

(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2009-110064
(P2009-110064A)

(43) 公開日 平成21年5月21日(2009.5.21)

(51) Int.Cl.	F I	テーマコード (参考)
G06N 5/04 (2006.01)	G06N 5/04 580A	5B075
G06F 17/30 (2006.01)	G06F 17/30 210D	
G06N 3/00 (2006.01)	G06F 17/30 220Z	
	G06N 3/00 560A	
	G06N 5/04 550N	

審査請求 未請求 請求項の数 6 OL (全 11 頁)

(21) 出願番号 特願2007-278893 (P2007-278893)
(22) 出願日 平成19年10月26日(2007.10.26)

(71) 出願人 000003078
株式会社東芝
東京都港区芝浦一丁目1番1号
(74) 代理人 110000235
特許業務法人 天城国際特許事務所
(72) 発明者 中田 康太
東京都港区芝浦一丁目1番1号 株式会社東芝内
(72) 発明者 櫻井 茂明
東京都港区芝浦一丁目1番1号 株式会社東芝内
(72) 発明者 折原 良平
東京都港区芝浦一丁目1番1号 株式会社東芝内
Fターム(参考) 5B075 ND03 NR12 QT10 UU40

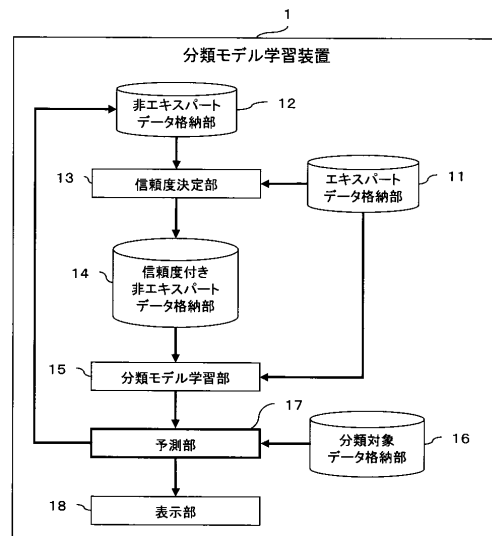
(54) 【発明の名称】 分類モデル学習装置および分類モデル学習方法

(57) 【要約】

【課題】 質の高い教師データが少量の状況であっても精度の良い分類モデルを構築する。

【解決手段】 ラベル付けの精度が所定の基準を満たす教師データをエキスパートデータとして格納するエキスパートデータ格納部 11 と、精度が不明の教師データを非エキスパートデータとして格納する非エキスパートデータ格納部 12 と、非エキスパートデータの信頼度をラベル付けが共通するエキスパートデータとの比較によって決定すると共に、非エキスパートデータに対して信頼度を付加する信頼度決定部 13 と、信頼度が付加された非エキスパートデータを格納する信頼度付き非エキスパートデータ格納部 14 と、エキスパートデータおよび信頼度付きの非エキスパートデータに基づいて所望のデータにラベル付けを行う分類モデルを学習する分類モデル学習部 15 と、を有することを特徴とする分類モデル学習装置。

【選択図】 図 1



【特許請求の範囲】

【請求項 1】

ラベル付けの精度が所定の基準を満たす教師データをエキスパートデータとして格納するエキスパートデータ格納部と、

前記ラベル付けの精度が不明の教師データを非エキスパートデータとして格納する非エキスパートデータ格納部と、

前記エキスパートデータ格納部および前記非エキスパートデータ格納部に接続され、前記非エキスパートデータの信頼度を前記ラベル付けが共通する前記エキスパートデータとの比較によって決定すると共に、前記非エキスパートデータに対して前記信頼度を付加する信頼度決定部と、

この信頼度決定部に接続され、前記信頼度が付加された非エキスパートデータを格納する信頼度付き非エキスパートデータ格納部と、

前記エキスパートデータ格納部および前記信頼度付き非エキスパートデータ格納部に接続され、前記エキスパートデータおよび前記信頼度が付加された非エキスパートデータに基づいて所望のデータに前記ラベル付けを行う分類モデルを学習する分類モデル学習部と、
を有することを特徴とする分類モデル学習装置。

10

【請求項 2】

前記信頼度決定部が、前記エキスパートデータおよび前記非エキスパートデータの各々を所定の規則に基づいて対応付けた座標の間の距離を求め、前記信頼度を前記距離に応じて付加することを特徴とする請求項 1 記載の分類モデル学習装置。

20

【請求項 3】

前記分類モデル学習部が、アンサンブル学習におけるデータ重みに対して前記信頼度決定部で付加された信頼度を反映させることにより前記分類モデルを学習することを特徴とする請求項 1 または請求項 2 記載の分類モデル学習装置。

【請求項 4】

ラベル付けの精度が所定の基準を満たしている教師データをエキスパートデータ、前記ラベル付けの精度が不明の教師データを非エキスパートデータとして格納するコンピュータが行う分類モデル学習方法であって、

前記非エキスパートデータの信頼度を前記ラベル付けが共通する前記エキスパートデータとの比較によって決定すると共に、前記非エキスパートデータに対して前記信頼度を付加する信頼度決定ステップと、

この信頼度決定ステップにおいて前記信頼度が付加された非エキスパートデータと前記エキスパートデータに基づいて所望のデータに前記ラベル付けを行う分類モデルを学習する分類モデル学習ステップと、

を有することを特徴とする分類モデル学習方法。

30

【請求項 5】

前記信頼度決定ステップにおいて、前記エキスパートデータおよび前記非エキスパートデータの各々を所定の規則に基づいて対応付けた座標の間の距離を求め、前記信頼度を前記距離に応じて付加することを特徴とする請求項 4 記載の分類モデル学習方法。

40

【請求項 6】

前記分類モデル学習ステップにおいて、アンサンブル学習におけるデータ重みに対して前記信頼度決定ステップにおいて付加された信頼度を反映させることにより前記分類モデルを学習することを特徴とする請求項 4 または請求項 5 記載の分類モデル学習方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、分類モデル学習装置および分類モデル学習方法に関する。

【背景技術】

【0002】

50

データマイニングにおいて重要な分野の一つに機械学習が挙げられる。機械学習は分類問題に使われることが多く、分類問題において計算機は人間のつけた評価を学習することで分類モデルを構築する。この分類モデルは画像認識、文字認識、テキスト分類など広い分野で多くの成果を上げている。

【0003】

機械学習には計算機に正しい判断を教える「教師データ」、つまり人間の手によって「ラベル」が付けられたデータが必要である。教師データをもとに計算機はどのような分類をすれば良いかを学習し、新しいデータに対して自動的に判断を下せるようになる。現代ではIT環境の発展により、大量で詳細な情報・データが機械学習に利用可能であり、これらを教師データとして用いればより正確な分類モデルの構築に繋がると期待されている。

10

【0004】

しかし、ここで「大量のデータ」から「大量の教師データ」を得る際のラベル付けが問題になっている。すなわち、得られたデータを教師データとして利用するためには、データに対して人間が判断したラベルを付与することが必要であるが、正確なラベル付けには、データが取られたドメインに対しての知識や経験などに基づく正確な判断が不可欠である。

【0005】

理想的にはこれらの条件を満たす対象分野のエキスパートがラベル付けを行うことが望ましいが、全てのデータのラベル付けをエキスパートに依頼することは金銭的・時間的にコストが高くなってしまふ。すなわち、現実的にはコストに制限がある場合がほとんどであり、その場合には非エキスパートがラベル付けを行い、低コストで教師データを得る。しかし、非エキスパートによる教師データには判断の不正確さから比較的多くの誤ったラベルが含まれてしまうことが考えられる。

20

【0006】

一般に、機械学習においては、教師データの取得に関する情報は用いられず、エキスパートによるラベル付けのような「良質の教師データ」と非エキスパートによるラベル付けのような「ノイズを含む教師データ」が混在する状況においても、全てのデータを同列に扱い、等しく学習に使用する。

【0007】

また、前述のコストの関係から、高コストであるエキスパートによる教師データは少量になりがちなのに対し、低コストである非エキスパートによる教師データは比較的大量に獲得できると考えられる。

30

【0008】

そのため、エキスパートによる少量の教師データと非エキスパートによる大量の教師データを従来どおり同列とみなして学習に使用した場合、非エキスパートデータに含まれるノイズが学習に大きく影響し、精度の良い分類モデルが構築できないケースが考えられる。

【0009】

一方、分類モデルを学習する際に、一部の教師データを選択的に使用して学習を行うことや、一部の教師データに重みを置いて学習を行うことは一般的に広く行われている。

40

【0010】

アンサンブル学習の代表的手法の一つであるAdaBoostもその一つである。AdaBoostは、学習データに対して重みを与えて学習器を生成し、その際に誤った分類をしたデータに対して重みを増して再度学習器を生成することを繰り返して複数の弱学習器を得て、それらの弱学習器の重みつき投票により分類を行う手法である（例えば特許文献1参照）。

【特許文献1】特開2002-133389号公報

【発明の開示】

【発明が解決しようとする課題】

50

【 0 0 1 1 】

しかしながら、従来技術は、あくまで所定のアルゴリズムに即した形で教師データに対してデータ重みをつけるものであり、教師データの精度の差異という学習過程を開始する前の知識・情報を含んだものではない。

【 0 0 1 2 】

したがって、例えばエキスパートによる少量の教師データと非エキスパートによる大量の教師データのような、質の異なる教師データを従来どおり同列として学習に使用した場合、質の劣る教師データに含まれるノイズが学習に大きく影響し、精度の良い分類モデルが構築できないという問題があった。

【 0 0 1 3 】

そこで、本発明は、従来技術の問題に鑑み、質の良い教師データが少量しか得られていない状況であっても精度の良い分類モデルの構築が可能になり、教師データ獲得に要するコストを削減可能な分類モデル学習装置および分類モデル学習方法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 】

【 0 0 1 4 】

本発明に係る分類モデル学習装置は、ラベル付けの精度が所定の基準を満たす教師データをエキスパートデータとして格納するエキスパートデータ格納部と、前記ラベル付けの精度が不明の教師データを非エキスパートデータとして格納する非エキスパートデータ格納部と、前記エキスパートデータ格納部および前記非エキスパートデータ格納部に接続され、前記非エキスパートデータの信頼度を前記ラベル付けが共通する前記エキスパートデータとの比較によって決定すると共に、前記非エキスパートデータに対して前記信頼度を付加する信頼度決定部と、この信頼度決定部に接続され、前記信頼度が付加された非エキスパートデータを格納する信頼度付き非エキスパートデータ格納部と、前記エキスパートデータ格納部および前記信頼度付き非エキスパートデータ格納部に接続され、前記エキスパートデータおよび前記信頼度が付加された非エキスパートデータに基づいて所望のデータに前記ラベル付けを行う分類モデルを学習する分類モデル学習部と、を有することを特徴とする。

【 0 0 1 5 】

本発明に係る分類モデル学習方法は、ラベル付けの精度が所定の基準を満たしている教師データをエキスパートデータ、前記ラベル付けの精度が不明の教師データを非エキスパートデータとして格納するコンピュータが行う分類モデル学習方法であって、前記非エキスパートデータの信頼度を前記ラベル付けが共通する前記エキスパートデータとの比較によって決定すると共に、前記非エキスパートデータに対して前記信頼度を付加する信頼度決定ステップと、この信頼度決定ステップにおいて前記信頼度が付加された非エキスパートデータと前記エキスパートデータに基づいて所望のデータに前記ラベル付けを行う分類モデルを学習する分類モデル学習ステップと、を有することを特徴とする。

【 発明の効果 】

【 0 0 1 6 】

本発明によれば、質の良い教師データが少量しか得られていない状況であっても精度の良い分類モデルの構築が可能であり、教師データ獲得に要するコストを削減可能な分類モデル学習装置および分類モデル学習方法が提供される。

【 発明を実施するための最良の形態 】

【 0 0 1 7 】

以下、本発明の実施形態について図面を用いて説明する。図 1 は、本発明の一実施形態に係る分類モデル学習装置 1 の全体構成例を示すブロック図である。同図に示されるように、本実施形態に係る分類モデル学習装置 1 は、エキスパートデータ格納部 1 1、非エキスパートデータ格納部 1 2、信頼度決定部 1 3、信頼度付き非エキスパートデータ格納部 1 4、分類モデル学習部 1 5、分類対象データ格納部 1 6、予測部 1 7、および表示部 1 8 から構成されている。

10

20

30

40

50

【0018】

エキスパートデータ格納部11は、エキスパートデータを格納する記憶装置である。「エキスパートデータ」とは、知識、経験が豊かな専門家によってラベル付けが行われており、ラベル付けの精度(信頼性)が高い教師データを示すものとする。

【0019】

非エキスパートデータ格納部12は、非エキスパートデータを格納する記憶装置である。「非エキスパートデータ」とは、ラベル付けは行われているが、その精度(信頼性)が不明確な教師データを示すものとする。

【0020】

信頼度決定部13は、エキスパートデータを参照することにより全ての非エキスパートデータの信頼度を決定し、信頼度付き非エキスパートデータ格納部14へ登録するプログラムである。

10

【0021】

信頼度付き非エキスパートデータ格納部14は、信頼度決定部13における処理によって信頼度が付与された非エキスパートデータ(以下、「信頼度付き非エキスパートデータ」という。)を格納する記憶装置である。

【0022】

分類モデル学習部15は、エキスパートデータと信頼度付き非エキスパートデータを用いて分類モデルを学習するプログラムである。

【0023】

20

分類対象データ格納部16は、新たに分類の対象となる、ラベルが付与されていないデータ(以下、「分類対象データ」という。)を格納する記憶装置である。

【0024】

予測部17は、分類モデル学習部15で得られた分類モデルを用いて分類対象データ格納部16に格納されているデータにラベル付けを行うプログラムである。尚、AdaBoostを用いた場合、予測部17での分類手法は、一般的なAdaBoostにおける手法と同様である。

【0025】

表示部18は、予測部17における予測結果を表示する表示装置である。

【0026】

30

以下、図2乃至図4に基づいて信頼度決定処理を詳細に説明する。信頼度決定部13は、非エキスパートデータを順次選択し、その非エキスパートデータの近傍のエキスパートデータを参照して信頼度を付与する。ここでは、2次元のデータを具体例として説明を行う。図2は、エキスパートデータおよび非エキスパートデータを2次元で具体的に表した図である。同図において、丸印はエキスパートデータ、四角印は非エキスパートデータを表し、各印の色はラベルを表している。これらの座標はデータを所定の規則に基づいて変換されたものである。例えば、電子メールの分類においては、多数の迷惑メールを解析することによって特徴語リストを予め作成しておき、この特徴語リストと受信メール本文内の単語を比較することで座標化を行う。具体的には、特徴語リストに含まれるN個の単語との比較結果を受信メール内に含まれる場合を1、含まれない場合を0として(1, 0, 1, ..., 1)のように表すことにより受信メールのデータをN次元に座標化することができる。ここでは、説明のためにメールを座標化したN次元のデータを擬似的に2次元で表しているものとする。すなわち、受信メール本文の内容が近似する場合には、座標も近似するので迷惑メールが否かのラベル付け等に用いることができる。

40

【0027】

図3は、信頼度決定部13における信頼度決定処理の具体例を示すフローチャートである。

【0028】

S301においては、信頼度が未設定の非エキスパートデータが存在するか否かを判定する。ここで、全ての非エキスパートデータに信頼度が付与されており、信頼度が未設定

50

の非エキスパートデータが無いと判定された場合は処理を終了する。これに対し、信頼度が未設定の非エキスパートデータが有ると判定された場合は、S302に進む。

【0029】

S302においては、非エキスパートデータ格納部12から信頼度が未設定の非エキスパートデータを1つ選択する。ここでは非エキスパートデータ具体例として、

【数1】

$$[\vec{x}_j, y_j] = [(1.0, 2.0), 1] \quad \dots (1)$$

【0030】

式(1)で示される非エキスパート2次元データの中でj番目のデータが選択されたものとする。

【0031】

S303においては、選択された非エキスパートデータの近傍のN個のエキスパートデータをエキスパートデータ格納部11から探索する。「近傍」とはデータの同士の近さを意味するため、素性の重みを考えた距離など問題に特化した距離の尺度を用いることも可能であるが、ここでは選択した非エキスパートデータとのユークリッド距離が所定の値の範囲内にあることを示すものとする。図4は、エキスパートデータと非エキスパートデータとのユークリッド距離を具体的に説明する図である。ここでは、点Nに位置する非エキスパートデータが選択され、点Nからのユークリッド距離が所定の値の範囲内であるエキスパートデータが探索されていることが示されている。

【0032】

S304においては、探索されたN個のエキスパートデータから、選択された非エキスパートデータの信頼度を決定する。尚、信頼度の上限値、下限値は任意に設定可能であるが、ここでは上限値を1、下限値を0とする。尚、エキスパートデータの信頼度は全て1とする。

【0033】

非エキスパートデータは、近傍のN個のエキスパートデータの中から自らと同じラベルを持つもののみから信頼度を受け取ることができ、受け取る信頼度はエキスパートデータへのユークリッド距離に応じた値となるように設定する。

【0034】

ここで、非エキスパートデータ (x_j, y_j) が選択されており、N個のエキスパートデータ (X_i, Y_i) ($i = 1, 2, \dots, N$)が探索されていたとすると、この信頼度は

【数2】

$$credits(x_j) = \frac{1}{N} \sum_{Y_i=y_j, i \in [1, N]} \frac{1}{l_j + 1} \quad \dots (2)$$

【0035】

式(2)で設定することができる。ここで l_j は、非エキスパートデータ (x_j, y_j) とi番目のエキスパートデータ X_i とのユークリッド距離である。

【0036】

したがって、式(2)は、参照したエキスパートデータとのラベルが等しく、かつ、ユークリッド距離が近い非エキスパートデータの信頼度が高くなることを示す。すなわち、対象の非エキスパートデータと参照したエキスパートデータとの距離が全て0であり、かつ、全てが同じラベルであった場合、信頼度は1となり、エキスパートデータと同じ信頼度となる。

【0037】

逆に、参照したエキスパートデータが全て異なるラベルであった場合、どのエキスパー

10

20

30

40

50

トデータからも信頼度を獲得できないため、対象の非エキスパートデータの信頼度は0になり、データが削除されたのと同様の状況となる。多くの非エキスパートデータは、0から1の間の値をとると考えられ、その値の大きさによって学習への影響度が異なってくる。

【0038】

エキスパートデータ的具体例として、 $N=3$ のときエキスパートデータ格納部11の中から

【数3】

$$[\vec{X}_{I_1}, Y_{I_1}] = [(1.0, 1.0), 1] \quad \dots (3)$$

10

$$[\vec{X}_{I_2}, Y_{I_2}] = [(3.0, 1.0), 0] \quad \dots (4)$$

$$[\vec{X}_{I_3}, Y_{I_3}] = [(-1.0, 2.0), 1] \quad \dots (5)$$

【0039】

式(3)~(5)で表されるエキスパートデータがS303において探索されているものとする、S304においては非エキスパートデータ(x_j, y_j)の信頼度を式(1)および式(3)~(5)を式(2)に代入し、

20

【数4】

$$c_j = \text{credits}(x_j)$$

$$= \frac{1}{3} \left(\frac{1}{\sqrt{(1.0-1.0)^2 + (2.0-1.0)^2 + 1}} + \frac{1}{\sqrt{(-1.0-1.0)^2 + (2.0-1.0)^2 + 1}} \right) \dots (6)$$

【0040】

式(6)により計算する。

30

【0041】

ここで、式(6)右辺の括弧内の第1項は式(3)で表されるエキスパートデータによって与えられる信頼度、第2項は式(5)で表されるエキスパートデータによって与えられる信頼度である。エキスパートデータ式(4)についてはラベル Y_{I_2} が非エキスパートデータ式(1)のラベル y_j と異なるため、信頼度は与えられない。

【0042】

S305において、S304で得られた信頼度 c_j を対象の非エキスパートデータに付加し、

【数5】

40

$$[(\vec{x}_1, y_1, c_1), \dots, (\vec{x}_j, y_j, c_j), \dots, (\vec{x}_n, y_n, c_n)] \quad \dots (7)$$

【0043】

式(7)の形で信頼度付き非エキスパートデータ格納部14に格納する。例えば、前述の式(1)で表される2次元データの場合は

【数6】

$$[\vec{x}_j, y_j, c_j] = [(1.0, 2.0), 1, 0.27] \quad \dots (8)$$

50

【 0 0 4 4 】

式 (8) の形で信頼度付き非エキスパートデータ格納部 1 4 に格納し、処理を終了する。尚、エキスパートデータの信頼度は常に 1 としているので、エキスパートデータは擬似的に

【 数 7 】

$$[(\vec{X}_1, Y_1, 1), \dots, (\vec{X}_i, Y_i, 1), \dots, (\vec{X}_N, Y_N, 1)] \dots (9)$$

【 0 0 4 5 】

式 (9) の形でエキスパートデータ格納部 1 1 に格納されているとみなすことができる。

10

【 0 0 4 6 】

図 5 は、分類モデル学習部 1 5 における分類モデル学習処理の具体例を示すフローチャートである。学習器については信頼度を反映する形のものであれば、どのような学習器でも機能すると考えられるが、ここではデータ重みに対する信頼度の組み込み易さを考慮して Ada Boost の手法に即した形で処理を行うものとする。尚、Bagging などの他の手法を用いても良い。

【 0 0 4 7 】

S501 においては、読み込まれた信頼度付き非エキスパートデータとエキスパートデータに、Ada Boost の手法に即して均等のデータ重み w_j を付ける。本発明では、Ada Boost における従来のデータ重み w_j に加え、信頼度決定部 1 3 で得られた信頼度 c_j が教師データに付加されているため、ここでは読み込まれたエキスパートデータ、非エキスパートデータは

20

【 数 8 】

$$[(\vec{x}_1, y_1, c_1, w_1), \dots, (\vec{x}_j, y_j, c_j, w_j), \dots, (\vec{x}_n, y_n, c_n, w_n)] \dots (10)$$

【 0 0 4 8 】

式 (10) の形で処理されるものとする。

【 0 0 4 9 】

S502 においては、非エキスパートデータに付与された信頼度 c_j をデータ重みに反映させる。ここでは、Ada Boost におけるデータ重み w_j に対して信頼度 c_j を反映させたデータ重み w'_j を

30

【 数 9 】

$$w'_j = c_j w_j \dots (11)$$

【 0 0 5 0 】

式 (11) により設定する。このように設定することにより、データ重み w_j が大きく学習に大きな影響を及ぼすと考えられる非エキスパートデータに関しても、その非エキスパートデータの信頼度 c_j が低ければデータ重み w'_j の値は小さくなり、非エキスパートデータに含まれる信頼度 c_j の低い教師データの影響を自然な形で小さくすることができる。

40

【 0 0 5 1 】

S503 においては、S502 で得られたデータ重み w'_j を用いて弱学習器を生成する。Ada Boost に用いられる弱学習器には決定木など様々なものが考えられる。

【 0 0 5 2 】

S504 においては、Ada Boost のアルゴリズムに従いデータ重みと弱学習器の性能に依るコスト関数の更新を行う。

【 0 0 5 3 】

S505 においては、終了条件を満たしているか否かを判定する。ここで、終了条件を

50

満たすと判定された場合にはS506へ進む。これに対し、終了条件を満たさないと判定された場合はS501に戻る。尚、一般的なAdaBoostの手法における終了条件は、弱学習器の数が所定数を満たすことである。例えばユーザが弱学習器を100個作るという設定にすれば、S501からS505を100回繰り返すことが終了条件である。

【0054】

S506においては、生成された弱学習器を組合せることにより精度の高い分類モデルである強学習器を生成し、処理を終了する。

【0055】

このように、教師データの精度の差異という学習過程を開始する前の知識を利用して非エキスパートデータに信頼度を付与し、分類モデルの学習に組み込むことで、エキスパートデータが少ない場合であっても精度の良い分類モデルを得ることができる。

10

【0056】

図6は、予測部17における分類処理の具体例を示すフローチャートである。

【0057】

S601においては、分類対象データ格納部16における分類対象データの有無を判定する。ここで、分類対象データが有ると判定された場合には、S602へ進む。これに対し、分類対象データが無いと判定された場合には、処理を終了する。

【0058】

S602においては、分類対象データ格納部16から分類対象データを1つ選択する。

【0059】

S603においては、選択した分類対象データを分類モデルに当てはめることにより、ラベル付けを行う。

20

【0060】

このように、分類モデル学習部15で得られた分類モデルを用いて分類対象データにラベル付けを行うことができる。

【0061】

尚、本発明は上記実施形態そのままに限定されるものではなく、実施段階ではその要旨を逸脱しない範囲で構成要素を変形して具体化できる。また、上記実施形態に開示されている複数の構成要素の適宜な組み合わせにより、種々の発明を形成できる。例えば、実施形態に示される全構成要素から幾つかの構成要素を削除してもよい。

30

【図面の簡単な説明】

【0062】

【図1】本発明の一実施形態に係る分類モデル学習装置1の全体構成例を示すブロック図。

【図2】本発明の一実施形態に係るエキスパートデータおよび非エキスパートデータを2次元で具体的に表した図。

【図3】本発明の一実施形態に係る信頼度決定部13における信頼度決定処理の具体例を示すフローチャート。

【図4】本発明の一実施形態に係るエキスパートデータと非エキスパートデータとのユークリッド距離を具体的に説明する図。

40

【図5】本発明の一実施形態に係る分類モデル学習部15における分類モデル学習処理の具体例を示すフローチャート。

【図6】本発明の一実施形態に係る予測部17における分類処理の具体例を示すフローチャート。

【符号の説明】

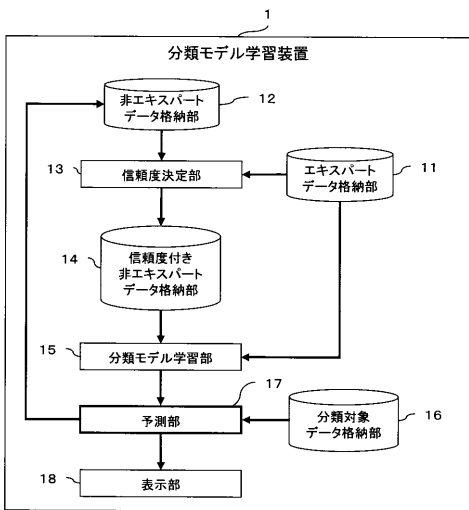
【0063】

- 1 ... 分類モデル学習装置、
- 11 ... エキスパートデータ格納部、
- 12 ... 非エキスパートデータ格納部、
- 13 ... 信頼度決定部、

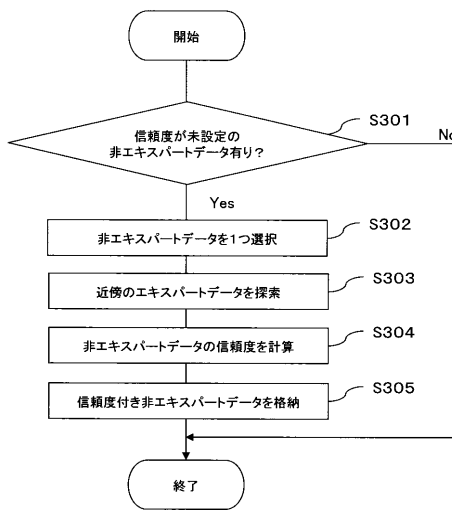
50

- 1 4 ... 信頼度付き非エキスパートデータ格納部、
- 1 5 ... 分類モデル学習部、
- 1 6 ... 分類対象データ格納部、
- 1 7 ... 予測部。

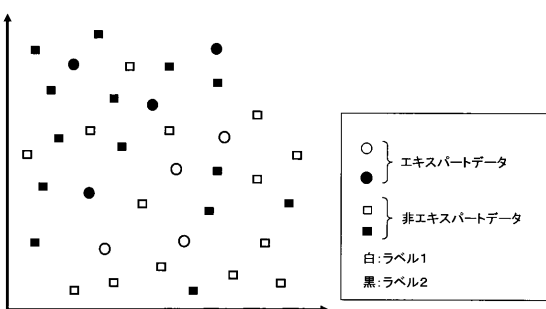
【 図 1 】



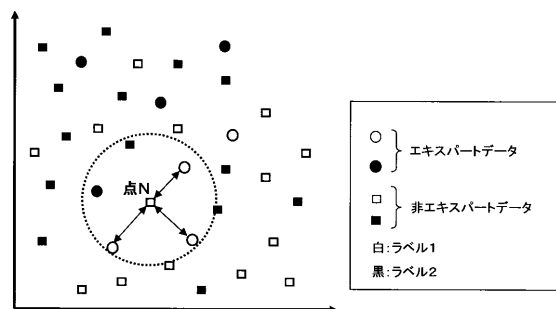
【 図 3 】



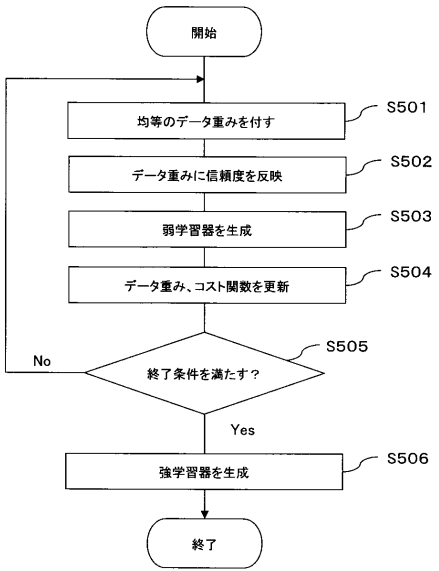
【 図 2 】



【 図 4 】



【図5】



【図6】

