

【公報種別】特許法第17条の2の規定による補正の掲載
 【部門区分】第6部門第3区分
 【発行日】平成24年9月13日(2012.9.13)

【公開番号】特開2011-34410(P2011-34410A)
 【公開日】平成23年2月17日(2011.2.17)
 【年通号数】公開・登録公報2011-007
 【出願番号】特願2009-181016(P2009-181016)
 【国際特許分類】

G 0 6 F 17/30 (2006.01)

【 F I 】

G 0 6 F 17/30 2 1 0 D

【手続補正書】

【提出日】平成24年7月27日(2012.7.27)

【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】

サンプル群をサンプルそれぞれの特徴量に基づいて複数のクラスタに分割するクラスタリング処理方法であって、

前記複数のクラスタはそれぞれ、M層($M = 2, \dots, K$)からなる複数の階層のいずれかに属し、

前記クラスタリング処理方法は、

処理対象である前記サンプルの特徴量を第1の階層に属するクラスタそれぞれの代表特徴量と比較した結果に基づいて、当該処理対象であるサンプルを前記第1の階層に属するクラスタに帰属させるサンプル帰属工程と、

第M-1の階層に属するクラスタを、第Mの階層に帰属させるか否かを判定する判定工程と、

前記判定工程で前記第M-1の階層に属するクラスタを前記第Mの階層に帰属させると判定した場合、前記第M-1の階層に属するクラスタを、前記第Mの階層に帰属させるクラスタ帰属工程と

を有し、

前記サンプルを順番に処理対象として、全ての前記サンプルが最も上位の階層である第Kの階層に属するクラスタのいずれかに帰属するまで、前記サンプル帰属工程と前記判定工程と前記クラスタ帰属工程とを繰り返し実行することを特徴とするクラスタリング処理方法。

【請求項2】

前記サンプル帰属工程は、

当該処理対象のサンプルの特徴量との距離が最も近い代表特徴量を有するクラスタを、前記第1の階層に属する複数のクラスタの中から探索する第1層クラスタ探索工程と、

前記第1層クラスタ探索工程にて探索したクラスタと前記処理対象のサンプルとの前記距離が第1の距離の閾値以下の場合、当該クラスタへ当該サンプルを帰属させ、前記距離が前記第1の距離の閾値より大きい場合は前記第1の階層に新たに定義するクラスタへ当該サンプルを帰属させる第1層クラスタ帰属工程と

を有することを特徴とする請求項1に記載のクラスタリング処理方法。

【請求項3】

前記クラスタ帰属工程は、

前記判定工程で前記第 M - 1 の階層に属するクラスタを前記第 M の階層に帰属させると判定した場合、前記第 M - 1 の階層に属する処理対象のクラスタの代表特徴量との距離が最も近い代表特徴量を有するクラスタを、前記第 M の階層に属する複数のクラスタの中から探索する第 M 層クラスタ探索工程と、

前記第 M 層クラスタ探索工程にて探索したクラスタと前記処理対象のクラスタとの前記距離が第 M の距離の閾値以下の場合には、探索した当該第 M の階層に属するクラスタに前記処理対象のクラスタを帰属させ、第 M の距離の閾値より大きい場合は第 M の階層に新たに定義するクラスタへ前記処理対象のクラスタを帰属させる第 M 層クラスタ帰属工程とを有することを特徴とする請求項 1 または 2 に記載のクラスタリング処理方法。

【請求項 4】

前記判定工程において、前記第 M - 1 の階層に属するクラスタが所定のクラスタ数の閾値を超えた場合に、前記第 M - 1 の階層に属するクラスタを前記第 M の階層に帰属させると判定することを特徴とする請求項 1 乃至 3 のいずれか一項に記載のクラスタリング処理方法。

【請求項 5】

前記判定工程において、所定のサンプルの数に対して前記第 1 層クラスタ帰属工程を行った際に前記第 1 の階層に属するクラスタを第 2 の階層に帰属させると判定することを特徴とする請求項 2 または 3 に記載のクラスタリング処理方法。

【請求項 6】

前記判定工程は、前記第 M - 1 の階層に属するクラスタ群に対して上位である前記第 M の階層への前記第 M 層クラスタ帰属工程を適用した際に、更に、前記第 M の階層に属するクラスタを前記第 M + 1 の階層に帰属させるか否かを判定することを特徴とする請求項 3 に記載のクラスタリング処理方法。

【請求項 7】

前記判定工程において、前記サンプルが帰属される確率が低い第 M - 1 の階層に属するクラスタを第 M の階層に帰属させると判定することを特徴とする請求項 1 乃至 3 のいずれか一項に記載のクラスタリング処理方法。

【請求項 8】

前記判定工程において、前記サンプルが帰属するクラスタとして出現頻度が低い第 M - 1 の階層に属するクラスタを第 M の階層に帰属させると判定することを特徴とする請求項 1 乃至 3 のいずれか一項に記載のクラスタリング処理方法。

【請求項 9】

前記判定工程において、新たに定義されてからの経過時間に基づいて、前記第 M - 1 の階層に属するクラスタを第 M の階層に帰属させると判定することを特徴とする請求項 1 乃至 3 のいずれか一項に記載のクラスタリング処理方法。

【請求項 10】

前記第 M の階層に属するクラスタに対する前記クラスタ数の閾値は、前記第 M - 1 の階層に属するクラスタに対する前記クラスタ数の閾値よりも大きいことを特徴とする請求項 4 乃至 9 のいずれか一項に記載のクラスタリング処理方法。

【請求項 11】

第 M - 1 の距離の閾値は、前記第 M の距離の閾値よりも小さいことを特徴とする請求項 3 乃至 10 のいずれか一項に記載のクラスタリング処理方法。

【請求項 12】

サンプル群をサンプルそれぞれの特徴量に基づいて複数のクラスタに分割するクラスタリング処理装置であって、

前記複数のクラスタはそれぞれ、M 層 (M = 2 , ... , K) からなる複数の階層のいずれかに属し、

前記クラスタリング処理装置は、

処理対象である前記サンプルの特徴量を第 1 の階層に属するクラスタそれぞれの代表特

微量と比較した結果に基づいて、当該処理対象であるサンプルを前記第 1 の階層に属するクラスタに帰属させるサンプル帰属手段と、

第 M - 1 の階層に属するクラスタを、第 M の階層に帰属させるか否かを判定する判定手段と、

前記判定手段で前記第 M - 1 の階層に属するクラスタを前記第 M の階層に帰属させると判定した場合、前記第 M - 1 の階層に属するクラスタを、前記第 M の階層に帰属させるクラスタ帰属手段と

を有し、

前記サンプルを順番に処理対象として、全ての前記サンプルが最も上位の階層である第 K の階層に属するクラスタのいずれかに帰属するまで、前記サンプル帰属手段と前記判定手段と前記クラスタ帰属手段とによる処理を繰り返し実行することを特徴とするクラスタリング処理装置。

【請求項 13】

コンピュータを、

処理対象であるサンプルの特徴量を、M 層 ($M = 2, \dots, K$) からなる複数の階層のいずれかに属するクラスタのうち、第 1 の階層に属するクラスタそれぞれの代表特徴量と比較した結果に基づいて、当該処理対象であるサンプルを前記第 1 の階層に属するクラスタに帰属させるサンプル帰属手段、

第 M - 1 の階層に属するクラスタを、第 M の階層に帰属させるか否かを判定する判定手段、

前記判定手段で前記第 M - 1 の階層に属するクラスタを前記第 M の階層に帰属させると判定した場合、前記第 M - 1 の階層に属するクラスタを、前記第 M の階層に帰属させるクラスタ帰属手段、

として機能させ、

前記サンプルを順番に処理対象として、全ての前記サンプルが最も上位の階層である第 K の階層に属するクラスタのいずれかに帰属するまで、前記サンプル帰属手段と前記判定手段と前記クラスタ帰属手段とによる処理を繰り返し実行させることを特徴とするプログラム。

【手続補正 2】

【補正対象書類名】明細書

【補正対象項目名】0009

【補正方法】変更

【補正の内容】

【0009】

上記課題を解決するために、本発明のクラスタリング処理方法は、以下の構成を有する。サンプル群をサンプルそれぞれの特徴量に基づいて複数のクラスタに分割するクラスタリング処理方法であって、前記複数のクラスタはそれぞれ、M 層 ($M = 2, \dots, K$) からなる複数の階層のいずれかに属し、前記クラスタリング処理方法は、処理対象である前記サンプルの特徴量を第 1 の階層に属するクラスタそれぞれの代表特徴量と比較した結果に基づいて、当該処理対象であるサンプルを前記第 1 の階層に属するクラスタに帰属させるサンプル帰属工程と、第 M - 1 の階層に属するクラスタを、第 M の階層に帰属させるか否かを判定する判定工程と、前記判定工程で前記第 M - 1 の階層に属するクラスタを前記第 M の階層に帰属させると判定した場合、前記第 M - 1 の階層に属するクラスタを、前記第 M の階層に帰属させるクラスタ帰属工程とを有し、前記サンプルを順番に処理対象として、全ての前記サンプルが最も上位の階層である第 K の階層に属するクラスタのいずれかに帰属するまで、前記サンプル帰属工程と前記判定工程と前記クラスタ帰属工程とを繰り返し実行することを特徴とする。