

確率的データ補間を用いたBCIのためのBoostingアルゴリズムの提案

林 勲[†] 鶴背 慎二^{††}

^{†, ††} 関西大学 総合情報学部

E-mail: [†] ihaya@cbii.kutc.kansai-u.ac.jp, ^{††} suraimuppoi@yahoo.co.jp

あらまし 近年, 脳からの情報により機械を制御する BCI(Brain-computer interface) や BMI(Brain-machine interface) の研究が盛んに行われている. 近赤外分光法 (NIRS) や脳波計測の装置により脳活動を計測し, 得られたデータのクラスター間の相違を判別モデルにより判別して機器を制御する. しかし, 従来の判別モデルでは, データクラスターを構成するのに多量のデータが必要であり, 機器環境の動的な変化には追従できない場合がある. 本論文では, Boosting アルゴリズムを用いて, 確率的データ補間による新たな BCI における判別法を提案する. Boosting 法におけるデータ付与の重みによるデータの選別ではなく, 脳活動データの特異性を考慮して定義された確率密度関数に応じて補間データを誤判別データの近傍に補充しデータを構成する. このデータ補間の Boosting 法により, 脳活動データが少ない場合や環境の変動にも適応できるロバスト性を有する判別関数を得ることができる. ここでは, 本手法を NIRS 計測装置を用いた脳活動計測の実験に適用し, その有用性を検討する.

キーワード BCI, Boosting アルゴリズム, データ補間

A Proposal of Boosting Algorithm for Brain-Computer Interface Using Probabilistic Data Interpolation

Isao HAYASHI[†] and Shinji TSURUSE^{††}

^{†, ††} Faculty of Informatics, Kansai University

E-mail: [†] ihaya@cbii.kutc.kansai-u.ac.jp, ^{††} suraimuppoi@yahoo.co.jp

Abstract Brain-computer interface(BCI) and brain-machine interface(BMI) have been come into the research limelight. The outer computer and machine are controlled by brain activity and the discriminated boundary, which are detected with near-infrared spectroscopy(NIRS) or electroencephalograph(EEG). However, in general, a large amount of activity data are necessary to determine the discriminated boundary in the conventional discriminant models. In this paper, we propose a new boosting algorithm for BCI using probabilistic data interpolation. In our model, interpolated data are generated by probabilistic distribution and assorted around errors instead of weights in the conventional Adaboost. By the interpolated data, the discriminated boundary is identified to control the outer machine effectively. We apply our method to arithmetic test with NIRS, and discuss the usefulness of our method.

Key words Brain-Computer Interface, Boosting Algorithm, Data Interpolation

1. はじめに

近年, 脳からの情報により機械を制御する BCI(Brain-computer interface) や BMI(brain-machine interface) の研究 [1-3] が盛んに行われている. 一般に, BCI や BMI で機器を制御するには, その機器を制御するのに必要な信号を脳活動として検出し, 制御行動ごとに判別式を同定して制御信号を出力する必要がある. 具体的には, 近赤外分光法 (NIRS) [4] や脳波計測 [5] の装置から脳活動信号を計測し, 得られたデータのクラスター間の相違を判別モデルで同定して, 脳活動信号を

判別モデルで判別して機器を制御する. この判別モデルとして, 線形判別式や SVM などを用いる場合が多い. しかし, 従来の判別モデルでは, データクラスターを構成するのに多量のデータが必要となる. また, BCI や BMI の動作環境下では, 環境が動的に変動する場合が多く, そのため, 過去に検出された脳活動データを用いて判別式を同定しても, その後の動的な環境変化に追従できない場合がある. 一方, 近年, パターン識別の分野において, 学習データを複数個に分割し, 各学習データに対する判別器を同定して, それらを統合的に組み合わせることにより全体の判別精度を向上させるアンサンブル学習 [6, 7] に対す

関心が高まっている．アンサンブル学習の一手法に Boosting アルゴリズムや Adaboost が提案されている [8]．Adaboost では，データに付与する重みを定義し，その重みにより選択的にデータを抽出して学習データを構成し，複数個の判別器を同定する．同定された判別器を統合的に組み合わせることによって，全体の判別精度を高めることが可能となる．

本論文では，Boosting アルゴリズムを用いて，確率的データ補間による新たな BCI における判別法を提案する．Adaboost のようなデータ付与の重みによるデータの選別ではなく，脳活動データの特異性を考慮して定義された確率密度関数により補間データを誤判別データの近傍に補充しデータを構成する．補間された脳信号データは，脳活動の時系列の特異性を考慮して，その誤判別データと同クラスターに分類される．Adaboost では，学習データ量は増加されないが，本手法では補間データを補填するので，データ数不足による判別器の信頼度低下を防ぐことができる．また，多量のデータを必要としないことから，環境が変動する場合でもモデル同定が可能となる．つまり，データ補間の Boosting 法により，NIRS や脳波計測装置において，脳活動データが少ない場合や環境の変動にも適応できるロバスト性を有する判別関数を得ることができる．ここでは，本手法を NIRS 計測装置を用いた脳活動計測の実験に適用し，その有用性を検討する．

2. Boosting アルゴリズム

アンサンブル学習とは，判別器を複数個用意し，各判別結果を統合することによりロバスト性の高い統合的な判別モデルを構築する一手法である．

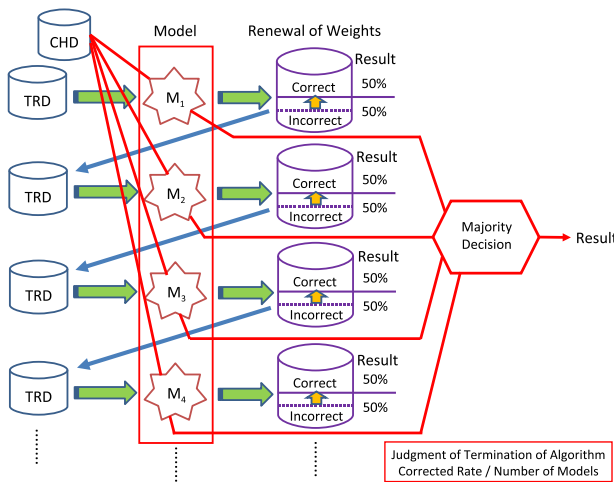


図 1 Adaboost

まず，Adaboost を説明する．Adaboost の概念図を図 1 に示す．Adaboost では，単純な識別器 $M_1, M_2, \dots, M_i, \dots$ を設定し，学習データ (TRD) に対して各モデルを同定して結果を算出する．判別率が 50% 以下になる場合には TRD のクラスを反転する．次ステップでのモデル M_i の TRD を構成する際に，全体のデータ量を変えずに，誤判別となったデータが全データ中で 50% 以上になるように増加させる．

いま，データ集合 D に第 s 番目のデータが存在し，その第 s 番目のデータに重み $w_i(s)$, $s = 1, 2, \dots, N$ が付与されているとする．まず，誤判別されたデータの総和 W_i を求める

$$W_i = \sum_{s=1}^N w_i(s) I(C(s) \neq O(s)) \quad (1)$$

ただし， $C(s)$ は TRD の教師クラス， $O(s)$ は判別器の出力であり，

$$I(C(s) \neq O(s)) = \begin{cases} 1 & ; \text{ if } C(s) \neq O(s) \\ 0 & ; \text{ if } C(s) = O(s) \end{cases} \quad (2)$$

である．

データ再構築の際には，重み $w_i(s)$ を次式で更新してデータを選択し，誤判別のデータを増加させる．

$$w_i(s) = \begin{cases} \frac{w_{i-1}(s)}{2(1-W_{i-1})} & ; C(s) \neq O(s) \\ \frac{w_{i-1}(s)}{2W_{i-1}} & ; C(s) = O(s) \end{cases} \quad (3)$$

ただし， $\sum_{s=1}^N w_i(s) = 1$ であり， $w_1(s) = 1/N$ である．

その後，評価用データ (CHD) を判別器に入力して認識率を得る．これらの手順を繰り返し，規定の繰り返し回数終了した時点で，これらの出力を多数決で勘案して統合結果を得る．

3. データ補間を用いた Boosting アルゴリズム

Adaboost は，データ選択の重みをデータに付与するが，提案するアルゴリズムでは，確率密度関数に応じた補間データを誤判別のデータ近傍に補充し，誤判別のデータを増加させる．

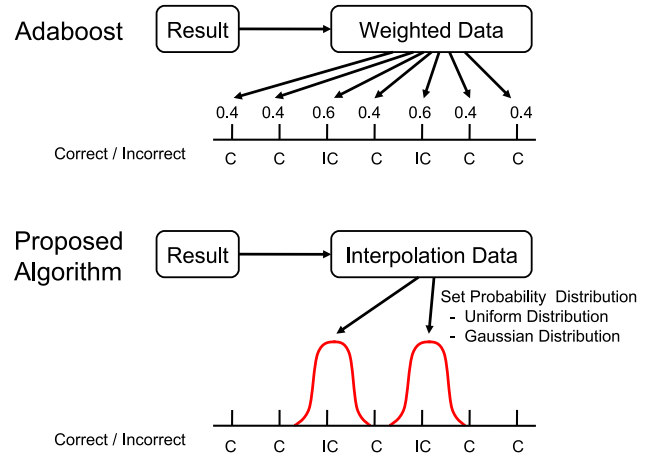


図 2 Selection of Error Data by Data Interpolation

図 2 に本手法のデータ選択法を説明する．いま，TRD において，誤判別されたデータが第 s 番目とする．この第 s 番目のデータの第 j 番目の属性値を $x_j(s)$ で表す．補間データ $x_j^{new}(s)$ は，属性値 $x_j(s)$ を平均とする確率密度関数 $f(x_j)$ から，その確率に基づいて選択される．

$$x_j^{new}(s) = \{x_j \in A \mid P(A) = \int_A f(x_j) dx_j\} \quad (4)$$

確率密度関数 $f(x_j)$ として，正規分布や一様分布などを設定

する．ここでは，下記の一様分布を用いる．

$$f(x_j) = \begin{cases} \frac{1}{x_j^{max} - x_j^{min}} & ; \text{ for } x_j^{min} \leq x_j \leq x_j^{max} \\ 0 & ; \text{ for } x_j < x_j^{min} \text{ or } x_j > x_j^{max} \end{cases} \quad (5)$$

ただし， x_j^{max} と x_j^{min} は，次の値とする．

$$x_j^{min} = \frac{3x_j(s) + x_j(s-1)}{4} \quad (6)$$

$$x_j^{max} = \frac{3x_j(s) + x_j(s+1)}{4} \quad (7)$$

ここで， $x_j(s-1)$ は第 $s-1$ 番目のデータ， $x_j(s+1)$ は第 $s+1$ 番目のデータである．

確率的データ補間法により，誤判別されたデータ付近に新たなデータが増加され，次ステップで同定される判別器は，より誤判別のデータに適合する判別器が得られる．Adaboost では，脳信号データの絶対量が変化しないのに対して，本手法は，脳信号データの絶対量を増やすことができるので，脳信号データの不足によるモデルの信頼性低下を防ぐことができる．また，BCI や BMI において，制御環境が動的に変動する場合でも判別モデルのためのデータ量が少なくても良いので，環境の変動に影響されない利点がある．

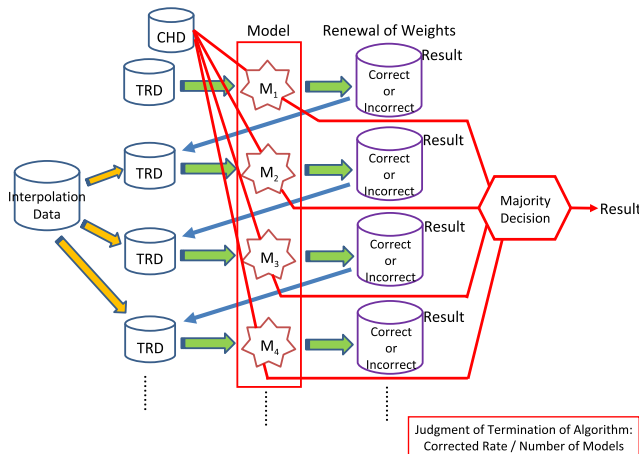


図 3 Proposed Boosting Algorithm

本手法の概念図を図 3 に示す．また，本手法のアルゴリズムを以下に示す．

Step 1 NIRS 計測装置の各計測点から W 個の脳信号データ D が得られたとし， D を学習に用いるための学習データ (D^{TRD} , W^{TRD} 個) と評価用に用いるための評価データ (D^{CHD} , W^{CHD} 個) に分割する．ここで， $W^{TRD} + W^{CHD} = W$ である．また， D から構成される補間データを D^{INT} とする．

Step 2 D^{TRD} を第 i 番目の判別器 M_i に入力し，結果 R_i の認識率 r_i^{TRD} を得る．

Step 3 D^{TRD} において，誤判別された第 s 番目のデータを抽出する．第 s 番目のデータの第 j 番目の属性値 $x_j(s)$ に対して，式 (5) ~ 式 (7) により，確率密度関数 $f(x_j)$ を設定し，補間データ $x_j^{new}(s)$ を発生させて， D^{INT} に保存する．

Step 4 結果 R_i において，正判別したデータと誤判別したデータとが同数になるように， D^{INT} からデータを補充する．ただし，補充データ数 d は

$$d \geq \frac{W}{2} - W^{TRD}(1 - r_i^{TRD}) \quad (8)$$

とする．

Step 5 D^{CHD} を M_1, M_2, \dots, M_i に適応して，結果の認識率 r_i^{CHD} を得る．

Step 6 Step2 から 5 までを繰り返し，あらかじめ決められたしきい値 θ に対して， $r_i^{CHD} \geq \theta$ を満足した時点，あるいは，繰り返し回数 M に対して， $i \geq M$ を満足した時点でアルゴリズムを終了する．

4. 計算タスクの NIRS 信号計測実験

本手法の有用性を検討するため，被験者に簡易計算のタスクを課し，NIRS 計測装置を用いて脳信号を計測した．NIRS 計測装置とは，透過性の高い近赤外光を外部から照射し，組織を透過してきた光の成分を分析して血液成分の吸光度変化量を測定する装置である．本研究では，提案手法の有用性を検討することが目的であり，その意味では，比較的少数の電極の方が手法の有用性を検討する際に確実であるので，2ch の簡易型 NIRS 計測装置 (エクセルオプメカトロニクス社，YN-502) を用いた．本装置は，周期 0.098s で，波長 770nm と 840nm の 2ch の電極を持ち，それぞれの電極で酸化ヘモグロビン (oxy-Hb) と還元ヘモグロビン (doxy-Hb) を計測するので，全体で 4 入力の信号データを構成する．



図 4 Experiment

NIRS 計測の対象タスクとして，小学一年生程度の桁上がりのない加算を被験者に課した．図 4 に実験風景を示し，図 5 に算術計算テストの問題を示す．計測実験は，暗幕で仕切られた空間内で行い，被験者には，計測前に年齢，性別，利き腕，健

| 繰り返り上の新しい値は | |
|-------------|---------|
| 1 + 6 = | 1 + 4 = |
| 4 + 0 = | 5 + 3 = |
| 8 + 1 = | 3 + 4 = |
| 8 + 0 = | 3 + 2 = |
| 2 + 5 = | 2 + 6 = |
| 6 + 3 = | 4 + 1 = |
| 0 + 2 = | 3 + 5 = |
| 4 + 4 = | 6 + 3 = |
| 6 + 3 = | 5 + 1 = |
| 3 + 5 = | 1 + 4 = |
| 5 + 0 = | 8 + 1 = |
| 2 + 1 = | 3 + 5 = |
| 2 + 6 = | 2 + 4 = |
| 3 + 0 = | 1 + 1 = |
| 9 + 0 = | 6 + 0 = |
| 3 + 0 = | 8 + 0 = |
| 3 + 0 = | 8 + 1 = |
| 7 + 1 = | 9 + 0 = |
| 9 + 0 = | 2 + 6 = |
| 4 + 3 = | 4 + 3 = |
| 0 + 6 = | 6 + 2 = |
| 5 + 2 = | 7 + 1 = |
| 8 + 1 = | 1 + 1 = |
| 7 + 0 = | 7 + 1 = |
| 5 + 1 = | 2 + 1 = |

図 5 Arithmetic Test

康状態，特殊な計算技能の訓練の有無等を確認して，特殊な計算技能を習得していない被験者のみを選択した．また，試行中のレストでは，背筋を伸ばし体の力を抜き，正面の注視点を注視することを伝え，さらに，タスク中では，作業に集中して，できる限り速く計算を行うことを伝えた．

計測実験のタスクは，前レストを 10s，算術計算の実行を 10s，後レスト区間の 30s を 1 試行として構成し，5 試行を 1 セットとする合計 250s で構成する．また，計算の開始と終了，1 試行の終了をベル音で知らせた．図 6 に算術計算実験のタイミングプロトコルを示す．

被験者は男 1 名，女 2 名の計 3 名とし，各被験者に対して数日間をかけて 12 セットを実施した．なお，国際 10-20 法に基づき，NIRS 計測点は，測定位置 F_{p1} と F_{p2} (大脳前頭前野域) の 2 箇所を設置した．採取したデータから，乱数により 1 人 10 試行ずつのデータを抽出し，脳信号のデータ集合 D を構成した．

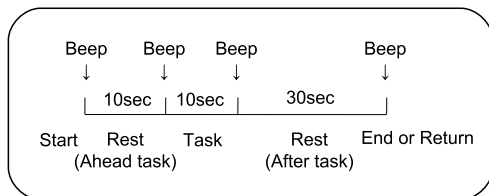


図 6 Timing Protocol of Experiment

5. 結果と考察

まず，NIRS により検出された被験者 C の脳血流量変化量の時系列の一例を図 7 と図 8 に示す．図 7 は右電極の酸化ヘモ

グロビン (oxy-Hb) と還元ヘモグロビン (doxy-Hb) を示し，図 8 は左電極の oxy-Hb と doxy-Hb を示す．脳血流量変化量は，データによって多少の誤差が存在するものの，タスク実行開始とともに，oxy-Hb の変化量の増加が始まり，doxy-Hb 変化量の減少が起こり，タスク実行終了の合図とともに緩やかに平常状態に戻っていくのがわかる．ここでは，このような特徴を持つ脳信号データを用いて，被験者がタスク実行中であるか，あるいは，レスト中であるかを判別器により同定する．

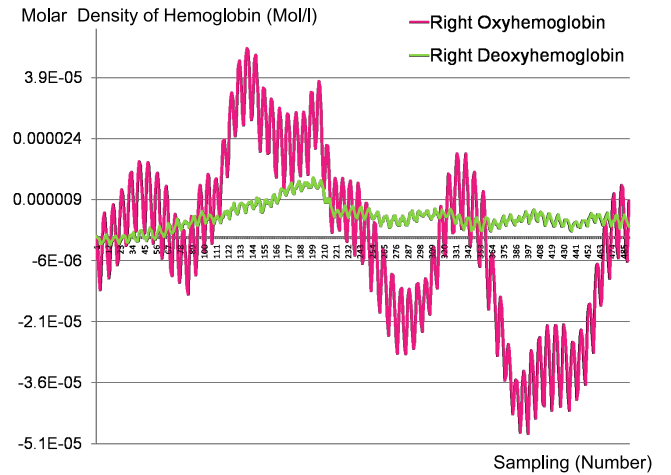


図 7 Waveforms of Right Chanel Detected by NIRS

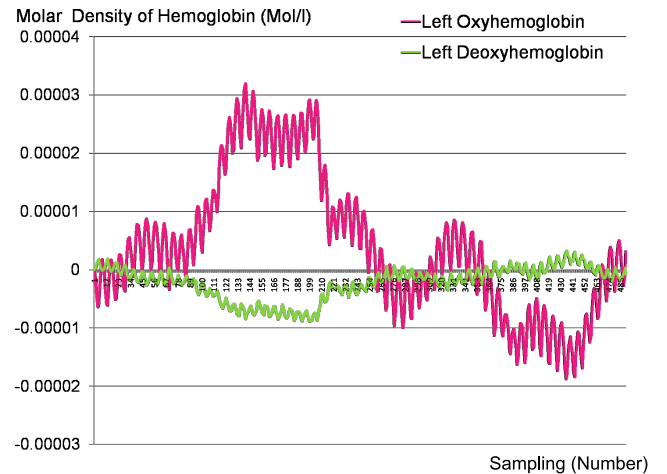


図 8 Waveforms of Left Chanel Detected by NIRS

さて，定式化した本手法のアルゴリズムの手順にしたがって，NIRS による脳信号データの同定の手順を説明しよう．ただし，判別器として決定木法の一つである REPtree を用い，終了判定は $M = 3$ とする回数による終了判定を採用した．

NIRS 計測装置の周期が $0.098s$ であり，1 試行が $50s$ であるので，全データ集合 D として， $W = 490$ 個の脳信号データが採取される．ただし，本実験では， $D = D^{TRD} = D^{CHD}$ とする．したがって， $W = W^{TRD} = W^{CHD} = 490$ となる．

まず， D_1^{TRD} を第 1 番目の判別器 M_1 に入力し，結果 R_1 の認識率 $r_1^{TRD} = r_1^{CHD} = 91.22\%$ を得る．次に，補間データを

作成する． R_1 における誤判別データ数は43個となり，式(5)～式(7)によって，各誤判別データの周辺に補間データ $x_j^{new}(s)$ を発生させ D^{INT} に保存する． D^{INT} から， R_1 における誤判別データに補間データを加えた個数が正判別データの個数と等しくなるまで，補間データを加える．この新たなデータ集合 D_2^{TRD} は， R_1 における補間データ数 d_1 が404個であるので， $W_2^{TRD} = 894$ 個となる．

D_2^{TRD} を第2番目の判別器 M_2 に入力し，結果 R_2 の認識率 $r_2^{CHD} = 92.24\%$ を得る．この結果，新たに補充するデータ数は $d_2 = 414$ となり， M_3 における D_3^{TRD} を $W_3^{TRD} = 904$ として構成する． M_3 による認識率は $r_3^{TRD} = 93.67\%$ となるが，終了判定の $M = 3$ により，アルゴリズムを終了する．

以上の経過から，それぞれの結果 R_1, R_2, R_3 が得られ，多数決原理により結果を統合して，最終の認識率 95.31% を得る．この各ステップでの補間データの数量と認識率を表1に示す．また，原データと補間データのデータ間の比較を表2に示す．補間データのデータ間距離は，原データの25.0%となっているので， $Rightoxy - Hb(r - oxy)$ では，それぞれ， $4.5e-06$ と $1.1e-06$ となっている．また，補間データのデータ間距離の最大値は，原データの平均値よりも，それぞれ， $0.8e-06 \sim 1.0e-07$ ほど小さく，密度が高くデータを補間していることがわかる．さらに，それぞれのデータ距離の最小値が 0.0 となっている．これは，時系列データにおいて，隣接するサンプリングデータの脳血流量変化量が同値になっていることを示している．

表1 Correct Rate and Number of Interpolation Data

| Model | Corrected Rate(%) | Interpolation Data(for next step) | TRD | CHD | Correct Rate of T.R.(%) |
|-------|-------------------|-----------------------------------|-----|-----|-------------------------|
| M_1 | 91.22 | 404 | 490 | 490 | 95.31 |
| M_2 | 92.24 | 414 | 894 | 490 | |
| M_3 | 93.67 | 428 | 904 | 490 | |

表2 Comparison between Original Data and Interpolation Data

| Data | Vari. | Min | Max | Average |
|--------------------|--------|-----------|-----------|-----------|
| Original Data | r-oxy | 0.0 | $1.1e-05$ | $4.5e-06$ |
| | r-doxy | 0.0 | $2.5e-06$ | $7.4e-07$ |
| | l-oxy | 0.0 | $1.7e-06$ | $5.7e-07$ |
| | l-doxy | $3.0e-08$ | $5.5e-06$ | $2.2e-06$ |
| Interpolation Data | r-oxy | 0.0 | $2.7e-06$ | $1.1e-06$ |
| | r-doxy | 0.0 | $6.3e-07$ | $1.8e-07$ |
| | l-oxy | 0.0 | $4.2e-07$ | $1.4e-07$ |
| | l-doxy | $7.5e-09$ | $1.4e-06$ | $5.4e-07$ |

次に，3名の被験者の解析結果について述べる．各ステップごとの平均認識率を表3に示す．被験者 B において， r_1^{CHD} の識別率は 93.69% であり， $r_2^{CHD} = 93.47\%$ ， $r_3^{CHD} = 93.33\%$ であるので，第1の判別モデルの認識率より下回っているが，統合の認識率は，どの判別モデルの認識率よりも 1.0% 以上で高い認識率を示している．この傾向は，被験者 C において，最も顕著に表れているが，これは判別モデルが各ステップごとに特

異なデータに適合し，統合結果として認識率が高まる Boosting の特徴を示している．すなわち，NIRS データに対する本手法の有用性を示している．

表3 Correct Rate for Subjects

| Subject | $r_1^{CHD}(\%)$ | $r_2^{CHD}(\%)$ | $r_3^{CHD}(\%)$ | 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 |
| Average | 92.13 | 91.62 | 92.06 | 93.10 |

最後に，本手法と REPtree との比較，および，他の Boosting 手法との比較を行った．比較の結果を表4に示す．本手法と REPtree との比較では， 0.97% の認識率の向上がみられ，有意差 0.05% 水準での t 検定において，有意 ($p = 0.01616$) な差が認められた．

次に，判別器を REPtree とし，終了判定を $M = 3$ とする同条件下において，他の Boosting 手法との比較を行った．被験者 A に対しては，Adaboost と比較して 0.3% の認識率の向上が見られ，Multiboost と比較しても 1.7% の向上が見られた．また，被験者 B に対しては，同様に，Adaboost と比較して 0.3% の向上が見られ，Multiboost と比較しても 1.7% の向上が見られた．しかし，被験者 C に対しては，Multiboost と比較して 1.2% の向上が見られたものの，Adaboost と比較して 0.33% の低下が見られた．3名の被験者に対する比較では，Multiboost に対して，有意差 0.05% 水準での t 検定において有意 ($p = 0.0006860$) な差が認められた．一方，Adaboost に対しては， 0.56% の向上が見られたが，有意差 0.05% 水準での t 検定において有意 ($p = 0.1578$) な差は認められなかった．

以上の結果から，本手法は，他の Boosting 手法と比較して同等以上の認識率を確保することができるといえる．さらに，本手法は，脳信号データが少量の場合でも適応できるので，BCI の実用化に向けて有用な手法ともいえる．

表4 Comparison between Proposed Model and Others

| Subject | Proposed Model(%) | 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 |
| Average | 93.10 | 92.54 | 92.14 | 92.13 |

6. おわりに

本論文では，Boosting アルゴリズムを用いて，確率的データ補間による新たな BCI における判別法を提案して定式化した．さらに，本手法を NIRS 計測装置を用いた脳活動計測の実験に適用し，他手法との比較により有用性を検討した．今後，選択確率分布の設定方法やロバスト性について，さらに議論する必要がある．

なお，本研究の一部は「文部科学省私立大学戦略的研究基

盤形成支援事業（平成 20 年度～平成 24 年度）」によって行われた。

文 献

- [1] M.A.Lebedev, J.M.Carmena, J.E.O'Doherty, M.Zaksenhouse, 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] L.R.Hochberg, M.D.Serruya, G.M.Friebs, J.A.Mukand, M.Saleh, A.H.Caplan, A.Branner, D.Chen, R.D.Penn, J.P.Donoghue, "Neuronal ensemble control of prosthetic devices by a human with tetraplegia," *Nature*, Vol.442, pp.164-173 (2006).
- [3] T.O.Zander, C.kothe, S.Welke, and M.Roetting: Enhancing Human-Machine Systems with Secondary Input from Passive Brain-Computer Interfaces, *Proceeding of 4th International BCI Workshop*, pp.44-49 (2008).
- [4] 新出, 坪根, 和田: 近赤外線分光法を用いた動作肢の判別, 信学技報, No.NC2007-44, pp.9-96 (2008)
- [5] 井上: 脳波からヒトの状態を探る, 日本知能情報ファジィ学会第 5 回脳と知覚研究部会ワークショップ特別講演 (2008)
- [6] 金森, 畑埜, 渡辺: ブースティング, 森北出版 (2006)
- [7] 中島, 荘司: ファジィアンサンブル識別器構築におけるデータ分割の効果, 第 25 回ファジィシステムシンポジウム講演論文集, No.3E2-01 (2009)
- [8] 村田, 金森, 竹ノ内: ブースティングと学習アルゴリズム: 三人寄れば文殊の知恵は本当か? 電子情報通信学会誌, Vol.88, No.9, PP.724-729 (2005)