaller than that of the conventional method. In Fig. 9, the conventional method has a distribution with two separated peaks; however, the proposed method has one off-centred peak on the right side. Thus, the difficulty of Fig. 9(a), acquired by the conventional method is 0.1384, and the value of Fig. 9(b), obtained from the proposed method is 0.0699. From these results, we can verify that the proposed eSVM achieves higher ensemble diversity than the conventional one, by using data partitioning according to the distance from the target to the other stimulus.

In order to verify the effect of the proposed method on classification accuracy, we show the classification accuracy of the proposed P300 speller in Table 1. This table compares the classification accuracy of the conventional and proposed methods in accordance with the increased number of iterations. Compared to the conventional method, the proposed method shows better performance in four repetitions, with respect to subject A, and exhibits better performance throughout the small number of repetitions with regard to subject B. These results were consistent with the previously described enhancement degree of diversity. The classification accuracy for subject A, with a relatively small improvement of ensemble diversity (difficulty is decreased by about 0.0016), changed little from that of the conventional algorithm. On the other hand, the classification accuracy for subject B, with significantly increased ensemble diversity (difficulty is decreased by about 0.0685), was enhanced by more than 4% when the number of repetitions was less than five (bold-face type in Table 1). In particular, the accuracy reached 70% by repeating the letter dataset only three times with subject B. Classification accuracy of 70% or more can most likely support a BCI speller application because verbal communications with the Language Support Program is known to be possible at that level [27].

Additionally, Table 1 indicates the classification accuracy of the proposed P300 speller, and an accuracy comparison with other P300 spellers. The algorithms we introduced as references are as follows: Salvatis and Sepulveda [8] used wavelet transformed signals as features and designed the ensemble structure using FLD instead of SVM. Perseh and Kiamini [9] used an ensemble of SVM and proposed a method that gives some weight to each SVM (algorithm 1), and another method that clusters the training signals in a new way (algorithm 2). El Dabbagh and Fakhr [10] used the sepa-
rated ERP signals as features and FLD as a classifier. The algorithms of [8,10] have a similar or lower performance than the conventional one. Although the conventional eSVM was applied in [9], the corresponding algorithm showed better performance than the conventional one. However, the improved performance was less than 70% when the number of repetitions was smaller than five.

When the number of repeated datasets is from four to five, the proposed algorithm has the highest performance above 70%. As shown in Table 1, the proposed method maintains a high and steadily increasing accuracy. While other algorithms achieve 70% or more classification accuracy by repeating the dataset for each letter over five times, the proposed method achieves 70% accuracy by only repeating the dataset four times for each letter. The conventional eSVM performs double averaging in repetition and partitioning indices as in [8]; however, the algorithm has low accuracy when the number of repeated datasets is small. On the other hand, the proposed algorithm used the data partitioning method to increase ensemble diversity. The increased diversity results in the classification accuracy of the proposed algorithm, even though the number of repeated datasets is small. This signifies that the number of repetitions can be reduced for entering a letter. As a result, the reduction of repetition number leads to the enhancement of the typing speed of a P300 speller.

6. Conclusion

This paper presents the enhanced eSVM for a P300 speller, which types words using brain signal activities. The proposed P300 speller has a simple method using data partitioning according to the distance between the target letter and non-target intensifications. This method raises the diversity of an ensemble method and results in enhanced accuracy using a small number of repetitions.

In order to analyse the proposed method, we confirmed the diversity improvement from the difficulty value. From this result, we demonstrated that the EEG can have different characteristics depending on the distance from the target letter to a non-target intensification for the P300 speller. In addition, from the accuracy improvement of the proposed method in the small number of repetitions, the proposed method can improve the letter-typing speed of the P300 speller.

For further research, we can consider the computational complexity problem about the eSVM. This problem comes from several SVMs, which involves solving a constrained quadratic programming problem and recursive channel selection in a classifier. One method to solve this problem is to find reducing training data, because SVM training is known to be a constrained quadratic programming problem of $O(n^3)$ time, where $n$ is the number of samples in the training data. Another method is to find other classifiers that have lower computational complexity and high accuracy for an improved P300 speller.

Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2017R1D1A1B04035230).

References


Table 1

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Subject</th>
<th>Number of repeated datasets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Proposed algorithm</td>
<td>A</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>40</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>29.5</td>
</tr>
<tr>
<td>Conventional algorithm</td>
<td>A</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>35</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>23.5</td>
</tr>
<tr>
<td>Ref. [8] algorithm</td>
<td>A</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>44</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>31</td>
</tr>
<tr>
<td>Ref. [9] algorithm 1</td>
<td>A</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>38</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>25</td>
</tr>
<tr>
<td>Ref. [9] algorithm 2</td>
<td>A</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>44</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>30.5</td>
</tr>
<tr>
<td>Ref. [10] algorithm</td>
<td>A</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>44</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>31</td>
</tr>
</tbody>
</table>


