

ファジィデータ補間によるブースティングアルゴリズムの提案と 脳コンピュータインタフェースへの応用

林 勲 関西大学大学院 総合情報学研究科
鶴背 慎二 パナソニックシステムネットワークス (株)

近年、脳からの情報により外部機器を制御する BCI や BMI の研究が盛んに行われている。これらのシステムでは、近赤外分光法 (NIRS) や脳波計測装置 (EEG) により脳活動信号を計測し、識別モデルにより信号クラスターを識別して外部機器を制御する。本論文では、ファジィデータ補間によるブースティングアルゴリズムを提案する。誤識別の領域を定義したメンバシップ関数により誤識別データの近傍に補間データを発生してデータの識別精度を向上させ、3 種類の評価基準の加重平均により補間データのクラスを決定する。本手法では、誤識別の周辺に補間データを発生させるので、誤識別データとその周辺に発生した補間データの全てのデータにより識別線が推定され、その識別線が滑らかで評価データの認識率が良い。ここでは、NIRS を模擬した数値例により本手法の特性を議論し、NIRS 計測装置による脳活動計測実験に適用し、その有用性を検討する。

キーワード：脳コンピュータインタフェース、ブースティングアルゴリズム、バーチャルデータ

1. はじめに

近年、脳からの情報により外部機器を制御する BCI (Brain-Computer Interface) や BMI (Brain-Machine Interface) の研究 [1, 2] が盛んに行われている。多くは近赤外分光法 (NIRS) [3] や脳波計測装置により脳活動信号 [4] を計測し、識別モデルにより信号クラスターを識別して外部機器を制御する。しかし、計測時には、外部環境が動的に変動する場合が多く、従来の識別モデルでは環境変化に追従できない場合がある。また、変動する計測データにおいて、より新しい時系列データを用いて識別モデルを推定する際には、そのデータ数が少ないため精度を確保できない場合がある。一方、パターン分類問題に対して、複数個の識別器を構成しそれらを統合的に組み合わせることで全体の識別精度を向上させるアンサンブル学習 [5, 6] に対する関心が高まっている。アンサンブル学習の一手法に AdaBoost [7] がある。AdaBoost では、識別器の推定値と実データとの正誤結果から誤識別データの重みを更新し、その重みを用いて選択的にデータを抽出して次段階の識別機の学習データを構成する。この操作を繰り返し、最終的に、それぞれの識別器による多数決原理により評価データに対する統合の最終結果を算出する。このように、重みによって選択された誤識別の学習データにより、データ空間において局所的に認識率が高い識別器を多数構成し、これらの複数の識別器を評価データに適用させることで、全体データの識別率を向上させる。

本論文では、データ補間によるアンサンブル学習の新たな識別法を提案する。メンバシップ関数によってデータを補間するので、この手法を pdi-Boosting (Possibilistic Data Interpolation-Boosting) と呼ぶ [8-11]。pdi-Boosting は、AdaBoost のように重みによるデータ構成法ではなく、データの特長から誤識別の領域を定義したメンバシップ関数により誤識別データの近傍にバーチャルデータ (補間データ) を発生して、データの

識別精度を向上させる。AdaBoost では、重みにより特定の誤識別データが次段階で選択されるので、識別器による識別推定線は、その特定の誤識別データのみの影響で推定線が変更される。一方、pdi-Boosting では、誤識別の周辺に補間データを発生させるので、誤識別データとその周辺に発生した補間データの全てのデータにより推定線が変更される。そのため、pdi-Boosting は AdaBoost と比較して推定線が滑らかで評価データの識別率が良い。また、AdaBoost では、データ量は増加されないが、pdi-Boosting では補間データを補填するので、データ数不足による識別器の信頼度低下を防ぐことができる。さらに、pdi-Boosting では、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の同定が可能である。

しかし、pdi-Boosting では、補間データがメンバシップ値が低くかつ誤識別データから離れている場合には、そのデータのクラス決定に誤りが生じ、識別率を低下させることがある。そこで、ここでは補間データの識別クラスを決定する改良型の pdi-Boosting も定式化する。補間データのクラス決定には次の評価基準を組み合わせる。補間データ発生時のメンバシップ関数による誤識別データの評価値 (E_1)、補間データの識別クラスへの所属度による識別クラスの評価値 (E_2)、補間データの近傍クラスへの所属度による近傍クラスの評価値 (E_3) の 3 種類である。最終的な補間データの識別クラスは、これらの評価基準の加重平均により決定する。補間データが誤識別データ付近で発生する場合には、誤識別データの評価値 E_1 が大きくなり、識別クラスを中心周辺で発生する場合には、識別クラスの評価値 E_2 が大きくなる。また、近傍クラスにより評価値 E_3 の影響も受ける。このクラス決定法により、補間データはクラスの付け替えが発生し、識別率が向上することが期待できる。

ここでは、まず pdi-Boosting の定式化を行い、数値例により pdi-Boosting の特性を議論する。その後、pdi-Boosting を NIRS 計測装置による脳活動計測実験に適

用し、その有用性を検討する。最後に、クラス決定を用いた改良型 pdi-Boosting のアルゴリズムを定式化し、その有用性を数値例により検証する。

2. pdi-Boosting の定式化

アンサンブル学習とは、複数個の識別器を用意し、各識別結果を統合することにより高い識別率を得る手法である。pdi-Boosting の概念図を図 1 に示し、AdaBoost と pdi-Boosting の違いを図 2 に示す。AdaBoost のアルゴリズムでは、まず、学習データ (TRD) を用いて識別器 M_1 を学習し、 TRD の識別率を算出する。次ステップでは、データに付与した重みを用いて全データから誤識別データを 50% 以上選択し、新たな TRD を構成して、識別器 M_2 により TRD の識別率を算出する。終了判定が満足されるまでこの一連の操作を繰り返し、最後に、評価用データ (CHD) を L 個の識別器 $M_1, M_2, \dots, M_i, \dots, M_L$ に入力して、多数決により最終結果を得る。一方、pdi-Boosting では、メンバシップ関数を用いてバーチャルデータ (補間データ) を誤識別データの近傍に補充し、 TRD を増加させる。 TRD を増加させることによって識別器の識別精度が向上し、その結果、最終結果の識別率も向上する。

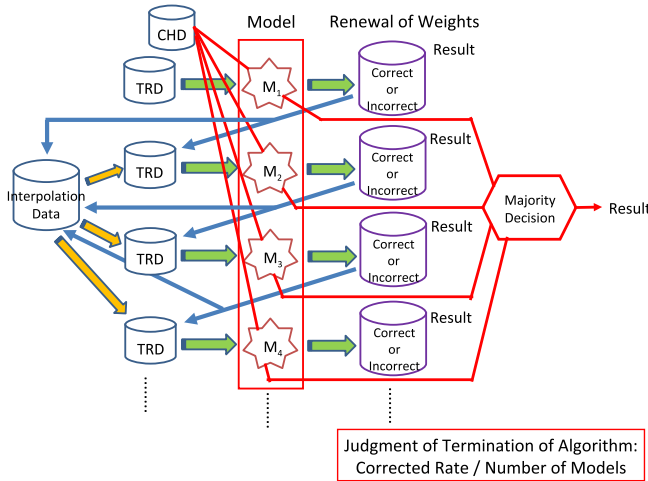


Fig. 1: pdi-Boosting Algorithm

さて、pdi-Boosting におけるデータの補間方法について説明しよう。 W 個のデータからなるデータ集合 D において、第 d 番目のデータを $\mathbf{x}(d) = (x_1(d), x_2(d), \dots, x_j(d), \dots, x_n(d))$ で表す。いま、学習データ D^{TRD} のデータ $\mathbf{x}(d)$ が誤識別されたとし、このデータを $\mathbf{x}^F(d)$ と記す。pdi-Boosting では、乱数により、ある実数 $h, 0 \leq h \leq 1$ が与えられると、ファジィ数 F のメンバシップ関数 $\mu_F(x_j)$ を用いて、補間データ $x_j^{int}(d)$ を次のように生成する。

$$x_j^{int}(d) = \{x_j \mid \mu_F(x_j) = h, \mu_F(x_j^F(d)) = 1\} \quad (1)$$

$$h \sim N(1, 1), \quad 0 \leq h \leq 1 \quad (2)$$

ただし、ファジィ数 F の中心は $x_j^F(d)$ である。

メンバシップ関数 $\mu_F(x_j)$ には、次のような L 関数や

正規分布、一様分布などがある。

$$\mu_F(x_j) = L\left(\frac{x_j - x_j^F(d)}{c}\right), \quad c \geq 0 \quad (3)$$

$$\mu_F(x_j) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_j - x_j^F(d))^2}{2\sigma^2}\right) \quad (4)$$

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

ただし、 c はメンバシップ関数の幅であり、 $L(x)$ 関数の例として $L(x) = \max(0, 1 - |x|)$ の三角型メンバシップ関数などがある。また、 σ は正規分布の標準偏差であり、一様分布の x_j^{max} と x_j^{min} は次の値である。

$$x_j^{min} = \frac{3x_j^F(d) + x_j^F(d_{min})}{4} \quad (6)$$

$$x_j^{max} = \frac{3x_j^F(d) + x_j^F(d_{max})}{4} \quad (7)$$

$$x_j^F(d_{min}) = \arg \min_d (x_j^F(d) - x_j(d)), \quad x_j^F(d) - x_j(d) \geq 0 \quad (8)$$

$$x_j^F(d_{max}) = \arg \min_d (x_j(d) - x_j^F(d)), \quad x_j(d) - x_j^F(d) \geq 0 \quad (9)$$

ここで、 x_j^{min} と x_j^{max} はそれぞれ $x_j^F(d)$ から隣接する x_j への内分点であるが、3 : 1 の比率は経験値によるものである。

pdi-Boosting のアルゴリズムを以下に示す。

- Step 1** 脳信号の離散データ D (個数: W 個) を学習データ D^{TRD} (個数: W^{TRD} 個) と評価データ D^{CHD} (個数: W^{CHD} 個) に分割する。また、 D から構成される補間データを D^{INT} で表す。
- Step 2** 第 i 番目の識別器 M_i に D^{TRD} を入力し、第 i 番目の結果 R_i の識別率 r_i^{TRD} を得る。
- Step 3** 誤識別された第 d 番目のデータを D^{TRD} から一時的に抽出する。誤識別データの第 j 番目の属性値 $x_j^F(d)$ に対して、メンバシップ関数 $\mu_F(x_j)$ により補間データ $x_j^{int}(d)$ を発生し、 D^{INT} に追加する。
- Step 4** 結果 R_i での正識別データと誤識別データが同数になるように、乱数により D^{INT} から $v \geq \frac{W}{2} - W^{TRD}(1 - r_i^{TRD})$ 個の補間データを取り出し D^{TRD} に加える。
- Step 5** $i = i + 1$ として Step 2 から 4 までを繰り返し、しきい値 θ に対して $r_i^{CHD} \geq \theta$ を満足した $K = i$ の時点、あるいは、識別器の個数 L と繰り返し回数 $K, K \leq L$ に対して $i \geq K$ を満足した時点でアルゴリズムを終了する。
- Step 6** $M_1, M_2, \dots, M_i, \dots, M_K$ に D^{CHD} を適用し、多数決により結果の識別率 r_K^{CHD} を得る。

3. pdi-Boosting の特性検証

脳信号データを模した数値データを作成し、pdi-Boosting の特性を検討した。ここでは、脳活動状態の 2 識別問題

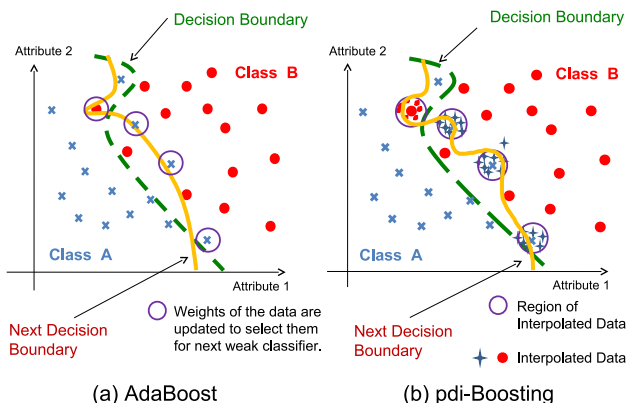


Fig. 2: AdaBoost and pdi-Boosting

を取り上げる。いま、脳信号の定常状態を0で表し、賦活状態を1で表す。また、外乱データを含む TRD と CHD を構成するため、標準偏差 s を変更して数種類の正規乱数を発生し、それぞれ500個の D^{TRD} と D^{CHD} を生成した。ここでは、補正データのメンバシップ関数 $\mu_F(x_j)$ は正規分布とし、識別器には REPTree、終了規範は繰り返し判定として、回数は $K = 3$ とした。

特性解析では、メンバシップ関数の標準偏差 σ と外乱データの標準偏差 s の組み合わせに対して、 TRD の個数 W^{TRD} を変更し、 $W^{CHD} = 500$ として10回の繰り返し試行により CHD の識別率を議論した。

$s : 0.2, 0.4, 0.6, 0.8, 1.0$
 $\sigma : 0.005, 0.01, 0.05, 0.1, 0.2, 0.6, 1.0$
 $W^{TRD} : 2, 5, 10, 20, 30, 50, 75, 100, 250, 350, 500$
 $W^{CHD} : 500$

外乱の標準偏差 s が0.4, 0.8, 1.0, メンバシップ関数の標準偏差 σ が0.05のときの pdi-Boosting と AdaBoost の識別率を図3に示す。なお、pdi-Boosting は実線で示し、AdaBoost は破線で示す。図3から、pdi-Boosting には次の特性があることがわかる。

- 1) データ数が極端に少ない場合でも AdaBoost と比較して CHD の識別率は低くない。
- 2) データ数が多い場合、識別率は AdaBoost よりも高く、データ個数による識別率の変動は小さい。
- 3) 外乱の標準偏差が大きい場合でも識別率の頑健性は高い。

1) において、図4に30個までの少量データでの外乱の標準偏差 $s = 0.8$ における pdi-Boosting と AdaBoost の識別率を示す。20個以下の CHD では、pdi-Boosting が AdaBoost よりも識別率が高いことがわかる。2) では、データ数が多い場合、pdi-Boosting は AdaBoost よりも識別率が高くかつ上昇している。また、 TRD が100個以上では、全ての外乱に対して、pdi-Boosting は AdaBoost よりも識別率が高い。3) では、外乱の標準偏差が大きいほど pdi-Boosting の識別率が高い。

特に、3) は重要である。図5に500個のデータに対して外乱を加えた場合の pdi-Boosting と AdaBoost の識別率の比較を示す。外乱の標準偏差 s が大きくなるにつれて、pdi-Boosting の識別率は低下している。しかし、 $s = 0.6$ 以上では、pdi-Boosting は AdaBoost よりも識別率が高い。これらの結果より、pdi-Boosting は外乱に対する頑健性が高いことがわかる。また、pdi-Boosting は AdaBoost よりも高い識別率を得ることがわかる。

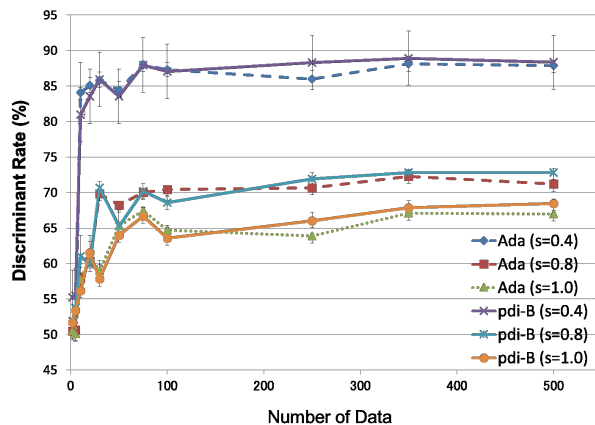


Fig. 3: Discriminant Rate by Changing Number of Data and S.D.

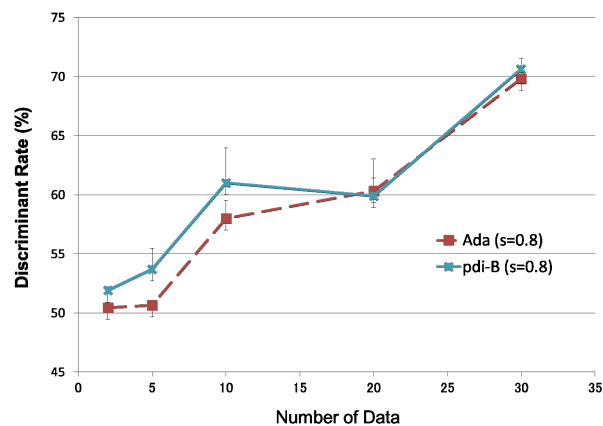


Fig. 4: Discriminant Rate of $s=0.8$

最後に、正規分布型メンバシップ関数の標準偏差 σ を変更し、補間データの発生領域との関係について検討を行った。誤識別データの位置を0.5として補間データを発生したときの分布状態を図6に示す。標準偏差 σ が大きくなるに伴い、補間データの発生領域が拡大しているが、 $\sigma = 1.0$ の場合でも約80%程度のデータが区間 $[0, 1]$ に含まれている。したがって、標準偏差が $\sigma = 1.0$ 程度では、補間データは誤識別データ付近に集中し、より広範囲な補間データが必要な場合には、さらに大きな標準偏差を与える必要があることがわかる。

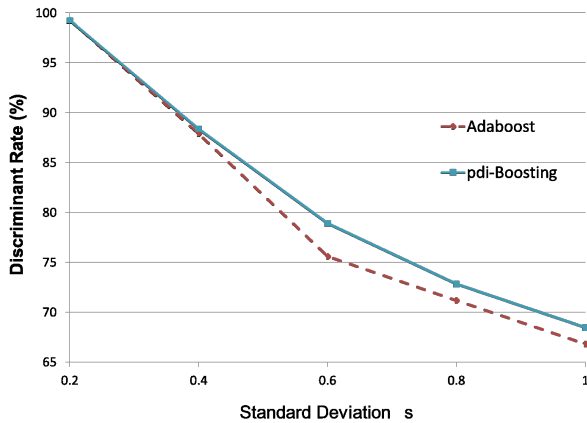


Fig. 5: Robustness of pdi-Boosting

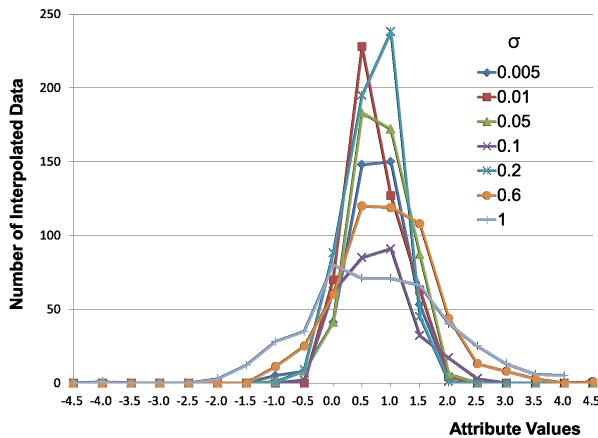


Fig. 6: Frequency Distribution of Interpolated Data

4. 計算タスクによる NIRS 脳信号計測

実際の脳活動データに対する pdi-Boosting の有用性を議論するため、被験者に簡易計算のタスクを課し、NIRS 計測装置で計測した被験者の脳信号を pdi-Boosting で識別した。ここでは、pdi-Boosting の有用性を検討することが目的であるので、識別程度を煩雑にすることなく、2プローブ測定装置である 2ch 簡易型 NIRS 計測装置 (エクセルオブメカトロニクス社, YN-502) を用いた。本装置は周期 $0.098s$ で波長 $770nm$ と $840nm$ の 2ch のプローブを持ち、それぞれのプローブで酸化ヘモグロビン (oxy-Hb) と還元ヘモグロビン (doxy-Hb) の変化を計測する。

被験者へのタスクとして、小学一年生程度の桁上がりのない加算を課した。図 7 に実験風景を示す。計測前に被験者の年齢、性別、利き腕、健康状態、特殊な計算技能訓練の有無等を確認し、特殊な計算技能を習得していない被験者を選択した。タスクでは、1 試行として前レストに $10s$ 、算術計算の実行時間に $10s$ 、後レストに $30s$ を設定し、5 試行を 1 セットとして休憩をとりなが

ら、各被験者に 12 セットを実施した。図 8 に算術計算タスクのタイミングプロトコルを示す。

計測のプローブ位置は国際 10-20 法の F_{p1} と F_{p2} (大脳前頭前野域) の 2 箇所とし、この 2 プローブで酸化ヘモグロビン (oxy-Hb) と還元ヘモグロビン (doxy-Hb) の変化を計測する。被験者は男性 1 名、女性 2 名の合計 3 名である。各被験者の 60 試行から正規乱数によって 10 試行の脳信号データを抽出し、30 試行を平均化して、NIRS 計測装置のプローブ数と周期 $0.098s$ によって、4 入力 1 クラスの約 490 個の脳信号データ集合 D を構成した。識別クラスは、被験者が算術計算状態か、あるいは、レスト状態かである。一例として、図 9 に被験者 C の左プローブの酸化ヘモグロビン (oxy-Hb) と還元ヘモグロビン (doxy-Hb) の変化量を示す。算術計算の開始 (前レストの終了) とともに、oxy-Hb の変化量が増加し doxy-Hb 変化量が減少し、算術計算の終了 (後レストの開始) 後に緩やかに平常状態に戻っていくのがわかる。



Fig. 7: Experiment Scenery

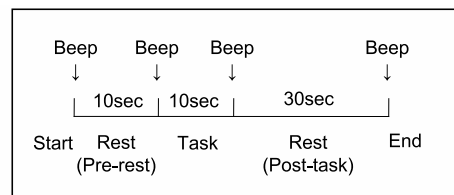


Fig. 8: Timing Protocol of Experiment

さて、pdi-Boosting を用いた脳信号データの識別について説明しよう。まず、補正データのメンバシップ関数 $\mu_F(x_j)$ には一様分布を用い、識別器には REPTree を用いる。終了規範の繰り返し回数は $K = 3$ とする。データ集合 D は $W = 490$ であり、 $D = D^{TRD} = D^{CHD}$ として $W = W^{TRD} = W^{CHD} = 490$ とする。

D_1^{TRD} を第 1 番目の識別器 M_1 に入力し、結果 R_1 の識別率 $r_1^{TRD} = r_1^{CHD} = 91.22\%$ を得る。 R_1 における誤識別データ数は 43 個であった。式 (1) と式 (2)、及び、式 (5)~式 (7) によって、各誤識別データの周辺にデータ $x_j^{int}(d)$ を発生させ、補間データとして D^{INT} に保存

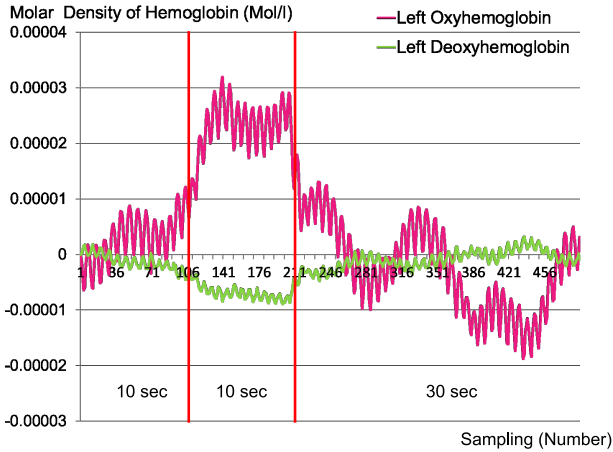


Fig. 9: Waveforms of Left Channel Detected by NIRS

する。誤識別データと補間データの合計個数が正識別データの個数と等しくなるまで、 D^{INT} から補間データを加える。 R_1 における補間データ数 v_1 は 404 個となり、新たなデータ集合 D_2^{TRD} は $W_2^{TRD} = 894$ 個となった。 D_2^{TRD} を第 2 番目の識別器 M_2 に入力し、結果 R_2 の識別率 $r_2^{CHD} = 92.24\%$ を得る。 R_2 における補間データ数は $v_2 = 414$ となり、 D_3^{TRD} は $W_3^{TRD} = 904$ となった。 R_3 では、 M_3 による識別率が $r_3^{TRD} = 93.67\%$ となり、終了判定 $M = 3$ によって学習が終了する。

多数決原理により識別器の結果 R_1 , R_2 , R_3 を統合し、最終結果の識別率として 95.31% を得る。各段階の補間データの個数と識別率を表 1 に示す。また、表 2 にそれぞれの被験者の認識率を示す。例えば、被験者 B の結果では、 $r_1^{CHD} = 93.69\%$, $r_2^{CHD} = 93.47\%$, $r_3^{CHD} = 93.33\%$ となり、各識別器の識別率は下がっていく傾向にあるが、最終の 3 つの識別器の統合の結果は 94.78% となり、それぞれの識別器よりも高い認識率となった。この傾向は、被験者 A や被験者 C , 被験者 D においても同様となった。例えば、被験者 C のタスク開始時と終了時の識別器の結果を図 10 に示す。識別器 M_1 は、タスク実行中と判断する範囲が広く、タスク実行の終了後もタスク実行と判断している。識別器 M_2 は、タスク実行と判断している範囲が狭い。識別器 M_3 は、タスク終了時の識別結果が悪い。これらの結果により、統合結果は、タスク前半とタスク中は M_3 を重視し、タスク後半では M_2 を重視し M_1 を加味している。それぞれの識別器を統合することにより、その統合の識別器が各データを補間的に識別し、さらに良い認識率を得る Boosting の特徴をよく表しているといえる。

最後に、pdi-Boosting と REPTree, 及び、他の Boosting 手法と認識率を比較する。各手法の被験者ごとの認識率の比較を表 3 に示す。それぞれの Boosting 手法と識別器である REPTree の比較では、AdaBoost の平均認識率は REPTree より 0.89% 高く、Mutiboost は 0.50% 低い。しかし、pdi-Boosting は REPTree より 0.91% 高い。

一方、pdi-Boosting と他の Boosting 手法の比較では、被験者 A に対して、pdi-Boosting の認識率は AdaBoost より 0.71% の向上が見られ、Mutiboost より 0.66% の向上が見られた。また、被験者 B に対して、AdaBoost

Table 1: Correct Rate and Number of Interpolation Data

Model	Corrected 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			

Table 2: Correct Rate for 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
D	94.57	91.65	94.24	95.42
Ave.	92.68	90.95	91.52	93.85

より 0.54% の向上が見られ、Multiboost より 1.21% の向上が見られた。しかし、被験者 C と被験者 D に対しては、Multiboost とはそれぞれ 1.2% , 1.85% だけ高いが、AdaBoost より 0.23% , 0.53% だけ低い。平均認識率の比較では、pdi-Boosting は Multiboost と比較して、 1.41% だけ高いが、有意差 0.05% 水準の t 検定ではわずかな差であるが有意 ($p = 0.064551$) な差は認められなかった。また、AdaBoost との比較では、 0.02% だけ高いが、有意差 0.05% 水準の t 検定では有意 ($p = 0.856851$) な差が認められなかった。さらに、pdi-Boosting と他の 3 手法との tukey 法による多重比較 [12] では、有意差 0.05% 水準で有意な差が認められなかった。

以上の結果から、認識率では、pdi-Boosting は他の Boosting 手法と比較して同等あるいは同等以上の識別率を得た。しかし、tukey 法の多重比較を用いた他手法との比較では、十分に良い結果が得られたとはいえなかった。今後はより多様なデータを用いて提案手法の特性をより詳細に検証する必要がある。

Table 3: Comparison between Proposed Model and Others

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
D	95.42	95.95	93.57	94.57
Ave.	94.26	94.24	92.85	93.35

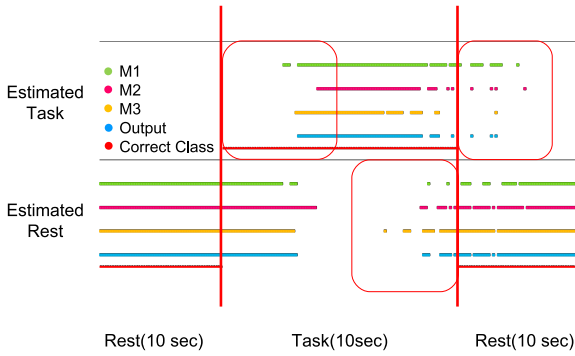


Fig. 10: Decision Boundary of Subject C

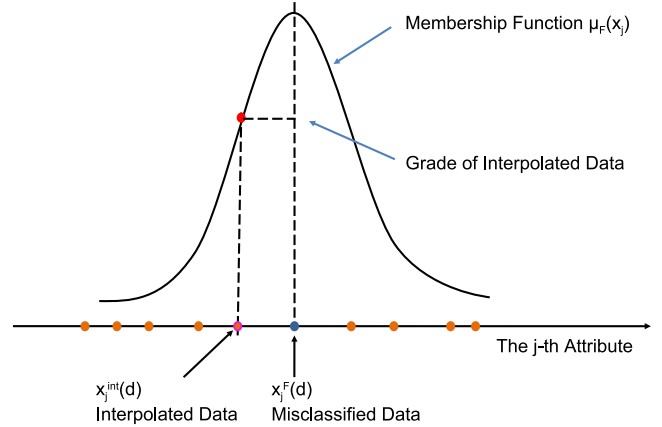


Fig. 11: Evaluation E_1

5. 改良型 pdi-Boosting の定式化

提案した pdi-Boosting では、誤識別データの周りに新たな補間データを発生することにより、認識率が向上した。しかし、補間データのクラスは必ずしも誤識別データと同じクラスとは限らない。そこで、ここでは、補間データの新たなクラス決定法を提案する。

いま、補間データ $x_j^{int}(d)$ が誤識別データ $x_j^F(d)$ から発生したとする。新たなクラス決定法では、補間データ $x_j^{int}(d)$ の新たなクラス k^* は、次の3つの評価基準：誤識別データの評価 (E_1)、識別クラスの評価 (E_2)、近傍クラスの評価 (E_3) から決定する。

(1) 誤識別データの評価 (E_1)

第 j 属性の評価値 E_{j1} は、次のメンバシップ値 $\mu_F(x_j^{int}(d))$ を用いて定義する (図 11 参照)。評価値 E_{j1} は補間データの誤識別データへの依存度を示し、この評価値 E_{j1} が小さい補間データほど、その誤識別データへの依存度が高い。

$$E_{j1}^k = \begin{cases} 1 - \mu_F(x_j^{int}(d)), & \text{for } k \in \{x_j^F(d)\} \\ \mu_F(x_j^{int}(d)), & \text{for } k \notin \{x_j^F(d)\} \end{cases}$$

(2) 識別クラスの評価 (E_2)

クラス k をもつ e 番目のデータを $x_j^k(e)$ 、クラス k の中心を x_c^k とするとき、第 j 属性の評価値 E_{j2} は補間データと各クラスの中心との距離を用いて定義する (図 12 参照)。この評価値 E_{j2} が小さい補間データほど、そのクラスへの依存度が高い。

$$E_{j2}^k = \frac{|x_j^{int}(d) - x_c^k| - \min_e |x_j^k(e) - x_c^k|}{\max_e |x_j^k(e) - x_c^k| - \min_e |x_j^k(e) - x_c^k|}, \text{ for } \forall d$$

(3) 近傍クラスの評価 (E_3)

第 j 属性の評価値 E_{j3} は、補間データと第 s 番目のデータ $x_j(s)$ や $x_j^k(e)$ との距離を用いて定義する (図 13

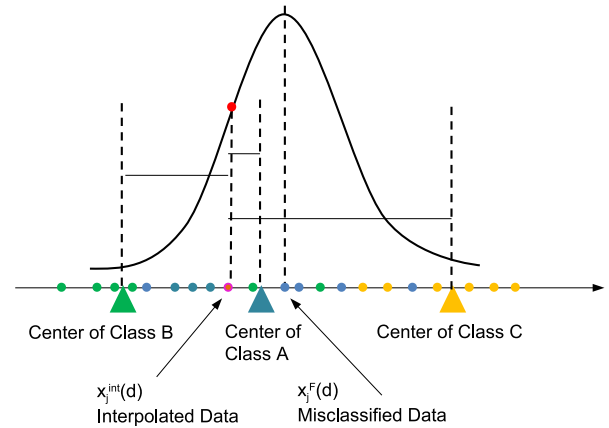


Fig. 12: Evaluation E_2

参照)。この評価値 E_{j3} が小さい補間データほど、その補間データに近いデータのクラスへの依存度が高い。

$$E_{j3}^k = \frac{\min_e |x_j^k(e) - x_j^{int}(d)| - \min_s |x_j(s) - x_j^{int}(d)|}{\max_s |x_j(s) - x_j^{int}(d)| - \min_s |x_j(s) - x_j^{int}(d)|}$$

for $\forall d$

これらの評価基準では、もし補間データが誤識別データ近傍で発生する場合には、その補間データのクラスへの評価 E_1 が高まり、あるクラスの中心近傍で発生する場合には、そのクラスへの評価 E_2 が高まる。一方、補間データの近傍データのクラスへの評価は E_3 で計算される。なお、補間データを発生させる生起関数は、確率に基づく正規分布ではなく可能性測度を表現するメンバシップ関数を想定している。したがって、データ個数に影響されることなく、出現した観測データは一つも排除せずすべての観測データから補間データを生成し、そ

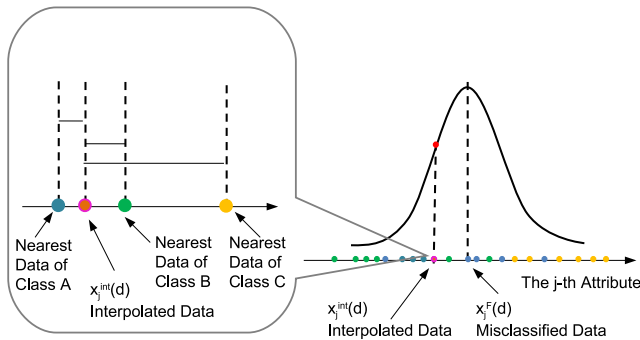


Fig. 13: Evaluation E_3

のクラスを付与する。

各 j 属性でこれらの 3 つの評価基準を統合し評価値 E_j^k を得て、 n 個の全属性を統合して全体の評価値 E^k を得る。 $x_j^{int}(d)$ のクラスは、評価値 E^k が最小となるクラス k^* として定義する。

$$k^* = \{k | \min_k E^k = \min_k \sum_{j=1}^n E_j^k\} \quad (10)$$

$$E_j^k = w_1 E_{j1}^k + w_2 E_{j2}^k + w_3 E_{j3}^k \quad (11)$$

ただし、 w_1, w_2, w_3 は各評価値の重みである。

改良型 pdi-Boosting のアルゴリズムを次のように定式化する。

- Step 1 脳信号の離散データ D (個数: W 個) を学習データ D^{TRD} (個数: W^{TRD} 個) と評価データ D^{CHD} (個数: W^{CHD} 個) に分割する。また、 D から構成される補間データを D^{INT} で表す。
- Step 2 第 i 番目の識別器 M_i に D^{TRD} を入力し、第 i 番目の結果 R_i の識別率 r_i^{TRD} を得る。
- Step 3 誤識別された第 d 番目のデータを D^{TRD} から一時的に抽出する。誤識別データの第 j 番目の属性値 $x_j^F(d)$ に対して、メンバシップ関数 $\mu_F(x_j)$ により補間データ $x_j^{int}(d)$ を発生させる。
- Step 4 式 (10) と (11) により、補間データ $x_j^{int}(d)$ のクラス k^* を求める。補間データ $x_j^{int}(d)$ を D^{INT} に追加する。
- Step 5 結果 R_i での正識別データと誤識別データが同数になるように、乱数により D^{INT} から $v \geq \frac{W}{2} - W^{TRD}(1 - r_i^{TRD})$ 個の補間データを取り出し D^{TRD} に加える。
- Step 6 $i = i + 1$ として Step 2 から 5 までを繰り返し、しきい値 θ に対して $r_i^{CHD} \geq \theta$ を満足した $K = i$ の時点、あるいは、識別器の個数 L と繰り返し回数 K , $K \leq L$ に対して $i \geq K$ を満足した時点でアルゴリズムを終了する。
- Step 7 $M_1, M_2, \dots, M_i, \dots, M_K$ に D^{CHD} を適用し、多数決により結果の識別率 r_K^{CHD} を得る。

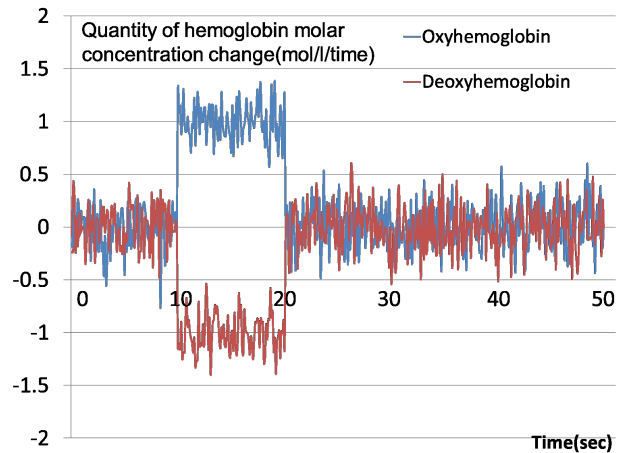


Fig. 14: Data of $s=0.2$

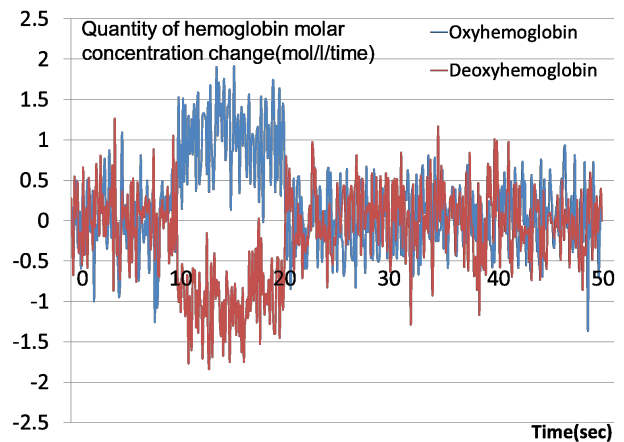


Fig. 15: Data of $s=0.4$

6. 改良型 pdi-Boosting の特性評価

改良型 pdi-Boosting のクラス決定の有用性を議論するため、脳信号を模した数値データの 2 識別問題を考える。いま、酸化ヘモグロビン (oxy-Hb) と還元ヘモグロビン (doxy-Hb) の NIRS 信号を模擬するため、サンプリング時間を 20ms として 10 秒間のデータを発生させる。ただし、NIRS 信号の定常状態を 0 で表し、酸化ヘモグロビン (oxy-Hb) の賦活状態を 1 で表し、還元ヘモグロビン (doxy-Hb) の賦活状態を -1 で表す。なお、標準偏差 s を変更した正規乱数により NIRS 信号データに外乱を与え、それぞれ 490 個の D^{TRD} と D^{CHD} を生成した。ただし、ここでは、補正データのメンバシップ関数 $\mu_F(x_j)$ は正規分布とし、識別器には REPTree、終了規範は繰り返し判定として、回数は $K = 3$ とした。

特性解析では、外乱データの標準偏差 s を変更し、 $W^{CHD} = 490$ として 10 回の繰り返し試行により CHD の識別率を議論した。図 14~17 に外乱の標準偏差 s を変化させた場合の脳信号データの例を示す。

$$w : w_1 = w_2 = w_3 = 1/3$$

$$s : 0.2, 0.4, 0.6, 0.8$$

$\sigma : 0.0001$

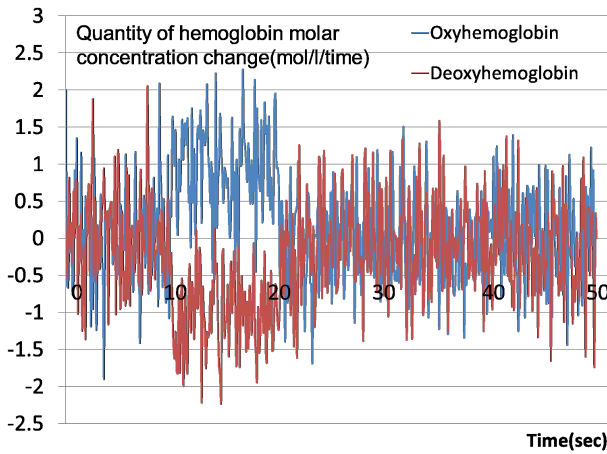


Fig. 16: Data of $s=0.6$

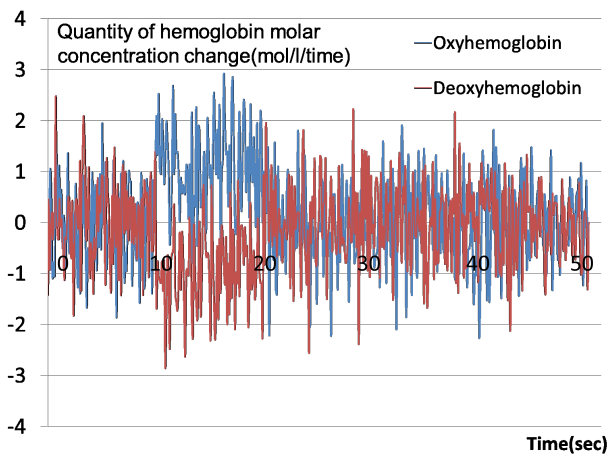


Fig. 17: Data of $s=0.8$

結果を表 4 に示す。改良型 pdi-Boosting は他手法より識別率が向上していることがわかる。特に、メンバー関数が一様分布である従来型 pdi-Boosting に対しては、平均識別率において 1.28% の向上が認められ、有意差 0.01% 水準の t 検定において、有意な差が認められた ($p = 0.000037$)。

注目すべきは標準偏差 $s = 0.8$ の場合である。繰り返し施行の第 1 段階 ($i = 1$) では、 $W^{TRD} = 866$ 個となった。当初の TRD の全データ数は 490 個であったので、差分の 376 個は補間データである。その中で、改良型アルゴリズムによりクラスを変更した補間データは 67.6% の 254 個に達した。具体的には、定常状態から賦活状態のクラスに変更された補間データは 73 個であり、賦活状態から定常状態に変更された補間データは 181 個である。このクラス変更により、改良型 pdi-Boosting の識別率が向上したと考えられる。

Table 4: Comparison of New pdi-Boosting with Other Methods

SD	Mod. pdi-B. (%)	pdi-B. (Uniform) (%)	Ada Boost (%)	Multi Boost (%)	REP Tree (%)
0.2	99.81	99.63	99.80	99.76	99.37
0.4	97.32	96.18	96.57	94.84	97.14
0.6	92.65	90.05	91.22	91.22	92.41
0.8	88.78	87.69	88.63	88.55	88.33
Ave.	94.64	93.38	94.06	93.59	92.11

7. おわりに

本論文では、pdi-Boosting を定式化し、数値データと NIRS 計測データにより、その有用性を検証した。また、3つの評価基準による補間データの識別クラスの決定方法を新たに提案し、その改良型 pdi-Boosting の有用性を示した。ただし、今後、提案手法を実際の NIRS データに適応するには、pdi-Boosting に学習機能を付加する必要がある。今後、学習機能の有用性を議論するとともに、クラス決定アルゴリズムの洗練化を行い、NIRS 計測や EEG 計測等の実際の脳信号データに適用して、その特性や有用性を議論する必要がある。

なお、本研究の一部は、関西大学研究拠点形成支援経費「脳認知ロボティクスによる橋梁診断スキームの構築」(2013年～2014年)と日本建設情報総合センター研究助成「インタラクティブ・ロボティクスによる橋梁診断スキームの構築」(2015年～2016年)の助成を得た。

参考文献

- [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] 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 Internatinal BCI Workshop*, pp.44-49 (2008).
- [3] 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)
- [4] 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)
- [5] 金森, 畑埜, 渡辺: ブースティング, 森北出版 (2006)
- [6] 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)

- [7] 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)
- [8] 林, 鶴背: 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, 信学技報, Vol.109, No.461, pp.303-308 (2010)
- [9] 林, 鶴背: 脳コンピュータインタフェースのためのpdi-Boostingの提案, 第26回ファジィシステムシンポジウム講演論文集, pp.288-291 (2010)
- [10] I.Hayashi, S.Tsuruse, J.Suzuki, R.T.Kozma: A Proposal for Applying pdi-Boosting to Brain-Computer Interfaces, *Proc. of 2012 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE2012) in 2012 IEEE World Congress on Computational Intelligence (WCCI2012)*, pp.635-640 (2012).
- [11] I.Hayashi, S.Tsuruse: An Evaluation of pdi-Boosting for Brain-Computer Interfaces, *Proc. of the 6th International Conference on Soft Computing and Intelligent Systems and the 13th International Symposium on Advanced Intelligent Systems (SCIS-ISIS2012)*, pp.1215-1220 (2012).
- [12] 金田, 笹井, 古橋: 統計・多変量解析とソフトコンピューティング - 超多自由度系解析を目指して-, 共立出版 (2012)

研究に従事。工学博士。米国電気電子学会 (IEEE), 日本知能情報ファジィ学会, 日本神経回路学会, 日本視覚学会, 日本基礎心理学会, システム制御情報学会等の会員。

鶴背 慎二 (つるせ しんじ) (非会員)

2008年関西大学総合情報学部卒業, 2012年関西大学大学院総合情報学研究科博士課程前期修了後, パナソニックシステムソリューションズジャパン (株)(現パナソニックシステムネットワークス (株))に入社。脳とロボットの相互結合モデルに関する研究に従事。

[連絡先]

林 勲 関西大学大学院 総合情報学研究科
〒569-1095 大阪府高槻市霊仙寺町2-1-1
tel. 072-690-2448
fax. 072-690-2491
e.mail ihaya@cpii.kutc.kansai-u.ac.jp

著者紹介

林 勲 (はやし いさお) (正会員)

1981年大阪府立大学工学部経営工学科卒業後, シャープ (株) 入社。1985年大阪府立大学大学院工学研究科経営工学専攻博士前期課程修了。松下電器産業 (株)(現パナソニック (株)) 中央研究所を経て, 1993年阪南大学商学部経営情報学科講師, 1997年経営情報学部教授, 2004年より関西大学総合情報学部総合情報学科教授。1997年南オーストラリア州立大学 KES 招聘研究員, 1999年米国ボストン大学大学院 CNS 招聘研究員。2010年米国ボストン大学大学院 CNS 招聘教授。神経回路モデルを用いた視覚モデル, ファジィインタフェースによる脳とロボットとの相互結合モデル, 動作解析とスポーツ戦略の

A Proposal of Boosting Algorithm by Possibilistic Data Interpolation and its Application to Brain-Computer Interface

Isao HAYASHI Graduate School of Informatics, Kansai University
Shinji TSURUSE Panasonic System Networks Co.,Ltd.

Abstract:

Brain-computer interface (BCI) and brain-machine interface (BMI) 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 propose a new boosting algorithm for BCI using a possibilistic data interpolation scheme. In our model, interpolated data is generated around classification errors using membership function, and the class attribute is decided by a rule with three kinds of criterions. By using the interpolated data, the discriminated boundary is shown to control the external machine effectively. We verify our boosting method with some numerical examples in which NIRS data is assumed to detect from subjects, and discuss the results.

Keywords:

Brain-Computer Interface, Boosting Algorithm, Possibilistic Data Interpolation.

Contact Address: Isao HAYASHI

Graduate School of Informatics, Kansai University
2-1-1, Ryozenji-cho, Takatsuki, Osaka 569-1095, JAPAN
Telephone: +81-72-690-2448
E.Mail: ihaya@cpii.kutc.kansai-u.ac.jp