

【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3月8日 (2012.3.8)

【公開番号】特開2011-134230(P2011-134230A)

【公開日】平成23年7月7日 (2011.7.7)

【年通号数】公開・登録公報2011-027

【出願番号】特願2009-294925(P2009-294925)

【国際特許分類】

G 0 6 F 19/00 (2011.01)

G 0 6 F 17/30 (2006.01)

G 0 6 Q 10/00 (2012.01)

【F I】

G 0 6 F 19/00 1 3 0

G 0 6 F 17/30 2 2 0 Z

G 0 6 F 17/30 1 7 0 A

G 0 6 F 17/60 1 5 0

【手続補正書】

【提出日】平成24年1月24日 (2012.1.24)

【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】請求項 1

【補正方法】変更

【補正の内容】

【請求項 1】

所定の確率分布モデルに基づいて、電子データである複数の文書それぞれを、予め定められた複数のトレンドクラスのいずれかに確率的に属させる計算を行うトレンド分析装置であって、

前記複数の文書、

前記確率分布モデルで使用され、前記文書毎に、確率的にその文書の属する前記トレンドクラスを推定するための第 1 の計算式、

前記確率分布モデルで使用され、前記文書を構成する単語に対応する概念であるトークン毎に割り当てられる単語を推定するためのスイッチ変数として、( 1 ) 前記複数の文書すべてについての単語生成分布に対応する値、( 2 ) 前記トレンドクラスについての単語生成分布に対応する値、( 3 ) 前記トレンドクラスに確率的に属する 1 以上のトピッククラスについての単語生成分布に対応する値、の 3 つの値のうちいずれかの値をとるか決定するための第 2 の計算式、および、

所定の計算終了条件、を記憶する記憶部と、

前記文書それぞれが属する前記トレンドクラス、および、前記トレンドクラスに属する 1 以上のトピッククラス、の初期値を設定する初期設定部と、

前記第 1 の計算式および前記スイッチ変数に基づいて、前記文書毎に、前記トレンドクラスそれぞれについて、その文書の属する確率を計算し、その確率が、0 以上 1 未満の乱数値よりも高い場合に、その文書がその前記トピッククラスに属するものと推定するトレンドクラス推定部と、

前記第 2 の計算式に基づいて、前記文書毎の前記トークン毎に、前記文書すべてについての単語生成分布を用いて計算した場合の正解の単語を選択する条件付確率と、前記トレンドクラスについての単語生成分布を用いて計算した場合の正解の単語を選択する条件付確率と、前記トピッククラスについての単語生成分布を用いて計算した場合の正解の単語を選択する条件付確率と、を計算して、そのうちの最大の数値が、0 以上 1 未満の乱数値

よりも高い場合に、その最大の数値を計算するのに使用した単語生成分布に対応するスイッチ変数を決定するスイッチ変数推定部と、

前記所定の計算終了条件を満たすまで、前記トレンドクラス推定部と前記スイッチ変数推定部とに、計算を繰り返させる収束判定部と、

前記収束判定部によって前記所定の計算終了条件が満たされたと判定されたとき、前記複数の文書それぞれが属する前記トレンドクラスを含む計算結果を出力する出力部と、

を備えることを特徴とするトレンド分析装置。

【手続補正 2】

【補正対象書類名】特許請求の範囲

【補正対象項目名】請求項 3

【補正方法】変更

【補正の内容】

【請求項 3】

所定の確率分布モデルに基づいて、電子データである複数の文書それぞれを、予め定められた複数のトレンドクラスのいずれかに確率的に属させる計算を行うトレンド分析装置によるトレンド分析方法であって、

前記トレンド分析装置は、

前記複数の文書、

前記確率分布モデルで使用され、前記文書毎に、確率的にその文書の属する前記トレンドクラスを推定するための第 1 の計算式、

前記確率分布モデルで使用され、前記文書を構成する単語に対応する概念であるトークン毎に割り当てられる単語を推定するためのスイッチ変数として、(1) 前記複数の文書すべてについての単語生成分布に対応する値、(2) 前記トレンドクラスについての単語生成分布に対応する値、(3) 前記トレンドクラスに確率的に属する 1 以上のトピッククラスについての単語生成分布に対応する値、の 3 つの値のうちいずれかの値をとるか決定するための第 2 の計算式、および、

所定の計算終了条件、を記憶する記憶部と、

初期設定部と、トレンドクラス推定部と、スイッチ変数推定部と、収束判定部と、出力部と、を備えており、

前記初期設定部は、前記文書それぞれが属する前記トレンドクラス、および、前記トレンドクラスに属する 1 以上のトピッククラス、の初期値を設定し、

前記トレンドクラス推定部は、前記第 1 の計算式および前記スイッチ変数に基づいて、前記文書毎に、前記トレンドクラスそれぞれについて、その文書の属する確率を計算し、その確率が、0 以上 1 未満の乱数値よりも高い場合に、その文書がその前記トピッククラスに属するものと推定し、

前記スイッチ変数推定部は、前記第 2 の計算式に基づいて、前記文書毎の前記トークン毎に、前記文書すべてについての単語生成分布を用いて計算した場合の正解の単語を選択する条件付確率と、前記トレンドクラスについての単語生成分布を用いて計算した場合の正解の単語を選択する条件付確率と、前記トピッククラスについての単語生成分布を用いて計算した場合の正解の単語を選択する条件付確率と、を計算して、そのうちの最大の数値が、0 以上 1 未満の乱数値よりも高い場合に、その最大の数値を計算するのに使用した単語生成分布に対応するスイッチ変数を決定し、

前記収束判定部は、前記所定の計算終了条件を満たすまで、前記トレンドクラス推定部と前記スイッチ変数推定部とに、計算を繰り返させ、

前記出力部は、前記収束判定部によって前記所定の計算終了条件が満たされたと判定されたとき、前記複数の文書それぞれが属する前記トレンドクラスを含む計算結果を出力する

ことを特徴とするトレンド分析方法。

【手続補正 3】

【補正対象書類名】明細書

【補正対象項目名】 0 0 0 8

【補正方法】 変更

【補正の内容】

【 0 0 0 8 】

前記課題を解決するために、本発明は、所定の確率分布モデルに基づいて、電子データである複数の文書それぞれを、予め定められた複数のトレンドクラスのいずれかに確率的に属させる計算を行うトレンド分析装置であって、複数の文書、確率分布モデルで使用され、文書毎に、確率的にその文書の属するトレンドクラスを推定するための第 1 の計算式、確率分布モデルで使用され、文書を構成する単語に対応する概念であるトークン毎に割り当てられる単語を推定するためのスイッチ変数として、( 1 ) 複数の文書すべてについての単語生成分布に対応する値、( 2 ) トレンドクラスについての単語生成分布に対応する値、( 3 ) トレンドクラスに確率的に属する 1 以上のトピッククラスについての単語生成分布に対応する値、の 3 つの値のうちいずれかの値をとるか決定するための第 2 の計算式、および、所定の計算終了条件、を記憶する記憶部と、文書それぞれが属するトレンドクラス、および、トレンドクラスに属する 1 以上のトピッククラス、の初期値を設定する初期設定部と、第 1 の計算式および前記スイッチ変数に基づいて、文書毎に、トレンドクラスそれぞれについて、その文書の属する確率を計算し、その確率が、0 以上 1 未満の乱数値よりも高い場合に、その文書がそのトピッククラスに属するものと推定するトレンドクラス推定部と、第 2 の計算式に基づいて、文書毎のトークン毎に、文書すべてについての単語生成分布を用いて計算した場合の正解の単語を選択する条件付確率と、トレンドクラスについての単語生成分布を用いて計算した場合の正解の単語を選択する条件付確率と、トピッククラスについての単語生成分布を用いて計算した場合の正解の単語を選択する条件付確率と、を計算して、そのうちの最大の数値が、0 以上 1 未満の乱数値よりも高い場合に、その最大の数値を計算するのに使用した単語生成分布に対応するスイッチ変数を決定するスイッチ変数推定部と、所定の計算終了条件を満たすまで、トレンドクラス推定部とスイッチ変数推定部とに、計算を繰り返させる収束判定部と、収束判定部によって所定の計算終了条件が満たされたと判定されたとき、複数の文書それぞれが属するトレンドクラスを含む計算結果を出力する出力部と、を備えることを特徴とする。

【手続補正 4】

【補正対象書類名】 明細書

【補正対象項目名】 0 0 0 9

【補正方法】 変更

【補正の内容】

【 0 0 0 9 】

かかる発明によれば、予め定められた複数のトレンドクラスを有する所定の確率分布モデルにおいて、スイッチ変数に応じた単語生成分布、第 1 の計算式、第 2 の計算式などを用いて、各文書が確率的に属するトレンドクラスを計算することができる。つまり、文書について、トピック毎の時間変化の速度の差も考慮した上でトレンド分析を行うことができる。