

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2009-75737  
(P2009-75737A)

(43) 公開日 平成21年4月9日(2009.4.9)

(51) Int.Cl. F 1 テーマコード (参考)  
**GO6N 5/04 (2006.01)** GO6N 5/04 580A  
**GO6N 3/00 (2006.01)** GO6N 3/00 560A

審査請求 未請求 請求項の数 32 O L (全 13 頁)

(21) 出願番号 特願2007-242419 (P2007-242419)  
 (22) 出願日 平成19年9月19日 (2007.9.19)

(71) 出願人 00004237  
 日本電気株式会社  
 東京都港区芝五丁目7番1号  
 (74) 代理人 100065385  
 弁理士 山下 穰平  
 (74) 代理人 100130029  
 弁理士 永井 道雄  
 (72) 発明者 寺本 礼仁  
 東京都港区芝五丁目7番1号 日本電気株式会社社内

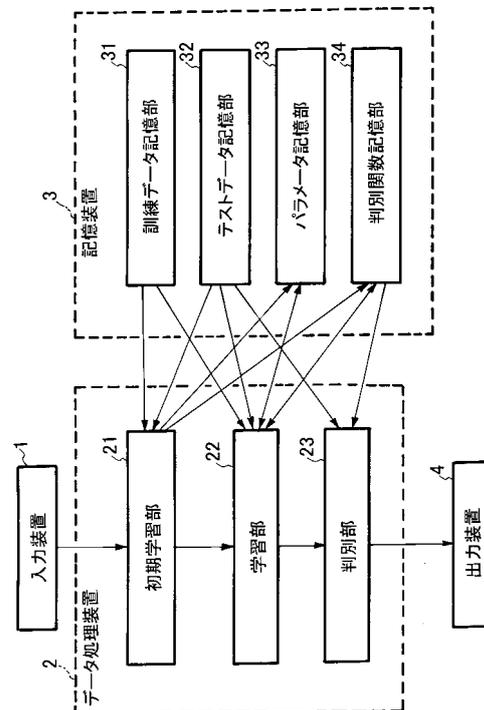
(54) 【発明の名称】 半教師あり学習方法、半教師あり学習装置及び半教師あり学習プログラム

(57) 【要約】

【課題】 下位学習機械として教師あり学習を適用可能な汎用的かつ学習パラメータの少ない実用的な高精度の分類器を提供する。

【解決手段】 訓練データを記憶する。ラベル無しデータおよびテストデータを記憶する。訓練データにより教師あり学習を行う。勾配を記憶する。訓練データとラベル無しデータおよびテストデータを結合し、教師あり学習を反復的に行う。学習された判別関数を記憶する。判別関数を用いてテストデータのラベルを予測する。

【選択図】 図1



**【特許請求の範囲】****【請求項 1】**

アンサンブル学習に基づく半教師あり学習方法において、  
記憶データ、テストデータ及びラベル無しデータを記憶するデータ記憶ステップと、  
前記データ記憶ステップにおいて記録した訓練データに基づいて教師あり学習を行う初期学習ステップと、  
前記初期学習ステップにおいて算出されたデータを記憶する算出結果記憶ステップと、  
前記データ記憶ステップにおいて記録した訓練データとラベル無しデータの属性を結合したデータ及び前記算出結果記憶ステップで記憶しているデータに基づいて教師あり学習を反復的に行う学習ステップと  
を備えることを特徴とする半教師あり学習方法。

10

**【請求項 2】**

請求項 1 に記載の半教師あり学習方法において、  
前記学習ステップにおける教師あり学習は、前記初期学習ステップによって算出された訓練データの勾配及びテストデータの勾配をラベルとして行われることを特徴とする半教師あり学習方法。

**【請求項 3】**

請求項 1 又は 2 に記載の半教師あり学習方法において、  
前記初期学習ステップにおける教師あり学習の方式がブースティング又は勾配ブースティングであることを特徴とする半教師あり学習方法。

20

**【請求項 4】**

請求項 1 乃至 3 の何れか 1 項に記載の半教師あり学習方法において、  
前記初期学習ステップにおける訓練データに関する損失関数が凸関数であることを特徴とする半教師あり学習方法。

**【請求項 5】**

請求項 1 乃至 4 の何れか 1 項に記載の半教師あり学習方法において、  
前記初期学習ステップにおける訓練データに関する損失関数が指数関数であることを特徴とする半教師あり学習方法。

**【請求項 6】**

請求項 1 乃至 5 の何れか 1 項に記載の半教師あり学習方法において、  
前記初期学習ステップにおける各ラウンドで学習された判別関数に縮小パラメータを乗ずる、初期正則化ステップを更に備えることを特徴とする半教師あり学習方法。

30

**【請求項 7】**

請求項 1 乃至 6 の何れか 1 項に記載の半教師あり学習方法において、  
前記初期学習ステップにおける下位学習機械が決定木であることを特徴とする半教師あり学習方法。

**【請求項 8】**

請求項 1 乃至 7 の何れか 1 項に記載の半教師あり学習方法において、  
前記学習ステップにおける教師あり学習の方式がブースティング又は勾配ブースティングであることを特徴とする半教師あり学習方法。

40

**【請求項 9】**

請求項 1 乃至 8 の何れか 1 項に記載の半教師あり学習方法において、  
前記学習ステップにおけるラベル無しデータとはテストデータであることを特徴とする半教師あり学習方法。

**【請求項 10】**

請求項 1 乃至 9 の何れか 1 項に記載の半教師あり学習方法において、  
前記学習ステップにおける訓練データに関する損失関数が凸関数であることを特徴とする半教師あり学習方法。

**【請求項 11】**

請求項 1 乃至 10 の何れか 1 項に記載の半教師あり学習方法において、

50

前記学習ステップにおける訓練データに関する損失関数が指数関数であることを特徴とする半教師あり学習方法。

【請求項 1 2】

請求項 1 乃至 1 1 の何れか 1 項に記載の半教師あり学習方法において、前記学習ステップにおいて、ラベル無しデータに対する下位学習機械の予測値の分散を最小化することを特徴とする半教師あり学習方法。

【請求項 1 3】

請求項 1 乃至 1 2 の何れか 1 項に記載の半教師あり学習方法において、前記学習ステップにおける各ラウンドで学習された判別関数に縮小パラメータを乗ずる、正則化ステップを更に備えることを特徴とする半教師あり学習方法。

10

【請求項 1 4】

請求項 1 乃至 1 3 の何れか 1 項に記載の半教師あり学習方法において、前記学習ステップにおける下位学習機械が決定木であることを特徴とする半教師あり学習方法。

【請求項 1 5】

請求項 1 乃至 1 4 の何れか 1 項に記載の半教師あり学習方法において、前記学習ステップにおいて得られた学習結果に基づいてテストデータのラベルを予測する予測ステップを更に備えることを特徴とする半教師あり学習方法。

【請求項 1 6】

アンサンブル学習に基づく半教師あり学習装置において、訓練データを記憶する訓練データ記憶部と、ラベル無しデータおよびテストデータを記憶するテストデータ記憶部と、訓練データにより教師あり学習を行う初期学習部と、勾配を記憶するパラメータ記憶部と、訓練データとラベル無しデータおよびテストデータを結合し、教師あり学習を反復的に行う学習部と、学習された判別関数を記憶する判別関数記憶部と、判別関数を用いてテストデータのラベルを予測する判別部を備えることを特徴とする半教師あり学習装置。

20

【請求項 1 7】

アンサンブル学習に基づく半教師あり学習装置において、記憶データ、テストデータ及びラベル無しデータを記憶するデータ記憶手段と、前記データ記憶手段において記録した訓練データに基づいて教師あり学習を行う初期学習手段と、前記初期学習手段において算出されたデータを記憶する算出結果記憶手段と、前記データ記憶手段において記録した訓練データとラベル無しデータの属性を結合したデータ及び前記算出結果記憶手段で記憶しているデータに基づいて教師あり学習を反復的に行う学習手段とを備えることを特徴とする半教師あり学習装置。

30

【請求項 1 8】

請求項 1 7 に記載の半教師あり学習装置において、前記学習手段における教師あり学習は、前記初期学習手段によって算出された訓練データの勾配及びテストデータの勾配をラベルとして行われることを特徴とする半教師あり学習装置。

40

【請求項 1 9】

請求項 1 7 又は 1 8 に記載の半教師あり学習装置において、前記初期学習手段における教師あり学習の方式がブースティング又は勾配ブースティングであることを特徴とする半教師あり学習装置。

【請求項 2 0】

請求項 1 7 乃至 1 9 の何れか 1 項に記載の半教師あり学習装置において、前記初期学習手段における訓練データに関する損失関数が凸関数であることを特徴とする半教師あり学習装置。

【請求項 2 1】

50

請求項 17 乃至 20 の何れか 1 項に記載の半教師あり学習装置において、  
前記初期学習手段における訓練データに関する損失関数が指数関数であることを特徴とする半教師あり学習装置。

【請求項 22】

請求項 17 乃至 21 の何れか 1 項に記載の半教師あり学習装置において、  
前記初期学習手段における各ラウンドで学習された判別関数に縮小パラメータを乗ずる、第 1 の正則化手段を更に備えることを特徴とする半教師あり学習装置。

【請求項 23】

請求項 17 乃至 22 の何れか 1 項に記載の半教師あり学習装置において、  
前記初期学習手段における下位学習機械が決定木であることを特徴とする半教師あり学習装置。

10

【請求項 24】

請求項 17 乃至 23 の何れか 1 項に記載の半教師あり学習装置において、  
前記学習手段における教師あり学習の方式がブースティング又は勾配ブースティングであることを特徴とする半教師あり学習装置。

【請求項 25】

請求項 17 乃至 24 の何れか 1 項に記載の半教師あり学習装置において、  
前記学習手段におけるラベル無しデータとはテストデータであることを特徴とする半教師あり学習装置。

【請求項 26】

請求項 17 乃至 25 の何れか 1 項に記載の半教師あり学習装置において、  
前記学習手段における訓練データに関する損失関数が凸関数であることを特徴とする半教師あり学習装置。

20

【請求項 27】

請求項 17 乃至 26 の何れか 1 項に記載の半教師あり学習装置において、  
前記学習手段における訓練データに関する損失関数が指数関数であることを特徴とする半教師あり学習装置。

【請求項 28】

請求項 17 乃至 27 の何れか 1 項に記載の半教師あり学習装置において、  
前記学習手段において、ラベル無しデータに対する下位学習機械の予測値の分散を最小化することを特徴とする半教師あり学習装置。

30

【請求項 29】

請求項 17 乃至 28 の何れか 1 項に記載の半教師あり学習装置において、  
前記学習手段における各ラウンドで学習された判別関数に縮小パラメータを乗ずる、第 2 の正則化手段を更に備えることを特徴とする半教師あり学習装置。

【請求項 30】

請求項 17 乃至 29 の何れか 1 項に記載の半教師あり学習装置において、  
前記学習手段における下位学習機械が決定木であることを特徴とする半教師あり学習装置。

【請求項 31】

アンサンブル学習に基づく半教師あり学習プログラムにおいて、  
訓練データを記憶する訓練データ記憶機能と、ラベル無しデータおよびテストデータを記憶するテストデータ記憶機能と、訓練データにより教師あり学習を行う初期学習機能と、勾配を記憶するパラメータ記憶機能と、訓練データとラベル無しデータおよびテストデータを結合し、教師あり学習を反復的に行う学習機能と、学習された判別関数を記憶する判別関数記憶機能と、判別関数を用いてテストデータのラベルを予測する判別機能をコンピュータに実現させることを特徴とする半教師あり学習プログラム。

40

【請求項 32】

アンサンブル学習に基づく半教師あり学習プログラムにおいて、  
記憶データ、テストデータ及びラベル無しデータを記憶するデータ記憶機能と、

50

前記データ記憶機能において記録した訓練データに基づいて教師あり学習を行う初期学習機能と、

前記初期学習機能において算出されたデータを記憶する算出結果記憶機能と、

前記データ記憶機能において記録した訓練データとラベル無しデータの属性を結合したデータ及び前記算出結果記憶機能で記憶しているデータに基づいて教師あり学習を反復的に行う学習機能と

をコンピュータに実現させることを特徴とする半教師あり学習プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

10

本発明は、アンサンブル学習を用いてラベル有りデータだけでなくラベル無しデータも学習することによる予測精度の高い半教師あり学習方法、半教師あり学習装置及び半教師あり学習プログラムに関する。

【背景技術】

【0002】

ラベル有りデータを訓練データとして学習機械に学習させたい一方で、テストデータのラベルを予測する学習方式は、教師あり学習と呼ばれている。

【0003】

教師あり学習の方法としては、ブースティング、バギング、サポートベクターマシンなどが有名であり、様々なデータにおいて適用されている。ブースティング、バギング、サポートベクターマシンについては、非特許文献1乃至3に記載されている。

20

【0004】

しかしながら、教師あり学習で高精度の分類器を構成するためには十分な量の訓練データが必要である。この点、一般的にはデータを人手でラベル付けするために、十分な量の訓練データを得るためには非常に時間や労力を要するという問題がある。また、得られたデータの分布に偏りが大きい場合は、テストデータに対する予測精度が低いという問題が指摘されている。

【0005】

これらの問題に対して、半教師あり学習と呼ばれる手法が提案されている。従来のブースティングやサポートベクターマシンの方法論を拡張することにより、半教師あり学習が

30

【0006】

ここで、半教師あり学習とは、訓練データだけでなく、ラベル無しデータあるいはテストデータの分布も考慮した学習方式を指し、訓練データが少数の場合であっても高精度の分類器を構成することを目的とする学習方式をいう。

【0007】

なお以下の説明では、ラベル無しデータおよびテストデータを単にテストデータと記載する。

【非特許文献1】Y Freund, RE Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, Journal of Computer and System Sciences, 1997, 23-27.

40

【非特許文献2】L. Breiman, Bagging Predictors, Machine Learning, 1996, 123-140.

【非特許文献3】J. P. Vert, K. Tsuda and B. Scholkopf, "A primer on kernel methods", in Kernel Methods in Computational Biology, MIT Press, 2004, 35-70.

【非特許文献4】F.d'Alche Buc, Y. Grandvalet, and C. Ambroise, Semi-supervised margin boost, Advances in Neural Information Processing Systems 14, MIT Press, 2002.

【非特許文献5】Y. Grandvalet, F.d'Alche Buc, and C. Ambroise, Boosting mixture models for semi-supervised learning, ICANN, Springer-Verlag, 2002, 41-48.

【非特許文献6】K. P. Bennett, A. Demiriz, and R. Macin, Exploiting unlabeled dat

50

a in ensemble methods, SIGKDD, ACM, 2002.

【非特許文献 7】O. Chapelle and A. Zien, Semi-supervised classification by low density separation, In Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics, 2005.

【非特許文献 8】J Friedman, Greedy Function Approximation: A Gradient Boosting Machine, The Annals of Statistics, 2001, 1189-1232.

【非特許文献 9】Boaz Leskes, The Value of Agreement, a New Boosting Algorithm, Master Thesis, Section Computational Science and Institute for Logic, Language & Computation, University of Amsterdam, 2005.

【非特許文献 10】A. Buja and Y. S. Lee, Data Mining Criteria for Tree-Based Regression and Classification, Proceedings of KDD 2001, 27-36. 10

【非特許文献 11】UCI Machine Learning Repository Content Summary [平成 19 年 9 月 12 日検索]、インターネット <http://mllearn.ics.uci.edu/MLSummary.html>

【発明の開示】

【発明が解決しようとする課題】

【0008】

もっとも、上記した非特許文献 4 乃至 7 に記載の方法では、以下に示すような問題点がある。

【0009】

まず、非特許文献 4 及び 5 に記載の方法は、ブースティングに用いる下位学習機械として半教師あり学習を行う学習機械が必要であり、従来の教師あり学習の手法を用いることができない。そのため、適用可能なデータ形式に制限があり汎用性に乏しいという問題がある。また、大規模データに対して、計算時間を要するという問題もある。 20

【0010】

一方、非特許文献 6 に記載の方法では、非特許文献 4, 5 において問題であったブースティングの下位学習機械として、教師あり学習の学習機械を用いることが可能である。しかし、学習の定式化に問題があるため、分類精度の向上効果が乏しく、性能が劣化する場合もあることが報告されている。

【0011】

また、非特許文献 7 に記載の方法では、サポートベクターマシンの目的関数にテストデータに関する損失関数を導入しているが、計算方法や最適化法が煩雑なため、多数の学習パラメータのチューニングや計算時間を要する問題がある。 30

【0012】

そこで、本発明は上述した問題点を鑑み、学習理論に基づき、ブースティングを用いてラベル無しデータに対する予測値の分散を最小化しつつ訓練データの損失関数も最小化することで、下位学習機械として教師あり学習を適用可能な汎用的かつ学習パラメータの少ない実用的な高精度の分類器を提供することを目的とする。

【課題を解決するための手段】

【0013】

本発明の第 1 の観点によれば、方法としてアンサンブル学習に基づく半教師あり学習方法において、記憶データ、テストデータ及びラベル無しデータを記憶するデータ記憶ステップと、前記データ記憶ステップにおいて記録した訓練データに基づいて教師あり学習を行う初期学習ステップと、前記初期学習ステップにおいて算出されたデータを記憶する算出結果記憶ステップと、前記データ記憶ステップにおいて記録した訓練データとラベル無しデータの属性を結合したデータ及び前記算出結果記憶ステップで記憶しているデータに基づいて教師あり学習を反復的に行う学習ステップとを備えることを特徴とする半教師あり学習方法が提供される。 40

【0014】

更に、本発明の第 2 の観点によれば、第 1 の装置としてアンサンブル学習に基づく半教師あり学習装置において、訓練データを記憶する訓練データ記憶部と、ラベル無しデータ 50

およびテストデータを記憶するテストデータ記憶部と、訓練データにより教師あり学習を行う初期学習部と、勾配を記憶するパラメータ記憶部と、訓練データとラベル無しデータおよびテストデータを結合し、教師あり学習を反復的に行う学習部と、学習された判別関数を記憶する判別関数記憶部と、判別関数を用いてテストデータのラベルを予測する判別部を備えることを特徴とする半教師あり学習装置が提供される。

【0015】

更に、本発明の第3の観点によれば、第2の装置としてアンサンブル学習に基づく半教師あり学習装置において、記憶データ、テストデータ及びラベル無しデータを記憶するデータ記憶手段と、前記データ記憶手段において記録した訓練データに基づいて教師あり学習を行う初期学習手段と、前記初期学習手段において算出されたデータを記憶する算出結果記憶手段と、前記データ記憶手段において記録した訓練データとラベル無しデータの属性を結合したデータ及び前記算出結果記憶手段で記憶しているデータに基づいて教師あり学習を反復的に行う学習手段とを備えることを特徴とする半教師あり学習装置が提供される。

10

【0016】

更に、本発明の第4の観点によれば、第1のプログラムとしてアンサンブル学習に基づく半教師あり学習プログラムにおいて、訓練データを記憶する訓練データ記憶機能と、ラベル無しデータおよびテストデータを記憶するテストデータ記憶機能と、訓練データにより教師あり学習を行う初期学習機能と、勾配を記憶するパラメータ記憶機能と、訓練データとラベル無しデータおよびテストデータを結合し、教師あり学習を反復的に行う学習機能と、学習された判別関数を記憶する判別関数記憶機能と、判別関数を用いてテストデータのラベルを予測する判別機能をコンピュータに実現させることを特徴とする半教師あり学習プログラムが提供される。

20

【0017】

更に、本発明の第5の観点によれば、第2のプログラムとしてアンサンブル学習に基づく半教師あり学習プログラムにおいて、記憶データ、テストデータ及びラベル無しデータを記憶するデータ記憶機能と、前記データ記憶機能において記録した訓練データに基づいて教師あり学習を行う初期学習機能と、前記初期学習機能において算出されたデータを記憶する算出結果記憶機能と、前記データ記憶機能において記録した訓練データとラベル無しデータの属性を結合したデータ及び前記算出結果記憶機能で記憶しているデータに基づいて教師あり学習を反復的に行う学習機能とをコンピュータに実現させることを特徴とする半教師あり学習プログラムが提供される。

30

【発明の効果】

【0018】

本発明によれば、訓練データだけでなく、ラベル無しデータの分布も学習することから、少数の訓練データから高精度の分類器を構成することが可能になる。

【発明を実施するための最良の形態】

【0019】

次に、本発明の実施形態について図面を用いて説明する。

【0020】

本発明は訓練データだけでなくテストデータも用いて、ブースティングによる学習を行い、下位学習機械として任意の教師あり学習の学習機械を利用することで、訓練データが少ない状況においても、高精度の分類器を構成する新しい方法である。以下に本発明を実施するための形態について、図面を参照して説明する。

40

【0021】

図1を参照すると、本発明の実施形態は、キーボード等の入力装置1と、プログラム制御により動作するデータ処理装置2と、情報を記憶する記憶装置3と、ディスプレイ装置や印刷装置等の出力装置4を備える。

【0022】

データ処理装置2は、初期学習部21と、学習部22と、判別部23を備える。

50

## 【0023】

初期学習部 2 1 は、訓練データのみを用いてブースティングによる学習を行い、テストデータの勾配を計算する部分である。学習部 2 2 は、訓練データ及びテストデータの勾配をラベルとしてブースティングによる学習を行い、判別関数を更新する部分である。判別部 2 3 は、学習された判別関数を用いてテストデータのラベルを予測する部分である。

## 【0024】

記憶装置 3 は、訓練データを格納する訓練データ記憶部 3 1 と、テストデータを格納するテストデータ記憶部 3 2 と、損失関数の勾配を格納するパラメータ記憶部 3 3 と、判別関数を格納する判別関数記憶部 3 4 を備える。

## 【0025】

次に図 1、図 2 及び図 3 を参照して、本発明を実施するための形態の動作について、説明する。

## 【0026】

まず、入力装置 1 によって実行指示が与えられ、訓練データ記憶部 3 1 及びテストデータ記憶部 3 2 からデータ処理装置 2 に訓練データとテストデータが入力される（図 2、ステップ A 1）。

## 【0027】

次に、初期学習部 2 1 により、訓練データを用いた判別関数  $F$  の教師あり学習が行われる（図 2、ステップ A 4）。なお初期学習部 2 1 の具体的な動作（ステップ A 2 及びステップ A 3）については下記する。

## 【0028】

そして、判別関数を反復的に更新し（図 2、ステップ A 5）、判別関数による訓練データの勾配を計算する（図 2、ステップ A 6）。

## 【0029】

その後、算出された勾配がパラメータ記憶部 3 3、判別関数が判別関数記憶部 3 4 に記憶される。

## 【0030】

図 2 を参照して、初期学習部 2 1 の具体的な動作について、説明する。

## 【0031】

まず、訓練データ記憶部 3 1 とテストデータ記憶部 3 2 から訓練データ及びテストデータがデータ処理装置 2 に入力される（図 2、ステップ A 1）。

## 【0032】

次に初期学習部 2 1 は、反復回数に  $t_1$ 、縮小パラメータに  $v_1$  を設定する（図 2、ステップ A 2）。また、反復回数を計数する指示パラメータ  $T$  を 1 に初期化する（図 2、ステップ A 3）。

## 【0033】

訓練データを用いた判別関数  $F$  の教師あり学習を行う（図 2、ステップ A 4）。

## 【0034】

ラウンド  $T - 1$  で得られた判別関数  $F_{T-1}$  に、学習により得られた判別関数  $F$  を加えることにより、判別関数  $F_T$  として更新する（図 2、ステップ A 5）。数式として、以下に示す。

## 【0035】

【数 1】

$$F_T = F_{T-1} + v_1 \cdot F$$

ここで、 $v_1$  ( $0 < v_1 < 1$ ) は縮小パラメータであり、オーバーフィッティングを防

10

20

30

40

50

ぐ正則化のために導入されている。1 が導入されていることにより汎化誤差を最小にするような  $F_T$  に収束することとなる。

【0036】

次に、判別関数  $F_T$  を用いて、損失関数を最小化する訓練データの勾配を求める（図2、ステップA6）。

【0037】

ここで、教師あり学習におけるブースティングの損失関数は、訓練データにおける誤り率を最小化することにより、テストデータの誤り率も同時に最小化することを目的として、設計されている。非特許文献8には、勾配ブースティングと呼ばれる方法で損失関数を判別関数により微分することで、損失関数の最小化方向を探索するという方法が記載されている。

10

【0038】

具体的な損失関数としては、以下の数式で表されるような関数がある。

【0039】

【数2】

$$L = \sum_{i=1}^N \exp(-y_i \cdot F(x_i))$$

20

ここで、 $L$  は損失関数、 $y_i$  は  $i$  番目のデータのクラスラベル、 $y_i \in (-1, 1)$ 、 $F(x_i)$  は判別関数であり、 $i$  番目のデータの属性  $x_i$  が与えられたときの出力を表す。損失関数は凸関数であるので、損失関数  $L$  を  $F(x_i)$  で微分し、負をとることによって、損失関数の最小化方向が得られる。

【0040】

【数3】

$$g_i = \frac{\partial L}{\partial F(x_i)} = y_i \cdot \exp(-y_i \cdot F(x_i))$$

30

$$(x_i, y_i) \in D_{labeled}$$

ここで、 $g_i$  はデータ  $x_i$  の勾配であり、 $D_{labeled}$  は訓練データの属性とラベルの組の集合である。

40

【0041】

次に、反復回数の指示パラメータ  $T$  に1を加える（図2、ステップA7）。そして  $T$  の値が予め定めた定数  $t$  に達すれば、学習を終了する。一方、定数  $t$  に達していない場合は、勾配をラベルとした判別関数の教師あり学習（図2、ステップA4）に戻る（図2、ステップA8）。

【0042】

次に、判別関数  $F_T$  を用いてテストデータの勾配を算出する（図2、ステップA9）。

【0043】

ここで、テストデータはラベル無しデータである。そのため、テストデータについては、訓練データと同じ損失関数や勾配を用いることができない。しかし、非特許文献9にお

50

いて、下位学習機械の集合から得られるラベル無しデータの予測値の分散を小さくすることにより、ブースティングの汎化性能が向上できることが示されている。すなわち、ラベル無しデータの予測値の分散を最小にするような勾配を、テストデータについて求めれば性能を向上させることが可能である。

【0044】

そこで、逐次的に分散を最小化することのできる勾配を以下のようにして導出する。下位学習機械の数が  $L$  と下位学習機械の数が  $L + 1$  における予測値の分散をそれぞれ、 $V_L$ 、 $V_{L+1}$  とすると、

【0045】

【数4】

$$V_L(x_i) = \frac{1}{L} \sum_{l=1}^L f_l(x_i)^2 - \left[ \frac{1}{L} \sum_{l=1}^L f_l(x_i) \right]^2$$

10

$$V_{L+1}(x_i) = \frac{1}{L+1} \sum_{l=1}^{L+1} f_l(x_i)^2 - \left[ \frac{1}{L+1} \sum_{l=1}^{L+1} f_l(x_i) \right]^2$$

20

と書ける。ここで、 $f_l$  は  $l$  番目に学習された判別関数であり、 $x_i$  はテストデータである。ここで、 $L + 1$  番目の下位学習機械で求めるべき勾配を  $g$  とすると、

【0046】

【数5】

$$V_{L+1}(x_i) = \frac{1}{L+1} \left( \sum_{l=1}^L f_l(x_i)^2 + g^2 \right) - \left[ \frac{1}{L+1} \sum_{l=1}^L f_l(x_i) + g \right]^2$$

30

と書き直すことができる。 $g$  は  $V = V_T - V_{T+1}$  を最大化するように求めればよい。そこで、 $g$  により  $V$  の微分が 0 になるように  $g$  を定める。

【0047】

【数6】

$$\frac{\partial \Delta V}{\partial g} = 0 \Rightarrow g = \frac{1}{L} \sum_{l=1}^L f_l(x_i)$$

40

ここで、 $V$  が  $g$  に関する 2 階微分が常に負であることから、凹関数であり、大域的な最大値を求めることができることが保証されている。また、上記の  $g$  を  $V$  に代入すると、 $V > 0$  であるので、単調に分散を減少させることができる。

【0048】

初期学習部 21 によって得られた訓練データとテストデータの勾配はパラメータ記憶部 33 に格納し、学習された判別関数  $F_T$  は判別関数記憶部 34 に格納する。

50

## 【0049】

初期学習部 2 1 は勾配ブースティングと同一のアルゴリズムであり、テストデータの勾配を複数の判別関数から予測値の分散を求めるために学習を行う。

## 【0050】

次に、学習部 2 2 の動作を図 3 を用いて説明する。

## 【0051】

まず、訓練データ記憶部 3 1 とテストデータ記憶部 3 2 から訓練データ及びテストデータが学習部 2 2 に入力される。また、初期学習部 2 1 によって得られた訓練データとテストデータの勾配がパラメータ記憶部 3 3 から学習部 2 2 に入力される（図 3、ステップ B 1）。

10

## 【0052】

反復回数  $t_2$ 、縮小パラメータ  $\alpha_2$  が設定される（図 3、ステップ B 2）。反復回数を計数する指示パラメータ  $T$  を 1 に初期化し（図 3、ステップ B 3）、訓練データとテストデータの勾配をラベルとして訓練データ及びテストデータを用いた判別関数  $F$  の教師あり学習を行う（図 3、ステップ B 4）。

## 【0053】

次に、ラウンド  $T - 1$  で得られた判別関数  $F_{T-1}$  に、学習により得られた判別関数  $F$  を加えることにより、判別関数  $F_T$  として更新する（図 3、ステップ B 5）。

得られた判別関数  $F_T$  を用いて、訓練データ及びテストデータの勾配を計算する（図 3、ステップ B 6）。

20

## 【0054】

初期学習部 2 1 と同様にして、訓練データ及びテストデータの勾配をそれぞれで求めた後、反復回数の指示パラメータ  $T$  に 1 を加える（図 3、ステップ B 7）。

## 【0055】

$T$  が予め定めた定数  $t_2$  に達すれば、学習を終了する。一方、達していない場合は、勾配をラベルとした判別関数の教師あり学習（図 3、ステップ B 4）に戻る（図 3、ステップ B 8）。学習された判別関数は判別関数記憶部 3 4 に格納する。

## 【0056】

判別学習部 2 3 では、テストデータをテストデータ記憶部 3 2 から入力し、判別関数記憶部 3 4 から判別関数を入力し、テストデータのラベルを予測する。

30

## 【0057】

ラベル情報としては、例えば医学・生物学分野の場合、疾患や薬効の有無、病態の進行度の他に生存時間などを用いることができる。教師付き学習の方法としては、例えばバギング、ブースティングなどのアンサンブル学習や、サポートベクターマシン、決定木、生存木を用いることができる。なお、上記したラベル情報や教師付き学習方法は例示であり、他のラベル情報や教師付き学習方法を利用することも可能である。

## 【0058】

そして、テストデータの予測結果は出力装置 4 から出力される。

## 【0059】

なお、半教師あり学習装置は、ハードウェア、ソフトウェア又はこれらの組合せにより実現することができる。

40

## 【実施例】

## 【0060】

次に、本発明の実施例について説明する。

## 【0061】

実施に用いるデータとして、糖尿病に関する臨床情報を機械学習ベンチマークデータの UC I レポジトリ（非特許文献 1 1 参照）から取得した。

## 【0062】

糖尿病発症の有無（属性名：diabetes）を診察された 768 人の患者のうち臨床情報 8 項目に基づき、性能評価を行った。768 人の患者のうち糖尿病と診断された者

50

は268人、発症していないと診断された者は500人である。臨床情報の属性における欠損値については、カテゴリーデータについては、最頻カテゴリー、数値データについては、中央値により補完した。

【0063】

本発明の学習パラメータとして、縮小パラメータ  $\gamma_1, \gamma_2 = 1$  とし、訓練データの損失関数として、勾配ブースティングのAdaboostタイプの指数関数を用いた。下位学習機械としては、決定木の1つであるCARTを用いた。CARTの詳細については、非特許文献10に記載されている。なお説明の便宜上、本発明の方法をSSBoost (Semi-Supervised Boosting) と記載する。

【0064】

本発明の対照法として、訓練データのみを用いた勾配ブースティングとCARTを用いた。なお、勾配ブースティングの詳細は非特許文献8に記載されている。便宜上、対照法をAdaboostと記載する。SSBoostはAdaboostと全く同様に訓練データについて学習を20回行い ( $t_1 = 20$ )、その後、20回テストデータの勾配も用いて学習を行った ( $t_2 = 20$ )。

【0065】

Adaboostの反復回数は40回とした。SSBoostとAdaboostは計40回の反復学習を行い、同等の条件下で、公平な性能比較を行った。性能比較の結果を図4に示す。

【0066】

図4に示された結果から、本発明の方法SSBoostはAdaboost及びCARTと比較して、分類性能が常に高いことが分かる。

【0067】

以上より、本発明を実施することにより下位学習機械として教師あり学習を適用可能な高精度の分類器を実現する事ができる。

【図面の簡単な説明】

【0068】

【図1】本発明の実施形態の基本的構成を表す図である。

【図2】本発明の実施形態における初期学習部の処理手順を示すフローチャートである。

【図3】本発明の実施形態における反復学習部の処理手順を示すフローチャートである。

【図4】本発明の実施形態の対照法として、訓練データのみを用いた勾配ブースティングとCARTを用いた性能比較の結果を表す図である。

【符号の説明】

【0069】

- 1 入力装置
- 2 データ処理装置
- 3 記憶装置
- 4 出力装置
- 21 初期学習部
- 22 学習部
- 23 判別手段
- 31 訓練データ記憶部
- 32 テストデータ記憶部
- 33 パラメータ記憶部
- 34 判別関数記憶部

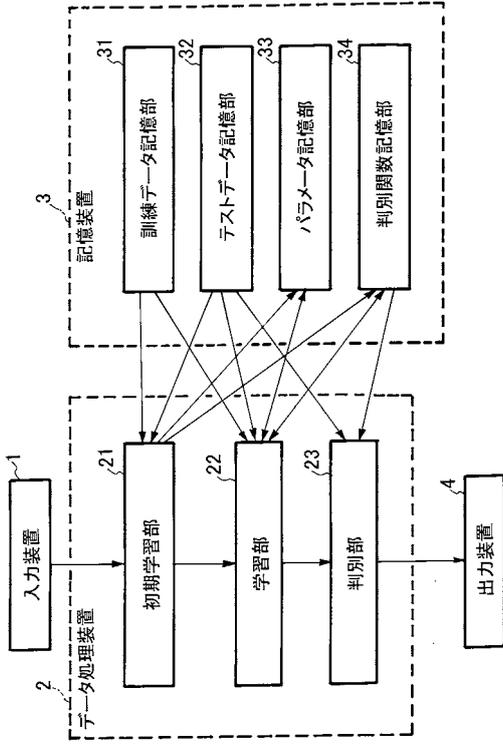
10

20

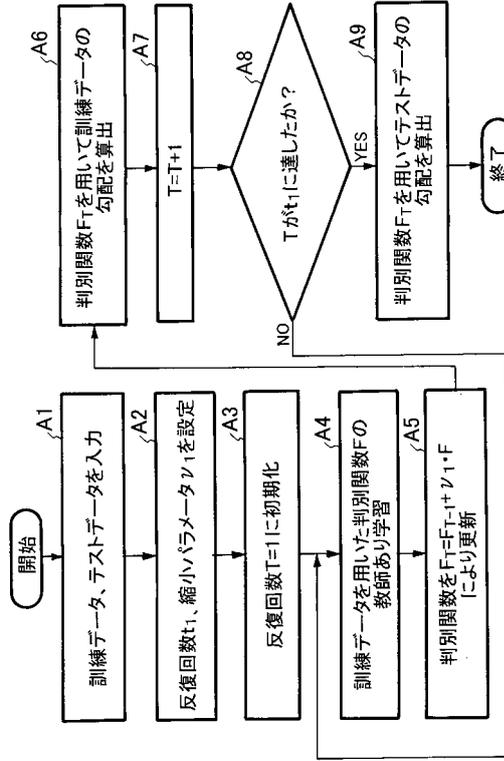
30

40

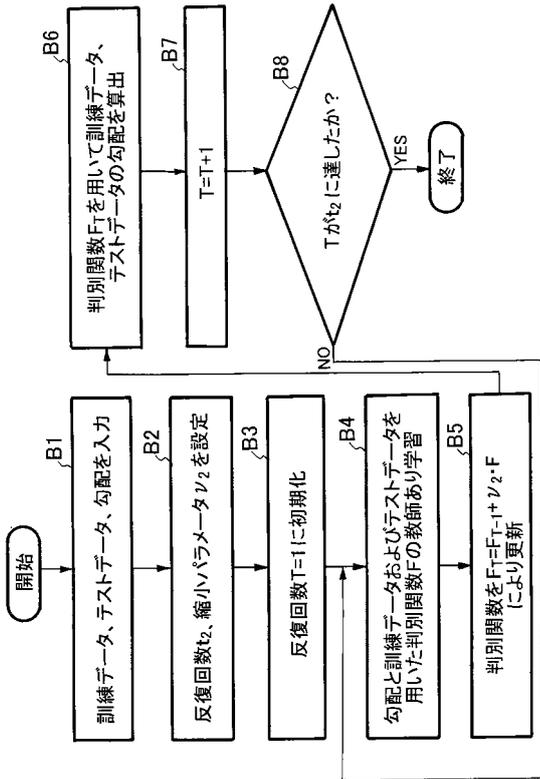
【図 1】



【図 2】



【図 3】



【図 4】

訓練データの割合 (%)	SBoost		
	CART	AdaBoost	SSBoost
20%	67.97	70.68	70.96
30%	67.96	72.33	72.9
40%	66.07	72.65	73.57
50%	64.12	72.95	73.79
60%	66.24	73.3	74.01
70%	64.32	72.87	73.79
80%	65.37	72.93	73.63
90%	65.39	73.95	74.58
100%	63.86	74.19	74.81