An Evaluation of pdi-Boosting for Brain-Computer Interfaces

Isao Hayashi
Department of Informatics, Kansai University
Takatsuki, Osaka 569–1095, Japan
Email: ihaya@cbii.kutc.kansai-u.ac.jp

Shinji Tsuruse
Panasonic System Solutions Japan Co.,Ltd.
Chuo Ward, Tokyo 104–0061, Japan
Email: suraimuppoi@yahoo.co.jp

***Abstract*—Brain-computer interface (BCI) technologies have recently entered the research limelight. In many such systems, external computers and machines are controlled by brain activity signals measured using near-infrared spectroscopy (NIRS) or electroencephalograph (EEG) devices. In this paper, we evaluate the boosting algorithm for BCI using simple numerical examples where we add various amounts of disturbances. In pdi-Boosting, interpolated data is generated around classification errors using a probability distribution function. By using the interpolated data, the discriminated boundary is shown to control the external machine effectively. We verify our boosting method with an experiment in which NIRS data is obtained from subjects performing a basic arithmetic task, and discuss the results.**

I. Introduction

Recently, many papers on brain-computer interfaces (BCIs) have been published in the literature [1]. In the BCIs, brain activity signals are measured using near infra-red spectroscopy (NIRS) [2]–[4] and electroencephalographic (EEG) devices [5]. Then the discriminant model is used to extract the signals discriminated boundaries in order to control an external computer. However, the external computer is unable to follow dynamic changes as the identified model uses prior brain activity data. In a different approach, ensemble learning models [6], [7] are applied to the pattern classification problems. The ensemble learning model consists of multiple weak classifiers, and the final output is determined by the result of a majority rule decision between the weak classifiers, done to improve overall discriminant accuracy. AdaBoost [8], [9] is a remarkable boosting method among ensemble learning models; it assigns high weights to the misclassified data and thus labels data so that the next iteration of the classifier model and focus on the misclassified data.

In this paper, we evaluate our boosting algorithm for BCI, which interpolates data around misclassified data using a given probability density function. We call our method pdi-Boosting (Probabilistic Data Interpolation-Boosting) [10]–[12]. The misclassified data for the classifier model is not labeled by weights as in AdaBoost. Instead, the discriminant model is characterized by the addition to the existing data of new data generated around the misclassified data using the given probability density function. Therefore, as pdi-Boosting generates interpolated data around the misclassified data, we can obtain a discriminant boundary with inherent robustness, and the external computer is able follow dynamic changes in the environment. First, we formulate the pdi-Boosting algorithm and discuss the characteristics of pdi-Boosting by simple numerical examples where we add various amounts of disturbances. Second, we apply our novel method to an experiment in which brain activity is measured using a NIRS device to demonstrate the effectiveness of pdi-Boosting.

II. Formulation of pdi-Boosting

AdaBoost [8] is an outstanding boosting method. In each iteration we select training data (TRD) from the set of misclassified data with high weights of over 50%, and then apply this data to a weak classifier in the consecutive iteration. After identifications are made by the weak classifier, the weights of the data are updated. After iterating the procedure sequentially, the final output is determined by majority rule decision of the multiple weak classifiers $M_1, M_2, \cdots, M_i, \cdots, M_L$, when the checking data (CHD) is given to these models.

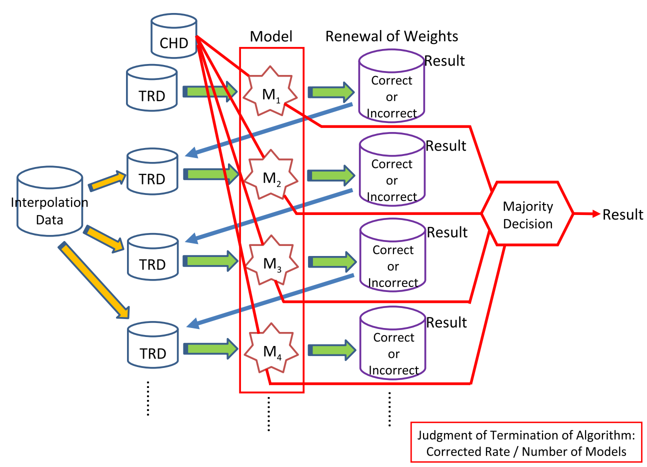

Fig. 1. Conceptual Diagram of pdi-Boosting

In the pdi-Boosting algorithm, new data is interpolated around the misclassified data using a probability density function instead of the updating of the weights as in AdaBoost. A conceptual diagram of pdi-Boosting is shown in Figure 1; the algorithm is as follows. Similar to AdaBoost, the final output is determined by a rule of majority decision using the multiple weak classifiers, when the checking data (CHD) is given to these models. However, the difference between AdaBoost and pdi-Boosting is that the amount of data in pdi-Boosting increases as compared to AdaBoost as shown in Figure 2.

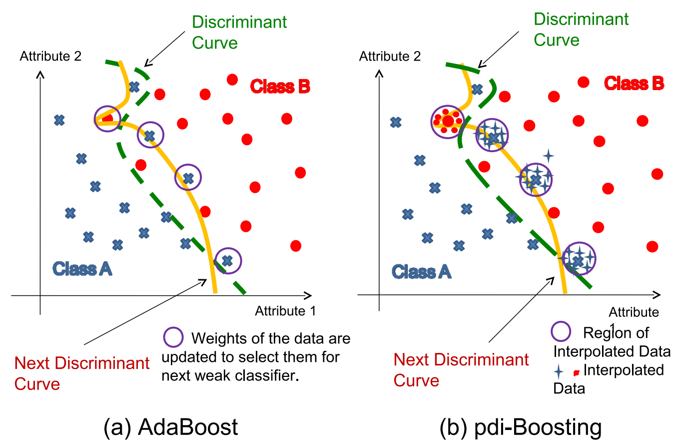

Fig. 2. Conceptual Difference between AdaBoost and pdi-Boosting

Assume that the misclassified data is given as the s-th data in TRD, and the j-th attribute of the s-th data is denoted by $x_j(s)$. The interpolated data $x_j^{new}(s)$ is generated by a probability density function $f(x_j)$ around the misclassified data $x_j(s)$ with mean value.

$$x_j^{new}(s) = \{x_j \in A \mid P(A) = \int_A f(x_j)dx_j\} \tag{1}$$

In general we choose the normal distribution function to be our probability density function $f(x_j)$ as follows:

$$f(x_j) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp(-\frac{(x_j - x_j(s))^2}{2\sigma_1^2}). \tag{2}$$

However, we may also adopt a uniform distribution as follows:

$$f(x_j) = \begin{cases} \frac{1}{x_j^{max} - x_j^{min}} & ; \ for\ x_j^{min} \le x_j \le x_j^{max} \\ 0 & ; \ for\ x_j < x_j^{min}\ or\ x_j > x_j^{max} \end{cases} \tag{3}$$

where x_j^{max} and x_j^{min} are defined as

$$x_j^{min} = \frac{3x_j(s) + x_j(s+1)}{4} \tag{4}$$

$$x_j^{max} = \frac{3x_j(s) + x_j(s-1)}{4} \tag{5}$$

and $x_j(s-1)$ and $x_j(s+1)$ denote the $(s-1)$-th and the $(s+1)$-th data, respectively.

We formulate the algorithm of pdi-Boosting as follows:

Step 1: Brain activity D of size W is measured at each electrode of the NIRS measurement device. The data D is divided to two data sets: the training data D^{TRD} with the size W^{TRD}, and the checking data D^{CHD} with the size W^{CHD}, where $W = W^{TRD} + W^{CHD}$. In addition, the interpolated data from D is denoted by D^{INT}.

Step 2: The training data D^{TRD} is given as input into the i-th weak classifier M_i. The recognition rate r_i^{TRD} is calculated and the result given as R_i.

Step 3: The s-th misclassified data is selected from D^{TRD}. With this s-th data, a new interpolated data $x_j^{new}(s)$ is generated around $x_j(s)$ of the j-th attribute by the probability density function $f(x_j)$ defined in equations (2) through (5), and this new data $x_j^{new}(s)$ is saved to D^{INT}.

Step 4: Interpolated data are extracted from D^{INT} until the number of misclassified data are the same as the number of correctly classified data, where the number of interpolated data d satisfies

$$d \ge \frac{W}{2} - W^{TRD}(1 - \frac{r_i^{TRD}}{100}). \tag{6}$$

Step 5: Let θ be the threshold value and M is the number of iterations. The algorithm terminates when either one of the conditions $r_i^{CHD} \ge \theta$ or $i \ge M$ is satisfied.

Step 6: We apply D^{CHD} to $M_1, M_2, \cdots, M_i$ to obtain the final discriminant result with recognition rate r_i^{CHD}.

Because new data is added around the misclassified data using a probability density function, and in each successive iteration the weak classifier fits the misclassified data closer than in the previous iteration, the final result will more closely approximate the given data.

III. Robustness of pdi-Boosting

When brain activity is measured using NIRS or EEG devices the observed data includes significant amounts of noise. To study the effect noisy NIRS data has on pdi-Boosting, we simulate examples of possible cases of noisy signals using numerical data. We give the parameters of the noise disturbances which we will add to the signal, and then discuss the robustness of pdi-Boosting with respect to the parameters of the noise added to the data.

We assume that oxyhemoglobin (oxy-Hb) and deoxyhemoglobin (doxy-Hb) data are normalized to lie in the range $[0, 1]$. We add four types of noise to the signals consisting of normally distributed random numbers with mean value m and standard deviation s as follows:

(A) Data generated with $m = 0$ and $s = 0.2$
(B) Data generated with $m = 0$ and $s = 0.4$
(C) Data generated with $m = 0$ and $s = 0.6$
(D) Data generated with $m = 0$ and $s = 0.8$

where the number of clusters is 2. We assume the value corresponding to the steady state to be 0, and that to the activation state to be ± 1.

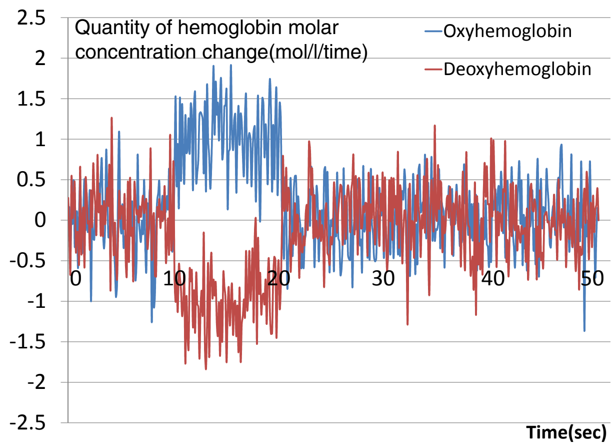

Fig. 3. Data Pattern B

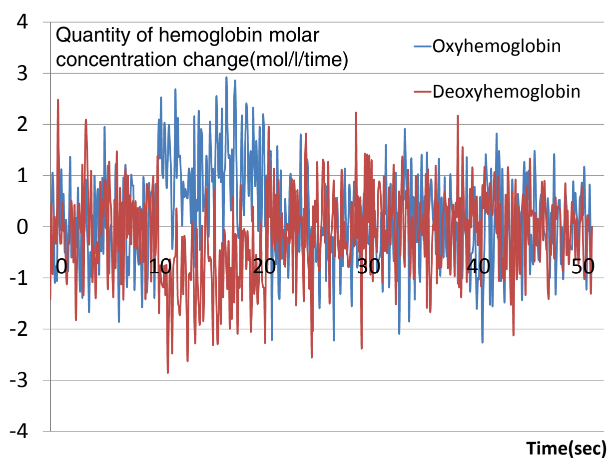

Fig. 4. Data Pattern D

The two patterns, $s = 0.4, 0.8$, are given in Figures 3 and 4, respectively. The observed data is chosen to lie in a $50s$ time interval, where the task begins $10s$ after the start of measurement, and the duration of the task is $10s$.

The results of pdi-Boosting are given in Table I. We show that the average recognition rate is 10 times that of the disturbance data. In addition, We choose the uniform distribution to be the probability density function $f(x_j)$, and we adopted REPtree as the weak classifier, which is a type of decision tree method. The termination rule for the algorithm is set to be at iteration number $M = 3$. In Table I, the recognition rate of pdi-Boosting is shown to be slightly higher than that of other methods with standard deviation s ranging from 0.2 to 0.6. On the other hand, the recognition rate of pdi-Boosting is clearly higher than other methods when the standard deviation is 0.8. Additionally, the difference in the recognition rate between pdi-Boosting and REPtree was about 10%. Looking at these results, pdi-Boosting is more accurate than other methods even though the significance of pdi-Boosting is not clear by a test. As a result, we have shown the higher robustness of pdi-Boosting for noisy data as compared to conventional Boosting methods and REPtree.

TABLE I
ACCURACY RATES FOR DATA PATTERNS WITH ADDED NOISE

Data	pdi-B. (%)	AdaBoost (%)	MultiBoost (%)	REPtree (%)
Pattern A	99.82	99.80	99.76	99.37
Pattern B	97.33	96.57	94.84	97.14
Pattern C	92.65	91.22	91.22	92.41
Pattern D	88.78	88.63	88.55	79.54
Ave.	94.64	94.06	93.59	92.11

IV. EVALUATION OF PDI-BOOSTING

We simulate examples of signals using numerical data to discuss the evaluation on pdi-Boosting, We assume that oxyhemoglobin (oxy-Hb) and deoxyhemoglobin (doxy-Hb) data are normalized to lie in the range $[0, 1]$, where the number of clusters is 2. We assume the value corresponding to the steady state to be 0, and that to the activation state to be 1. We add five types of noise to the signals consisting of normally distributed random numbers of 500 with standard deviation s as follows:

Number of data: $2, 5, 10, 20, 30, 50, 75, 100, 250, 350, 500$
Standard deviation $\sigma = 0.005, 0.01, 0.05, 0.1, 0.2, 0.6, 1.0$
Standard deviation $s = 0.2, 0.4, 0.6, 0.8, 1.0$

We choose the normal distribution function to be the probability density function $f(x_j)$, and we adopted REPtree as the weak classifier, which is a type of decision tree method. The termination rule for the algorithm is set to be at iteration number $M = 3$. An evaluation of pdi-Boosting is shown as the following results.

1) The recognition rate does not decrease even though the number of data is extremely small. In addition, when the number of the data is large, the recognition rate is high and its variance is small.
2) The recognition rate in case of the learning data is high even if the number of data is few.
3) When noise is large, the robustness of the recognition rate is high.

As the first result, the recognition rate of pdi-Boosting was higher than AdaBoost even though the number of data is extremely less than ten. In addition, when the number of the data is large, the recognition rate of pdi-Boosting was higher than AdaBoost. As the second result, the recognition rate of pdi-Boosting in case of the learning data was higher than

AdaBoost even if the number of data is few. As the third result, the larger the noise, the higher the recognition rate of pdi-Boosting compared with AdaBoost.

We must pay in particular attention to the third result. The recognition rate of pdi-Boosting and AdaBoost is shown in Figure 5 when the standard deviation s for generating noise was changed with 0.4, 0.8, and 1.0. In Figure 5, the recognition rate of pdi-Boostingis is drawn in solid line, and AdaBoost is drawn in dashed line. The cognitive rate is the mean value of ten times of trials. The larger the noise, the lower the recognition rate of pdi-Boosting and AdaBoost. The variance of recognition rate is extremely large in the case of the number of data less than 50. However, the recognition rate of pdi-Boosting is higher than AdaBoost, when the number of data is over 100. In addition, The recognition rate of pdi-Boosting increases in proportion to the number of data. From these results, pdi-Boosting has a robustness as to noise compared with AdaBoost. As a result, we must notice that the recognition rate of pdi-Boosting is higher than AdaBoost.

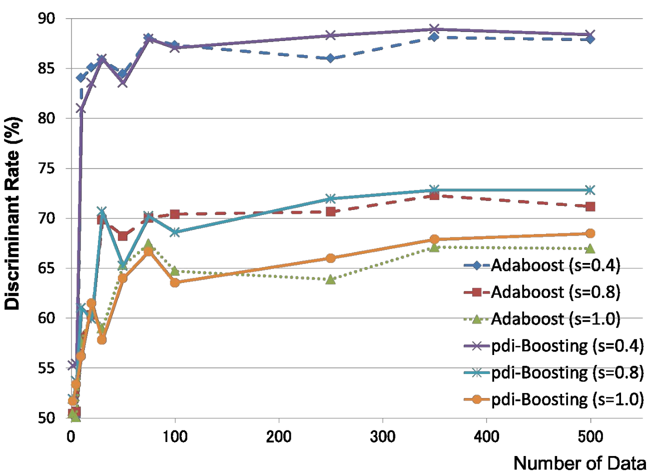

Fig. 5. Discriminant Rate by Changing Number of Data and S.D.

We discuss the recognition rate when the standard deviation σ for generating probability distribution function $f(X_j)$ was changed. When we assume a position of misclassified data at 0.5, we show the frequency distribution of the interpolation data in the figure 6. The larger the standard deviation σ, the wider the domain of interpolation data. In the case of $\sigma = 1.0$, around 80% of data are included in interval of $[0, 1]$. Therefore, we notice that we must define larger standard deviation, when more extensive interpolation data are necessary.

V. NIRS Signal Measurement Experiment of Calculation Task

As an application of pdi-Boosting to NIRS measurement tasks, we performed the following experiment. We assigned subjects the task of performing a simple arithmetic calculation, and measured brain activity during the task using a NIRS device. We used the 2ch NIRS device (YN-502, Excel of

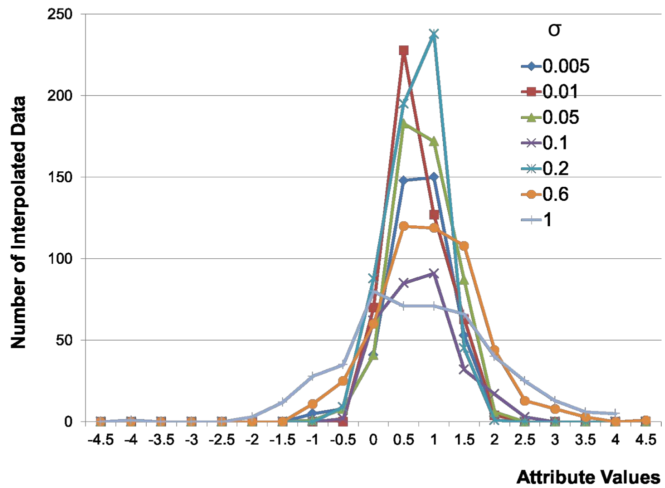

Fig. 6. Frequency Distribution of Interpolated Data

Mechatronics Company, Tokyo, Japan). This device is able to measure the oxyhemoglobin (oxy-Hb) and deoxyhemoglobin (doxy-Hb) at two electrodes with the wavelength $770nm$ and $840nm$ with period $0.098s$. Our brain signal data sets consist of the above four attributes measurable by our NIRS device.

We gave a simple test consisting of arithmetic addition problems to our subjects as the experimental task. The test consisted of a simple addition problem that does not require carrying of decimals, similar to elementary school first grade level problems. Figure 7 shows the environment of experiment. We confirmed the ability of three subjects to perform the task, which is the subject's age, sex, dominant arm, health condition, and their ability to perform specific calculations before experiment.

The experiment consisted of 5 trials that we called a set, and was 250 seconds in total. A trial is 50 seconds which consists of a 10 second pre-rest, a 10 second interval allotted for our simple arithmetic calculation test and 30 seconds of a post-task resting. We show the experiment protocol of the arithmetic calculation test in Figure 8.

To serve as our three subjects, we selected one man and two women, and performed 12 sets of trials with each subject extended over a period of several days. In the trials, we placed the two electrodes of the NIRS device on prefrontal areas F_{p1} and F_{p2} as described in the ten-twenty electrode system of the American Electroencephalographic Society. The discrete-valued data sets D of brain signals were created from a randomly selected 10 trials by each of the three subjects.

VI. Results and Discussion

In Figures 9 and 10 we show the change in cerebral blood flow of subject C, which is the brain activity signal measured by NIRS equipment. Figure 9 shows the change in flow of oxy-Hb and doxy-Hb measured at the right side electrode, and Figure 10 shows it that of the left side electrode. In Figure 10, we see that when the task is started at $10s$, the flow of oxy-Hb

Fig. 7. Experiment

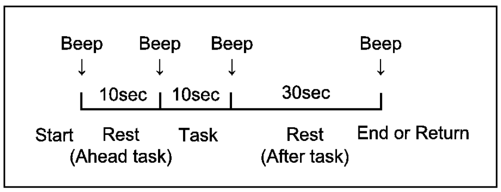

Fig. 8. Timing Protocol of Experiment

increases, and the flow of doxy-Hb decreases. After the task finishes at $20s$, both cerebral blood flow rates gradually return to the normal steady state.

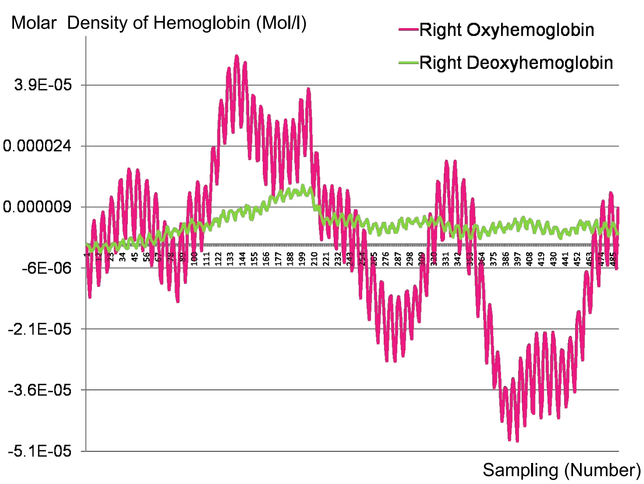

Fig. 9. Cerebral Blood Flow Change of Right Side Electrode

Next, we explain the model for cerebral blood flow changes using the pdi-Boosting algorithm. We choose the uniform distribution to be the probability density function $f(x_j)$, and we use REPtree as our weak classifier. The termination rule for the algorithm is set to be at iteration number $M = 3$. We

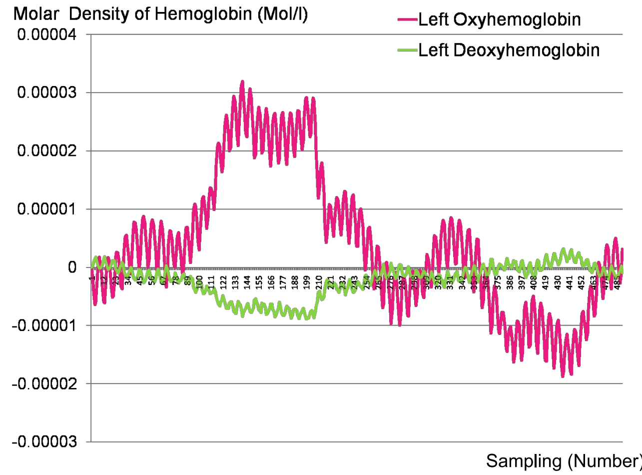

Fig. 10. Cerebral Blood Flow Change of Left Side Electrode

set the size of data set D to be $W = 490$. However, we also set $W = W^{TRD} = W^{CHD} = 490$ and $D = D^{TRD} = D^{CHD}$ as we did not have ample data in this experiment.

First, we fed the data set D_1^{TRD} into the first weak classifier M_1 and obtained the recognition rate $r_1^{TRD} = r_1^{CHD} = 91.22\%$ as result R_1. The interpolated data d_1 is randomly selected from D^{INT} and applied to the next weak classifier, until the number of the interpolated data becomes equal to the number of correctly identified data. The number of data points in D_2^{TRD} is $W_2^{TRD} = 894$ because the number of interpolated data in d_1 is 404.

In the next step, we input D_2^{TRD} into the second weak classifier M_2 and obtain the recognition rate $r_2^{CHD} = 92.24\%$ as result R_2. The recognition rate of M_3 is $r_3^{CHD} = 93.67\%$, and the algorithm stops running finished by the termination rule for $M = 3$.

From the three results R_1, R_2, and R_3 the majority rule yields a final result of 95.31% for the recognition rate. The number of interpolated data points as well as the recognition rates are summarized in Table II.

TABLE II
ACCURACY RATES AND NUMBER OF INTERPOLATION DATA

Model	Accuracy Rate(%)	Inter. Data (for next step)	TRD	CHD
M_1	91.22	404	490	490
M_2	92.24	414	894	490
M_3	93.67	428	904	490
pdi-B. (TRD)	95.31			

Next, we discuss the recognition rates for the three individual subjects. We show that the average recognition rate for each step in Table III. In the case of subject B, the recognition rate in the three consecutive steps are respectively 93.69%, 93.47%, and 93.33%. However, the total recognition rate determined by the majority rule is 94.78%, higher by a full 1.0% than the recognition rates in the individual three steps. This phenomena is apparent in case of the other two subjects as well. The results are remarkable as the higher recognition rates in the total recognition rate show the effectiveness of pdi-Boosting.

TABLE III
ACCURACY RATE FOR INDIVIDUAL SUBJECTS

Sub.	$r_1^{CHD}(\%)$	$r_2^{CHD}(\%)$	$r_3^{CHD}(\%)$	T.R.(%)
A	90.33	90.29	90.51	91.04
B	93.69	93.47	93.33	94.78
C	92.13	88.37	88.01	94.16
Ave.	92.13	91.62	92.06	93.10

Finally, we compare pdi-Boosting with REPtree and other conventional Boosting algorithms. The comparison results are summarized in Table IV. The comparison of pdi-Boosting and REPtree, shows that the recognition rate of pdi-Boosting is only 0.97% higher than than that of REPTree on average for all three subjects. In addition, the recognition rate of pdi-Boosting shows a significant difference ($p = 0.01616$) as compared with REPTree by the t-test with significance level 0.05%.

On the other hand, comparing pdi-Boosting with the other boosting methods, AdaBoost and MultiBoost, the recognition rate of pdi-Boosting is only 0.3% higher than that of AdaBoost, and 1.7% higher than that of MultiBoost in subject A. In subject C, the recognition rate of pdi-Boosting is only 1.2% higher than MultiBoost, but the recognition rate became lower by a small 0.33% than for AdaBoost. Unfortunately, the recognition rate of pdi-Boosting couldn't show a significant difference as compared with other boosting methods by the multiple comparison of tukey method. However, the recognition rate of pdi-Boosting shows a significant difference ($p = 0.0006860$) compared with MultiBoost by the t-test with significance level 0.05%. The recognition rate of pdi-Boosting is only 0.56% higher than AdaBoost, but we could not show a significant difference ($p = 0.1578$) as compared to AdaBoost by the t-test with significance level 0.05%.

TABLE IV
COMPARISON OF THE PROPOSED AND EXISTING MODELS

Sub.	pdi-B. (%)	AdaBoost (%)	MultiBoost (%)	REPtree (%)
A	91.04	90.33	90.38	90.33
B	94.78	94.24	93.57	93.69
C	94.16	94.39	92.96	92.92
Ave.	93.10	92.54	92.14	92.13

As a result, we may conclude that recognition rates using pdi-Boosting are higher than those of other Boosting methods. Therefore, the use of the proposed pdi-Boosting algorithm is advantageous in practical BCI applications.

VII. CONCLUSION

In this paper, we evaluated our classification method based on a boosting algorithm using a probabilistic data interpolation scheme. In addition, we verified our method in an experiment in which brain activity is measured using a NIRS device, and discussed the effectiveness of our new approach by comparing its performance to that of conventional boosting algorithms. In future work, we plan to discuss how to optimize the probability density functions used, and how to apply this method to a range of practical BCI problems.

ACKNOWLEDGMENT

This work was supported in part by the Ministry of Education, Culture, Sports, Science, and Technology of Japan under Strategic Project to Support the Formation of Research Bases at Private Universities 2008-2012, in addition the Organization for Research and Development of Innovative Science and Technology (ORDIST) of Kansai University.

REFERENCES

[1] M.A.Lebedev, J.M.Carmera, J.E.O'Doherty, M.Zacksenhouse, C.S.Henriquez, J.C.Principe, and M.A.L.Nicolelis, "Cortical ensemble adaptation to represent velocity of an artificial actuator controlled by a brain-machine interface," *Journal of Neuroscience*, Vol.25, No.19, pp.4681-4693 (2005).

[2] M.Wolf, G.Morren, D.Haensse, T.Karen, U.Wolf, J.C.Fauchere, and H.U.Bucher: "Near Infrared Spectroscopy to Study the Brain: An Overview," *Opto-Electronics Review*, Vol.16, No.4, pp.413-419 (2008).

[3] R.Sitaram, H.Zhang, C.Guan, M.Thulasidas, Y.Hoshi, A.Ishikawa, K.Shimizu, and N.Birbaumer: "Temporal Classification of Multichannel Near-Infrared Spectroscopy Signals of Motor Imagery for Developing a Brain-Computer Interface," *Neuroimage*, Vol.34, No.4, pp.1416-1427 (2007).

[4] W.Niide, T.Tsubone, and Y.Wada: "Discrimination of moving limb with near-infrared spectroscopy," *IEICE Technical Report, Neurocomputing*, Vol.107, No.542, pp.191-196 (2008) (in Japanese).

[5] T.Yamaguchi, K.Nagata, Q.T.Pham, G.Pfurtscheller, and K.Inoue: "Pattern Recognition of EEG Signal during Motor Imagery by Using SOM," *International Journal of Innovative Computing, Information and Control*, Vol.4, No.10, pp.2617-2630 (2008)

[6] Y.Freund: "Boosting a Weak Learning Algorithm by Majority," *Information and Computation*, Vol.121, No.2, pp.256-285 (1995).

[7] T.Nakashima, and Y.Shoji: "The Effect of Data Partition in Constructing Fuzzy Ensemble Classifiers," *Proceedings of the 25th Fuzzy System Symposium*, No.3E2-01 (2009) (in Japanese).

[8] Y.Freund, and R.E.Schapire: "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *Journal of Computer and System Sciences*, Vol.55, No.1, pp.119-139 (1997).

[9] R.E.Schapire: "The Boosting Approach to Machine Learning: An Overview," In D.D.Denison, M.H.Hansen, C.Holmes, B.Mallick, and B.Yu, editors: *Nonlinear Estimation and Classification*, Springer (2003).

[10] I.Hayashi, and S.Tsuruse: "A Proposal of Boosting Algorithm for Brain-Computer Interface Using Probabilistic Data Interpolation," *IEICE Technical Report, Neurocomputing*, Vol.109, No.461, pp.303-308 (2010) (in Japanese).

[11] I.Hayashi, and S.Tsuruse: "A Proposal of Boosting Algorithm by Probabilistic Data Interpolation for Brain-Computer Interface," *Proceedings of the 26th Fuzzy System Symposium*, pp.288-291 (2010) (in Japanese).

[12] Isao Hayashi, Shinji Tsuruse, Junichi Suzuki, and Robert Thijs Kozma: "A Proposal for Applying pdi-Boosting to Brain-Computer Interfaces," *Proceedings of 2012 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE2012) in 2012 IEEE World Congress on Computational Intelligence (WCCI2012)*, pp.635-640 (2012).