

脳コンピュータインタフェースのための pdi-Boosting の効率化

An Efficiency of Boosting Algorithm by Probabilistic Data Interpolation for Brain-Computer Interface

林 勲

I.Hayashi

関西大学大学院

Kansai University

鶴背 慎二

S.Tsuruse

パナソニックシステムソリューションズジャパン (株)

Panasonic System Solutions Japan 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 novel boosting algorithm for BCI using a probabilistic data interpolation scheme. In our model, interpolated data is generated around classification errors using a probability distribution 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.

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, 2] が盛んに行われている。一般的には、近赤外分光法 (NIRS) [3] や脳波計測 [4] 装置から脳活動信号を計測し、得られた信号クラスター間の相違を判別モデルを用いて同定し機器を制御する。しかし、計測時に外部環境が動的に変動する場合が多く、判別モデルを用いて制御しても、環境変化に追従できない場合がある。

一方、パターン識別の研究分野において、学習データに対して複数個の判別器を構成し、それらを統合的に組み合わせる全体の判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、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する。複数個の判別器による推定値を多数決原理により統合して最終結果を算出する。偏りのデータを構成し、それに適合した判別器を学習することで、全体の識別率を向上することができる。一方、著者らが提案している pdi-Boosting (Probabilistic Data Interpolation-Boosting) [8, 9] では、学習データにより複数の判別器を同定し、誤判別データが存在した場合に、その誤判別データ付近に補間データを発生して判別器の識別率を向上させる。補間データを付与することにより、学習データが少ない場合や外的環境の変化に強く、複雑なデータに対する頑健性に優れている [10]。しかし、正規乱数により補間データを発生させるため、補間データが誤判別データ付近に発生しない場合には、その誤判別データのクラス決定に誤りが生じ、かえって識別率を低下させる場合もある。

本論文では、補間データの識別クラスを決定する新た

な pdi-Boosting のアルゴリズムを提案する。補間データのクラスは、次の評価基準を組み合わせる。補間データを発生させる確率密度関数による誤判別データの評価値 E_1 、補間データの識別クラスへの距離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指標である近傍クラスの評価値 E_3 の 3 種類の評価基準を重み平均化し統合的に補間データの識別クラスを決定する。補間データが誤判別データ付近で発生する場合には、誤判別データの識別クラスへの依存度 E_1 が高まり、誤判別データの周辺で発生する場合には、識別クラスへの依存度 E_2 が高まる。これらのクラス決定法により、補間データは誤判別データのクラスから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、クラス決定を用いた新たな 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討する。

2. pdi-Boosting

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頑健性の高い判別モデルを構築する手法である。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 では、判別器 $M_1, M_2, \dots, M_i, \dots$ 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。ただし、次ステップでのモデルのための TRD を構成する際に、各データの重みを用いて、誤判別となったデータが全データ中で 50% 以上になるようにデータを選択する。重みは順次更新され、判別器による同定が繰り返される。最終的には、評価用データ (CHD) を判別器に入力し、多数決で出力を計算して最終結果を得る。一方、pdi-Boosting では、確率密度関

数を用いて補間データを誤判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り、判別器の識別精度が向上し、最終的な統合結果の識別率も向上する。

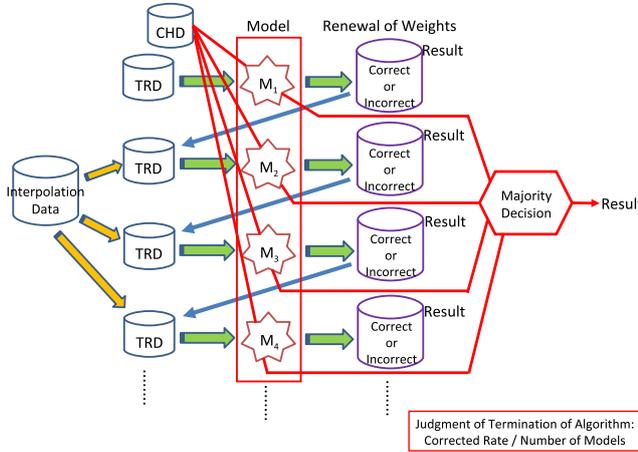


Fig. 1: pdi-Boosting Algorithm

処理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) を学習データ D^{TRD} (個数: W^{TRD}) と評価データ D^{CHD} (個数: W^{CHD}) に分割する。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 の識別率 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、確率密度関数 $f(x_j) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp(-\frac{(x_j - x_j^F(s))^2}{2\sigma_1^2})$ により、補間データ $x_j^{int}(s)$ を発生し、 D^{INT} に含める。
- Step 4** 結果 R_i において、正識別データと誤識別データが同数になるように、 D^{INT} から $d \geq \frac{W}{2} - W^{TRD}(1 - r_i^{TRD})$ の補間データを取り出し、 D^{TRD} に加える。
- Step 5** Step2 から 4 までを繰り返し、しきい値 θ に対して $r_i^{CHD} \geq \theta$ を満足した時点、あるいは、繰り返し回数 K 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** D^{CHD} を M_1, M_2, \dots, M_i に適用して、多数決により結果の識別率 r_i^{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 の特性を検証する。いま、2 個の識別問題として、脳信号の定常状態を 0 で表し、賦活状態を 1 で表した。外乱として 500 個の数値データに正規乱数 σ_2 を付与しデータ集合とした。REPTree を判別器として用い、繰り返し回数は $K = 3$ として、下記のパラメータを変化させた 385 個の組み合わせに対して、それぞれ 10 回の繰り返しにより識別率を算出した。

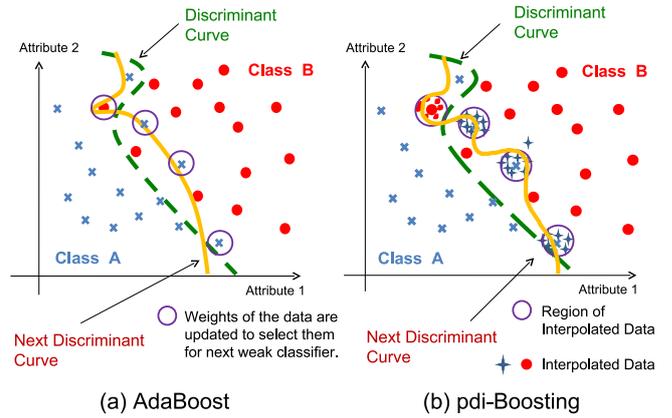


Fig. 2: AdaBoost and pdi-Boosting

データ数: 2, 5, 10, 20, 30, 50, 75, 100, 250, 350, 500
 補間 σ_1 : 0.005, 0.01, 0.05, 0.1, 0.2, 0.6, 1.0
 外乱 σ_2 : 0.2, 0.4, 0.6, 0.8, 1.0

pdi-Boosting には次の特性がある。

- 1) データ数が極端に少ない場合でも評価データの識別率が低下しない。また、データ数が大きい場合には識別率の分散が小さくて高い。
- 2) 学習データの識別率はデータ数の少ない場合でも高い。
- 3) 外乱が大きい場合でも識別率の頑健性が高い。

1) では、データ数が 10 個以下の極端に少ない場合でも、pdi-Boosting は AdaBoost よりも識別率が高かった。また、データ数が大きい場合、pdi-Boosting は、AdaBoost と比較して識別率が高い。2) では、学習データを評価データとした場合、データ数の少ない場合では pdi-Boosting は AdaBoost より高い認識率を得た。3) では、外乱が大きいほど、pdi-Boosting は AdaBoost より識別率が高かった。

特に、3) は重要である。外乱の標準偏差 $\sigma_2 = 0.4, 0.8, 1.0$ における pdi-Boosting と AdaBoost の識別率を図 3 に示す。pdi-Boosting が実線で、AdaBoost が破線で描かれている。pdi-Boosting も AdaBoost も、外乱の量が大きくなるにつれて識別率が低下し、学習データが 50 個以下の場合に識別率の変動が特に大きい。しかし、学習データが 100 個以上の場合、どの外乱に対しても、pdi-Boosting の識別率は AdaBoost よりも高い。また、AdaBoost は、学習データが十分に確保できる場合でも、識別率は安定しないが、pdi-Boosting の識別率は学習データの個数と比例して上昇している。この傾向は、外乱量が大きい $\sigma_2 = 1.0$ において最も顕著に見られる。これらの結果から、pdi-Boosting は AdaBoost と比較して、外乱に対する頑健性に高いことがわかる。また、学習データ数が十分に確保できる場合には、pdi-Boosting は AdaBoost よりも安定した高い識別率を得ることがわかる。

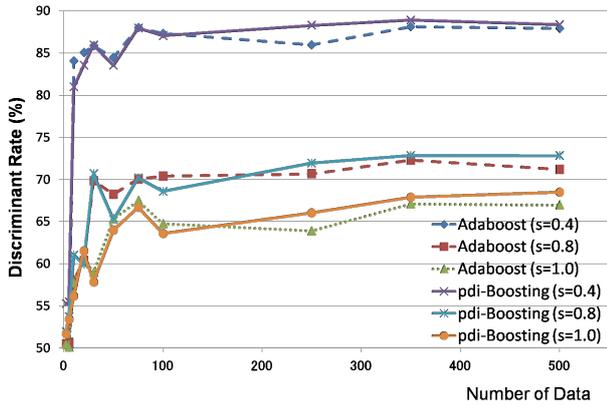


Fig. 3: Discriminant Rate by Changing Number of Data and S.D.

補間データを発生する正規分布の標準偏差 σ_1 を変化させた場合の識別率を検証した。誤判別の位置を 0.5 とし、補間データの度数分布を図 4 に示す。標準偏差 σ_1 の変化によって補間データの発生領域が拡大しているが、 $\sigma_1 = 1.0$ の場合でも、80% 程度のデータが区間 $[0, 1]$ に含まれている。したがって、この程度の標準偏差の変化では、識別率には大きな影響を与えず、より広範囲の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。

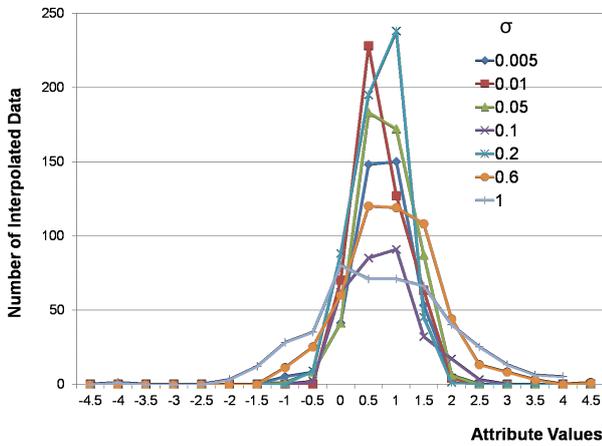


Fig. 4: Frequency Distribution of Interpolated Data

4. 識別クラスの決定

補間データのクラスを決定する新たなクラス決定法を提案する。いま、補間データ $x_j^{int}(s)$ が誤判別データ $x_j^F(s)$ から発生した場合、この $x_j^{int}(s)$ のクラスは、次の評価値 E_k が最小となるクラス k^* をもつ。

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

$$E_j^k = w_1 E_{j1}^k + w_2 E_{j2}^k + w_3 E_{j3}^k \quad (2)$$

ただし、 E_1 は誤判別データの評価値、 E_2 は識別クラスの評価値、 E_3 は近傍クラスの評価値を表す指標であり、 w_1, w_2, w_3 はそれぞれの重みである。

3つの評価基準 E_{j1}, E_{j2}, E_{j3} は次のように定義される。

(1) 誤判別データの評価値

評価値 E_{j1} は、補間データの誤判別データへの依存度を発生確率密度関数 $f(x_j)$ を用いて定義する (図 5 参照)。評価値 E_{j1} が小さい補間データほど、その誤判別データへの依存度が高いことを表している。

$$E_{j1}^k = \begin{cases} P(x_j^{int}(s)) & x_j^F \in k \\ 1 - P(x_j^{int}(s)) & x_j^F \notin k \end{cases}$$

$$P(x_j^{int}(s)) = \int_{x_j^F(s)}^{x_j^{int}(s)} f(x_j) dx_j$$

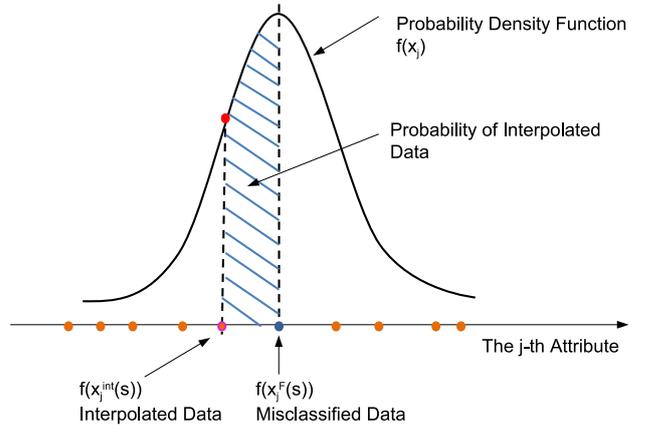


Fig. 5: Evaluation E1

(2) 識別クラスの評価値

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 6 参照)。評価値 E_{j2} が小さい補間データほど、そのクラスにより依存していると仮定し、補間データのクラスを決定する。

$$E_{j2}^k = \frac{|x_j^{int}(s) - x_c^k| - \min_i |x_i^k - x_c^k|}{\max_i |x_i^k - x_c^k| - \min_i |x_i^k - x_c^k|}$$

ただし、 x_c^k は、クラス k の中心である。

(3) 近傍クラスの評価値

評価値 E_{j3} は、各クラスにおいて補間データに最も近いデータ x_j^N との距離を用いて定義する (図 7 参照)。補間データの近傍にあるデータのクラスを用いて、その補間データのクラスを決定する。評価値 E_{j3} が小さい補間データほど、近傍データのクラスに依存していると仮定し、補間データのクラスを決定する。

$$E_{j3}^k = \frac{|x_j^N - x_j^{int}(s)| - \min_i |x_i^k - x_j^{int}(s)|}{\max_i |x_i^k - x_j^{int}(s)| - \min_i |x_i^k - x_j^{int}(s)|}$$

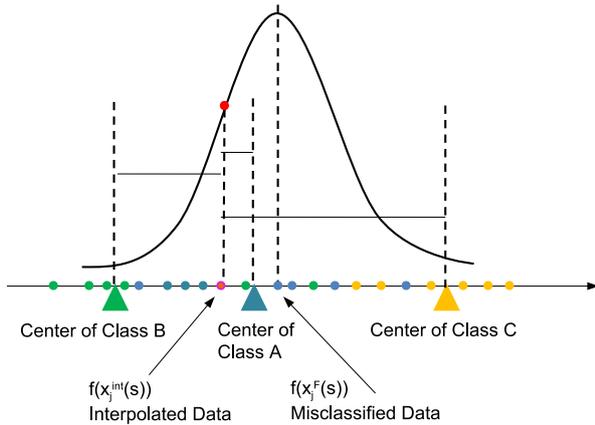


Fig. 6: Evaluation E2

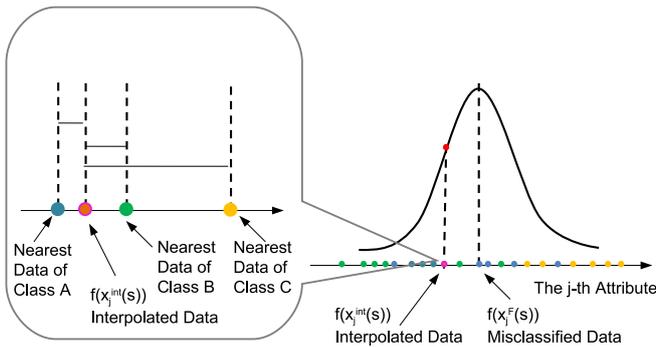


Fig. 7: Evaluation E3

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 を得て、 n 個の属性を合計して全体の評価値 E を得る。評価値 E が最も小さくなるクラス k^* をその補間データのクラスとして決定する。補間データが誤判別データ付近で発生する場合には、誤判別データの評価値 E_1 が高まり、クラスの中心付近に補間データが発生する場合には、識別クラスの評価値 E_2 が高まる。一方、補間データ付近の識別クラスの評価は評価値 E_3 で計算される。

5. 改良型 pdi-Boosting の定式化

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) を学習データ D^{TRD} (個数: W^{TRD}) と評価データ D^{CHD} (個数: W^{CHD}) に分割する。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 の識別率 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 $x_j^F(s)$ に対して、確率密度関数 $f(x_j) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp(-\frac{(x_j - x_j^F(s))^2}{2\sigma_1^2})$ により、補間データ $x_j^{int}(s)$ を発生させる。

Step 4 補間データ $x_j^{int}(s)$ のクラス k^* を式 (1) と (2) により求める。補間データ $x_j^{int}(s)$ を D^{INT} に含める。

Step 5 結果 R_i において、正識別データと誤識別データが同数になるように、 D^{INT} から $d \geq \frac{W}{2} - W^{TRD}(1 - r_i^{TRD})$ の補間データを取り出し、 D^{TRD} に加える。

Step 6 Step2 から 5 までを繰り返し、しきい値 θ に対して、 $r_i^{CHD} \geq \theta$ を満足した時点、あるいは、繰り返し回数 G に対して、 $i \geq G$ を満足した時点でアルゴリズムを終了する。

Step 7 D^{CHD} を M_1, M_2, \dots, M_i に適用して、多数決により結果の識別率 r_i^{CHD} を得る。

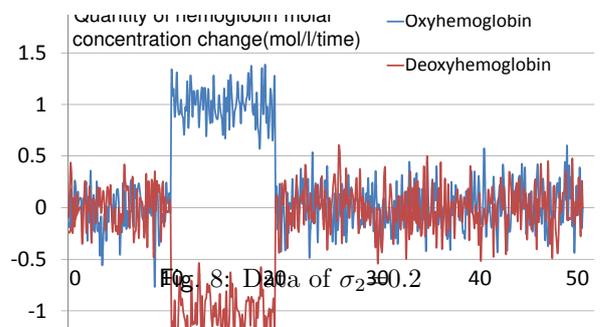
6. 改良型 pdi-Boosting の特性評価

脳信号データの数値データを作成し、改良型 pdi-Boosting の特性を検証した。いま、脳信号の定常状態を 0 で表し、賦活状態を 1 で表した 2 個の識別問題に対して、外乱として 490 個の数値データに標準偏差 σ_2 の正規乱数を付与した。判別器には REPTree を用い、繰り返し回数を $G = 3$ とした。4 種類の正規分布による外乱 σ_2 を加えた数値データの一部を図 8~11 に示す。

重み: $w_1 = w_2 = w_3 = 1/3$

補間 σ_1 : 0.0001

外乱 σ_2 : 0.2, 0.4, 0.6, 0.8



結果を表 1 に示す。識別率は各 10 回の数値データの平均値である。本手法との比較では、大きな識別率の差は認められなかったが、識別率は向上している。特に、発生確率密度関数を一様分布とした従来型の pdi-Boosting

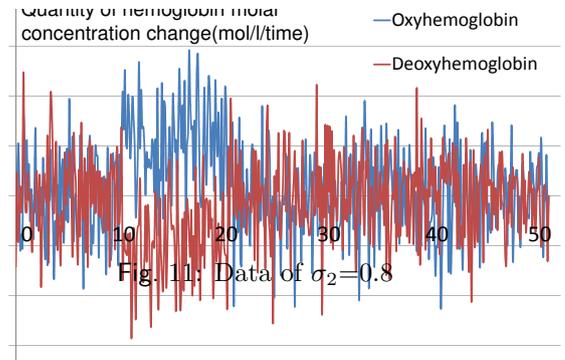
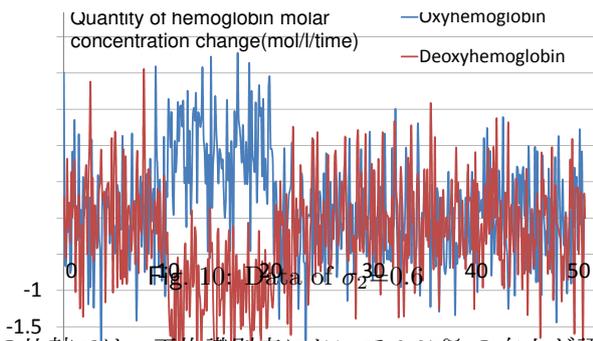
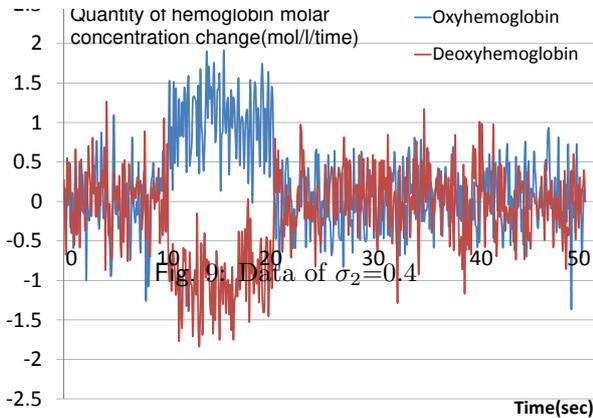


Table 1: Comparison of New pdi-Boosting with Other Methods

SD	New 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.	94.92	94.65	94.06	93.59	92.11

との比較では、平均識別率において0.27%の向上が認められた。この識別率の向上は、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 ($p = 0.000037$)。

標準偏差 $\sigma_2 = 0.8$ の外乱データの結果では、490個の全データに対して本手法を適用した場合、 $G = 1$ において全データ数は866個となった。クラス決定法を採用しない従来型 pdi-Boosting と比較して、この差分の補充された376個の補間データの中で、クラスを変更した補間データは67%に達した。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個であった。このように、クラス決定アルゴリズムを適用することにより、クラスの変更が頻繁に行われ、その結果、識別率が向上している。

7. おわりに

本論文では、pdi-Boosting の特性を検証し、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提案した。また、本手法をNIRS計測を模した数値データに適用し、本手法の有用性を示した。今後、クラス決定アルゴリズムの洗練化を行うとともに、NIRS計測やEEG計測等で観測した脳信号に対して、本手法を適用しその

有用性を検証する必要がある。

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「セキュアライフ創出のための安全知循環ネットワークに関する研究」(2008年~2012年)と関西大学先端科学技術推進機構社会基盤情報学応用研究グループ「社会基盤情報学に基づくメンテナンス技術開発」(2010年~2012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，莊司：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).

[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