【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】令和4年1月25日(2022.1.25)

【国際公開番号】WO2020/161845
【出願番号】特願2020-570278(P2020-570278)
【国際特許分類】
G06F 16/906(2019.01)
【FI】
G06F 16/906

【手続補正書】
【提出日】令和1年12月26日(2019.12.26)
【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】
複数の要素データを分類するクラスタリング装置において、
前記複数の要素データを記憶するデータ記憶部と、
分類結果を評価する評価値を算出する評価値算出部と、
設定されるクラスタ数の複数のクラスタに前記複数の要素データを分類することで得られる前記評価値に基づき、最適なクラスタ数の前記複数のクラスタに前記複数の要素データを分類する一括処理部と、
追加される新規の要素データを、前記一括処理部によって前記複数の要素データが分類されている前記複数のクラスタのうちの前記新規の要素データに最も近いクラスタに分類する更新処理部と、
前記更新処理部が前記新規の要素データを分類することで得られる前記評価値と、当該新規の要素データを分類する前の前記評価値とを比較することによって、前記新規の要素データの分類後の分類結果の妥当性を判断する判断部と
を備えることを特徴とするクラスタリング装置。
【請求項２】
前記評価値算出部は、
クラスタ内の要素データの分散の程度を示す各々の前記クラスタについての第１指標値を当該クラスタの要素データのデータ数に基づく第１の値で規格化した値の各前記クラスタの第１総和値である内的結合度と、クラスタ間の距離の指標となる各々の前記クラスタについての第２指標値の第２総和値をクラスタ数に基づく第２の値で規格化した外的分離度とをそれぞれ求め、前記内的結合度と前記外的分離度とを変数とする所定の演算式から前記評価値を算出する
ことを特徴とする請求項１に記載のクラスタリング装置。
【請求項３】
前記評価値算出部は、
Ｘ：要素データ
Ｎｃ：クラスタ数
Ｃｉ：クラスタＩＤが「ｉ」のクラスタ（ｉ＝１，２・・・Ｎｃ）
ｎｉ：クラスタＣｉの要素データの個数
Ｖｉ：クラスタＣｉの重心
ＧＧ：全データの基準点

ｄ（Ｘ，Ｖｉ）：要素データＸと重心Ｖｉとの間のデータ重心間距離またはデータ重心間距離の二乗
ｄ（Ｖｉ，ＧＧ）：重心Ｖｉと基準点ＧＧとの間の重心距離または重心距離の二乗
として、以下の式（Ａ）で表される前記評価値または式の逆数で表される前記評価値を算出する
ことを特徴とする請求項１に記載のクラスタリング装置。
【数１】

$$評価値 = \frac{SBS/Nc}{SWD} \quad \cdots (A)$$

但し、$SWD = \sum_{i=1}^{Nc} \left(\frac{1}{n_i} \sum_{X \in C_i} d(X,V_i)\right)$

$SBS = \sum_{i=1}^{Nc} (n_i \cdot d(V_i,GG))$

【請求項４】
前記評価値算出部は、
Ｘ：要素データ
Ｎｃ：クラスタ数
Ｃｉ：クラスタＩＤが「ｉ」のクラスタ（ｉ＝１，２・・・Ｎｃ）
ｎｉ：クラスタＣｉの要素データの個数
Ｖｉ：クラスタＣｉの重心
ＧＧ：全データの基準点
ｄ（Ｘ，Ｖｉ）：要素データＸと重心Ｖｉとの間のデータ重心間距離またはデータ重心間距離の二乗
ｄ（Ｖｉ，Ｖｊ）：１のクラスタＣｉの重心Ｖｉと他のクラスタＣｊの重心Ｖｊとの間のクラスタ間距離またはクラスタ間距離の二乗
として、以下の式（Ｂ）で表される前記評価値または式の逆数で表される前記評価値を算出する
ことを特徴とする請求項１に記載のクラスタリング装置。
【数２】

$$評価値 = \frac{SBS/Nc}{SWD} \quad \cdots (B)$$

但し、$SWD = \sum_{i=1}^{Nc} \left(\frac{1}{n_i} \sum_{X \in C_i} d(X,V_i)\right)$

$SBS = \sum_{i=1}^{Nc} \min_{i \neq j} d(V_i,V_j)$

【請求項５】
前記一括処理部は、前記複数の要素データの分類されるクラスタ情報を保持するレジスタ部とデータ計算を行なう計算回路とを含み、
前記評価値算出部は、前記一括処理部が前記複数の要素データを分類する際に得られる前記評価値を計算するためのデータを保持する評価レジスタ部と前記評価値を計算するロジック部とを有し、
電源供給が独立して制御される複数のパワードメインが形成され、少なくとも前記データ記憶部と、前記計算回路と、前記更新処理部と、前記ロジック部とが異なる前記パワードメインに分けられるとともに、これら前記パワードメインとは別の前記パワードメインに前記レジスタ部及び評価レジスタ部が分けられている
ことを特徴とする請求項１ないし４のいずれか１項に記載のクラスタリング装置。
【請求項６】
前記一括処理部は、前記複数の要素データの分類されるクラスタＩＤ及びクラスタ重心間

距離を保持するレジスタユニットと、データ計算を行なう計算回路とを含み、
前記更新処理部は、前記複数のクラスタの各々の重心を記憶する重心メモリと、前記複数のクラスタのうちから前記最も近いクラスタを特定し、前記新規の要素データを前記最も近いクラスタに分類する近隣探索回路部とを含み、
前記評価値算出部は、前記一括処理部が前記複数の要素データを分類する際に得られる前記評価値を計算するためのデータを保持する評価レジスタ部と前記評価値を計算するロジック部とを有し、
電源供給が独立して制御される複数のパワードメインが形成され、少なくとも前記データ記憶部と、前記計算回路と、前記重心メモリと、前記近隣探索回路部と、前記ロジック部とが互いに異なる前記パワードメインに分けられるとともに、これらの前記パワードメインとは別の１の前記パワードメインに前記レジスタユニット及び評価レジスタ部が分けられている
ことを特徴とする請求項１ないし４のいずれか１項に記載のクラスタリング装置。
【請求項７】
前記判断部は、分類結果が妥当ではない場合に、前記一括処理部による前記新規の要素データを含む全ての要素データの分類を実行させることを特徴とする請求項１ないし６のいずれか１項に記載のクラスタリング装置。
【請求項８】
前記データ記憶部は、不揮発性メモリで構成されることを特徴とする請求項１ないし７のいずれか１項に記載のクラスタリング装置。
【請求項９】
１または複数の新規の要素データの追加に応答して、前記新規の要素データを含む複数の要素データを分類した状態にするクラスタリング部と、
分類結果を評価する評価値を算出する評価値算出部と
を備え、
前記評価値算出部は、
Ｘ：要素データ
Ｎｃ：クラスタ数
Ｃｉ：クラスタＩＤが「ｉ」のクラスタ（ｉ＝１，２・・・Ｎｃ）
ｎｉ：クラスタＣｉの要素データの個数
Ｖｉ：クラスタＣｉの重心
ＧＧ：全データの基準点
ｄ（Ｘ，Ｖｉ）：要素データＸと重心Ｖｉとの間のデータ重心間距離またはデータ重心間距離の二乗
ｄ（Ｖｉ，ＧＧ）：重心Ｖｉと基準点ＧＧとの間の重心距離または重心距離の二乗
として、以下の式（Ａ）で表される前記評価値または式の逆数で表される前記評価値を算出する
ことを特徴とするクラスタリング装置。
【数３】

$$評価値 = \frac{SBS/Nc}{SWD} \quad \cdots (A)$$

$$但し、SWD = \sum_{i=1}^{Nc} \left(\frac{1}{n_i} \sum_{X \in C_i} d(X, V_i) \right)$$

$$SBS = \sum_{i=1}^{Nc} \left(n_i \cdot d(V_i, GG) \right)$$

【請求項１０】
１または複数の新規の要素データの追加に応答して、前記新規の要素データを含む複数の要素データを分類した状態にするクラスタリング部と、
分類結果を評価する評価値を算出する評価値算出部と

を備え、
前記評価値算出部は、
Ｘ：要素データ
Ｎｃ：クラスタ数
Ｃｉ：クラスタＩＤが「ｉ」のクラスタ（ｉ＝１，２・・・Ｎｃ）
ｎｉ：クラスタＣｉの要素データの個数
Ｖｉ：クラスタＣｉの重心
ＧＧ：全データの基準点
ｄ（Ｘ，Ｖｉ）：要素データＸと重心Ｖｉとの間のデータ重心間距離またはデータ重心間距離の二乗
ｄ（Ｖｉ，Ｖｊ）：１のクラスタＣｉの重心Ｖｉと他のクラスタＣｊの重心Ｖｊとの間のクラスタ間距離またはクラスタ間距離の二乗
として、以下の式（Ｂ）で表される前記評価値または式の逆数で表される前記評価値を算出する
ことを特徴とするクラスタリング装置。
【数４】

$$\text{評価値} = \frac{SBS/Nc}{SWD} \quad \cdots (B)$$

$$\text{但し、} SWD = \sum_{i=1}^{Nc} \left(\frac{1}{n_i} \sum_{x \in C_i} d(X,V_i) \right)$$

$$SBS = \sum_{i=1}^{Nc} \min_{i \neq j} d(V_i,V_j)$$

【請求項１１】
複数の要素データを分類するクラスタリング装置において、
前記複数の要素データを記憶するデータ記憶部と、
分類結果を評価する評価値を算出する評価値算出部と、
設定されるクラスタ数の複数のクラスタに前記複数の要素データを分類することで得られる前記評価値に基づき、最適なクラスタ数の前記複数のクラスタに前記複数の要素データを分類する一括処理部と、
追加される新規の要素データを、前記一括処理部によって前記複数の要素データが分類されている前記複数のクラスタのうちの前記新規の要素データに最も近いクラスタに分類する更新処理部と、
前記更新処理部が前記新規の要素データを分類することで得られる前記評価値に基づき、前記新規の要素データの分類後の分類結果の妥当性を判断する判断部と
を備え、
前記一括処理部は、前記複数の要素データの分類されるクラスタ情報を保持するレジスタ部とデータ計算を行なう計算回路とを含み、
前記評価値算出部は、前記一括処理部が前記複数の要素データを分類する際に得られる前記評価値を計算するためのデータを保持する評価レジスタ部と前記評価値を計算するロジック部とを有し、
電源供給が独立して制御される複数のパワードメインが形成され、少なくとも前記データ記憶部と、前記計算回路と、前記更新処理部と、前記ロジック部とが異なる前記パワードメインに分けられるとともに、これら前記パワードメインとは別の前記パワードメインに前記レジスタ部及び評価レジスタ部が分けられている
ことを特徴とするクラスタリング装置。
【請求項１２】
複数の要素データを分類するクラスタリング装置において、
前記複数の要素データを記憶するデータ記憶部と、
分類結果を評価する評価値を算出する評価値算出部と、

設定されるクラスタ数の複数のクラスタに前記複数の要素データを分類することで得られる前記評価値に基づき、最適なクラスタ数の前記複数のクラスタに前記複数の要素データを分類する一括処理部と、
追加される新規の要素データを、前記一括処理部によって前記複数の要素データが分類されている前記複数のクラスタのうちの前記新規の要素データに最も近いクラスタに分類する更新処理部と、
前記更新処理部が前記新規の要素データを分類することで得られる前記評価値に基づき、前記新規の要素データの分類後の分類結果の妥当性を判断する判断部と
を備え、
前記一括処理部は、前記複数の要素データの分類されるクラスタＩＤ及びクラスタ重心間距離を保持するレジスタユニットと、データ計算を行なう計算回路とを含み、
前記更新処理部は、前記複数のクラスタの各々の重心を記憶する重心メモリと、前記複数のクラスタのうちから前記最も近いクラスタを特定し、前記新規の要素データを前記最も近いクラスタに分類する近隣探索回路部とを含み、
前記評価値算出部は、前記一括処理部が前記複数の要素データを分類する際に得られる前記評価値を計算するためのデータを保持する評価レジスタ部と前記評価値を計算するロジック部とを有し、
電源供給が独立して制御される複数のパワードメインが形成され、少なくとも前記データ記憶部と、前記計算回路と、前記重心メモリと、前記近隣探索回路部と、前記ロジック部とが互いに異なる前記パワードメインに分けられるとともに、これらの前記パワードメインとは別の１の前記パワードメインに前記レジスタユニット及び評価レジスタ部が分けられている
ことを特徴とするクラスタリング装置。
【請求項１３】
クラスタ数を変化させながら、全ての要素データを用いた分類を行なうことで得られる各クラスタ数についての分類結果を評価する評価値から最適なクラスタ数を求め、求めた前記最適なクラスタ数で前記要素データを分類した分類結果を得る一括処理工程と、
前記一括処理工程後に新規の要素データが追加された際に、前記新規の要素データを既存のクラスタのいずれかに分類し、分類後に前記評価値を求める更新処理工程と、
前記更新処理工程で前記新規の要素データの分類後に得られる前記評価値と、当該新規の要素データの分類前の前記評価値とを比較することによって、前記更新処理工程による前記新規の要素データの分類後の分類結果の妥当性を判断する判断工程と
を有することを特徴とするクラスタリング方法。
【請求項１４】
１または複数の新規の要素データの追加に応答して、新規の要素データを含む複数の要素データを分類した状態にするクラスタリング工程と、
分類結果を評価する評価値を算出する評価値算出工程と
を有し、
前記評価値算出工程は、
Ｘ：要素データ
Ｎｃ：クラスタ数
Ｃ$_{i}$：クラスタＩＤが「ｉ」のクラスタ（ｉ＝１，２・・・Ｎｃ）
ｎ$_{i}$：クラスタＣ$_{i}$の要素データの個数
Ｖ$_{i}$：クラスタＣ$_{i}$の重心
ＧＧ：全データの基準点
ｄ（Ｘ，Ｖ$_{i}$）：要素データＸと重心Ｖ$_{i}$との間のデータ重心間距離またはデータ重心間距離の二乗
ｄ（Ｖ$_{i}$，ＧＧ）：重心Ｖ$_{i}$と基準点ＧＧとの間の重心距離または重心距離の二乗
として、以下の式（Ａ）で表される前記評価値または式の逆数で表される前記評価値を算出する

ことを特徴とするクラスタリング方法。
【数５】

$$評価値 = \frac{SBS/Nc}{SWD} \quad \cdots (A)$$

$$但し、SWD = \sum_{i=1}^{Nc} \left(\frac{1}{n_i} \sum_{x \in C_i} d(X,V_i)\right)$$

$$SBS = \sum_{i=1}^{Nc} (n_i \cdot d(V_i,GG))$$

【請求項１５】
１または複数の新規の要素データの追加に応答して、新規の要素データを含む複数の要素データを分類した状態にするクラスタリング工程と、
分類結果を評価する評価値を算出する評価値算出工程と
を有し、
前記評価値算出工程は、
Ｘ：要素データ
Ｎｃ：クラスタ数
Ｃ ｉ：クラスタＩＤが「ｉ」のクラスタ（ｉ＝１，２・・・Ｎｃ）
ｎ ｉ：クラスタＣ ｉの要素データの個数
Ｖ ｉ：クラスタＣ ｉの重心
ＧＧ：全データの基準点
ｄ（Ｘ，Ｖ ｉ）：要素データＸと重心Ｖ ｉとの間のデータ重心間距離またはデータ重心間距離の二乗
ｄ（Ｖ ｉ，Ｖ ｊ）：１のクラスタＣ ｉの重心Ｖ ｉと他のクラスタＣ ｊの重心Ｖ ｊとの間のクラスタ間距離またはクラスタ間距離の二乗
として、以下の式（Ｂ）で表される前記評価値または式の逆数で表される前記評価値を算出する
ことを特徴とするクラスタリング方法。
【数６】

$$評価値 = \frac{SBS/Nc}{SWD} \quad \cdots (B)$$

$$但し、SWD = \sum_{i=1}^{Nc} \left(\frac{1}{n_i} \sum_{x \in C_i} d(X,V_i)\right)$$

$$SBS = \sum_{i=1}^{Nc} \min_{i \neq j} d(V_i,V_j)$$