

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7251643号
(P7251643)

(45)発行日 令和5年4月4日(2023.4.4)

(24)登録日 令和5年3月27日(2023.3.27)

(51)国際特許分類 F I
G 0 6 N 20/00 (2019.01) G 0 6 N 20/00 1 3 0
G 0 6 N 20/00 1 6 0

請求項の数 6 (全16頁)

(21)出願番号	特願2021-546125(P2021-546125)	(73)特許権者	000004226 日本電信電話株式会社 東京都千代田区大手町一丁目5番1号
(86)(22)出願日	令和1年9月18日(2019.9.18)	(74)代理人	100107766 弁理士 伊東 忠重
(86)国際出願番号	PCT/JP2019/036651	(74)代理人	100070150 弁理士 伊東 忠彦
(87)国際公開番号	WO2021/053776	(74)代理人	100124844 弁理士 石原 隆治
(87)国際公開日	令和3年3月25日(2021.3.25)	(72)発明者	岩田 具治 東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内
審査請求日	令和3年12月24日(2021.12.24)	審査官	赤穂 州一郎

最終頁に続く

(54)【発明の名称】 学習装置、学習方法及びプログラム

(57)【特許請求の範囲】

【請求項1】

ラベルが付与された第1のデータの集合とラベルが付与されていない第2のデータの集合とを入力として、偽陽性率が所定の範囲である場合における評価指標を表す所定の目的関数の値と、前記目的関数のパラメータに関する微分値とを計算する計算手段と、
前記計算手段により計算された前記目的関数の値と前記微分値とを用いて、前記目的関数の値を最大化又は最小化するように前記パラメータを更新する更新手段と、
を有することを特徴とする学習装置。

【請求項2】

前記第1のデータの集合には、正例を示すラベルが付与された正例データと、負例を示すラベルが付与された負例データとが含まれ、

前記評価指標は部分AUCであり、

前記目的関数は、前記正例データと前記負例データとから計算される第1の部分AUCと、前記正例データと前記第2のデータとから計算される第2の部分AUCと、前記負例データと前記第2のデータとから計算される第3の部分AUCとの重み付け和で表される、ことを特徴とする請求項1に記載の学習装置。

【請求項3】

前記目的関数には、前記パラメータを有し、かつ、分類対象のデータが入力された場合に前記分類対象のデータが正例に分類されるスコアを出力する分類器が含まれ、

前記第1の部分AUCは、前記正例データのスコアが、偽陽性率が所定の範囲にある前

10

20

記負例データのスコアよりも高い場合に、高くなり、

前記第 2 の部分 AUC は、前記正例データのスコアが、前記分類器により負例であると分類される前記第 2 のデータのうち偽陽性率が所定の範囲にある前記第 2 のデータのスコアよりも高い場合に、高くなり、

前記第 3 の部分 AUC は、前記分類器により正例であると分類される前記第 2 のデータのスコアが、偽陽性率が所定の範囲にある前記負例データのスコアよりも高い場合に、高くなる、ことを特徴とする請求項 2 に記載の学習装置。

【請求項 4】

所定の終了条件を満たすか否かを判定する判定手段を有し、

前記学習装置は、

前記判定手段により終了条件を満たすと判定されるまで、前記計算手段による前記目的関数の値及び前記微分値の計算と、前記更新手段による前記パラメータの更新とを繰り返す、ことを特徴とする請求項 1 乃至 3 の何れか一項に記載の学習装置。

【請求項 5】

ラベルが付与された第 1 のデータの集合とラベルが付与されていない第 2 のデータの集合とを入力として、偽陽性率が所定の範囲である場合における評価指標を表す所定の目的関数の値と、前記目的関数のパラメータに関する微分値とを計算する計算手順と、

前記計算手順で計算された前記目的関数の値と前記微分値とを用いて、前記目的関数の値を最大化又は最小化するように前記パラメータを更新する更新手順と、

をコンピュータが実行することを特徴とする学習方法。

【請求項 6】

コンピュータを、請求項 1 乃至 4 の何れか一項に記載の学習装置における各手段として機能させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、学習装置、学習方法及びプログラムに関する。

【背景技術】

【0002】

二値分類と呼ばれるタスクが知られている。二値分類とは、データが与えられた場合に、このデータを正例又は負例のいずれかに分類するタスクのことである。

【0003】

二値分類の分類性能を評価する評価指標として部分的 AUC (p AUC : partial area under the ROC curve) が知られている。p AUC を最大化することにより、偽陽性率を低く保ったまま、分類性能を上げることが可能になる。

【0004】

p AUC を最大化する手法が従来から提案されている (例えば、非特許文献 1 参照) 。また、半教師あり学習法により AUC を最大化する手法も従来から提案されている (例えば、非特許文献 2 参照) 。

【先行技術文献】

【非特許文献】

【0005】

【文献】 Naonori Ueda, Akinori Fujino, "Partial AUC Maximization via Nonlinear Scoring Functions", arXiv:1806.04838, 2018

Akinori Fujino, Naonori Ueda, "A Semi-Supervised AUC Optimization Method with Generative Models", ICDM, 2016

【発明の概要】

【発明が解決しようとする課題】

【0006】

しかしながら、例えば上記の非特許文献 1 で提案されている手法では、大量のラベルあ

10

20

30

40

50

りデータを用意する必要がある。一方で、例えば上記の非特許文献2で提案されている手法では、半教師あり学習法によってラベルなしデータも活用することができるが、AUC全体を最大化するため、特定の偽陽性率に特化して分類性能を向上させることができない。

【0007】

本発明の実施の形態は、上記の点に鑑みてなされたもので、特定の偽陽性率における分類性能を向上させることを目的とする。

【課題を解決するための手段】

【0008】

上記目的を達成するため、本発明の実施の形態における学習装置は、ラベルが付与された第1のデータの集合とラベルが付与されていない第2のデータの集合とを入力として、偽陽性率が所定の範囲である場合における評価指標を表す所定の目的関数の値と、前記目的関数のパラメータに関する微分値とを計算する計算手段と、前記計算手段により計算された前記目的関数の値と前記微分値とを用いて、前記目的関数の値を最大化又は最小化するように前記パラメータを更新する更新手段と、を有することを特徴とする。

10

【発明の効果】

【0009】

特定の偽陽性率における分類性能を向上させることができる。

【図面の簡単な説明】

【0010】

【図1】本発明の実施の形態における学習装置及び分類装置の機能構成の一例を示す図である。

20

【図2】本発明の実施の形態における学習処理の一例を示すフローチャートである。

【図3】本発明の実施の形態における学習装置及び分類装置のハードウェア構成の一例を示す図である。

【発明を実施するための形態】

【0011】

以下、本発明の実施の形態について説明する。本発明の実施の形態では、ラベルありデータとラベルなしデータとが与えられた場合に、特定の偽陽性率における分類性能を向上させることが可能な学習装置10について説明する。また、当該学習装置10によって学習された分類器によりデータを分類する分類装置20についても説明する。なお、ラベルとは、このラベルが付与されたデータが正例又は負例のいずれであるかを示す情報（つまり、正解を示す情報）である。

30

【0012】

<理論的構成>

まず、本発明の実施の形態の理論的構成について説明する。入力データとして、正例を示すラベルが付与されたデータ（以降、「正例データ」とも表す。）の集合

【0013】

【数1】

40

$$\mathcal{P} = \{\mathbf{x}_m^P\}_{m=1}^{M_P}$$

と、負例を示すラベルが付与されたデータ（以降、「負例データ」とも表す。）の集合

【0014】

【数2】

50

$$\mathcal{N} = \{\mathbf{x}_m^N\}_{m=1}^{M_N}$$

と、ラベルなしデータの集合

【 0 0 1 5 】

【 数 3 】

10

$$\mathcal{U} = \{\mathbf{x}_m^U\}_{m=1}^{M_U}$$

20

とが与えられるものとする。ここで、各データは、例えば、D次元の特徴ベクトルである。ただし、各データはベクトルに限られず、任意の形式のデータ（例えば、系列データ、画像データ、集合データ等）であってもよい。

【 0 0 1 6 】

このとき、本発明の実施の形態では、偽陽性率が から の範囲のときの分類性能が高くなるように分類器を学習する。なお、 及び は予め与えられた任意の値（ただし、 $0 < < 1$ ）である。

【 0 0 1 7 】

本発明の実施の形態では学習対象の分類器を $s(x)$ と表す。分類器 $s(x)$ としては、任意の分類器を用いることが可能である。例えば、分類器 $s(x)$ としてニューラルネットワーク等を用いることが可能である。また、分類器 $s(x)$ はデータ x が正例に分類されるスコアを出力するものとする。つまり、データ x は、そのスコアが高いほど正例に分類されやすいものとする。

30

【 0 0 1 8 】

ここで、p A U C は偽陽性率が から の範囲のときの分類性能を示す評価指標である。本発明の実施の形態では、正例データと負例データとを用いて計算した p A U C と、正例データとラベルなしデータとを用いて計算した p A U C と、負例データとラベルなしデータとを用いて計算した p A U C とを利用して、分類器 $s(x)$ を学習する。なお、p A U C は評価指標の一例であって、p A U C の代わりに、特定の偽陽性率での分類性能を示す他の評価指標が用いられてもよい。

40

【 0 0 1 9 】

正例データと負例データとを用いて計算される p A U C は、正例データのスコアが、偽陽性率が から の範囲にある負例データのスコアよりも高い場合に、高い値となる。正例データと負例データとを用いて計算される p A U C は、例えば、以下の式（1）により計算可能である。

【 0 0 2 0 】

【 数 4 】

50

$$\widehat{\text{pAUC}}(\alpha, \beta) = \frac{1}{(\beta - \alpha)M_P M_N} \sum_{\mathbf{x}_m^P \in \mathcal{P}} \left[(j_\alpha - \alpha M_N) I(s(\mathbf{x}_m^P) > s(\mathbf{x}_{(j_\alpha)}^N)) \right. \\ \left. + \sum_{j=j_\alpha+1}^{j_\beta} I(s(\mathbf{x}_m^P) > s(\mathbf{x}_{(j)}^N)) + (\beta M_N - j_\beta) I(s(\mathbf{x}_m^P) > s(\mathbf{x}_{(j_\beta+1)}^N)) \right] \quad (1)$$

10

ここで、 $I(\cdot)$ は指示関数であり、

【 0 0 2 1 】

【数 5】

$$j_\alpha = \lceil \alpha M_N \rceil, j_\beta = \lfloor \beta M_N \rfloor$$

20

である。また、

【 0 0 2 2 】

【数 6】

$$\mathbf{x}_{(j)}^N$$

30

は負例データをスコアの降順に並べたときの j 番目の負例データを表す。

【 0 0 2 3 】

正例データとラベルなしデータとを用いて計算される p A U C は、正例データのスコアが、負例と推定されるラベルなしデータの中で偽陽性率が から の範囲にあるラベルなしデータのスコアよりも高い場合に、高い値となる。正例データとラベルなしデータとを用いて計算される p A U C は、例えば、以下の式 (2) により計算可能である。

40

【 0 0 2 4 】

【数 7】

50

$$\widehat{\text{pAUC}}_{\text{PU}}(\theta_{\text{P}} + \alpha\theta_{\text{N}}, \theta_{\text{P}} + \beta\theta_{\text{N}}) = \frac{1}{(\beta - \alpha)\theta_{\text{N}}M_{\text{P}}M_{\text{U}}} \sum_{\mathbf{x}_m^{\text{P}} \in \mathcal{P}} \left[(k_{\bar{\alpha}} - \bar{\alpha}M_{\text{U}})I(s(\mathbf{x}_m^{\text{P}}) > s(\mathbf{x}_{(k_{\bar{\alpha}})}^{\text{U}})) \right. \\ \left. + \sum_{k=k_{\bar{\alpha}}+1}^{k_{\bar{\beta}}} I(s(\mathbf{x}_m^{\text{P}}) > s(\mathbf{x}_{(k)}^{\text{U}})) + (\bar{\beta}M_{\text{U}} - k_{\bar{\beta}})I(s(\mathbf{x}_m^{\text{P}}) > s(\mathbf{x}_{(k_{\bar{\beta}}+1)}^{\text{U}})) \right] \quad (2)$$

10

ここで、
【 0 0 2 5 】
【 数 8 】

$$\bar{\alpha} = \theta_{\text{P}} + \alpha\theta_{\text{N}}, \bar{\beta} = \theta_{\text{P}} + \beta\theta_{\text{N}}, k_{\bar{\alpha}} = \lceil \bar{\alpha}M_{\text{U}} \rceil, k_{\bar{\beta}} = \lfloor \bar{\beta}M_{\text{U}} \rfloor$$

20

であり、 θ_{N} はラベルなしデータの中の負例の割合である。また、
【 0 0 2 6 】
【 数 9 】

$$\mathbf{x}_{(k)}^{\text{U}}$$

30

はラベルなしデータをスコアの降順に並べたときの k 番目のラベルなしデータを表す。
【 0 0 2 7 】

負例データとラベルなしデータとを用いて計算される p A U C は、正例と推定されるラベルなしデータのスコアが、偽陽性率が α から β の範囲にある負例データのスコアよりも高い場合に、高い値となる。負例データとラベルなしデータとを用いて計算される p A U C は、例えば、以下の式 (3) により計算可能である。

40

【 0 0 2 8 】
【 数 1 0 】

50

$$\begin{aligned}
\widehat{\text{pAUC}}_{\text{NU}}((0, \theta_P), (\alpha, \beta)) &= \frac{1}{(\beta - \alpha)\theta_P M_U M_N} \left[(j_\alpha - \alpha M_N) \sum_{k=0}^{k_{\theta_P}} I(s(\mathbf{x}_{(k)}^U) > s(\mathbf{x}_{(j_\alpha)}^N)) \right. \\
&+ \sum_{k=0}^{k_{\theta_P}} \sum_{j=j_\alpha+1}^{j_\beta} I(s(\mathbf{x}_{(k)}^U) > s(\mathbf{x}_{(j)}^N)) + (\beta M_N - j_\beta) \sum_{k=0}^{k_{\theta_P}} I(s(\mathbf{x}_{(k)}^U) > s(\mathbf{x}_{(j_\beta+1)}^N)) \\
&+ (\theta_P M_U - k_{\theta_P}) \sum_{j=j_\alpha+1}^{j_\beta} I(s(\mathbf{x}_{(k_{\theta_P}+1)}^U) > s(\mathbf{x}_{(j)}^N)) \\
&\left. + (\theta_P M_U - k_{\theta_P})(\beta M_N - j_\beta) I(s(\mathbf{x}_{(k_{\theta_P}+1)}^U) > s(\mathbf{x}_{(j_\beta+1)}^N)) \right] \quad (3)
\end{aligned}$$

10

ここで、 p はラベルなしデータの中の正例の割合である。また、
【 0 0 2 9 】
【 数 1 1 】

20

$$k_{\theta_P} = \lfloor \theta_P M_U \rfloor$$

30

である。

【 0 0 3 0 】

そして、正例データと負例データとを用いて計算される pAUC と、正例データとラベルなしデータとを用いて計算される pAUC と、負例データとラベルなしデータとを用いて計算される pAUC との重み付け和が最大化されるように、分類器 $s(x)$ のパラメータを更新することで、当該分類器 $s(x)$ を学習する。例えば、以下の式(4)に示す L を目的関数として、確率的勾配降下法等の既知の最適化手法を用いることで、当該目的関数 L の値が最大化されるように分類器 $s(x)$ のパラメータを更新することができる。

【 0 0 3 1 】

【 数 1 2 】

40

$$L = \lambda_1 \widehat{\text{pAUC}}(\alpha, \beta) + \lambda_2 \widehat{\text{pAUC}}_{\text{PU}}(\theta_P + \alpha\theta_N, \theta_P + \beta\theta_N) + \lambda_3 \widehat{\text{pAUC}}_{\text{NU}}((0, \theta_P), (\alpha, \beta)) \quad (4)$$

50

ここで、上記の式(4)の第1項は正例データと負例データとを用いて計算される $pAUC$ 、第2項は正例データとラベルなしデータとを用いて計算される $pAUC$ 、第3項は負例データとラベルなしデータとを用いて計算される $pAUC$ である。また、

【0032】

【数13】

～
.

10

はステップ関数を滑らかな関数(つまり、微分可能な関数)に近似したものを表す。ステップ関数の滑らかな近似としては、例えば、シグモイド関数等を用いることができる。

【0033】

また、 α_1 、 α_2 、 α_3 は非負値のハイパーパラメータである。これらのハイパーパラメータは、例えば、分類器 $s(x)$ の学習に利用したデータセット中の開発データが最大になるものを選択することができる。

20

【0034】

なお、上記の式(4)に示す目的関数 L に対して、更に、正則化項や教師なし学習項等が加えられてもよい。

【0035】

以上によって学習された分類器 $s(x)$ を用いることで、本発明の実施の形態では、特定の偽陽性率においてデータ x の分類性能を向上させることが可能となる。なお、本発明の実施の形態は正例データの集合と負例データの集合とラベルなしデータの集合とが与えられる場合について説明するが、例えば、正例データの集合とラベルなしデータの集合とが与えられる場合、負例データの集合とラベルなしデータの集合とが与えられる場合についても同様に適用可能である。正例データの集合とラベルなしデータの集合とが与えられる場合には上記の式(4)に示す目的関数 L は第2項のみとなり、負例データの集合とラベルなしデータの集合とが与えられる場合には上記の式(4)に示す目的関数 L は第3項のみとなる。

30

【0036】

また、本発明の実施の形態は、 $pAUC$ を多クラスの場合に拡張する方法を採用することで、多クラス分類問題にも同様に適用可能である。

【0037】

<機能構成>

以降では、本発明の実施の形態における学習装置10及び分類装置20の機能構成について、図1を参照しながら説明する。図1は、本発明の実施の形態における学習装置10及び分類装置20の機能構成の一例を示す図である。

40

【0038】

図1に示すように、本発明の実施の形態における学習装置10は、読込部101と、目的関数計算部102と、パラメータ更新部103と、終了条件判定部104と、記憶部105とを有する。

【0039】

記憶部105は、各種データを記憶する。記憶部105に記憶されている各種データには、例えば、分類器 $s(x)$ の学習に利用されるデータの集合(つまり、例えば、正例データの集合、負例データの集合、ラベルなしデータの集合)、目的関数のパラメータ(例えば、上記の式(4)に示す目的関数 L のパラメータ)等がある。

50

【 0 0 4 0 】

読込部 1 0 1 は、記憶部 1 0 5 に記憶されている正例データの集合と負例データの集合とラベルなしデータの集合とを読み込む。なお、読込部 1 0 1 は、例えば、正例データの集合と負例データの集合とラベルなしデータの集合とを所定のサーバ装置等から取得（ダウンロード）することで読み込んでよい。

【 0 0 4 1 】

目的関数計算部 1 0 2 は、読込部 1 0 1 により読み込んだ正例データの集合と負例データの集合とラベルなしデータの集合とを用いて、所定の目的関数（例えば、上記の式（4）に示す目的関数 L 等）の値とそのパラメータ（つまり、分類器 $s(x)$ のパラメータ）に関する微分値とを計算する。

10

【 0 0 4 2 】

パラメータ更新部 1 0 3 は、目的関数計算部 1 0 2 により計算された目的関数の値と微分値とを用いて、目的関数の値が高く（又は低く）なるようにパラメータを更新する。

【 0 0 4 3 】

終了条件判定部 1 0 4 は、所定の終了条件を満たすか否かを判定する。終了条件判定部 1 0 4 により終了条件を満たすと判定されるまで、目的関数計算部 1 0 2 による目的関数値及び微分値の計算とパラメータ更新部 1 0 3 によるパラメータの更新とが繰り返し実行される。これにより、分類器 $s(x)$ のパラメータが学習される。学習済みの分類器 $s(x)$ のパラメータは、例えば、任意の通信ネットワークを介して、分類装置 2 0 に送信される。

20

【 0 0 4 4 】

なお、終了条件としては、例えば、繰り返し回数が所定の回数を超えたこと、繰り返しの前後で目的関数値の変化量が所定の第 1 の閾値以下となったこと、更新の前後でパラメータの変化量が所定の第 2 の閾値以下となったこと等が挙げられる。

【 0 0 4 5 】

また、図 1 に示すように、本発明の実施の形態における分類装置 2 0 は、分類部 2 0 1 と、記憶部 2 0 2 とを有する。

【 0 0 4 6 】

記憶部 2 0 2 は、各種データを記憶する。記憶部 2 0 2 に記憶されている各種データには、例えば、学習装置 1 0 によって学習された分類器 $s(x)$ のパラメータ、この分類器 $s(x)$ により分類される分類対象のデータ x 等がある。

30

【 0 0 4 7 】

分類部 2 0 1 は、学習済みの分類器 $s(x)$ を用いて、記憶部 2 0 2 に記憶されているデータ x を分類する。すなわち、分類部 2 0 1 は、例えば、学習済みの分類器 $s(x)$ によりデータ x のスコアを算出した上で、このスコアにより当該データ x を正例又は負例のいずれかに分類する。なお、分類部 2 0 1 は、例えば、スコアが所定の第 3 の閾値以上である場合は正例、そうでない場合は負例と分類すればよい。これにより、特定の偽陽性率においてデータ x を高い精度で分類することができる。

【 0 0 4 8 】

なお、図 1 に示す学習装置 1 0 及び分類装置 2 0 の機能構成は一例であって、他の構成であってもよい。例えば、学習装置 1 0 と分類装置 2 0 とが一体で実現されていてもよい。

40

【 0 0 4 9 】

< 学習処理の流れ >

以降では、学習装置 1 0 が分類器 $s(x)$ を学習する学習処理について、図 2 を参照しながら説明する。図 2 は、本発明の実施の形態における学習処理の一例を示すフローチャートである。

【 0 0 5 0 】

まず、読込部 1 0 1 は、記憶部 1 0 5 に記憶されている正例データの集合と負例データの集合とラベルなしデータの集合とを読み込む（ステップ S 1 0 1 ）。

【 0 0 5 1 】

50

次に、目的関数計算部 102 は、上記のステップ S101 で読み込んだ正例データの集合と負例データの集合とラベルなしデータの集合とを用いて、所定の目的関数（例えば、上記の式（4）に示す目的関数 L 等）の値とそのパラメータに関する微分値とを計算する（ステップ S102）。

【0052】

次に、パラメータ更新部 103 は、上記のステップ S102 で計算された目的関数値及び微分値を用いて、当該目的関数値が高く（又は低く）なるようにパラメータを更新する（ステップ S103）。

【0053】

次に、終了条件判定部 104 は、所定の終了条件を満たすか否かを判定する（ステップ S104）。終了条件を満たすと判定されなかった場合はステップ S102 に戻る。一方で、終了条件を満たすと判定された場合は学習処理を終了する。

10

【0054】

以上により、上記のステップ S102～ステップ S103 の繰り返しによって分類器 $s(x)$ のパラメータが更新され、当該分類器 $s(x)$ が学習される。これにより、分類装置 20 は、学習済みの分類器 $s(x)$ を用いて、特定の偽陽性率においてデータ x を高い精度で分類することができる。

【0055】

<評価>

以降では、本発明の実施の形態の評価について説明する。本発明の実施の形態を評価するため、評価指標を $pAUC$ として、9つのデータセットを用いて評価した。なお、 $pAUC$ の値が高いほど、分類性能が高いことを表す。

20

【0056】

また、本発明の実施の形態の手法を $Ours$ として、比較手法を以下とした。

【0057】

- ・ CE : クロスエントロピーロスを最小化する従来分類手法
- ・ MA : AUC を最大化する従来分類手法
- ・ MPA : $pAUC$ を最大化する従来分類手法
- ・ SS : AUC を最大化する従来半教師あり分類手法
- ・ SSR : ラベル比率を用いて AUC を最大化する従来半教師あり分類手法
- ・ pSS : $pAUC$ を最大化する従来半教師あり分類手法
- ・ $pSSR$: ラベル比率を用いて $pAUC$ を最大化する従来半教師あり分類手法

30

このとき、 $\alpha = 0$, $\beta = 0.1$ とした場合における $Ours$ と各比較手法との $pAUC$ を以下の表 1 に示す。なお、Average は各データセットで算出された $pAUC$ の平均を表す。

【0058】

40

50

【表 1】

	CE	MA	MPA	SS	SSR	pSS	pSSR	Ours
Anthyroid	0.227	0.236	0.384	0.399	0.422	0.258	0.457	0.388
Cardiotocography	0.464	0.473	0.493	0.420	0.450	0.467	0.393	0.527
InternetAds	0.540	0.570	0.565	0.496	0.464	0.527	0.446	0.580
KDDCup99	0.880	0.868	0.874	0.837	0.832	0.867	0.802	0.884
PageBlocks	0.528	0.518	0.593	0.599	0.599	0.553	0.568	0.598
Pima	0.057	0.118	0.188	0.179	0.130	0.127	0.118	0.206
SpamBase	0.408	0.438	0.461	0.422	0.393	0.435	0.416	0.484
Waveform	0.270	0.253	0.288	0.268	0.281	0.305	0.226	0.306
Wilt	0.100	0.195	0.594	0.648	0.403	0.260	0.703	0.681
Average	0.386	0.408	0.493	0.474	0.442	0.422	0.459	0.517

また、 $\alpha = 0$, $\beta = 0.3$ とした場合における Ours と各比較手法との pAUC を以下の表 2 に示す。

【 0 0 5 9 】

10

20

30

40

50

【表 2】

	CE	MA	MPA	SS	SSR	pSS	pSSR	Ours
Anthyroid	0.442	0.436	0.517	0.516	0.445	0.428	0.506	0.503
Cardiotocography	0.680	0.705	0.698	0.661	0.665	0.686	0.637	0.725
InternetAds	0.664	0.697	0.695	0.629	0.631	0.621	0.590	0.672
KDDCup99	0.949	0.941	0.944	0.929	0.914	0.943	0.904	0.961
PageBlocks	0.679	0.677	0.717	0.746	0.744	0.729	0.753	0.727
Pima	0.255	0.324	0.387	0.384	0.364	0.327	0.346	0.355
SpamBase	0.698	0.690	0.691	0.663	0.627	0.662	0.617	0.687
Waveform	0.624	0.619	0.598	0.571	0.548	0.595	0.500	0.609
Wilt	0.326	0.440	0.813	0.803	0.687	0.539	0.790	0.845
Average	0.591	0.614	0.673	0.656	0.625	0.614	0.627	0.676

また、 $\alpha = 0.1$, $\beta = 0.2$ とした場合における Ours と各比較手法との pAUC を以下の表 3 に示す。

【 0 0 6 0 】

10

20

30

40

50

【表 3】

	CE	MA	MPA	SS	SSR	pSS	pSSR	Ours
Anthyroid	0.480	0.469	0.526	0.537	0.459	0.454	0.456	0.510
Cardiotocography	0.729	0.750	0.752	0.697	0.685	0.746	0.601	0.761
InternetAds	0.697	0.734	0.729	0.611	0.637	0.663	0.558	0.724
KDDCup99	0.982	0.977	0.982	0.967	0.956	0.973	0.963	0.988
PageBlocks	0.713	0.718	0.751	0.784	0.782	0.776	0.708	0.763
Pima	0.294	0.353	0.388	0.425	0.404	0.376	0.337	0.447
SpamBase	0.764	0.760	0.775	0.713	0.688	0.727	0.623	0.768
Waveform	0.708	0.695	0.626	0.536	0.594	0.683	0.522	0.654
Wilt	0.341	0.462	0.700	0.854	0.714	0.567	0.858	0.865
Average	0.634	0.658	0.692	0.681	0.658	0.663	0.625	0.720

上記の表 1 ~ 表 3 に示されるように、本発明の実施の形態の手法 (Ours) が他の比較手法よりも多くのデータセットで高い分類性能を達成していることがわかる。

【0061】

<ハードウェア構成>

最後に、本発明の実施の形態における学習装置 10 及び分類装置 20 のハードウェア構成について、図 3 を参照しながら説明する。図 3 は、本発明の実施の形態における学習装置 10 及び分類装置 20 のハードウェア構成の一例を示す図である。なお、学習装置 10 及び分類装置 20 は同様のハードウェア構成で実現されるため、以降では、主に、学習装置 10 のハードウェア構成について説明する。

【0062】

図 3 に示すように、本発明の実施の形態における学習装置 10 は、入力装置 301 と、表示装置 302 と、外部 I/F 303 と、通信 I/F 304 と、プロセッサ 305 と、メモリ装置 306 とを有する。これら各ハードウェアは、それぞれがバス 307 を介して通信可能に接続されている。

【0063】

入力装置 301 は、例えばキーボードやマウス、タッチパネル等であり、ユーザが各種操作を入力するのに用いられる。表示装置 302 は、例えばディスプレイ等であり、学習装置 10 の処理結果等を表示する。なお、学習装置 10 は、入力装置 301 及び表示装置 302 の少なくとも一方を有していなくてもよい。

【0064】

外部 I/F 303 は、外部装置とのインターフェースである。外部装置には、記録媒体 303a 等がある。学習装置 10 は、外部 I/F 303 を介して、記録媒体 303a の読み取りや書き込み等を行うことができる。記録媒体 303a には、例えば、学習装置 10 が有する各機能部 (例えば、読込部 101、目的関数計算部 102、パラメータ更新部 103 及び終了条件判定部 104 等) を実現する 1 以上のプログラム等が記録されていてもよ

10

20

30

40

50

い。

【0065】

記録媒体303aには、例えば、CD (Compact Disc)、DVD (Digital Versatile Disk)、SDメモリカード (Secure Digital memory card)、USB (Universal Serial Bus)メモリカード等がある。

【0066】

通信I/F304は、学習装置10を通信ネットワークに接続するためのインタフェースである。学習装置10が有する各機能部を実現する1以上のプログラムは、通信I/F304を介して、所定のサーバ装置等から取得(ダウンロード)されてもよい。

【0067】

プロセッサ305は、例えばCPU (Central Processing Unit)やGPU (Graphics Processing Unit)等であり、メモリ装置306等からプログラムやデータを読み出して処理を実行する演算装置である。学習装置10が有する各機能部は、メモリ装置306等に格納されている1以上のプログラムがプロセッサ305に実行させる処理により実現される。なお、分類装置20が有する各機能部(例えば、分類部201等)についても、同様に、メモリ装置306等に格納されている1以上のプログラムがプロセッサ305に実行させる処理により実現される。

【0068】

メモリ装置306は、例えばHDD (Hard Disk Drive)やSSD (Solid State Drive)、RAM (Random Access Memory)、ROM (Read Only Memory)、フラッシュメモリ等であり、プログラムやデータが格納される記憶装置である。学習装置10が有する記憶部105は、メモリ装置306等により実現される。なお、分類装置20が有する記憶部202についても、同様に、メモリ装置306等により実現される。

【0069】

本発明の実施の形態における学習装置10及び分類装置20は、図3に示すハードウェア構成を有することにより、上述した各種処理を実現することができる。なお、図3に示すハードウェア構成は一例であって、学習装置10は、他のハードウェア構成を有していてもよい。例えば、学習装置10や分類装置20は、複数のプロセッサ305を有していてもよいし、複数のメモリ装置306を有していてもよい。

【0070】

本発明は、具体的に開示された上記の実施の形態に限定されるものではなく、請求の範囲の記載から逸脱することなく、種々の変形や変更等が可能である。

【符号の説明】

【0071】

10	学習装置
20	分類装置
101	読込部
102	目的関数計算部
103	パラメータ更新部
104	終了条件判定部
105	記憶部
201	分類部
202	記憶部

10

20

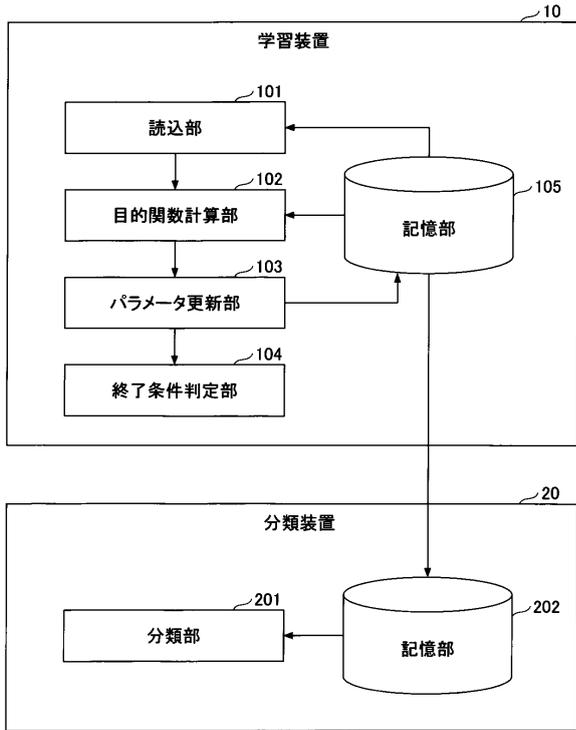
30

40

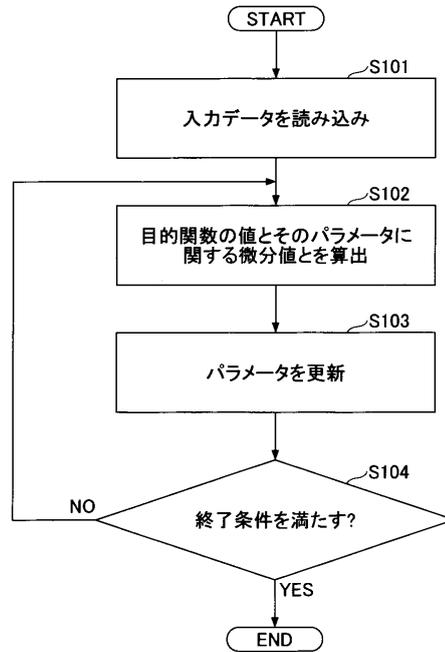
50

【図面】

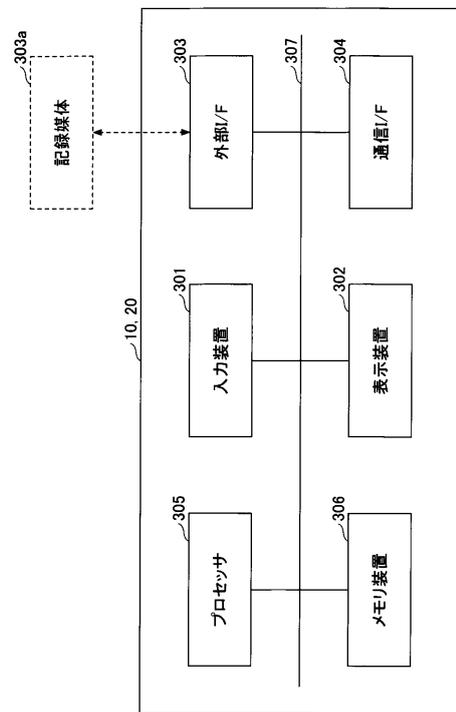
【図 1】



【図 2】



【図 3】



10

20

30

40

50

フロントページの続き

- (56)参考文献 特開 2 0 1 7 - 1 0 2 5 4 0 (J P , A)
特開 2 0 1 8 - 4 9 3 5 5 (J P , A)
特開 2 0 0 9 - 2 8 2 6 8 6 (J P , A)
特開 2 0 2 0 - 8 5 5 8 3 (J P , A)
特開 2 0 1 5 - 2 3 0 5 7 0 (J P , A)
中国特許出願公開第 1 0 9 3 4 4 8 6 9 (C N , A)
SAKAI, Tomoya 他, Semi-Supervised AUC Optimization based on Positive-Unlabeled Learning (正例とラベルなしデータからの, 電子情報通信学会技術研究報告, 日本, 一般社団法人電子情報通信学会, 2017年11月02日, 第117巻 第293号, pp.39-46 (IBISML2017-40) , ISSN 0913-5685
- (58)調査した分野 (Int.Cl., D B 名)
G 0 6 N 2 0 / 0 0