

(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2013-164863

(P2013-164863A)

(43) 公開日 平成25年8月22日(2013.8.22)

(51) Int.Cl.

G06N 5/04 (2006.01)
G06N 3/00 (2006.01)

F 1

G 06 N 5/04 5 8 O A
G 06 N 3/00 5 5 O C

テーマコード (参考)

審査請求 有 請求項の数 11 O L (全 33 頁)

(21) 出願番号 特願2013-90989 (P2013-90989)
(22) 出願日 平成25年4月24日 (2013.4.24)
(62) 分割の表示 特願2007-281035 (P2007-281035)
の分割
原出願日 平成19年10月29日 (2007.10.29)

(71) 出願人 000002185
ソニー株式会社
東京都港区港南1丁目7番1号
(74) 代理人 100082131
弁理士 稲本 義雄
(74) 代理人 100121131
弁理士 西川 孝
(72) 発明者 小林 由幸
東京都港区港南1丁目7番1号 ソニー株
式会社内

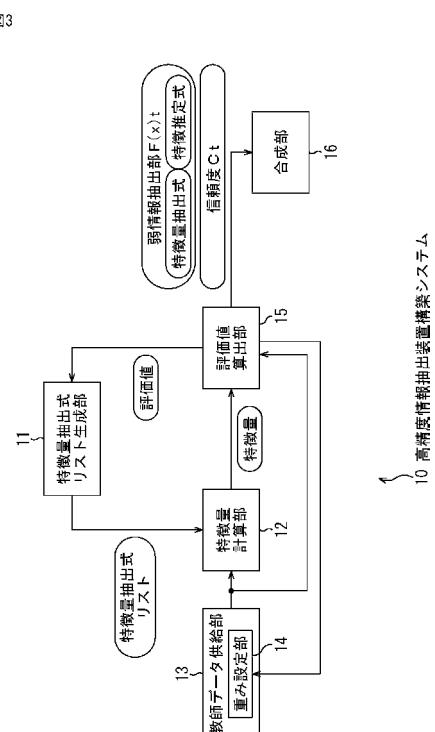
(54) 【発明の名称】情報処理装置、情報処理方法、およびプログラム

(57) 【要約】

【課題】入力データの種類に拘わらずアンサンブル学習における弱情報抽出部を自動的に生成することによって、より高精度の情報抽出装置を自動的に構築する。

【解決手段】高精度情報抽出装置構築システム10は、特徴量抽出式リストを生成する特徴量抽出式リスト生成部11、各特徴量抽出式により教師データの特徴量を計算する特徴量計算部12、教師データを供給する教師データ供給部13、計算された教師データの特徴量と教師データとに基づいて情報抽出式を機械学習により生成するとともに各特徴抽出式の評価値を算出する評価値算出部15、および、評価値算出部15から出力されるT個の弱情報抽出部F(X)tとそれに対応する信頼度Ctを用いて高精度情報抽出装置を構築する合成部16から構成される。本開示は、入力データから、入力データの特徴量を示す情報を高精度に抽出する情報抽出装置を構築する場合に適用できる。

【選択図】図3



【特許請求の範囲】**【請求項 1】**

入力データの特徴を示す情報を抽出する複数の情報抽出部の出力を合成して、前記情報抽出部よりも高精度で前記入力データの特徴を示す情報を抽出する高精度情報抽出部を構築するアンサンブル学習における前記情報抽出部を生成する情報処理装置において、

複数の演算子から成る特微量抽出式を複数含む特微量抽出式リストを、前世代の前記特微量抽出式リストに含まれる複数の特微量抽出式を遺伝子とみなし、前記特微量抽出式の評価値に基づいた遺伝的アルゴリズムを用いて前世代の前記特微量抽出式リストを更新することにより生成する特微量抽出式リスト生成手段と、

前記特微量抽出式リストに含まれる各特微量抽出式に、実データ、前記実データの特徴を示す情報、および重みからなる教師データの実データを入力して、前記実データに対応する複数の特微量を計算する特微量計算手段と、

計算された前記教師データの前記実データに対応する前記複数の特微量から、前記教師データの前記実データに対応する前記情報を推定する機械学習により、前記特微量抽出式リストに含まれる各特微量抽出式にそれぞれ対応する情報抽出部候補を生成するとともに、生成した前記情報抽出部候補によって正しく推定された教師データの重みを用いて各特微量抽出式にそれぞれ対応する評価値を算出し、最終世代の前記特微量抽出式リストに含まれる特微量抽出式のうちで前記評価値が最も良い特微量抽出式に対応する前記情報抽出部候補を、前記アンサンブル学習における1つの前記情報抽出部に決定して、決定した前記情報抽出部によって間違って推定された教師データの重みを用いて、決定した前記情報抽出部の信頼度を算出する評価値算出手段と、

決定された前記情報抽出部の前記信頼度を用いて、教師データの重みを更新する更新手段と

を含むことを特徴とする情報処理装置。

【請求項 2】

複数の前記情報抽出部を、前記情報抽出部の前記信頼度に基づいて合成することにより、前記高精度情報抽出部を構築する合成手段を

さらに含むことを特徴とする請求項1に記載の情報処理装置。

【請求項 3】

入力データの特徴を示す情報を抽出する複数の情報抽出部の出力を合成して、前記情報抽出部よりも高精度で前記入力データの特徴を示す情報を抽出する高精度情報抽出部を構築するアンサンブル学習における前記情報抽出部を生成する情報処理装置の情報処理方法において、

複数の演算子から成る特微量抽出式を複数含む第1世代の特微量抽出式リストをランダムに生成し、

前記特微量抽出式リストに含まれる各特微量抽出式に、実データ、前記実データの特徴を示す情報、および重みからなる教師データの実データを入力して、前記実データに対応する複数の特微量を計算し、

計算された前記教師データの前記実データに対応する前記複数の特微量から、前記教師データの前記実データに対応する前記情報を推定する機械学習により、前記特微量抽出式リストに含まれる各特微量抽出式にそれぞれ対応する情報抽出部候補を生成するとともに、生成した前記情報抽出部候補によって正しく推定された教師データの重みを用いて各特微量抽出式にそれぞれ対応する評価値を算出し、

前世代の前記特微量抽出式リストに含まれる複数の特微量抽出式を遺伝子とみなし、前記特微量抽出式の前記評価値に基づいた遺伝的アルゴリズムを用いて前世代の前記特微量抽出式リストを更新し、

最終世代の前記特微量抽出式リストに含まれる特微量抽出式のうちで前記評価値が最も良い特微量抽出式に対応する前記情報抽出部候補を、前記アンサンブル学習における1つの前記情報抽出部に決定して、

決定した前記情報抽出部によって間違って推定された教師データの重みを用いて、決定

10

20

30

40

50

した前記情報抽出部の信頼度を算出し、

決定された前記情報抽出部の前記信頼度を用いて、教師データの重みを更新する
ステップを含むことを特徴とする情報処理方法。

【請求項 4】

入力データの特徴を示す情報を抽出する複数の情報抽出部の出力を合成して、前記情報抽出部よりも高精度で前記入力データの特徴を示す情報を抽出する高精度情報抽出部を構築するアンサンブル学習における前記情報抽出部を生成する情報処理装置の制御用のプログラムであって、

複数の演算子から成る特徴量抽出式を複数含む第1世代の特徴量抽出式リストをランダムに生成し、10

前記特徴量抽出式リストに含まれる各特徴量抽出式に、実データ、前記実データの特徴を示す情報、および重みからなる教師データの実データを入力して、前記実データに対応する複数の特徴量を計算し、

計算された前記教師データの前記実データに対応する前記複数の特徴量から、前記教師データの前記実データに対応する前記情報を推定する機械学習により、前記特徴量抽出式リストに含まれる各特徴量抽出式にそれぞれ対応する情報抽出部候補を生成するとともに、生成した前記情報抽出部候補によって正しく推定された教師データの重みを用いて各特徴量抽出式にそれぞれ対応する評価値を算出し、

前世代の前記特徴量抽出式リストに含まれる複数の特徴量抽出式を遺伝子とみなし、前記特徴量抽出式の前記評価値に基づいた遺伝的アルゴリズムを用いて前世代の前記特徴量抽出式リストを更新し、20

最終世代の前記特徴量抽出式リストに含まれる特徴量抽出式のうちで前記評価値が最も良い特徴量抽出式に対応する前記情報抽出部候補を、前記アンサンブル学習における1つの前記情報抽出部に決定して、

決定した前記情報抽出部によって間違って推定された教師データの重みを用いて、決定した前記情報抽出部の信頼度を算出し、

決定された前記情報抽出部の前記信頼度を用いて、教師データの重みを更新する
ステップを含む処理を情報処理装置のコンピュータに実行させることを特徴とするプログラム。

【請求項 5】

入力データの特徴を示す情報を抽出する複数の情報抽出部の出力を合成して、前記情報抽出部よりも高精度で前記入力データの特徴を示す情報を抽出する高精度情報抽出部を構築するアンサンブル学習における前記情報抽出部を生成する情報処理装置において、30

実データ、および前記実データの特徴を示す情報からなる教師データをランダムに選択する選択手段と、

複数の演算子から成る特徴量抽出式を複数含む特徴量抽出式リストを、前世代の前記特徴量抽出式リストに含まれる複数の特徴量抽出式を遺伝子とみなし、前記特徴量抽出式の評価値に基づいた遺伝的アルゴリズムを用いて前世代の前記特徴量抽出式リストを更新することにより生成する特徴量抽出式リスト生成手段と、

前記特徴量抽出式リストに含まれる各特徴量抽出式に、前記教師データの実データを入力して、前記実データに対応する複数の特徴量を計算する特徴量計算手段と、40

各特徴量抽出式にそれぞれ対応する評価値として、各特徴量抽出式を用いて計算された前記教師データの前記実データに対応する前記複数の特徴量を用いて、前記教師データの前記実データに対応する前記情報を推定した場合の精度を算出し、最終世代の前記特徴量抽出式リストに含まれる特徴量抽出式のうちで前記評価値が最も良い特徴量抽出式を用いて計算された前記教師データの前記実データに対応する前記複数の特徴量から、前記教師データの前記実データに対応する前記情報を推定する機械学習により、前記アンサンブル学習における1つの前記情報抽出部を生成する評価値算出手段と

を含むことを特徴とする情報処理装置。

【請求項 6】

10

20

30

40

50

複数の前記情報抽出部を合成することにより、前記高精度情報抽出部を構築する合成手段を

さらに含むことを特徴とする請求項 5 に記載の情報処理装置。

【請求項 7】

入力データの特徴を示す情報を抽出する複数の情報抽出部の出力を合成して、前記情報抽出部よりも高精度で前記入力データの特徴を示す情報を抽出する高精度情報抽出部を構築するアンサンブル学習における前記情報抽出部を生成する情報処理装置の情報処理方法において、

実データ、および前記実データの特徴を示す情報からなる教師データをランダムに選択し、

複数の演算子から成る特徴量抽出式を複数含む第 1 世代の特徴量抽出式リストをランダムに生成し、

前記特徴量抽出式リストに含まれる各特徴量抽出式に、前記教師データの実データを入力して、前記実データに対応する複数の特徴量を計算し、

各特徴量抽出式にそれぞれ対応する評価値として、各特徴量抽出式を用いて計算された前記教師データの前記実データに対応する前記複数の特徴量を用いて、前記教師データの前記実データに対応する前記情報を推定した場合の精度を算出し、

前世代の前記特徴量抽出式リストに含まれる複数の特徴量抽出式を遺伝子とみなし、前記特徴量抽出式の前記評価値に基づいた遺伝的アルゴリズムを用いて前世代の前記特徴量抽出式リストを更新し、

最終世代の前記特徴量抽出式リストに含まれる特徴量抽出式のうちで前記評価値が最も良い特徴量抽出式を用いて計算された前記教師データの前記実データに対応する前記複数の特徴量から、前記教師データの前記実データに対応する前記情報を推定する機械学習により、前記アンサンブル学習における 1 つの前記情報抽出部を生成する

ステップを含むことを特徴とする情報処理方法。

【請求項 8】

入力データの特徴を示す情報を抽出する複数の情報抽出部の出力を合成して、前記情報抽出部よりも高精度で前記入力データの特徴を示す情報を抽出する高精度情報抽出部を構築するアンサンブル学習における前記情報抽出部を生成する情報処理装置の制御用のプログラムであって、

実データ、および前記実データの特徴を示す情報からなる教師データをランダムに選択し、

複数の演算子から成る特徴量抽出式を複数含む第 1 世代の特徴量抽出式リストをランダムに生成し、

前記特徴量抽出式リストに含まれる各特徴量抽出式に、前記教師データの実データを入力して、前記実データに対応する複数の特徴量を計算し、

各特徴量抽出式にそれぞれ対応する評価値として、各特徴量抽出式を用いて計算された前記教師データの前記実データに対応する前記複数の特徴量を用いて、前記教師データの前記実データに対応する前記情報を推定した場合の精度を算出し、

前世代の前記特徴量抽出式リストに含まれる複数の特徴量抽出式を遺伝子とみなし、前記特徴量抽出式の前記評価値に基づいた遺伝的アルゴリズムを用いて前世代の前記特徴量抽出式リストを更新し、

最終世代の前記特徴量抽出式リストに含まれる特徴量抽出式のうちで前記評価値が最も良い特徴量抽出式を用いて計算された前記教師データの前記実データに対応する前記複数の特徴量から、前記教師データの前記実データに対応する前記情報を推定する機械学習により、前記アンサンブル学習における 1 つの前記情報抽出部を生成する

ステップを含む処理を情報処理装置のコンピュータに実行させることを特徴とするプログラム。

【発明の詳細な説明】

【技術分野】

10

20

30

40

50

【0001】

本開示は、情報処理装置、情報処理方法、およびプログラムに関し、特に、例えば、楽曲データなどの入力データから当該入力データの特徴を示す情報を高精度に抽出できる高精度情報抽出アルゴリズムを自動的に構築するようにした情報処理装置、情報処理方法、およびプログラムに関する。

【背景技術】**【0002】**

従来、楽曲データや画像データなどを入力データとして、前記入力データの情報（入力データが楽曲データである場合、速さ、明るさ、にぎやかさ等）を出力することができるアルゴリズムを自動的に構築する発明が提案されている（例えば、特許文献1参照）。 10

【0003】

また、入力データから当該入力データの特徴を示す情報を高精度に抽出できるアルゴリズムを構築する方法として、アンサンブル学習と称する手法が知られている。

【0004】

アンサンブル学習は、複数の教師データを用いて精度の低い情報抽出装置（以下、弱情報抽出部(weak learner)と称する）を複数生成し、生成した複数の弱情報抽出部による出力を組み合わせることによって高精度の情報抽出装置を得る手法である。

【0005】

アンサンブル手法の例としては、ブースティング(boosting)とバッギング(bagging)の2種類を挙げることができる。 20

【0006】

ブースティングでは、各教師データに重み付けを行い、全ての教師データを用いて1つの弱情報抽出部を生成する。そして生成した弱情報抽出部によって正しく情報が抽出された教師データについては重みを下げ、間違って情報が抽出された教師データについては重みを上げることにより、教師データの重みを更新する。また、重みが更新された全ての教師データを用いてさらにもう1つの弱情報抽出部を生成する。以下同様の処理を繰り返すことにより、複数の弱情報抽出部を生成し、生成した複数の弱情報抽出部の出力をそれぞれの重み付けで組み合わせることによって高精度の情報抽出装置を得る。

【0007】

バッギングでは、全ての教師データから所定数をランダムにサンプリングして教師データ群を生成し、教師データ群を用いて1つの弱情報抽出部を生成する。この処理を繰り返すことにより、複数の弱情報抽出部を生成し、生成した複数の弱情報抽出部の出力を組み合わせることによって高精度の情報抽出装置を得る。 30

【先行技術文献】**【特許文献】****【0008】**

【特許文献1】米国特許出願公報 U S 2 0 0 4 / 0 1 8 1 4 0 1 A 1

【発明の概要】**【発明が解決しようとする課題】****【0009】**

上述したブースティングやバッギングなどアンサンブル手法では、より多くの弱情報抽出部を生成して、その出力を組み合わせることによって、より高精度の情報抽出装置を得ることができる。

【0010】

しかしながら、入力データの種類に拘わらずアンサンブル学習における弱情報抽出部を自動的に生成する方法は従来確立されておらず、弱情報抽出部を人手によって生成する必要があった。したがって、弱情報抽出部の数を増やして情報抽出装置の精度を所望のレベルまで到達させることができた。

【0011】

本開示はこのような状況に鑑みてなされたものであり、入力データの種類に拘わらずア 50

ンサンプル学習における弱情報抽出部を自動的に生成することによって、より高精度の情報抽出装置を自動的に構築できるようとするものである。

【課題を解決するための手段】

【0012】

本開示の第1の側面である情報処理装置は、入力データの特徴を示す情報を抽出する複数の情報抽出部の出力を合成して、前記情報抽出部よりも高精度で前記入力データの特徴を示す情報を抽出する高精度情報抽出部を構築するアンサンブル学習における前記情報抽出部を生成する情報処理装置において、複数の演算子から成る特微量抽出式を複数含む特微量抽出式リストを、前世代の前記特微量抽出式リストに含まれる複数の特微量抽出式を遺伝子とみなし、前記特微量抽出式の評価値に基づいた遺伝的アルゴリズムを用いて前世代の前記特微量抽出式リストを更新することにより生成する特微量抽出式リスト生成手段と、前記特微量抽出式リストに含まれる各特微量抽出式に、実データ、前記実データの特徴を示す情報、および重みからなる教師データの実データを入力して、前記実データに対応する複数の特微量を計算する特微量計算手段と、計算された前記教師データの前記実データに対応する前記複数の特微量から、前記教師データの前記実データに対応する前記情報を推定する機械学習により、前記特微量抽出式リストに含まれる各特微量抽出式にそれぞれ対応する情報抽出部候補を生成するとともに、生成した前記情報抽出部候補によって正しく推定された教師データの重みを用いて各特微量抽出式にそれぞれ対応する評価値を算出し、最終世代の前記特微量抽出式リストに含まれる特微量抽出式のうちで前記評価値が最も良い特微量抽出式に対応する前記情報抽出部候補を、前記アンサンブル学習における1つの前記情報抽出部に決定して、決定した前記情報抽出部によって間違って推定された教師データの重みを用いて、決定した前記情報抽出部の信頼度を算出する評価値算出手段と、決定された前記情報抽出部の前記信頼度を用いて、教師データの重みを更新する更新手段とを含むことを特徴とする。

10

20

30

40

【0013】

本開示の第1の側面である情報処理装置は、複数の前記情報抽出部を、前記情報抽出部の前記信頼度に基づいて合成することにより、前記高精度情報抽出部を構築する合成手段をさらに含むことができる。

【0014】

本開示の第1の側面である情報処理方法は、入力データの特徴を示す情報を抽出する複数の情報抽出部の出力を合成して、前記情報抽出部よりも高精度で前記入力データの特徴を示す情報を抽出する高精度情報抽出部を構築するアンサンブル学習における前記情報抽出部を生成する情報処理装置の情報処理方法において、複数の演算子から成る特微量抽出式を複数含む第1世代の特微量抽出式リストをランダムに生成し、前記特微量抽出式リストに含まれる各特微量抽出式に、実データ、前記実データの特徴を示す情報、および重みからなる教師データの実データを入力して、前記実データに対応する複数の特微量を計算し、計算された前記教師データの前記実データに対応する前記複数の特微量から、前記教師データの前記実データに対応する前記情報を推定する機械学習により、前記特微量抽出式リストに含まれる各特微量抽出式にそれぞれ対応する情報抽出部候補を生成するとともに、生成した前記情報抽出部候補によって正しく推定された教師データの重みを用いて各特微量抽出式にそれぞれ対応する評価値を算出し、前世代の前記特微量抽出式リストに含まれる複数の特微量抽出式を遺伝子とみなし、前記特微量抽出式の前記評価値に基づいた遺伝的アルゴリズムを用いて前世代の前記特微量抽出式リストを更新し、最終世代の前記特微量抽出式リストに含まれる特微量抽出式のうちで前記評価値が最も良い特微量抽出式に対応する前記情報抽出部候補を、前記アンサンブル学習における1つの前記情報抽出部に決定して、決定した前記情報抽出部によって間違って推定された教師データの重みを用いて、決定した前記情報抽出部の信頼度を算出し、決定された前記情報抽出部の前記信頼度を用いて、教師データの重みを更新するステップを含むことを特徴とする。

【0015】

本開示の第1の側面であるプログラムは、入力データの特徴を示す情報を抽出する複数

50

の情報抽出部の出力を合成して、前記情報抽出部よりも高精度で前記入力データの特徴を示す情報を抽出する高精度情報抽出部を構築するアンサンブル学習における前記情報抽出部を生成する情報処理装置の制御用のプログラムであって、複数の演算子から成る特徴量抽出式を複数含む第1世代の特徴量抽出式リストをランダムに生成し、前記特徴量抽出式リストに含まれる各特徴量抽出式に、実データ、前記実データの特徴を示す情報、および重みからなる教師データの実データを入力して、前記実データに対応する複数の特徴量を計算し、計算された前記教師データの前記実データに対応する前記複数の特徴量から、前記教師データの前記実データに対応する前記情報を推定する機械学習により、前記特徴量抽出式リストに含まれる各特徴量抽出式にそれぞれ対応する情報抽出部候補を生成するとともに、生成した前記情報抽出部候補によって正しく推定された教師データの重みを用いて各特徴量抽出式にそれぞれ対応する評価値を算出し、前世代の前記特徴量抽出式リストに含まれる複数の特徴量抽出式を遺伝子とみなし、前記特徴量抽出式の前記評価値に基づいた遺伝的アルゴリズムを用いて前世代の前記特徴量抽出式リストを更新し、最終世代の前記特徴量抽出式リストに含まれる特徴量抽出式のうちで前記評価値が最も良い特徴量抽出式に対応する前記情報抽出部候補を、前記アンサンブル学習における1つの前記情報抽出部に決定して、決定した前記情報抽出部によって間違って推定された教師データの重みを用いて、決定した前記情報抽出部の信頼度を算出し、決定された前記情報抽出部の前記信頼度を用いて、教師データの重みを更新するステップを含む処理を情報処理装置のコンピュータに実行させることを特徴とする。

10

20

30

40

50

【0016】

本開示の第1の側面においては、複数の演算子から成る特徴量抽出式を複数含む第1世代の特徴量抽出式リストがランダムに生成され、前記特徴量抽出式リストに含まれる各特徴量抽出式に、実データ、前記実データの特徴を示す情報、および重みからなる教師データの実データが入力されて、前記実データに対応する複数の特徴量が計算される。また、計算された前記教師データの前記実データに対応する前記複数の特徴量から、前記教師データの前記実データに対応する前記情報を推定する機械学習により、前記特徴量抽出式リストに含まれる各特徴量抽出式にそれぞれ対応する情報抽出部候補が生成されるとともに、生成された前記情報抽出部候補によって正しく推定された教師データの重みを用いて各特徴量抽出式にそれぞれ対応する評価値が算出される。さらに、前世代の前記特徴量抽出式リストに含まれる複数の特徴量抽出式を遺伝子とみなし、前記特徴量抽出式の前記評価値に基づいた遺伝的アルゴリズムを用いて前世代の前記特徴量抽出式リストが更新される。そして、最終世代の前記特徴量抽出式リストに含まれる特徴量抽出式のうちで前記評価値が最も良い特徴量抽出式に対応する前記情報抽出部候補が、前記アンサンブル学習における1つの前記情報抽出部に決定され、決定された前記情報抽出部によって間違って推定された教師データの重みを用いて、決定された前記情報抽出部の信頼度が算出され、決定された前記情報抽出部の前記信頼度を用いて、教師データの重みが更新される。

【0017】

本開示の第2の側面である情報処理装置は、入力データの特徴を示す情報を抽出する複数の情報抽出部の出力を合成して、前記情報抽出部よりも高精度で前記入力データの特徴を示す情報を抽出する高精度情報抽出部を構築するアンサンブル学習における前記情報抽出部を生成する情報処理装置において、実データ、および前記実データの特徴を示す情報からなる教師データをランダムに選択する選択手段と、複数の演算子から成る特徴量抽出式を複数含む特徴量抽出式リストを、前世代の前記特徴量抽出式リストに含まれる複数の特徴量抽出式を遺伝子とみなし、前記特徴量抽出式の評価値に基づいた遺伝的アルゴリズムを用いて前世代の前記特徴量抽出式リストを更新することにより生成する特徴量抽出式リスト生成手段と、前記特徴量抽出式リストに含まれる各特徴量抽出式に、前記教師データの実データを入力して、前記実データに対応する複数の特徴量を計算する特徴量計算手段と、各特徴量抽出式にそれぞれ対応する評価値として、各特徴量抽出式を用いて計算された前記教師データの前記実データに対応する前記複数の特徴量を用いて、前記教師データの前記実データに対応する前記情報を推定した場合の精度を算出し、最終世代の前記特

徴量抽出式リストに含まれる特徴量抽出式のうちで前記評価値が最も良い特徴量抽出式を用いて計算された前記教師データの前記実データに対応する前記複数の特徴量から、前記教師データの前記実データに対応する前記情報を推定する機械学習により、前記アンサンブル学習における1つの前記情報抽出部を生成する評価値算出手段とを含むことを特徴とする。

【0018】

本開示の第2の側面である情報処理装置は、複数の前記情報抽出部を合成することにより、前記高精度情報抽出部を構築する合成手段をさらに含むことができる。

【0019】

本開示の第2の側面である情報処理方法は、入力データの特徴を示す情報を抽出する複数の情報抽出部の出力を合成して、前記情報抽出部よりも高精度で前記入力データの特徴を示す情報を抽出する高精度情報抽出部を構築するアンサンブル学習における前記情報抽出部を生成する情報処理装置の情報処理方法において、実データ、および前記実データの特徴を示す情報からなる教師データをランダムに選択し、複数の演算子から成る特徴量抽出式を複数含む第1世代の特徴量抽出式リストをランダムに生成し、前記特徴量抽出式リストに含まれる各特徴量抽出式に、前記教師データの実データを入力して、前記実データに対応する複数の特徴量を計算し、各特徴量抽出式にそれぞれ対応する評価値として、各特徴量抽出式を用いて計算された前記教師データの前記実データに対応する前記複数の特徴量を用いて、前記教師データの前記実データに対応する前記情報を推定した場合の精度を算出し、前世代の前記特徴量抽出式リストに含まれる複数の特徴量抽出式を遺伝子とみなし、前記特徴量抽出式の前記評価値に基づいた遺伝的アルゴリズムを用いて前世代の前記特徴量抽出式リストを更新し、最終世代の前記特徴量抽出式リストに含まれる特徴量抽出式のうちで前記評価値が最も良い特徴量抽出式を用いて計算された前記教師データの前記実データに対応する前記複数の特徴量から、前記教師データの前記実データに対応する前記情報を推定する機械学習により、前記アンサンブル学習における1つの前記情報抽出部を生成するステップを含むことを特徴とする。

10

20

30

40

【0020】

本開示の第2の側面であるプログラムは、入力データの特徴を示す情報を抽出する複数の情報抽出部の出力を合成して、前記情報抽出部よりも高精度で前記入力データの特徴を示す情報を抽出する高精度情報抽出部を構築するアンサンブル学習における前記情報抽出部を生成する情報処理装置の制御用のプログラムであって、実データ、および前記実データの特徴を示す情報からなる教師データをランダムに選択し、複数の演算子から成る特徴量抽出式を複数含む第1世代の特徴量抽出式リストをランダムに生成し、前記特徴量抽出式リストに含まれる各特徴量抽出式に、前記教師データの実データを入力して、前記実データに対応する複数の特徴量を計算し、各特徴量抽出式にそれぞれ対応する評価値として、各特徴量抽出式を用いて計算された前記教師データの前記実データに対応する前記複数の特徴量を用いて、前記教師データの前記実データに対応する前記情報を推定した場合の精度を算出し、前世代の前記特徴量抽出式リストに含まれる複数の特徴量抽出式を遺伝子とみなし、前記特徴量抽出式の前記評価値に基づいた遺伝的アルゴリズムを用いて前世代の前記特徴量抽出式リストを更新し、最終世代の前記特徴量抽出式リストに含まれる特徴量抽出式のうちで前記評価値が最も良い特徴量抽出式を用いて計算された前記教師データの前記実データに対応する前記複数の特徴量から、前記教師データの前記実データに対応する前記情報を推定する機械学習により、前記アンサンブル学習における1つの前記情報抽出部を生成するステップを含む処理を情報処理装置のコンピュータに実行させることを特徴とする。

【0021】

本開示の第2の側面においては、実データ、および前記実データの特徴を示す情報からなる教師データがランダムに選択され、複数の演算子から成る特徴量抽出式を複数含む第1世代の特徴量抽出式リストがランダムに生成され、前記特徴量抽出式リストに含まれる各特徴量抽出式に、前記教師データの実データが入力されて、前記実データに対応する複

50

数の特徴量が計算される。また、各特徴量抽出式にそれぞれ対応する評価値として、各特徴量抽出式を用いて計算された前記教師データの前記実データに対応する前記複数の特徴量を用いて、前記教師データの前記実データに対応する前記情報を推定した場合の精度が算出される。さらに、前世代の前記特徴量抽出式リストに含まれる複数の特徴量抽出式を遺伝子とみなし、前記特徴量抽出式の前記評価値に基づいた遺伝的アルゴリズムを用いて前世代の前記特徴量抽出式リストが更新される。さらに、最終世代の前記特徴量抽出式リストに含まれる特徴量抽出式のうちで前記評価値が最も良い特徴量抽出式を用いて計算された前記教師データの前記実データに対応する前記複数の特徴量から、前記教師データの前記実データに対応する前記情報を推定する機械学習により、前記アンサンブル学習における1つの前記情報抽出部が生成される。

10

【発明の効果】

【0022】

本開示の第1および第2の側面によれば、入力データの種類に拘わらずアンサンブル学習における弱情報抽出部を自動的に生成することができる。

【0023】

また本開示の第1および第2の側面によれば、入力データの種類に拘わらずより高精度の情報抽出装置を自動的に構築できる。

【図面の簡単な説明】

【0024】

【図1】アンサンブル学習による高精度情報抽出装置を説明する図である。

20

【図2】図1の高精度情報抽出装置における弱情報抽出部の構成を説明する図である。

【図3】本開示の第1の実施の形態である高精度情報抽出装置構築システムの構成例を示すブロック図である。

【図4】特徴量抽出式の例を示す図である。

【図5】特徴量抽出式の構成を説明する図である。

【図6】特徴量抽出式リストの例を示す図である。

【図7】遺伝的アルゴリズムを説明するための図である。

【図8】教師データのデータ構造を示す図である。

【図9】図3の特徴量計算部によって計算される特徴量の例を示す図である。

【図10】特徴量抽出式に対応する情報推定式(閾値)、および評価値の決定方法を説明するための図である。

30

【図11】図3の高精度情報抽出装置構築システムによる動作を説明するフローチャートである。

【図12】図11のステップS2の処理を説明するフローチャートである。

【図13】図11のステップS5の処理を説明するフローチャートである。

【図14】本開示の第2の実施の形態である高精度情報抽出装置構築システムの構成例を示すブロック図である。

【図15】図14の高精度情報抽出装置構築システムによる動作を説明するフローチャートである。

【図16】図15のステップS62の処理を説明するフローチャートである。

40

【図17】コンピュータの構成例を示すブロック図である。

【発明を実施するための形態】

【0025】

以下、本開示を適用した具体的な実施の形態について、図面を参照しながら詳細に説明する。

【0026】

本開示を適用した高精度情報抽出装置構築システムは、図1に示すような、入力データXの特徴を示す情報を出力する複数の弱情報抽出部2-1乃至2-T、および、弱情報抽出部2-1乃至2-Tの出力を組み合わせることによって、入力データXの特徴を示す情報を高精度で出力する合成部3から成る高精度情報装置1を自動的に構築するものである

50

。

【0027】

なお、入力データXは、多次元のデータであればよく、その種類は任意である。例えば、時間の次元とチャンネルの次元を有する楽曲データ、X次元とY次元と画素の次元を有する画像データ、画像データに時間の次元を加えた動画像データなどを入力データXとすることができる。

【0028】

高精度情報装置1が出力する入力データXに対する情報としては、例えば、入力データXが楽曲データである場合、当該楽曲データの明暗（明るいか、明るくないかの2値情報）、または、明るさ（その値によって明るさの程度を示す数値情報。例えば、0から5までの値として、0は全く明るくない。5は非常に明るいとする）などを挙げることができる。勿論、上に挙げた例以外を入力データXとその情報とすることができます。

10

【0029】

図1の弱情報抽出部2-1は、図2に示すように、特徴量抽出式と情報推定式から構成されている。特徴量抽出式は、入力データXに対して所定の演算を行い、演算結果として1次元の値を出力する。情報推定式は、特徴量抽出式の出力である1次元の値から、入力データXの特徴を示す情報を推定する。ここで、情報推定式は、例えば、その出力（すなわち、弱情報抽出部2-1の出力）を、+1または-1の2値情報とする場合には、特徴量抽出式の出力である1次元の値と比較するための閾値からなる判別式とすることができる。また例えば、その出力を、所定の範囲の数値とする場合には、特徴量抽出式の出力である1次元の値を入力とする線形結合式とすることができる。

20

【0030】

図1の弱情報抽出部2-2乃至2-Tについても、弱情報抽出部2-1と同様である。

【0031】

次に、本開示の第1の実施の形態である高精度情報抽出装置構築システム10の構成例について、図3を参照して説明する。この高精度情報抽出装置構築システム10は、複数の教師データを用いたアンサンブル学習のブースティングによって高精度情報抽出装置1を構築するものである。

【0032】

この高精度情報抽出装置構築システム10は、m本の特徴量抽出式からなる特徴量抽出式リストを生成、更新する特徴量抽出式リスト生成部11、生成された各特徴量抽出式に教師データの実データを代入して特徴量を計算する特徴量計算部12、教師データを特徴量計算部12と評価値算出部15に供給する教師データ供給部13、特徴量計算部12によって計算された教師データに対応する特徴量と教師データに基づいて情報抽出式を機械学習により生成するとともに特徴量抽出式リストを構成する各特徴抽出式の評価値を算出する評価値算出部15、および、評価値算出部15から出力されるT個の弱情報抽出部F(X)_tとそれに対応する信頼度C_tを用いて高精度情報抽出装置1を構築する合成部16から構成される。

30

【0033】

特徴量抽出式リスト生成部11は、第1世代の特徴量抽出式リストを構成するm本の特徴量抽出式をランダムに生成し、生成した第1世代の特徴量抽出式リストを特徴量計算部12に供給する。

40

【0034】

ここで、特徴量抽出式リスト生成部11によって生成される特徴量抽出式について、図4を参照して説明する。図4A乃至図4Dは、それぞれ特徴量抽出式の例を示している。

【0035】

特徴量抽出式には、左端に入力データの種類が記述され、入力データの種類の右側には、1種類以上のオペレータ（演算子）が演算される順序に従って記述される。各オペレータには、適宜、処理対称軸とパラメータが含まれる。

【0036】

50

オペレータの種類としては、平均値(Mean)、高速フーリエ変換(FFT)、標準偏差(St Dev)、出現率(Ratio)、ローパスフィルタ(LPF)、ハイパスフィルタ(HPF)、絶対値(ABS)、2乗(Sqr)、平方根(Sqrt)、正規化(Normalize)、微分(Differential)、積分(Integrate)、最大値(MaxIndex)、不偏分散(UVariance)、ダウンサンプリング(DownSampling)などを挙げることができる。なお、決定されたオペレータによっては処理対称軸が固定されていることがあるので、その場合、パラメータに固定されている処理対称軸を採用する。また、パラメータを必要とするオペレータが決定された場合、パラメータもランダムまたは予め設定されている値に決定する。

【0037】

例えれば、図4Aに示された特徴量抽出式の場合、12TomesMが入力データであり、32#Differential, 32#MaxIndex, 16#LPF_1;0.861, 16#UVarianceそれがオペレータである。
また、各オペレータ中の32#, 16#などは処理対称軸を示している。

【0038】

ここで、12TomesMはモノラルのPCM(pulse coded modulation sound source)波形データを時間軸に沿って音程解析したものであることを示しており、48#はチャンネル軸、32#は周波数軸と音程軸、16#は時間軸を示している。オペレータ中の0.861はローパスフィルタ処理におけるパラメータであり、例えば透過させる周波数の閾値を示している。

【0039】

なお、第1世代の特徴量抽出式リストを構成する各特徴量抽出式の入力データの種類は入力データXと同じもの、オペレータの数と種類はランダムに決定されるが、各特徴量抽出式を生成する際の制約として、図5に示すように、複数のオペレータに対応する演算が順次実行されるにつれて、演算結果の保有次元数が順次減少し、特徴量抽出式の最終的な演算結果がスカラになるか、あるいはその次元数が1となるようになされている。

【0040】

図4A乃至図4Dに示された例から明らかなように、特徴量抽出式によって計算される特徴量は、例えば、楽曲データに対するテンポ、画像データに対する画素のヒストグラムなどのように、既存の概念で有意義と判断される値になるわけではない。すなわち、特徴量抽出式によって計算される特徴量は、単に入力データを特徴量抽出式に代入したときの演算結果に過ぎないものでよい。

【0041】

以下、特徴量抽出式リスト生成部11によって生成される特徴量抽出式リストは、図6に示すように、m本の特徴量抽出式f1乃至fmによって構成されているものとする。図6の例において、特徴量抽出式f1乃至fmの入力データであるWavMはモノラルのPCM波形データであり、保有次元は時間軸とチャンネル軸である。

【0042】

図3に戻る。特徴量抽出式リスト生成部11はまた、第2世代以降の特徴量抽出式リストを、前世代の特徴量抽出式リストを遺伝的アルゴリズム(GA:genetic algorism)に従つて更新することによって生成し、生成した特徴量抽出式リストを特徴量計算部12に供給する。

【0043】

ここで、遺伝的アルゴリズムとは、現世代の遺伝子から、選択処理、交差処理、突然変異処理、およびランダム生成処理により、次世代の遺伝子を生成するアルゴリズムを指す。具体的には、特徴量抽出式リストを構成する複数の各特徴量抽出式を遺伝子とみなし、現世代の特徴量抽出式リストを構成する複数の特徴量抽出式の評価値に応じて選択処理、交差処理、突然変異処理、およびランダム生成処理を行い、次世代の特徴量抽出式リストを生成する。

【0044】

すなわち、例えは図7に示すように、選択処理では、現世代の特徴量抽出式リストを構成する複数の特徴量抽出式のうち、評価値の高い特徴量抽出式f2を選択して次世代の特徴量抽出式リストに含める。交差処理では、現世代の特徴量抽出式リストを構成する複数

10

20

30

40

50

の特徴量抽出式のうち、評価値の高い複数の特徴量抽出式 f_2 と f_5 を交差させて（組み合わせて）特徴量抽出式を生成し、次世代の特徴量抽出式リストに含める。

【0045】

突然変異処理では、現世代の特徴量抽出式リストを構成する複数の特徴量抽出式のうち、評価値の高い特徴量抽出式 f_2 を部分的に突然変異させて（変更して）特徴量抽出式を生成し、次世代の特徴量抽出式リストに含める。ランダム生成処理では、新たな特徴量抽出式をランダムに生成して次世代の特徴量抽出式リストに含める。

【0046】

図3に戻る。特徴量計算部12は、特徴量抽出式リスト生成部11から供給された特徴量抽出式リストを構成する各特徴量抽出式 f_1 乃至 f_m に、教師データ供給部13から供給される複数の教師データ T_i の実データ D_i を代入し、教師データ T_i に対する特徴量を計算し、計算した特徴量を評価値算出部15に供給する。
10

【0047】

ここで、教師データ供給部13から供給される教師データ T_i について説明する。図8は、教師データ T_i のデータ構造を示している。

【0048】

総数 L 個の教師データ T_i ($i = 1, 2, \dots, L$) は、入力データ X と同じ種類のデータである実データ D_i 、実データ D_i に対応する情報 I_i を有している。情報 I_i は、例えば、実データ D_i の明るさを示すものであり、（明るいか (+1)、明るくないか (-1) の2値情報）であってもよいし、明るさ（その値によって明るさの程度を示す数値情報。例えば、0から5までの値として、0は全く明るくない。5は非常に明るいとする）であってもよい。以下、情報 I_i は、+1または-1の2値情報であるとする。
20

【0049】

さらに、教師データ T_i には重み W_i が設定されている。初期状態において、各教師データ T_i の重み W_i は均一であり、 L 個の教師データ T_i の重み W_i の合計が 1 と成るように正規化されている。

【0050】

教師データ供給部13は、重み設定部14を内蔵する。重み設定部14は、各教師データ T_i の重み W_i を、評価値算出部15から供給される、構築された弱情報抽出部 $F(X)_t$ によって各教師データ T_i が正しく判別されたか否かを示す情報と、弱情報抽出部 $F(X)_t$ に対応する信頼度 C_t に基づいて更新する。
30

【0051】

具体的には、弱情報抽出部 $F(X)_t$ によって正しく判別された教師データ T_i については現状の重み W_i を $\text{EXP}(-C_t)$ 倍とし、弱情報抽出部 $F(X)_t$ によって間違って判別された教師データ T_i については、現状の重み W_i を $\text{EXP}(C_t)$ 倍とする。さらに、 L 個の教師データ T_i の重み W_i の合計が 1 と成るように正規化する。

【0052】

図3に戻る。上述したように、教師データ T_i の数は L 、特徴量抽出式リストを構成する特徴量抽出式の数は m であるので、特徴量計算部12では、図9に示すように、($L \times m$) 個の特徴量が算出されることになる。以下、特徴量抽出式 f_j ($j = 1, 2, \dots, m$) に、教師データ T_i ($i = 1, 2, \dots, L$) の実データ D_i を代入して計算された特徴量を $f_j[T_i]$ と記述する。なお、図9には、特徴量 $f_j[T_i]$ の具体的な値が記載されている。
40

【0053】

評価値算出部15は、 L 個の教師データ T_i と L 個の特徴量 $f_1[T_i]$ とに基づき、特徴量抽出式 f_1 に対応する情報推定式を決定するとともに、特徴量抽出式 f_1 の評価値を算出する。なお、ここで情報推定式は、その出力を2値情報とするので、情報推定式として、特徴量 $f_1[T_i]$ と比較する閾値を有する判別式が決定される。

【0054】

この判別式は、例えば、

特徴量 $f_1[T_i] >$ 閾値 情報 = - 1
 特徴量 $f_1[T_i]$ 閾値 情報 = + 1

とする。

【0055】

具体的には、図10に示すように、横軸には特徴量 $f_1[T_i]$ の値、縦軸は重み W_i をかけた教師データの分布を示すグラフに、 L 個の教師データ T_i を情報 I_i の値 (+ 1 または - 1) に応じて 2 つに分類してプロットする。そして、情報推定式（閾値による判別式）によって、情報 I_i が正しく判別された教師データ T_i の重み W_i の合計が最大となるように閾値を決定する。すなわち、図10に示された左側の山の分布（教師データ T_i の情報 I_i の値が + 1 である分布）のうちの閾値よりも左側の面積と、図10に示された右側の山の分布（教師データ T_i の情報 I_i の値が - 1 である分布）のうちの閾値よりも右側の面積との合計が最大となるように閾値を決定する。10

【0056】

そして、決定した情報推定式（閾値による判別式）によって、情報 I_i が正しく判別された教師データ T_i の重み W_i の合計を、特徴量抽出式 f_1 の評価値とする。

【0057】

同様に、評価値算出部15は、特徴量抽出式 f_2 乃至 f_m にそれぞれ対応する情報推定式を決定するとともに、特徴量抽出式 f_2 乃至 f_m の評価値を算出する。20

【0058】

ここで算出された特徴量抽出式 f_1 乃至 f_m の評価値は、特徴量抽出式リスト生成部11に供給されて、次世代の特徴量抽出式リストの生成に利用される。

【0059】

さらに、評価値算出部15は、所定の終了条件が満たされた数世代後（最終世代とする）の特徴量抽出式リストを構成する特徴量抽出式 f_1 乃至 f_m のうち、最も評価値の良い特徴量抽出式 f_i とそれに対応する情報推定式から 1 つの弱情報抽出部 $F(X)_t$ を構築する。さらに、構築した弱情報抽出部 $F(X)_t$ の信頼度 C_t を算出する。

$$\text{信頼度 } C_t = 1 / 2 \log((1 - E) / E)$$

【0060】

ここで、 E は構築した弱情報抽出部 $F(X)_t$ のエラー率であり、弱情報抽出部 $F(X)_t$ によって、情報 I_i が間違って判別された教師データ T_i の重み W_i の合計値（図10の例では、左側の山の分布（教師データ T_i の情報 I_i の値が + 1 である分布）のうちの閾値よりも右側の面積と、図10に示された右側の山の分布（教師データ T_i の情報 I_i の値が - 1 である分布）のうちの閾値よりも左側の面積との合計）である。30

【0061】

ここで構築された弱情報抽出部 $F(X)_t$ とそれに対応する信頼度 C_t は合成部16に供給される。また、構築された弱情報抽出部 $F(X)_t$ によって各教師データ T_i が正しく判別されたか否かを示す情報と、弱情報抽出部 $F(X)_t$ 対応する信頼度 C_t が教師データ供給部13の重み設定部14に供給される。

【0062】

合成部16は、評価値算出部15から供給されている 1 個以上の弱情報抽出部 $F(X)_t$ と、それぞれに対応する信頼度 C_t を用いた次式に従い、高精度情報抽出装置1（図1）を構築する。40

$$\text{Sign}(C_t \cdot F(X)_t)$$

【0063】

すなわち、合成部16では、複数の弱情報抽出部 $F(X)_t$ の出力がその信頼度 C_t によって重み付けられて加算される。そして、その総和の符号が正であるならば、入力データ X の情報 I が + 1、その総和の符号が負であるならば、入力データ X の情報 I が - 1 とされる、高精度情報抽出装置1が構築される。

【0064】

次に、高精度情報抽出装置構築システム10による動作について、図11のフローチャ

ートを参照して説明する。

【0065】

ステップS1において、教師データ供給部13は、内蔵する重み設定部14により、予め用意されているL個の教師データT_iの各重みW_iを均等な値1/Lに初期化し、第1世代の教師データT_iとして特徴量計算部12および評価値算出部15に供給する。

【0066】

ステップS2において、第1世代の教師データT_iに対応する弱情報抽出部F(X)₁が生成される。ステップS2の処理について、図12のフローチャートを参照して詳述する。

【0067】

ステップS11において、特徴量抽出式リスト生成部11は、m本の特徴量抽出式をランダムに生成し、生成したm本の特徴量抽出式からなる第1世代の特徴量抽出式リストを特徴量計算部12に供給する。

【0068】

ステップS12において、特徴量計算部12は、特徴量抽出式リスト生成部11から供給された特徴量抽出式リストを構成するm本の特徴量抽出式f₁乃至f_mを順に1本ずつ注目する特徴量抽出式リストループを開始する。

【0069】

ステップS13において、特徴量計算部12は、注目している特徴量抽出式f_jに、L個の教師データT_iを代入してL個の特徴量f_j[T_i]を計算し、評価値算出部15に出力する。ステップS14において、評価値算出部15は、L個の教師データT_iとL個の特徴量f_j[T_i]とに基づき、特徴量抽出式f_jに対応する情報推定式を決定し、ステップS15において、特徴量抽出式f_jの評価値を算出する。ステップS12乃至S15の処理により、注目している特徴量抽出式f_jに対応する情報推定式と評価値が得られたことになる。

【0070】

ステップS16において、特徴量計算部12は、特徴量抽出式リスト生成部11から供給された特徴量抽出式リストを構成するm本の特徴量抽出式f₁乃至f_mのうち、注目していないものが残っているか否かを判定する。そして、注目していない特徴量抽出式が残っていると判定された場合、処理はステップS12に戻り、ステップS12乃至S16の処理が繰り返される。

【0071】

そして、ステップS16において、特徴量抽出式リスト生成部11から供給された特徴量抽出式リストを構成するm本の特徴量抽出式f₁乃至f_mのうち、注目していないものが残っていないと判定された場合、m本の特徴量抽出式f₁乃至f_mにそれぞれ対応する情報推定式と評価値が得られたことになるので、処理はステップS17に進められる。

【0072】

ステップS17において、評価値算出部15は、所定の終了条件を満たしているか否かを判定する。ここで、所定の終了条件としては、例えば、m本の特徴量抽出式f₁乃至f_mにそれぞれ対応する評価値のうち、最も良い評価値の値が所定の数世代の間、向上していないことなどとすることができます。

【0073】

ステップS17において、所定の終了条件を満たしていないと判定された場合、処理はステップS18に進められる。

【0074】

ステップS18において、評価値算出部15は、現世代の特徴量抽出式リストを構成するm本の特徴量抽出式f₁乃至f_mにそれぞれ対応する評価値を特徴量抽出式リスト生成部11に供給する。特徴量抽出式リスト生成部11は、現世代の特徴量抽出式リストを遺伝的アルゴリズムに従って更新することにより、次世代の特徴量抽出式リストを生成して特徴量計算部12に供給する。

10

20

30

40

50

【0075】

具体的には、遺伝的アルゴリズムの選択処理として、現世代の特徴量抽出式リストを構成する複数の特徴量抽出式のうち、評価値の良い m_s 本の特徴量抽出式を選択して次世代の特徴量抽出式リストに含める。また、遺伝的アルゴリズムの交差処理として、現世代の特徴量抽出式リストを構成する複数の特徴量抽出式のうち、評価値の良いものほど選択され易いように重み付けをして 2 本の特徴量抽出式を選択し、選択した 2 本の特徴量抽出式を交差させる（組み合わせる）ことにより、 $m \times$ 本の特徴量抽出式を生成し、次世代の特徴量抽出式リストに含める。

【0076】

さらに、遺伝的アルゴリズムの突然変異処理として、現世代の特徴量抽出式リストを構成する複数の特徴量抽出式のうち、評価値の良いものほど選択され易いように重み付けをして 1 本の特徴量抽出式を選択し、選択した 1 本の特徴量抽出式を部分的に突然変異させる（変更する）ことにより、 m_m 本の特徴量抽出式を生成し、次世代の特徴量抽出式リストに含める。さらにより、遺伝的アルゴリズムのランダム生成処理として、新たに m_r ($= m - m_s - m_x - m_m$) 本の特徴量抽出式をランダムに生成して次世代の特徴量抽出式リストに含める。

10

【0077】

以上のように、次世代の特徴量抽出式リストが生成されて特徴量計算部 12 に供給された後、処理はステップ S12 に戻り、ステップ S12 乃至 S18 の処理が繰り返される。そして、ステップ S17 において、所定の終了条件を満たしていると判定された場合、処理はステップ S19 に進められる。

20

【0078】

ステップ S19 において、評価値算出部 15 は、現世代、すなわち、最終世代の特徴量抽出式リストを構成する m 本の特徴量抽出式 f_1 乃至 f_m のうち、評価値が最も良い特徴量抽出式 f_i とそれに対応する情報推定式から 1 つ目の弱情報抽出部 $F(X)_1$ を構築し、構築した弱情報抽出部 $F(X)_1$ の信頼度 C_1 を算出する。さらに、評価値算出部 15 は、構築した 1 つ目の弱情報抽出部 $F(X)_1$ とそれに対応する信頼度 C_1 を合成部 16 に供給する。またさらに、評価値算出部 15 は、構築した弱情報抽出部 $F(X)_1$ によって第 1 世代の各教師データ T_i が正しく判別されたか否かを示す情報と、弱情報抽出部 $F(X)_1$ 対応する信頼度 C_1 を教師データ供給部 13 の重み設定部 14 に供給する。

30

【0079】

以上のように、第 1 世代の教師データ T_i に対応する 1 つ目の弱情報抽出部 $F(X)_1$ が構築され、その信頼度 C_1 が算出された後、処理は図 11 のステップ S3 に戻される。

【0080】

ステップ S3 において、合成部 16 は、現時点までに評価値算出部 15 から供給されている 1 個以上の弱情報抽出部 $F(X)_t$ とそれに対応する信頼度 C_t を合成することにより、高精度情報抽出装置 1 を構築する。

40

【0081】

ステップ S4 において、合成部 16 は、所定の終了条件を満たしているか否かを判定する。ここで、所定の終了条件としては、例えば、所定の数 T だけ弱情報抽出部 $F(X)_t$ とそれに対応する信頼度 C_t が評価値算出部 15 から供給されていること、高精度情報抽出装置 1 が所望の精度に達していること、またはユーザから終了が指示されたことのうち、少なくとも 1 つが満たされていることなどとすることができる。

【0082】

ステップ S4 において、所定の終了条件を満たしていないと判定された場合、処理はステップ S5 に進められる。

【0083】

ステップ S5 において、教師データ供給部 13 の重み設定部 14 は、各教師データ T_i の重み W_i を、ステップ S19 の処理で評価値算出部 15 から供給された、現世代の教師データ T_i に対応する弱情報抽出部 $F(X)_t$ によって各教師データ T_i が正しく判別さ

50

れたか否かを示す情報と、弱情報抽出部 $F(X)_t$ に対応する信頼度 C_t に基づいて更新する。

【0084】

ステップ S 5 の処理について、図 13 のフローチャートを参照して詳述する。

【0085】

ステップ S 4 1において、重み設定部 14 は、L 個の教師データ T_i を順に 1 つずつ注目する教師データループを開始する。

【0086】

ステップ S 4 2において、重み設定部 14 は、注目する教師データ T_i が、現世代の教師データ T_i に対応する弱情報抽出部 $F(X)_t$ によって正しく判別されたか否かを判定する。
10

【0087】

ステップ S 4 2において、正しく判別されたと判定された場合、処理はステップ S 4 3 に進められる。ステップ S 4 3において、重み設定部 14 は、注目する教師データ T_i の現状の重み W_i を $\text{EXP}(-C_t)$ 倍することにより更新する。この後、処理はステップ S 4 5 に進められる。

【0088】

反対に、ステップ S 4 2において、正しく判別されていない（間違って判別されている）と判定された場合、処理はステップ S 4 4 に進められる。ステップ S 4 4において、重み設定部 14 は、注目する教師データ T_i の現状の重み W_i を $\text{EXP}(C_t)$ 倍することにより更新する。この後、処理はステップ S 4 5 に進められる。
20

【0089】

ステップ S 4 5において、重み設定部 14 は、L 個の教師データ T_i のうち、注目していないものが残っているか否かを判定し、注目していないものが残っていると判定した場合、ステップ S 4 1 に戻って、ステップ S 4 1 乃至 4 5 の処理を繰り返す。

【0090】

そして、ステップ S 4 5において、L 個の教師データ T_i のうち、注目していないものが残っていないと判定された場合、L 個の教師データ T_i にそれぞれ対応する重み W_i を全て更新したので、処理はステップ S 4 6 に進められる。

【0091】

ステップ S 4 6において、重み設定部 14 は、L 個の教師データ T_i にそれぞれ対応する更新した重み W_i の合計が 1 となるように正規化する。
30

【0092】

以上のようにして重み W_i が更新された L 個の教師データ T_i が、次世代の教師データ T_i として評価値算出部 15 に供給される。なお、2 世代以降の教師データ T_i は、実データ D_i および情報 I_i に変更はなく、重み W_i だけが更新されているので、次世代の教師データ T_i として、更新された重み W_i だけを評価値算出部 15 に供給するようにしてもよい。

【0093】

この後、処理は図 11 のステップ S 2 に戻り、ステップ S 2 乃至 S 5 の処理が繰り返される。この繰り返しにより、教師データ T_i が更新された世代数と同じ数だけ、弱情報抽出部 $F(X)_t$ とそれに対応する信頼度 C_t が合成部 16 に供給され、合成部 16 により、徐々に制度が高められた高精度情報抽出装置 1 が構築されることになる。
40

【0094】

そして、ステップ S 4において、所定の終了条件を満たしていると判定された場合、処理はステップ S 5 に進められる。ステップ S 5において、合成部 16 は、直前のステップ S 3 の処理で構築した高精度情報抽出装置 1 を、最終的な高精度情報抽出装置 1 として出力する。

【0095】

以上で、本開示の第 1 の実施の形態である高精度情報抽出装置構築システム 10 の動作

10

20

30

40

50

説明を終了する。

【0096】

次に、本開示の第2の実施の形態である高精度情報抽出装置構築システム50の構成例について、図14を参照して説明する。この高精度情報抽出装置構築システム50は、複数の教師データを用いたアンサンブル学習のバッギングによって高精度情報抽出装置1を構築するものである。

【0097】

なお、この高精度情報抽出装置構築システム50と、図3に示された本開示の第1の実施の形態である、アンサンブル学習のブースティングによって高精度情報抽出装置1を構築する高精度情報抽出装置構築システム10との間で共通する構成要素については同一の符号を付しているので、その説明は適宜省略する。

10

【0098】

この高精度情報抽出装置構築システム50は、m本の特徴量抽出式からなる特徴量抽出式リストを生成、更新する特徴量抽出式リスト生成部11、生成された各特徴量抽出式に教師データの実データを代入して特徴量を計算する特徴量計算部12、教師データを特徴量計算部12と評価値算出部53に供給する教師データ供給部51、特徴量計算部12によって計算された教師データに対応する特徴量と教師データとに基づいて情報抽出式を機械学習により生成するとともに特徴量抽出式リストを構成する各特徴抽出式の評価値を算出する評価値算出部53、および、評価値算出部15から出力されるT個の弱情報抽出部F(X)_tとそれに対応する信頼度C_tを用いて高精度情報抽出装置1を構築する合成部54から構成される。

20

【0099】

特徴量抽出式リスト生成部11は、第1世代の特徴量抽出式リストをランダムに生成する。また、特徴量抽出式リスト生成部11は、第2世代以降の特徴量抽出式リストを、前世代の特徴量抽出式リストを遺伝的アルゴリズムにしたがって更新することにより生成する。生成された特徴量抽出式リストは特徴量計算部12に供給される。

30

【0100】

特徴量計算部12は、特徴量抽出式リスト生成部11から供給された特徴量抽出式リストを構成する各特徴量抽出式f₁乃至f_mに、教師データ供給部51から供給される、教師データT_i(i=1, 2, ..., J)の実データD_iを代入し、教師データT_iに対する特徴量を計算し、計算した特徴量を評価値算出部53に供給する。

30

【0101】

ここで、教師データ供給部51から供給される教師データT_iについて説明する。

【0102】

教師データ供給部51は、内蔵するランダムサンプリング部52により、総数L個の教師データT_iの中からJ個をランダムに選択して教師データ群を生成し、特徴量計算部12および評価値算出部53に供給する。なお、教師データT_iは、入力データXと同じ種類のデータである実データD_i、実データD_iに対応する情報I_iを有している。情報I_iは、例えば、実データD_iの明るさを示すものであり、(明るいか(+1)、明るくないか(-1)の2値情報)であってもよいし、明るさ(その値によって明るさの程度を示す数値情報。例えば、0から5までの値として、0は全く明るくない。5は非常に明るいとする)であってもよい。以下、情報I_iは、+1または-1の2値情報であるとする。

40

【0103】

さらに、教師データT_iには重みW_iが設定されている。初期状態において、各教師データT_iの重みW_iは均一であり、L個の教師データT_iの重みW_iの合計が1と成るように正規化されている。

【0104】

上述したように、教師データT_iの数はJ、特徴量抽出式リストを構成する特徴量抽出式の数はmであるので、特徴量計算部12では、(J×m)個の特徴量が算出されることになる。

50

【0105】

評価値算出部53は、J個の教師データT_iとJ個の特徴量f_{1[Ti]}とに基づき、特徴量抽出式f₁の評価値を算出する。

【0106】

なお、情報推定式によって推定しようとする情報I_iが2値情報である場合には、評価値として、J個の教師データT_iとJ個の特徴量f_{1[Ti]}とのFDR(fisher discriminant ratio)を算出する。

$$FDR = ((X \text{ の平均}) - (Y \text{ の平均}))^2 / ((X \text{ の標準偏差}) - (Y \text{ の標準偏差}))$$

ただし、Xは教師データT_i、Yは特徴量f_{1[Ti]}を示すものとする。

【0107】

また、情報推定式によって推定しようとする情報I_iが数値情報である場合には、評価値として、J個の教師データT_iとJ個の特徴量f_{1[Ti]}とのPearsonの相関係数rを算出する。

$$r = (X \text{ と } Y \text{ の共分散}) / ((X \text{ の標準偏差}) \times (Y \text{ の標準偏差}))$$

ただし、Xは教師データT_i、Yは特徴量f_{1[Ti]}を示すものとする。

【0108】

同様に、評価値算出部53は、特徴量抽出式f₂乃至f_mの評価値も算出する。ここで算出された特徴量抽出式f₁乃至f_mの評価値は、特徴量抽出式リスト生成部11に供給されて、次世代の特徴量抽出式リストの生成に利用される。

【0109】

さらに、評価値算出部53は、所定の終了条件が満たされた数世代後(最終世代とする)の特徴量抽出式リストを構成するm本の特徴量抽出式f₁乃至f_mのうち、評価値が最も良い特徴量抽出式f_iの計算結果であるJ個の特徴量f_{1[Ti]}とJ個の教師データT_iとを用いた機械学習により、評価値が最も良い特徴量抽出式f_iに対応する情報推定式を決定する。

【0110】

なお、情報推定式は、その出力が2値情報である場合には線形判別などの機械学習によって判別式が決定される。また、その出力が数値情報である場合には線形回帰などの機械学習によって線形結合式が決定される。

【0111】

またさらに、評価値算出部53は、評価値が最も良い特徴量抽出式f_iとそれに対応する情報推定式から1つの弱情報抽出部F(X)_tを構築して合成部54に供給する。

【0112】

合成部54は、評価値算出部53から供給されている1個以上の弱情報抽出部F(X)_tから次式に従って高精度情報抽出装置1(図1)を構築する。

$$(F(X)_t) / t$$

【0113】

すなわち、合成部54では、複数の弱情報抽出部F(X)_tの出力の平均値を出力とする高精度情報抽出装置1が構築されることになる。

【0114】

次に、高精度情報抽出装置構築システム50による動作について、図15のフローチャートを参照して説明する。

【0115】

ステップS61において、教師データ供給部51のランダムサンプリング部52は、予め用意されているL個の教師データT_iの中から、ランダムにJ個の教師データT_iを選択して教師データ群を生成し、第1世代の教師データ群として特徴量計算部12および評価値算出部53に供給する。

【0116】

ステップS62において、第1世代の教師データ群に対応する弱情報抽出部F(X)₁が生成される。ステップS62の処理について、図16のフローチャートを参照して詳述

10

20

30

40

50

する。

【0117】

ステップS71において、特徴量抽出式リスト生成部11は、m本の特徴量抽出式をランダムに生成し、生成したm本の特徴量抽出式からなる第1世代の特徴量抽出式リストを特徴量計算部12に供給する。

【0118】

ステップS72において、特徴量計算部12は、特徴量抽出式リスト生成部11から供給された特徴量抽出式リストを構成するm本の特徴量抽出式f1乃至fmを順に1本ずつ注目する特徴量抽出式リストループを開始する。

【0119】

ステップS73において、特徴量計算部12は、注目している特徴量抽出式fjに、現世代の教師データ群を構成するJ個の教師データTiを代入してJ個の特徴量fj[Ti]を計算し、評価値算出部53に出力する。ステップS74において、評価値算出部53は、J個の教師データTiとJ個の特徴量fj[Ti]とに基づき、特徴量抽出式fjに対応する評価値を算出する。ステップS73乃至S74の処理により、注目している特徴量抽出式fjの価値が得られたことになる。

【0120】

ステップS76において、特徴量計算部12は、特徴量抽出式リスト生成部11から供給された特徴量抽出式リストを構成するm本の特徴量抽出式f1乃至fmのうち、注目していないものが残っているか否かを判定する。そして、注目していない特徴量抽出式が残っていると判定された場合、処理はステップS72に戻り、ステップS72乃至S76の処理が繰り返される。

【0121】

そして、ステップS76において、特徴量抽出式リスト生成部11から供給された特徴量抽出式リストを構成するm本の特徴量抽出式f1乃至fmのうち、注目していないものが残っていないと判定された場合、m本の特徴量抽出式f1乃至fmにそれぞれ対応する評価値が得られたことになるので、処理はステップS76に進められる。

【0122】

ステップS76において、評価値算出部53は、所定の終了条件を満たしているか否かを判定する。ここで、所定の終了条件としては、例えば、m本の特徴量抽出式f1乃至fmにそれぞれ対応する評価値のうち、最も良い評価値の値が所定の数世代の間、向上していないことなどとすることができます。

【0123】

ステップS76において、所定の終了条件を満たしていないと判定された場合、処理はステップS77に進められる。

【0124】

ステップS77において、評価値算出部53は、現世代の特徴量抽出式リストを構成するm本の特徴量抽出式f1乃至fmにそれぞれ対応する評価値を特徴量抽出式リスト生成部11に供給する。特徴量抽出式リスト生成部11は、現世代の特徴量抽出式リストを遺伝的アルゴリズムに従って更新することにより、次世代の特徴量抽出式リストを生成して特徴量計算部12に供給する。

【0125】

以上のように、生成された次世代の特徴量抽出式リストが特徴量計算部12に供給された後、処理はステップS72に戻り、ステップS72乃至S77の処理が繰り返される。そして、ステップS76において、所定の終了条件を満たしていると判定された場合、処理はステップS78に進められる。

【0126】

ステップS78において、評価値算出部53は、現世代、すなわち、最終世代の特徴量抽出式リストを構成するm本の特徴量抽出式f1乃至fmのうち、評価値が最も良い特徴量抽出式fiの計算結果であるJ個の特徴量f1[Ti]とJ個の教師データTiとを用い

10

20

30

40

50

た機械学習により、評価値が最も良い特徴量抽出式 f_i に対応する情報推定式を決定する。また、評価値算出部 53 は、評価値が最も良い特徴量抽出式 f_i とそれに対応する情報推定式から 1 つ目の弱情報抽出部 $F(X)_1$ を構築して合成部 54 に供給する。

【0127】

以上のように、第 1 世代の教師データ群に対応する 1 つ目の弱情報抽出部 $F(X)_1$ が構築され後、処理は図 15 のステップ S63 に戻される。

【0128】

ステップ S63 において、合成部 54 は、現時点までに評価値算出部 53 から供給されている 1 個以上の弱情報抽出部 $F(X)_t$ を合成する（出力の平均を演算する）ことにより、高精度情報抽出装置 1 を構築する。

10

【0129】

ステップ S64 において、合成部 54 は、所定の終了条件を満たしているか否かを判定する。ここで、所定の終了条件としては、例えば、所定の数 T だけ弱情報抽出部 $F(X)_t$ が評価値算出部 53 から供給されていること、高精度情報抽出装置 1 が所望の精度に達していること、またはユーザから終了が指示されたことのうち、少なくとも 1 つが満たされていることなどとすることができます。

【0130】

ステップ S64 において、所定の終了条件を満たしていないと判定された場合、処理はステップ S61 に戻される。そして、ステップ S61 乃至 64 の処理が繰り返される。

20

【0131】

この繰り返し毎、順次、次世代以降の教師データ群が生成されて、それに対応する弱情報抽出部 $F(X)_t$ が合成部 54 に供給され、合成部 54 により、徐々に制度が高められた高精度情報抽出装置 1 が構築されることになる。

【0132】

そして、ステップ S64 において、所定の終了条件を満たしていると判定された場合、処理はステップ S65 に進められる。ステップ S65 において、合成部 54 は、直前のステップ S63 の処理で構築した高精度情報抽出装置 1 を、最終的な高精度情報抽出装置 1 として出力する。

【0133】

以上で、本開示の第 2 の実施の形態である高精度情報抽出装置構築システム 50 の動作説明を終了する。

30

【0134】

以上説明したように、本開示を適用した高精度情報抽出装置構築システム 10 および 50 によれば、アンサンブル学習における弱情報抽出部を、遺伝的アルゴリズムを用いて生成することができる。

【0135】

また、本開示を適用した高精度情報抽出装置構築システム 10 および 50 によれば、任意の種類の入力データ X から高い精度で情報を抽出することができる高精度情報抽出装置を構築することができる。

40

【0136】

ところで、上述した一連の処理は、ハードウェアにより実行することもできるし、ソフトウェアにより実行することもできる。一連の処理をソフトウェアにより実行する場合には、そのソフトウェアを構成するプログラムが、専用のハードウェアに組み込まれているコンピュータ、または、各種のプログラムをインストールすることで、各種の機能を実行することが可能な、例えば汎用のパーソナルコンピュータなどに、プログラム記録媒体からインストールされる。

【0137】

図 13 は、上述した一連の処理をプログラムにより実行するコンピュータのハードウェアの構成例を示すプロック図である。

【0138】

50

このコンピュータ 100において、CPU (Central Processing Unit) 101, ROM (Read Only Memory) 102, RAM (Random Access Memory) 103は、バス 104により相互に接続されている。

【0139】

バス 104には、さらに、入出力インターフェース 105が接続されている。入出力インターフェース 105には、キーボード、マウス、マイクロホンなどよりなる入力部 106、ディスプレイ、スピーカなどよりなる出力部 107、ハードディスクや不揮発性のメモリなどよりなる記憶部 108、ネットワークインターフェースなどよりなる通信部 109、磁気ディスク、光ディスク、光磁気ディスク、或いは半導体メモリなどの着脱可能な記録媒体 111を駆動するドライブ 110が接続されている。

10

【0140】

以上のように構成されるコンピュータでは、CPU 101が、例えば、記憶部 108に記憶されているプログラムを、入出力インターフェース 105およびバス 104を介して、RAM 103にロードして実行することにより、上述した一連の処理が行われる。

【0141】

なお、コンピュータが実行するプログラムは、本明細書で説明する順序に沿って時系列に処理が行われるプログラムであっても良いし、並列に、あるいは呼び出しが行われたとき等の必要なタイミングで処理が行われるプログラムであっても良い。

【0142】

また、プログラムは、1台のコンピュータにより処理されるものであってもよいし、複数のコンピュータによって分散処理されるものであってもよい。さらに、プログラムは、遠方のコンピュータに転送されて実行されるものであってもよい。

20

【0143】

また、本明細書において、システムとは、複数の装置により構成される装置全体を表すものである。

【0144】

なお、本開示の実施の形態は、上述した実施の形態に限定されるものではなく、本開示の要旨を逸脱しない範囲において種々の変更が可能である。

【符号の説明】

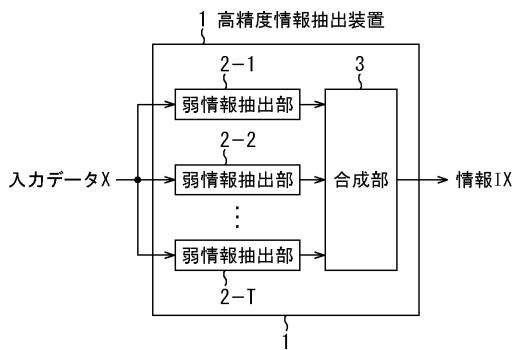
【0145】

10 高精度情報抽出装置構築システム, 11 特徴量抽出式リスト生成部, 12
特徴量計算部, 13 教師データ供給部, 14 重み設定部, 15 評価値算出部,
16 合成部, 50 高精度情報抽出装置構築システム, 51 教師データ供給部,
52 ランダムサンプリング部, 53 評価値算出部, 54 合成部, 1
01 CPU, 111 記録媒体

30

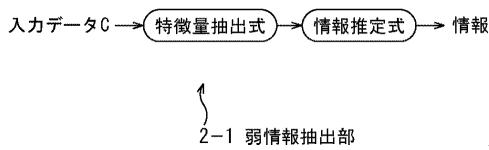
【図1】

図1



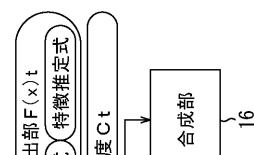
【図2】

図2



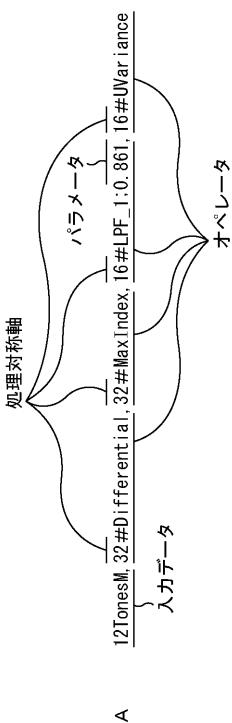
【図3】

図3



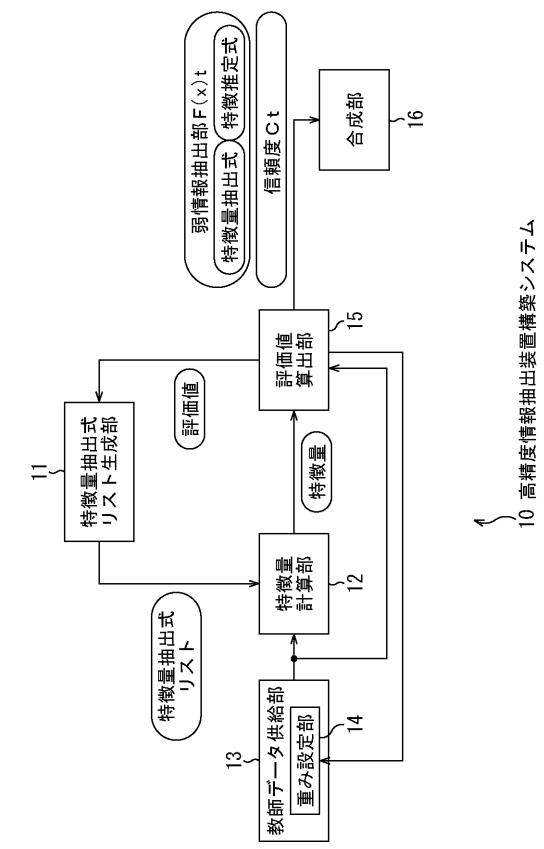
【図4】

図4



【図5】

図5



10 高精度情報抽出装置構築システム

【図6】

図6 特微量抽出式リスト

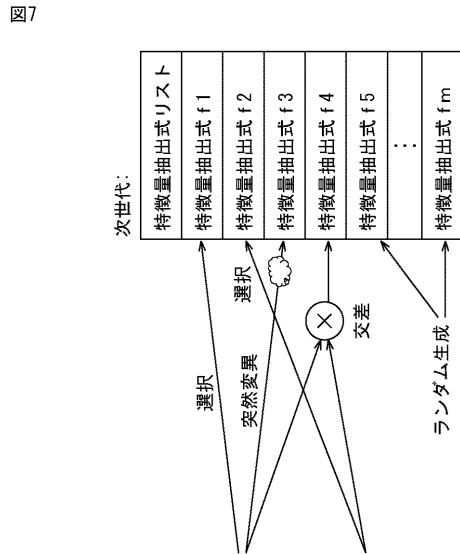
特微量抽出式 f1	WavM, 16#DownSampling_To:27. 999268, 16#NormalizeEach, 16#Skewness
特微量抽出式 f2	WavM, 16#DownSampling_To:62. 265602, 16#StdDev
特微量抽出式 f3	WavM, 16#HPF_1:0. 215608, 16#LogAxis0CH:9. 528571, 16#Order, 16#Entropy
特微量抽出式 f4	WavM, 16#Histogram:33. 312595, 16#StdDev
特微量抽出式 f5	WavM, 16#Histogram:39. 602398, 16#Skewness

【図8】

図8

教師データ T	実データ D	情報 I (明暗: 2値)	情報 I (明るさ: 1~5)	重み W
教師データ T1	実データ D1	情報 I1 (明暗 = 1)	情報 I1 (明るさ) = 1	W1 = 1/L
教師データ T2	実データ D2	情報 I2 (明暗) = -1	情報 I2 (明るさ) = 5	W2 = 1/L
:	:	:	:	:
教師データ TL	実データ DL	情報 I L (明暗) = -1	情報 I L (明るさ) = -1	W L = 1/L

【図7】



【図9】

図9

現世代:

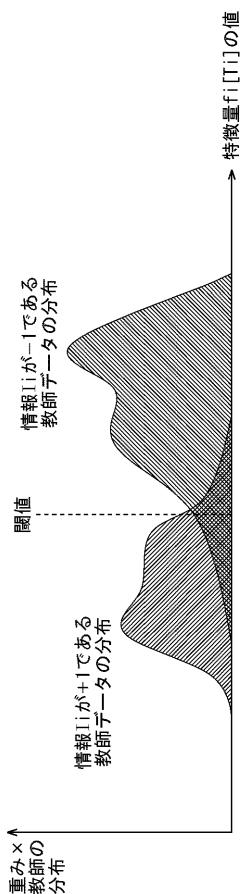
	特微量抽出式リスト	評価値
特微量抽出式 f1	0.3	
特微量抽出式 f2	0.8	
特微量抽出式 f3	0.2	
特微量抽出式 f4	0.5	
特微量抽出式 f5	0.6	
⋮	⋮	
特微量抽出式 fm	0.5	

次世代:

	式 f1	式 f2	…	式 fm
教師データ1	0.142565	0.431047		0.713684
教師データ2	0.104266	0.753864		0.931711
教師データ3	0.868273	0.051845		0.720371
⋮	⋮	⋮	⋮	⋮
教師データ L	0.322262	0.172228		0.90102

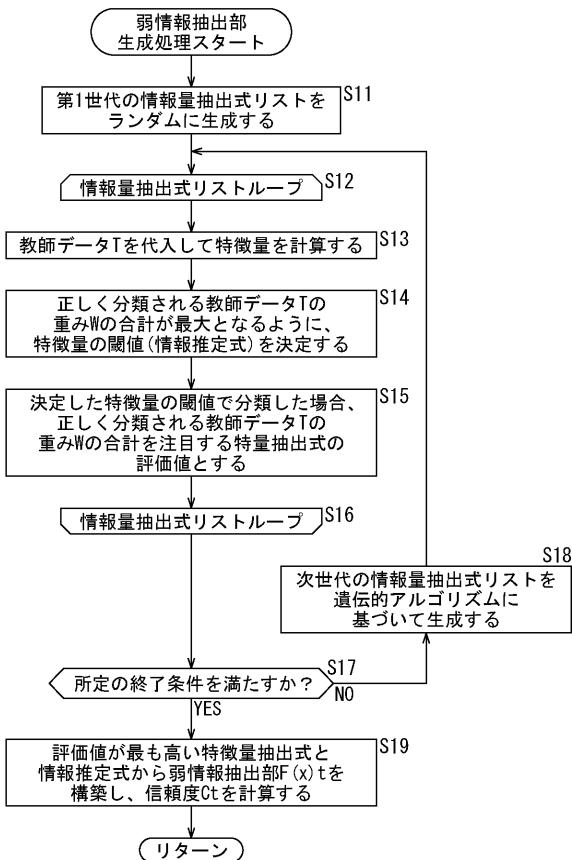
【図10】

図10



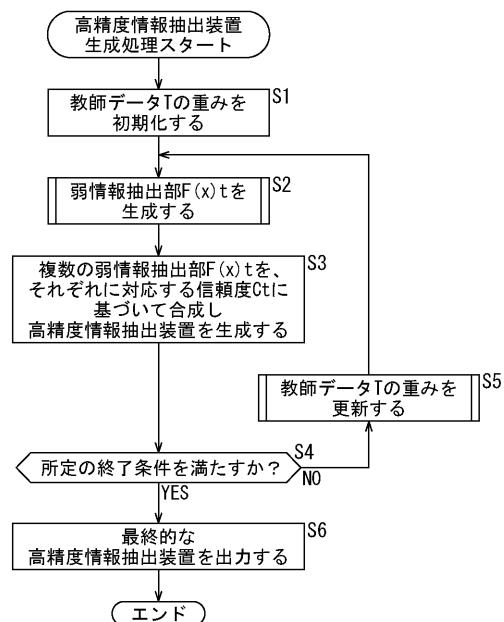
【図12】

図12



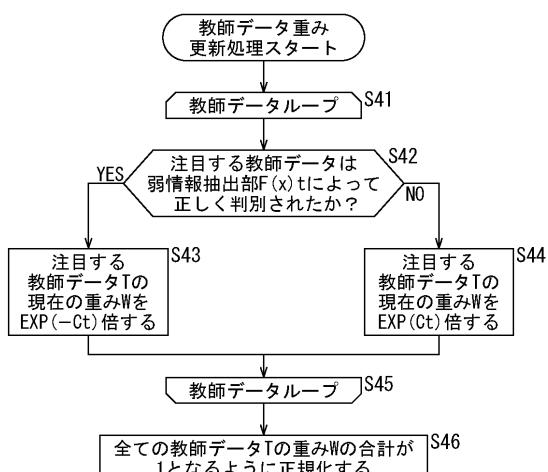
【図11】

図11



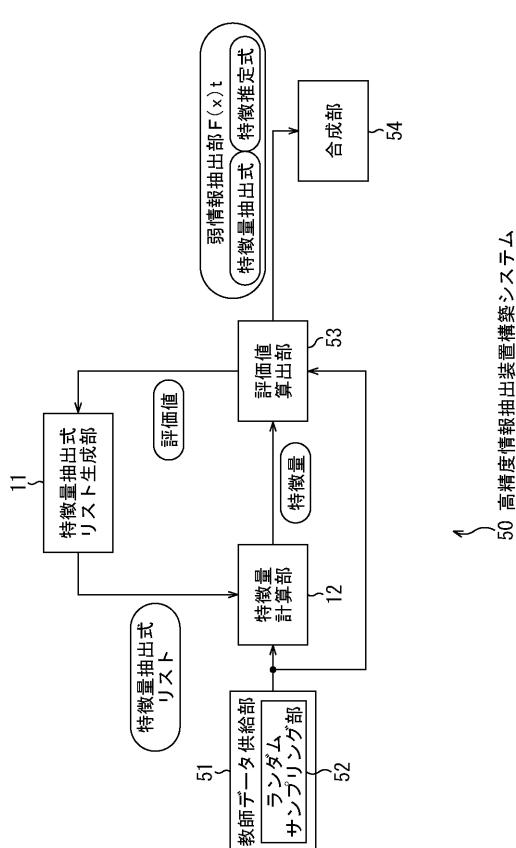
【図13】

図13



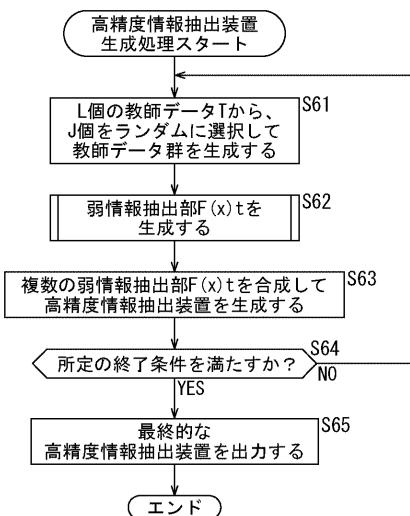
【図14】

図14



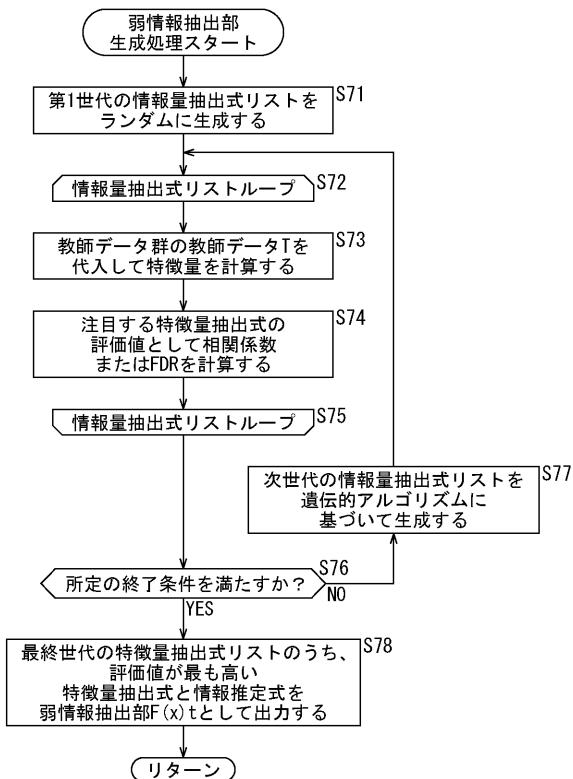
【図15】

図15



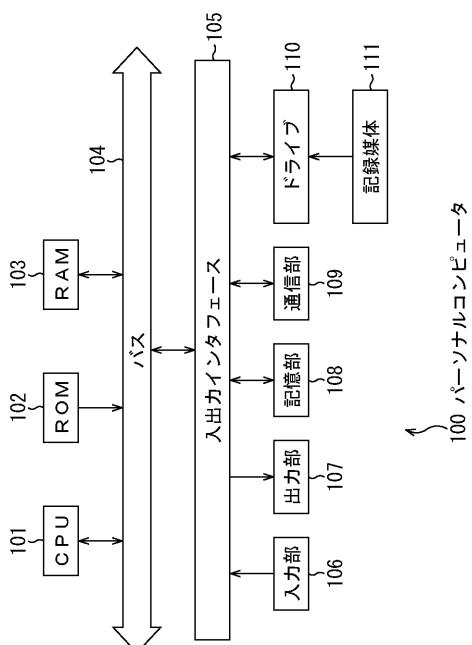
【図16】

図16



【図17】

図17



【手続補正書】

【提出日】平成25年5月22日(2013.5.22)

【手続補正1】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項1】

複数の次元軸を保有する入力データの特徴を示す情報を第一の精度で抽出する複数の第一の情報抽出部の出力を合成して、第二の精度で前記入力データの特徴を示す情報を抽出する第二の情報抽出部を構築するアンサンブル学習における前記第一の情報抽出部を生成する情報処理装置において、

前記入力データに所定の演算を行う複数のオペレータを組み合わせた特徴量抽出式を複数含む特徴量抽出式リストを、遺伝的アルゴリズムを用いて前世代の前記特徴量抽出式リストを更新することにより生成する特徴量抽出式リスト生成部と、

前記特徴量抽出式リストに含まれる各特徴量抽出式に、教師データの実データを入力して、前記実データに対応する複数の特徴量を計算する特徴量計算部と、

計算された前記複数の特徴量から、前記特徴量抽出式リストに含まれる各特徴量抽出式にそれぞれ対応する情報抽出部候補を生成するとともに、生成した前記情報抽出部候補によって正しく推定された教師データの重みを用いて各特徴量抽出式にそれぞれ対応する評価値を算出し、最終世代の前記特徴量抽出式リストに含まれる特徴量抽出式のうちで前記評価値が最も良い特徴量抽出式に対応する前記情報抽出部候補を、前記アンサンブル学習における二つの前記第一の情報抽出部に決定して、決定した前記第一の情報抽出部によって間違って推定された教師データの重みを用いて、決定した前記第一の情報抽出部の信頼度を算出する算出部と、

前記信頼度を用いて、教師データの重みを更新する更新部と
を備え、

前記オペレータは、前記入力データの保有する次元軸のうち、任意の次元軸を前記所定の演算処理の対象に指定するための処理対象軸指定情報を含む
情報処理装置。

【請求項2】

前記入力データは、時間の次元とチャンネルの次元を有する楽曲データ、X次元とY次元と画素の次元を有する画像データ、または画像データに時間の次元を加えた動画像データである

請求項1に記載の情報処理装置。

【請求項3】

前記入力データが楽曲データである場合、前記第二の情報抽出部は、前記楽曲データの特徴を示す情報として明暗または明るさを抽出する

請求項2に記載の情報処理装置。

【請求項4】

前記特徴量抽出式リスト生成部は、前記特徴量抽出式リストを構成する複数の各特徴量抽出式を遺伝子とみなし、現世代の特徴量抽出式リストを構成する複数の特徴量抽出式の評価値に応じて選択処理、交差処理、突然変異処理、およびランダム生成処理を行い、次世代の特徴量抽出式リストを生成する

請求項1に記載の情報処理装置。

【請求項5】

前記特徴量抽出式リスト生成部は、前記突然変異処理として、現世代の特徴量抽出式リストを構成する複数の特徴量抽出式のうち、評価値の高い前記特徴量抽出式を部分的に突然変異させて特徴量抽出式を生成し、次世代の特徴量抽出式リストに含める

請求項 4 に記載の情報処理装置。

【請求項 6】

複数の次元軸を保有する入力データの特徴を示す情報を第一の精度で抽出する複数の第一の情報抽出部の出力を合成して、第二の精度で前記入力データの特徴を示す情報を抽出する第二の情報抽出部を構築するアンサンブル学習における前記第一の情報抽出部を生成する情報処理装置の情報処理方法において、

前記情報処理装置による、

前記入力データに所定の演算を行う複数のオペレータを組み合わせた特微量抽出式を複数含む特微量抽出式リストに含まれる各特微量抽出式に、教師データの実データを入力して、前記実データに対応する複数の特微量を計算する特微量計算ステップと、

計算された前記複数の特微量から、前記特微量抽出式リストに含まれる各特微量抽出式にそれぞれ対応する情報抽出部候補を生成するとともに、生成した前記情報抽出部候補によって正しく推定された教師データの重みを用いて各特微量抽出式にそれぞれ対応する評価値を算出する第一の算出ステップと、

算出された前記評価値に基づいた遺伝的アルゴリズムを用いて前世代の前記特微量抽出式リストを更新する更新ステップと、

最終世代の前記特微量抽出式リストに含まれる特微量抽出式のうちで前記評価値が最も良い特微量抽出式に対応する前記情報抽出部候補を、前記アンサンブル学習における一つの前記第一の情報抽出部に決定して、決定した前記第一の情報抽出部によって間違って推定された教師データの重みを用いて、決定した前記第一の情報抽出部の信頼度を算出する第二の算出ステップと、

前記信頼度を用いて、教師データの重みを更新する更新ステップと
を含み、

前記オペレータは、前記入力データの保有する次元軸のうち、任意の次元軸を前記所定の演算処理の対象に指定するための処理対象軸指定情報を含む
情報処理方法。

【請求項 7】

複数の次元軸を保有する入力データの特徴を示す情報を第一の精度で抽出する複数の第一の情報抽出部の出力を合成して、第二の精度で前記入力データの特徴を示す情報を抽出する第二の情報抽出部を構築するアンサンブル学習における前記第一の情報抽出部を生成するコンピュータを、

前記入力データに所定の演算を行う複数のオペレータを組み合わせた特微量抽出式を複数含む特微量抽出式リストを、遺伝的アルゴリズムを用いて前世代の前記特微量抽出式リストを更新することにより生成する特微量抽出式リスト生成部と、

前記特微量抽出式リストに含まれる各特微量抽出式に、教師データの実データを入力して、前記実データに対応する複数の特微量を計算する特微量計算部と、

計算された前記複数の特微量から、前記特微量抽出式リストに含まれる各特微量抽出式にそれぞれ対応する情報抽出部候補を生成するとともに、生成した前記情報抽出部候補によって正しく推定された教師データの重みを用いて各特微量抽出式にそれぞれ対応する評価値を算出し、最終世代の前記特微量抽出式リストに含まれる特微量抽出式のうちで前記評価値が最も良い特微量抽出式に対応する前記情報抽出部候補を、前記アンサンブル学習における一つの前記第一の情報抽出部に決定して、決定した前記第一の情報抽出部によって間違って推定された教師データの重みを用いて、決定した前記第一の情報抽出部の信頼度を算出する算出部と、

前記信頼度を用いて、教師データの重みを更新する更新部と
して機能させ、

前記オペレータは、前記入力データの保有する次元軸のうち、任意の次元軸を前記所定の演算処理の対象に指定するための処理対象軸指定情報を含む
プログラム。

【請求項 8】

複数の次元軸を保有する入力データの特徴を示す情報を第一の精度で抽出する複数の第一の情報抽出部の出力を合成して、第二の精度で前記入力データの特徴を示す情報を抽出する第二の精度情報抽出部を構築するアンサンブル学習における前記第一の情報抽出部を生成する情報処理装置において、

教師データをランダムに選択する選択部と、

前記入力データに所定の演算を行う複数のオペレータを組み合わせた特徴量抽出式を複数含む特徴量抽出式リストを、遺伝的アルゴリズムを用いて前世代の前記特徴量抽出式リストを更新することにより生成する特徴量抽出式リスト生成部と、

前記特徴量抽出式リストに含まれる各特徴量抽出式に、前記教師データの実データを入力して、前記実データに対応する複数の特徴量を計算する特徴量計算部と、

各特徴量抽出式にそれぞれ対応する評価値として、各特徴量抽出式を用いて計算された前記教師データの前記実データに対応する前記複数の特徴量を用いて、前記教師データの前記実データに対応する前記情報を推定した場合の精度を算出し、最終世代の前記特徴量抽出式リストに含まれる特徴量抽出式のうちで前記評価値が最も良い特徴量抽出式を用いて計算された前記教師データの前記実データに対応する前記複数の特徴量から、前記教師データの前記実データに対応する前記情報を推定する機械学習により、前記アンサンブル学習における二つの前記第一の情報抽出部を生成する評価値算出部と

を含む情報処理装置。

【請求項 9】

複数の前記第一の情報抽出部を合成することにより、前記第二の情報抽出部を構築する合成部を

さらに含む請求項 8 に記載の情報処理装置。

【請求項 10】

複数の次元軸を保有する入力データの特徴を示す情報を第一の精度で抽出する複数の第一の情報抽出部の出力を合成して、第二の精度で前記入力データの特徴を示す情報を抽出する第二の精度情報抽出部を構築するアンサンブル学習における前記第一の情報抽出部を生成する情報処理装置の情報処理方法において、

前記情報処理装置による、

教師データをランダムに選択し、

前記入力データに所定の演算を行う複数のオペレータを組み合わせた特徴量抽出式を複数含む第1世代の特徴量抽出式リストをランダムに生成し、

前記特徴量抽出式リストに含まれる各特徴量抽出式に、前記教師データの実データを入力して、前記実データに対応する複数の特徴量を計算し、

各特徴量抽出式にそれぞれ対応する評価値として、各特徴量抽出式を用いて計算された前記教師データの前記実データに対応する前記複数の特徴量を用いて、前記教師データの前記実データに対応する前記情報を推定した場合の精度を算出し、

前記特徴量抽出式の前記評価値に基づいた遺伝的アルゴリズムを用いて前世代の前記特徴量抽出式リストを更新し、

最終世代の前記特徴量抽出式リストに含まれる特徴量抽出式のうちで前記評価値が最も良い特徴量抽出式を用いて計算された前記教師データの前記実データに対応する前記複数の特徴量から、前記教師データの前記実データに対応する前記情報を推定する機械学習により、前記アンサンブル学習における二つの前記第一の情報抽出部を生成する

ステップを含む情報処理方法。

【請求項 11】

複数の次元軸を保有する入力データの特徴を示す情報を第一の精度で抽出する複数の第一の情報抽出部の出力を合成して、第二の精度で前記入力データの特徴を示す情報を抽出する第二の精度情報抽出部を構築するアンサンブル学習における前記第一の情報抽出部を生成する情報処理装置の制御用のプログラムであって、

教師データをランダムに選択し、

前記入力データに所定の演算を行う複数のオペレータを組み合わせた特徴量抽出式を

複数含む第1世代の特徴量抽出式リストをランダムに生成し、

前記特徴量抽出式リストに含まれる各特徴量抽出式に、前記教師データの実データを入力して、前記実データに対応する複数の特徴量を計算し、

各特徴量抽出式にそれぞれ対応する評価値として、各特徴量抽出式を用いて計算された前記教師データの前記実データに対応する前記複数の特徴量を用いて、前記教師データの前記実データに対応する前記情報を推定した場合の精度を算出し、

前記特徴量抽出式の前記評価値に基づいた遺伝的アルゴリズムを用いて前世代の前記特徴量抽出式リストを更新し、

最終世代の前記特徴量抽出式リストに含まれる特徴量抽出式のうちで前記評価値が最も良い特徴量抽出式を用いて計算された前記教師データの前記実データに対応する前記複数の特徴量から、前記教師データの前記実データに対応する前記情報を推定する機械学習により、前記アンサンブル学習における二つの前記第一の情報抽出部を生成する

ステップを含む処理を情報処理装置のコンピュータに実行させるプログラム。

【手続補正2】

【補正対象書類名】明細書

【補正対象項目名】0012

【補正方法】変更

【補正の内容】

【0012】

本開示の第一の側面である情報処理装置は、複数の次元軸を保有する入力データの特徴を示す情報を第一の精度で抽出する複数の第一の情報抽出部の出力を合成して、第二の精度で前記入力データの特徴を示す情報を抽出する第二の情報抽出部を構築するアンサンブル学習における前記第一の情報抽出部を生成する情報処理装置において、前記入力データに所定の演算を行う複数のオペレータを任意に組み合わせた特徴量抽出式を複数含む特徴量抽出式リストを、遺伝的アルゴリズムを用いて前世代の前記特徴量抽出式リストを更新することにより生成する特徴量抽出式リスト生成部と、前記特徴量抽出式リストに含まれる各特徴量抽出式に、教師データの実データを入力して、前記実データに対応する複数の特徴量を計算する特徴量計算部と、計算された前記複数の特徴量から、前記特徴量抽出式リストに含まれる各特徴量抽出式にそれぞれ対応する情報抽出部候補を生成するとともに、生成した前記情報抽出部候補によって正しく推定された教師データの重みを用いて各特徴量抽出式にそれぞれ対応する評価値を算出し、最終世代の前記特徴量抽出式リストに含まれる特徴量抽出式のうちで前記評価値が最も良い特徴量抽出式に対応する前記情報抽出部候補を、前記アンサンブル学習における二つの前記第一の情報抽出部に決定して、決定した前記第一の情報抽出部によって間違って推定された教師データの重みを用いて、決定した前記第一の情報抽出部の信頼度を算出する算出部と、前記信頼度を用いて、教師データの重みを更新する更新部とを備え、前記オペレータは、前記入力データの保有する次元軸のうち、任意の次元軸を前記所定の演算処理の対象に指定するための処理対象軸指定情報を含む。

【手続補正3】

【補正対象書類名】明細書

【補正対象項目名】0013

【補正方法】変更

【補正の内容】

【0013】

前記入力データは、時間の次元とチャンネルの次元を有する楽曲データ、X次元とY次元と画素の次元を有する画像データ、または画像データに時間の次元を加えた動画像データとすることができます。

前記入力データが楽曲データである場合、前記第二の情報抽出部は、前記楽曲データの特徴を示す情報として明暗または明るさを抽出することができる。

前記特徴量抽出式リスト生成部は、前記特徴量抽出式リストを構成する複数の各特徴量

抽出式を遺伝子とみなし、現世代の特徴量抽出式リストを構成する複数の特徴量抽出式の評価値に応じて選択処理、交差処理、突然変異処理、およびランダム生成処理を行い、次世代の特徴量抽出式リストを生成することができる。

前記特徴量抽出式リスト生成部は、前記突然変異処理として、現世代の特徴量抽出式リストを構成する複数の特徴量抽出式のうち、評価値の高い前記特徴量抽出式を部分的に突然変異させて特徴量抽出式を生成し、次世代の特徴量抽出式リストに含めることができる

。

【手続補正4】

【補正対象書類名】明細書

【補正対象項目名】0014

【補正方法】変更

【補正の内容】

【0014】

本開示の第一の側面である情報処理方法は、複数の次元軸を保有する入力データの特徴を示す情報を第一の精度で抽出する複数の第一の情報抽出部の出力を合成して、第二の精度で前記入力データの特徴を示す情報を抽出する第二の情報抽出部を構築するアンサンブル学習における前記第一の情報抽出部を生成する情報処理装置の情報処理方法において、前記情報処理装置による、前記入力データに所定の演算を行う複数のオペレータを任意に組み合わせた特徴量抽出式を複数含む特徴量抽出式リストに含まれる各特徴量抽出式に、教師データの実データを入力して、前記実データに対応する複数の特徴量を計算する特徴量計算ステップと、計算された前記複数の特徴量から、前記特徴量抽出式リストに含まれる各特徴量抽出式にそれぞれ対応する情報抽出部候補を生成するとともに、生成した前記情報抽出部候補によって正しく推定された教師データの重みを用いて各特徴量抽出式にそれぞれ対応する評価値を算出する第一の算出ステップと、算出された前記評価値に基づいた遺伝的アルゴリズムを用いて前世代の前記特徴量抽出式リストを更新する更新ステップと、最終世代の前記特徴量抽出式リストに含まれる特徴量抽出式のうちで前記評価値が最も良い特徴量抽出式に対応する前記情報抽出部候補を、前記アンサンブル学習における一つの前記第一の情報抽出部に決定して、決定した前記第一の情報抽出部によって間違って推定された教師データの重みを用いて、決定した前記第一の情報抽出部の信頼度を算出する第二の算出ステップと、前記信頼度を用いて、教師データの重みを更新する更新ステップとを含み、前記オペレータは、前記入力データの保有する次元軸のうち、任意の次元軸を前記所定の演算処理の対象に指定するための処理対象軸指定情報を含む。

【手続補正5】

【補正対象書類名】明細書

【補正対象項目名】0015

【補正方法】変更

【補正の内容】

【0015】

本開示の第一の側面であるプログラムは、複数の次元軸を保有する入力データの特徴を示す情報を第一の精度で抽出する複数の第一の情報抽出部の出力を合成して、第二の精度で前記入力データの特徴を示す情報を抽出する第二の情報抽出部を構築するアンサンブル学習における前記第一の情報抽出部を生成するコンピュータを、前記入力データに所定の演算を行う複数のオペレータを任意に組み合わせた特徴量抽出式を複数含む特徴量抽出式リストを、遺伝的アルゴリズムを用いて前世代の前記特徴量抽出式リストを更新することにより生成する特徴量抽出式リスト生成部と、前記特徴量抽出式リストに含まれる各特徴量抽出式に、教師データの実データを入力して、前記実データに対応する複数の特徴量を計算する特徴量計算部と、計算された前記複数の特徴量から、前記特徴量抽出式リストに含まれる各特徴量抽出式にそれぞれ対応する情報抽出部候補を生成するとともに、生成した前記情報抽出部候補によって正しく推定された教師データの重みを用いて各特徴量抽出式にそれぞれ対応する評価値を算出し、最終世代の前記特徴量抽出式リストに含まれる特

徴量抽出式のうちで前記評価値が最も良い特徴量抽出式に対応する前記情報抽出部候補を、前記アンサンブル学習における一つの前記第一の情報抽出部に決定して、決定した前記第一の情報抽出部によって間違って推定された教師データの重みを用いて、決定した前記第一の情報抽出部の信頼度を算出する算出部と、前記信頼度を用いて、教師データの重みを更新する更新部として機能させ、前記オペレータは、前記入力データの保有する次元軸のうち、任意の次元軸を前記所定の演算処理の対象に指定するための処理対象軸指定情報を含む。

【手続補正6】

【補正対象書類名】明細書

【補正対象項目名】0016

【補正方法】変更

【補正の内容】

【0016】

本開示の第一の側面においては、入力データに所定の演算を行う複数のオペレータを任意に組み合わせた特徴量抽出式を複数含む特徴量抽出式リストに含まれる各特徴量抽出式に、教師データの実データが入力されて、前記実データに対応する複数の特徴量が計算される。また、計算された前記複数の特徴量から、前記特徴量抽出式リストに含まれる各特徴量抽出式にそれぞれ対応する情報抽出部候補が生成されるとともに、生成した前記情報抽出部候補によって正しく推定された教師データの重みを用いて各特徴量抽出式にそれぞれ対応する評価値が算出される。さらに、算出された前記評価値に基づいた遺伝的アルゴリズムを用いて前世代の前記特徴量抽出式リストが更新される。そして、最終世代の前記特徴量抽出式リストに含まれる特徴量抽出式のうちで前記評価値が最も良い特徴量抽出式に対応する前記情報抽出部候補が、アンサンブル学習における一つの前記第一の情報抽出部に決定され、決定された前記第一の情報抽出部によって間違って推定された教師データの重みを用いて、決定された前記第一の情報抽出部の信頼度が算出され、前記信頼度を用いて、教師データの重みが更新される。

【手続補正7】

【補正対象書類名】明細書

【補正対象項目名】0017

【補正方法】変更

【補正の内容】

【0017】

本開示の第二の側面である情報処理装置は、複数の次元軸を保有する入力データの特徴を示す情報を第一の精度で抽出する複数の第一の情報抽出部の出力を合成して、第二の精度で前記入力データの特徴を示す情報を抽出する第二の精度情報抽出部を構築するアンサンブル学習における前記第一の情報抽出部を生成する情報処理装置において、教師データをランダムに選択する選択部と、前記入力データに所定の演算を行う複数のオペレータを組み合わせた特徴量抽出式を複数含む特徴量抽出式リストを、遺伝的アルゴリズムを用いて前世代の前記特徴量抽出式リストを更新することにより生成する特徴量抽出式リスト生成部と、前記特徴量抽出式リストに含まれる各特徴量抽出式に、前記教師データの実データを入力して、前記実データに対応する複数の特徴量を計算する特徴量計算部と、各特徴量抽出式にそれぞれ対応する評価値として、各特徴量抽出式を用いて計算された前記教師データの前記実データに対応する前記複数の特徴量を用いて、前記教師データの前記実データに対応する前記情報を推定した場合の精度を算出し、最終世代の前記特徴量抽出式リストに含まれる特徴量抽出式のうちで前記評価値が最も良い特徴量抽出式を用いて計算された前記教師データの前記実データに対応する前記複数の特徴量から、前記教師データの前記実データに対応する前記情報を推定する機械学習により、前記アンサンブル学習における一つの前記第一の情報抽出部を生成する評価値算出部とを含む。

【手続補正8】

【補正対象書類名】明細書

【補正対象項目名】0018

【補正方法】変更

【補正の内容】

【0018】

本開示の第二の側面である情報処理装置は、複数の前記第一の情報抽出部を合成することにより、前記第二の情報抽出部を構築する合成部をさらに含むことができる。

【手続補正9】

【補正対象書類名】明細書

【補正対象項目名】0019

【補正方法】変更

【補正の内容】

【0019】

本開示の第二の側面である情報処理方法は、複数の次元軸を保有する入力データの特徴を示す情報を第一の精度で抽出する複数の第一の情報抽出部の出力を合成して、第二の精度で前記入力データの特徴を示す情報を抽出する第二の精度情報抽出部を構築するアンサンブル学習における前記第一の情報抽出部を生成する情報処理装置の情報処理方法において、前記情報処理装置による、教師データをランダムに選択し、前記入力データに所定の演算を行う複数のオペレータを組み合わせた特徴量抽出式を複数含む第1世代の特徴量抽出式リストをランダムに生成し、前記特徴量抽出式リストに含まれる各特徴量抽出式に、前記教師データの実データを入力して、前記実データに対応する複数の特徴量を計算し、各特徴量抽出式にそれぞれ対応する評価値として、各特徴量抽出式を用いて計算された前記教師データの前記実データに対する前記複数の特徴量を用いて、前記教師データの前記実データに対する前記情報を推定した場合の精度を算出し、前記特徴量抽出式の前記評価値に基づいた遺伝的アルゴリズムを用いて前世代の前記特徴量抽出式リストを更新し、最終世代の前記特徴量抽出式リストに含まれる特徴量抽出式のうちで前記評価値が最も良い特徴量抽出式を用いて計算された前記教師データの前記実データに対する前記複数の特徴量から、前記教師データの前記実データに対する前記情報を推定する機械学習により、前記アンサンブル学習における二つの前記第一の情報抽出部を生成するステップを含む。

【手続補正10】

【補正対象書類名】明細書

【補正対象項目名】0020

【補正方法】変更

【補正の内容】

【0020】

本開示の第二の側面であるプログラムは、複数の次元軸を保有する入力データの特徴を示す情報を第一の精度で抽出する複数の第一の情報抽出部の出力を合成して、第二の精度で前記入力データの特徴を示す情報を抽出する第二の精度情報抽出部を構築するアンサンブル学習における前記第一の情報抽出部を生成する情報処理装置の制御用のプログラムであって、教師データをランダムに選択し、前記入力データに所定の演算を行う複数のオペレータを組み合わせた特徴量抽出式を複数含む第1世代の特徴量抽出式リストをランダムに生成し、前記特徴量抽出式リストに含まれる各特徴量抽出式に、前記教師データの実データを入力して、前記実データに対応する複数の特徴量を計算し、各特徴量抽出式にそれぞれ対応する評価値として、各特徴量抽出式を用いて計算された前記教師データの前記実データに対する前記複数の特徴量を用いて、前記教師データの前記実データに対する前記情報を推定した場合の精度を算出し、前記特徴量抽出式の前記評価値に基づいた遺伝的アルゴリズムを用いて前世代の前記特徴量抽出式リストを更新し、最終世代の前記特徴量抽出式リストに含まれる特徴量抽出式のうちで前記評価値が最も良い特徴量抽出式を用いて計算された前記教師データの前記実データに対する前記複数の特徴量から、前記教師データの前記実データに対する前記情報を推定する機械学習により、前記アンサンブル

ル学習における二つの前記第一の情報抽出部を生成するステップを含む処理を情報処理装置のコンピュータに実行させる。

【手続補正11】

【補正対象書類名】明細書

【補正対象項目名】0021

【補正方法】変更

【補正の内容】

【0021】

本開示の第二の側面においては、教師データがランダムに選択され、入力データに所定の演算を行う複数のオペレータを組み合わせた特徴量抽出式を複数含む第1世代の特徴量抽出式リストがランダムに生成され、前記特徴量抽出式リストに含まれる各特徴量抽出式に、前記教師データの実データが入力されて、前記実データに対応する複数の特徴量が計算される。また、各特徴量抽出式にそれぞれ対応する評価値として、各特徴量抽出式を用いて計算された前記教師データの前記実データに対応する前記複数の特徴量を用いて、前記教師データの前記実データに対応する前記情報を推定した場合の精度が算出される。さらに、前記特徴量抽出式の前記評価値に基づいた遺伝的アルゴリズムを用いて前世代の前記特徴量抽出式リストが更新される。さらに、最終世代の前記特徴量抽出式リストに含まれる特徴量抽出式のうちで前記評価値が最も良い特徴量抽出式を用いて計算された前記教師データの前記実データに対応する前記複数の特徴量から、前記教師データの前記実データに対応する前記情報を推定する機械学習により、前記アンサンブル学習における二つの前記第一の情報抽出部が生成される。