(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第5258140号
(P5258140)

(45)発行日　平成25年8月7日(2013.8.7)　(24)登録日　平成25年5月2日(2013.5.2)

(51)Int.Cl.　F I
G06F 17/30 (2006.01)　G06F 17/30 220Z
G06F 17/30 340A
G06F 17/30 340B
G06F 17/30 350C

請求項の数 11　(全 17 頁)

(21)出願番号　特願2003-544641 (P2003-544641)
(86)(22)出願日　平成14年10月22日(2002.10.22)
(65)公表番号　特表2005-509954 (P2005-509954A)
(43)公表日　平成17年4月14日(2005.4.14)
(86)国際出願番号　PCT/IB2002/004413
(87)国際公開番号　WO2003/042879
(87)国際公開日　平成15年5月22日(2003.5.22)
審査請求日　平成17年10月21日(2005.10.21)
審判番号　不服2010-16461 (P2010-16461/J1)
審判請求日　平成22年7月22日(2010.7.22)
(31)優先権主張番号　10/014,180
(32)優先日　平成13年11月13日(2001.11.13)
(33)優先権主張国　米国(US)

(73)特許権者　590000248
コーニンクレッカ　フィリップス　エレクトロニクス　エヌ　ヴィ
オランダ国　５６５６　アーエー　アインドーフェン　ハイテック　キャンパス　５
(74)代理人　100087789
弁理士　津軽　進
(74)代理人　100122769
弁理士　笛田　秀仙
(74)代理人　100163821
弁理士　柴田　沙希子
(72)発明者　グッタ　スリニヴァス　ヴィ　アール
オランダ国　５６５６　アーアー　アインドーフェン　プロフ　ホルストラーン　６

最終頁に続く

(54)【発明の名称】アイテムの推薦器においてアイテムの近さを評価するための方法及び装置

(57)【特許請求の範囲】
【請求項１】
推薦器において用いられる方法であって、少なくとも１つのシンボル的なフィーチャを有しているアイテムを、１つ又は複数のグループに割り当て、前記推薦器のユーザのための視聴履歴を生成する方法において、前記推薦器が、
前記アイテムを第三者が利用した履歴を集めるステップと、
前記履歴における前記アイテムについて、前記アイテムと前記グループのそれぞれにおける少なくとも１つのアイテムとの対応するシンボル的なフィーチャ値間の距離を計算するステップであって、前記距離は、前記シンボル的なフィーチャ値の取り得る値それぞれに対応する、すべての実例のクラス分けの全体的な類似性に基づくものであるステップと、
前記アイテムと前記グループのそれぞれにおける少なくとも１つのアイテムとの近さを決定する各前記フィーチャ値間の前記距離を総計するステップと、
前記アイテムを最小距離値に関連する前記グループに割り当てるステップと、
前記グループに基づいて、前記視聴履歴を生成するステップと、
を実施する、方法。
【請求項２】
前記距離を計算する前記ステップは、前記シンボル的なフィーチャ間の前記距離を計算するために変更されたバリューディファレンスメトリック技法を用いる、請求項１に記載の方法。

【請求項３】
特定のシンボル的なフィーチャに関する２つの値Ｖ１とＶ２との間の距離δは、
δ（Ｖ１，Ｖ２）＝　Σ｜Ｃ１ｉ／Ｃ１－Ｃ２ｉ／Ｃ２｜
によって与えられ、ここで、Ｃ１ｉは、Ｖ１がクラスｉに分類された回数であり、Ｃ１は、Ｖ１がデータセットに出現した合計回数であり、Ｃ２ｉは、Ｖ２がクラスｉに分類された回数であり、Ｃ２は、Ｖ２がデータセットに出現した合計回数である、請求項１に記載の方法。
【請求項４】
前記アイテムは番組であり、クラスｉは「視聴」クラス及び「非視聴」クラスであり、特定のシンボル的なフィーチャに関する２つの値Ｖ１とＶ２との間の前記距離δは、
【数１】

$$\delta(V1, V2)=\left|\frac{C1\_watched}{C1\_total}-\frac{C2\_watched}{C2\_total}\right|+\left|\frac{C1\_not\_watched}{C1\_total}-\frac{C2\_not\_watched}{C2\_total}\right|$$

によって与えられ、ここで、Ｃ１_ｗａｔｃｈｅｄはＶ１が「視聴」クラスに分類された回数、Ｃ１＿ｎｏｔ＿ｗａｔｃｈｅｄはＶ１が「非視聴クラス」に分類された回数、Ｃ１＿ｔｏｔａｌはＶ１がデータセットに出現した合計回数であり、Ｃ２_ｗａｔｃｈｅｄはＶ２が「視聴」クラスに分類された回数、Ｃ２＿ｎｏｔ＿ｗａｔｃｈｅｄはＶ２が「非視聴クラス」に分類された回数、Ｃ２＿ｔｏｔａｌは、Ｖ２がデータセットに出現した合計回数である、請求項１に記載の方法。
【請求項５】
前記グループのそれぞれにおける前記少なくとも１つのアイテムは、予め計算されたグループの平均的なアイテムである、請求項１に記載の方法。
【請求項６】
前記アイテムは番組である、請求項１に記載の方法。
【請求項７】
前記アイテムはコンテントである、請求項１に記載の方法。
【請求項８】
前記アイテムは製品である、請求項１に記載の方法。
【請求項９】
推薦器において用いられるシステムであって、少なくとも１つのシンボル的なフィーチャを有しているアイテムを、１つ又は複数のグループに割り当て、前記推薦器のユーザのための視聴履歴を生成するシステムにおいて、前記推薦器が、
前記アイテムを第三者が利用した履歴を集める手段と、
前記履歴における前記アイテムについて、前記アイテムと前記グループのそれぞれにおける少なくとも１つのアイテムとの対応するシンボル的なフィーチャ値間の距離を計算する手段であって、前記距離は、前記シンボル的なフィーチャ値の取り得る値それぞれに対応する、すべての実例のクラス分けの全体的な類似性に基づくものである手段と、
前記アイテムと前記グループのそれぞれにおける少なくとも１つのアイテムとの近さを決定する各前記フィーチャ値の前記距離を総計する手段と、
前記アイテムを最小距離値に関連する前記グループに割り当てる手段と、
前記グループに基づいて、前記視聴履歴を生成する手段と、
を有する、システム。
【請求項１０】

推薦器において用いられるシステムであって、少なくとも１つのシンボル的なフィーチャを有しているアイテムを、１つ又は複数のグループに割り当て、前記推薦器のユーザのための視聴履歴を生成するシステムにおいて、コンピュータ読み取り可能なコードを記憶するためのメモリと、前記メモリに機能的に結合されたプロセッサとを有しており、前記プロセッサが、

前記アイテムを第三者が利用した履歴を集め、

前記履歴における前記アイテムについて、前記アイテムと前記グループのそれぞれにおける少なくとも１つのアイテムとの対応するシンボル的なフィーチャ値間の距離を、前記シンボル的なフィーチャ値の取り得る値それぞれに対応する、すべての実例のクラス分けの全体的な類似性に基づいて計算し、

前記アイテムと前記グループのそれぞれにおける少なくとも１つのアイテムとの近さを決定する各前記フィーチャ値間の前記距離を総計し、

前記アイテムを最小距離値に関連する前記グループに割り当て、

前記グループに基づいて、前記視聴履歴を生成する

システム。

【請求項１１】

プログラム可能な装置がコンピュータプログラムを実行するときに請求項９に記載のシステムとして機能することを可能にするコンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、「Method and Apparatus for Partitioning a Plurality of Items into Groups of Similar Items in a Recommender of Such Items」というタイトルの米国特許出願（代理人整理番号ＵＳ０１０５６８）、「Method and Apparatus for Generating A Stereotypical Profile for Recommending Items of Interest Using Item-Based Clustering」というタイトルの米国特許出願（代理人整理番号ＵＳ０１０５６９）、「Method and Apparatus for Recommending Items of Interest Based on Preferences of a Selected Third Party」というタイトルの米国特許出願（代理人整理番号ＵＳ０１０５７２）、「Method and Apparatus for Recommending Items of Interest Based on Stereotype Preferences of Third Parties」というタイトルの米国特許出願（代理人整理番号ＵＳ０１０５７５）及び「Method and Apparatus for Generating a Stereotypical Profile for Recommending Items of Interest Using Feature-Based Clustering」というタイトルの米国特許出願（代理人整理番号ＵＳ０１０５７６）に関連し、これらの各出願は、本願と同時に出願され、本発明の譲受人に譲渡され、参照によってここに盛り込まれる。

【０００２】

本発明は、例えばテレビ番組のような興味のあるアイテムを推薦するための方法及び装置に関し、より具体的には、ユーザの購入又は視聴履歴が利用できるようになる前に興味のある番組及び他のアイテムを推薦するための技法に関する。

【背景技術】

【０００３】

テレビの視聴者に利用可能なチャネルの数は、このようなチャネル上で利用可能な番組コンテントの多様性とともに増加しているので、テレビの視聴者が興味のあるテレビ番組を識別することはますます骨の折れることになっている。電子番組ガイド（ＥＰＧ）は、例えばタイトル、時間、日付及びチャネルによって利用可能なテレビ番組を識別するとともに、個人化された好みに従って利用可能なテレビ番組がサーチされ又は分類されることを許すことによって興味のある番組の識別を容易にする。

【０００４】

興味のあるテレビ番組及びその他のアイテムを推薦するための多数の推薦ツールが提案され又は提言されている。テレビ番組推薦ツールは、例えば、特定の視聴者にとって興味のありうる推薦された番組の組を得るために視聴者の好みを電子番組ガイドに適用する。

概して、テレビ番組推薦ツールは、黙示的若しくは明示的な技法を使用して又は前述のなんらかの組み合わせを使用して視聴者の好みを取得する。黙示的なテレビ番組推薦ツールは、視聴者の視聴履歴から導き出される情報に基づいて目立たないやり方でテレビ番組推薦を生成する。他方、明示的なテレビ番組推薦ツールは、視聴者プロファイルを導き出し推薦を生成するために、例えばタイトル、ジャンル、俳優、チャネル及び日付／時間のような番組属性に関する好みについて明示的に視聴者に尋ねる。

【発明の開示】

【発明が解決しようとする課題】

【０００５】

今日利用可能な推薦ツールは、ユーザが興味のあるアイテムを識別するのを助けるが、それらは多くの制約を受け、これらの制約は、克服される場合にはこのような推薦ツールの利便性及び性能を大幅に改善することができる。例えば、包括的に、明示的な推薦ツールは初期化するのが非常に長たらしく退屈であり、それぞれの新しいユーザは粒度の粗いレベルで彼らの好みを特定する非常に詳細な調査に応える必要がある。黙示的なテレビ番組推薦ツールは、視聴の振る舞いを観察することによって目立たないようにプロファイルを導き出すが、それらは正確になるために長い時間を必要とする。加えて、このような黙示的なテレビ番組推薦ツールは、いかなる推薦をし始めるためにも視聴履歴の少なくとも最小量を必要とする。こうして、このような黙示的なテレビ番組推薦ツールは、推薦ツールが最初に取得されるときにはいかなる推薦もすることができない。

【０００６】

従って、十分に個人化された視聴履歴が利用できるようになる前に目立たないように例えばテレビ番組のようなアイテムを推薦することができる方法及び装置が必要とされている。加えて、第三者の視聴習慣に基づいて所与のユーザのための番組推薦を生成するための方法及び装置が必要とされている。

【課題を解決するための手段】

【０００７】

概して、例えばテレビ番組推薦のようなユーザにとって興味のあるアイテムを推薦するための方法及び装置が開示される。本発明の一つの見地によれば、例えばユーザが初めて推薦器を取得するときのように、ユーザの視聴履歴又は購入履歴が利用できるようになる前に推薦が生成される。最初に、１又は複数の第三者からの視聴履歴又は購入履歴が、特定のユーザにとって興味のあるアイテムを推薦するために使用される。

【０００８】

第三者の視聴又は購入履歴は、代表的な視聴者により選択されたアイテムの一般的なパターンを反映する典型的（stereotype）なプロファイルを生成するために処理される。それぞれの典型的なプロファイルは、何らかの態様で互いに類似するアイテム（データポイント）のクラスタである。ユーザは、自分の興味に最も近いアイテムにより自分のプロファイルを初期化するために、興味のある１つ又は複数の典型を選択する。

【０００９】

クラスタリングルーチンは、１つのクラスタにおけるポイント（例えばテレビ番組）が他のいかなるクラスタよりもそのクラスタの平均に近くなるように、第三者の視聴又は購入履歴（データセット）をクラスタに分ける。更に平均計算ルーチンが、クラスタのシンボル的（象徴的、symbolic）な平均を計算するために開示される。テレビ番組のような所与のデータポイントが、各クラスタの平均を使用して、データポイントと各クラスタとの間の距離に基づいてクラスタに割り当てられる。

【００１０】

開示される距離計算ルーチンは、所与のテレビ番組と所与のクラスタの平均との間の距離に基づいて、テレビ番組の各クラスタに対する近さを評価する。計算された距離メトリックは、クラスタの範囲を決定するためにサンプルデータセットにおけるさまざまな標本間の区別を定量化する。バリューディファレンスメトリック（value difference metric、ＶＤＭ）技法又はその変形したものが、２つのテレビ番組間のフィーチャ値の間の距離

を計算するために用いられる。既知の変更されたＶＤＭ（modified VDM、ＭＶＤＭ）技法によれば、特定のフィーチャに関する２つの値の間の距離δは、下式によって与えられる。

$$\delta(V1,V2)=\Sigma\left|C1i/C1-C2i/C2\right|^{r}$$

ここで、Ｖ１及びＶ２は、考慮中のフィーチャに関する２つの可能な値である。例示的な実施例の番組推薦環境において、興味のあるクラスは、「視聴（watched）」及び「非視聴（not-watched）」である。概して、開示される距離計算ルーチンは、すべての分類について同じ相対頻度で値が生じる場合、それらの値を類似するものと識別する。こうして、２つの値は、それらがすべての可能な分類について類似する尤度（likelihoods）を与える場合は類似する。

【発明を実施するための最良の形態】

【００１１】

本発明のより完全な理解及び本発明の他の特徴及び利点は、以下の詳細な説明及び図面を参照することにより得られる。

【００１２】

図１は、本発明によるテレビ番組推薦器１００を示す。図１に示すように、例示的なテレビ番組推薦器１００は、特定の視聴者にとって興味のある番組を識別するために、図２に関連して以下に記述される番組データベース２００における番組を評価する。推薦される番組の組は、例えば良く知られたオンスクリーン表示技法を使用するセットトップ端末／テレビジョン（図示せず）を使用して視聴者に対して提示されることができる。本発明は、ここでテレビ番組推薦のコンテキストにおいて説明されているが、本発明は、例えば視聴履歴又は購入履歴のようなユーザの振る舞いの評価に基づくいかなる自動的に生成される推薦にも適用されることができる。

【００１３】

本発明の１つの特徴によれば、テレビ番組推薦器１００は、例えばユーザが最初にテレビ番組推薦器１００を取得するときのように、ユーザの視聴履歴１４０が利用できるようになる前に、テレビ番組推薦を生成することができる。図１に示すように、テレビ番組推薦器１００は、まず、特定のユーザにとって興味のある番組を推薦するために１又は複数の第三者からの視聴履歴１３０を用いる。概して、第三者の視聴履歴１３０は、比較的大きい母集団を代表する例えば年齢、収入、性別及び教育のようなデモグラフィックスを有する１つ又は複数のサンプル母集団の視聴習慣に基づく。

【００１４】

図１に示すように、第三者の視聴履歴１３０は、所与の母集団によって視聴される番組及び視聴されない番組の組からなる。視聴される番組の組は、所与の母集団によって実際に視聴される番組を観察することによって得られる。視聴されない番組の組は、例えば番組データベース２００における番組をランダムに抽出することによって得られる。他の変形例において、視聴されない番組の組は、「An Adaptive Sampling Technique for Selecting Negative Examples for Artificial Intelligence Applications」というタイトルの米国特許出願第０９／８１９，２８６号（２００１年3月２８日出願）の教示に従って得られる。前述の米国特許出願は本発明の譲受人に譲渡され、その内容はここに参照によって盛り込まれる。

【００１５】

本発明の別の特徴によれば、テレビ番組推薦器１００は、代表的な視聴者によって視聴されるテレビ番組の一般的なパターンを反映する典型的なプロファイルを生成するために第三者の視聴履歴１３０を処理する。以下に更に記述されるように、典型的なプロファイルは、何らかの態様で互いに類似するテレビ番組（データポイント）のクラスタである。こうして、所与のクラスタは、特定のパターンを示す第三者の視聴履歴１３０からのテレビ番組の特定のセグメントに対応する。

【００１６】

第三者の視聴履歴１３０は、本発明に従って、ある特定のパターンを示す番組のクラス

タを提供するために処理される。そののち、ユーザは、最も関連する１つ又は複数の典型を選択することができ、それによって、自分の興味に最も近い番組により自分のプロファイルを初期化することができる。典型的なプロファイルは、それらの記録パターンに依存して、それぞれの個々のユーザの特定の個人の視聴振る舞いに対して適応し、発展し、フィードバックが番組に与えられる。一実施例において、番組スコアを決定する際、ユーザ自身の視聴履歴１４０からの番組には、第三者の視聴履歴１３０からの番組よりも高い重みを与えることができる。

【００１７】

テレビ番組推薦器１００は、中央処理ユニット（ＣＰＵ）のようなプロセッサ１１５並びにＲＡＭ及び／又はＲＯＭのようなメモリ１２０を有する、パーソナルコンピュータ又はワークステーションのようないかなるコンピューティング装置としても具体化されることができる。テレビ番組推薦器１００はまた、例えばセットトップ端末又はディスプレイ（図示せず）における特定用途向け集積回路（ＡＳＩＣ）として具体化されることもできる。加えて、テレビ番組推薦器１００は、例えばカリフォルニア州サニーベールのＴｉｖｏ社から市販されているＴｉｖｏ（Ｒ）システム並びに「Method and Apparatus for Recommending Television Programming Using Decision Trees」というタイトルの米国特許出願第０９／４６６，４０６号（１９９９年１２月１７日出願）、「Bayesian TV Show Recommender」というタイトルの米国特許出願第０９／４９８，２７１号（２０００年２月４日出願）、「Three-Way Media Recommendation Method and System」というタイトルの米国特許出願第０９／６２７，１３９号（２０００年７月２７日出願）に添付の明細書に記載されているテレビ番組推薦器又はそれらの組み合わせのような、いかなる利用可能なテレビ番組推薦器としても具体化されることができる。前述の各々の米国特許出願の内容は本発明の特徴及び機能を実行するために変形されて参照によってここに盛り込まれる。

【００１８】

図１に示され、図２乃至図８に関連して以下に更に記述されるように、テレビ番組推薦器１００は、番組データベース２００、典型的プロファイルプロセス３００、クラスタリングルーチン４００、平均計算ルーチン５００、距離計算ルーチン６００及びクラスタ性能評価ルーチン８００を含む。概して、番組データベース２００は、よく知られた電子番組ガイドとして具体化されることができ、所与の時間間隔において利用可能なそれぞれの番組に関する情報を記録する。典型的プロファイルプロセス３００は、（ｉ）代表的な視聴者によって視聴されるテレビ番組の一般的なパターンを反映する典型的なプロファイルを生成するために第三者の視聴履歴１３０を処理し、（ｉｉ）ユーザが最も関連する１つ又は複数の典型を選択し、それによって自分のプロファイルを初期化することを可能にし、（ｉｉｉ）選択された典型に基づいて推薦を生成する。

【００１９】

クラスタリングルーチン４００は、１つのクラスタにおけるポイント（テレビ番組）が他のいかなるクラスタよりもそのクラスタの平均（重心）に近くなるように第三者視聴履歴１３０（データセット）をクラスタに分けるために、典型的プロファイルプロセス３００によって呼び出される。クラスタリングルーチン４００は、クラスタのシンボル的な平均を計算するために平均計算ルーチン５００を呼び出す。距離計算ルーチン６００は、所与のテレビ番組と所与のクラスタの平均と間の距離に基づいてそれぞれのクラスタに対するテレビ番組の近さを評価するために、クラスタリングルーチン４００によって呼び出される。最後に、クラスタリングルーチン４００は、クラスタを生成するためのストップ基準が満たされたときを決定するために、クラスタリング性能評価ルーチン８００を呼び出す。

【００２０】

図２は、図１の番組データベース（電子番組ガイド）２００からのサンプルテーブルである。上述したように、番組データベース２００は、所与の時間間隔において利用可能であるそれぞれの番組ごとの情報を記録する。図２に示すように、番組データベース２００は、例えばレコード２０５乃至２２０のような複数のレコードを含み、それぞれのレコー

ドは、所与の番組に関連する。それぞれの番組について、番組データベース２００は、フィールド２４０及び２４５においてそれぞれ番組に関連する日付／時間及びチャネルを示す。加えて、それぞれの番組に関するタイトル、ジャンル及び俳優は、フィールド２５０、２５５及び２７０においてそれぞれ識別される。番組の持続期間及び説明のようなよく知られた付加的なフィーチャ（図示せず）もまた、番組データベース２００に含められることができる。

【００２１】

図３は、本発明の特徴を取り入れる典型的プロファイルプロセス３００の例示的な実現例を示すフローチャートである。上述したように、典型的プロファイルプロセス３００は、（ｉ）代表的な視聴者によって視聴されるテレビ番組の一般的なパターンを反映する典型的なプロファイルを生成するために第三者の視聴履歴１３０を処理し、（ｉｉ）ユーザが最も関連する１つ又は複数の典型を選択し、それによって自分のプロファイルを初期化することを可能にし、（ｉｉｉ）選択された典型に基づいて推薦を生成する。第三者の視聴履歴１３０の処理は、オフラインで実施されてもよく、例えば工場で実施されてもよく、テレビ番組推薦器１００は、ユーザによる選択のために生成された典型的なプロファイルをインストールされて、ユーザに提供されることができることに留意する。

【００２２】

こうして、図３に示すように、典型的プロファイルプロセス３００は、まずステップ３１０の間に、第三者の視聴履歴１３０を集める。そののち、ステップ３２０の間に、典型的プロファイルプロセス３００は、典型的なプロファイルに対応する番組のクラスタを生成するために、図４に関連して以下に記述されるクラスタリングルーチン４００を実行する。以下に更に記述されるように、例示的なクラスタリングルーチン４００は、視聴履歴データセット１３０に対して、例えば「ｋ-means（ｋ平均）」クラスタルーチンのような非監督型データクラスタリングアルゴリズムを用いることができる。上述したように、クラスタリングルーチン４００は、１つのクラスタにおけるポイント（テレビ番組）が他のいかなるクラスタよりもそのクラスタの平均（重心）に近くなるように、第三者視聴履歴１３０（データセット）をクラスタに分ける。

【００２３】

ステップ３３０の間、典型的プロファイルプロセス３００は、それぞれの典型的なプロファイルを特徴付ける１つ又は複数のラベルをそれぞれのクラスタに割り当てる。１つの例示的な実施例において、クラスタの平均は、クラスタ全体について代表的なテレビ番組になり、平均番組のフィーチャは、クラスタにラベルをつけるために使用されることができる。例えば、テレビ番組推薦器１００は、ジャンルがそれぞれのクラスタについて優勢な又は規定するフィーチャであるように構成されることができる。

【００２４】

ステップ３４０の間、ラベルをつけられた典型的なプロファイルは、ユーザの興味に最も近い１つ又は複数の典型的なプロファイルの選択のために、それぞれのユーザに対して提示される。それぞれの選択されたクラスタを構成する番組は、当該典型の「一般的な視聴履歴」と考えられることができ、それぞれのクラスタについて典型的なプロファイルを構築するために使用されることができる。こうして、ステップ３５０の間、選択された典型的なプロファイルからの番組からなる視聴履歴がユーザのために生成される。最後に、ステップ３６０の間、前のステップで生成された視聴履歴が、番組推薦を得るために番組推薦器に適用される。番組推薦器は、当業者にとって明らかであるであるように、上述した番組推薦器のようないかなる通常の番組推薦器としてここで変形されて具体化されてもよい。プログラム制御は、ステップ３７０の間に終了する。

【００２５】

図４は、本発明の特徴を取り入れるクラスタリングルーチン４００の例示的な実現例を示すフローチャートである。上述したように、ステップ３２０の間、クラスタリングルーチン４００は、１つのクラスタにおけるポイント（テレビ番組）が他のいかなるクラスタよりもそのクラスタの平均（重心）に近くなるように第三者視聴履歴１３０（データセッ

ト）をクラスタに分けるために、典型的プロファイルプロセス３００によって呼び出される。概して、クラスタリングルーチンは、サンプルデータセットにおける標本のグループを見つける監督されないタスクに注目する。本発明は、ｋ－ｍｅａｎｓクラスタリングアルゴリズムを使用して、データセットをｋ個のクラスタに分ける。以下に記述されるように、クラスタリングルーチン４００への２つの主なパラメータは、（ｉ）図６に関連して以下に記述される最も近いクラスタを見つけるための距離メトリック及び（ｉｉ）生成すべきクラスタの数ｋである。

【００２６】

標本データの更なるクラスタリングが分類の精度のいかなる改善も与えないときに安定したｋに達するという条件で、例示的なクラスタリングルーチン４００は、ｋの動的な値を使用する。加えて、クラスタサイズは、空のクラスタが記録されるところまでインクリメントされる。こうして、クラスタの自然なレベルに達するとき、クラスタリングは止まる。

【００２７】

図４に示すように、クラスタリングルーチン４００は、まずステップ４１０の間に、ｋ個のクラスタを確立する。例示的なクラスタリングルーチン４００は、クラスタの最小数、例えば2を選択することから始まる。この固定の数について、クラスタリングルーチン４００は、視聴履歴データセット１３０全体を処理し、いくつかの繰り返しを通して、安定していると考えられうる２つのクラスタに達する（すなわちアルゴリズムが別の繰り返しを通るとしても、いかなる番組もあるクラスタから別のクラスタに移動しない）。ステップ４２０の間、現在のｋ個のクラスタは、１つ又は複数の番組により初期化される。

【００２８】

１つの例示的な実現例において、ステップ４２０の間、クラスタは、第三者視聴履歴１３０から選択されるシード番組により初期化される。クラスタを初期化するための番組は、ランダムに又は逐次的に選択されることができる。逐次的な実現例において、クラスタは、視聴履歴１３０における１番目の番組から始まる番組を用いて又は視聴履歴１３０におけるランダムなポイントから始まる番組を用いて初期化されることができる。更に他の変形例において、それぞれのクラスタを初期化する番組の数が変えられてもよい。最後に、クラスタは、第三者視聴履歴１３０における番組からランダムに選択されたフィーチャ値からなる１つ又は複数の「仮定的な（hypothetical）」番組を用いて初期化されてもよい。

【００２９】

そののち、ステップ４３０の間、クラスタリングルーチン４００は、それぞれのクラスタの現在の平均を計算するために、図５に関連して後述される平均計算ルーチン５００を始める。ステップ４４０の間、クラスタリングルーチン４００は、第三者視聴履歴１３０におけるそれぞれの番組の各クラスタに対する距離を決定するために、図６に関連して記述される距離計算ルーチン６００を実行する。ステップ４６０の間、視聴履歴１３０におけるそれぞれの番組は、最も近いクラスタに割り当てられる。

【００３０】

ステップ４７０の間、番組があるクラスタから別のクラスタへ移動したかどうか決定するためのテストが実施される。番組があるクラスタから別のクラスタに移動したとステップ４７０の間に決定される場合、プログラム制御はステップ４３０に戻り、クラスタの安定した組が識別されるまで上述した態様で続く。しかしながら、いかなる番組もあるクラスタから別のクラスタにステップ４７０の間に移動しなかったと決定される場合、プログラム制御はステップ４８０に進む。

【００３１】

ステップ４８０の間、特定された性能基準が満たされたか又は空のクラスタが識別されるか（総称して「ストップ基準」と呼ぶ）を決定するための他のテストが実施される。ステップ４８０の間にストップ基準が満たされていないと決定される場合、ステップ４８５の間にｋの値がインクリメントされ、プログラム制御はステップ４２０に戻り、上述した

態様で続く。しかしながら、ステップ４８０の間にストップ基準が満たされたと決定される場合、プログラム制御が終了する。ストップ基準の評価は、図８に関連して以下に更に記述される。

【００３２】

例示的なクラスタリングルーチン４００は、ただ１つのクラスタに番組を配置し、それゆえクリスプクラスタと呼ばれるものを生成する。他の変形例は、特定の標本（テレビ番組）が部分的に多くのクラスタに属することを可能にするファジークラスタリングを用いる。ファジークラスタリング方法において、テレビ番組は、テレビ番組がクラスタ平均にどれくらい近いかを表す重みを割り当てられる。重みは、クラスタ平均からのテレビ番組の距離の逆二乗に依存しうる。単一のテレビ番組に関連するすべてのクラスタ重みの合計は、１００％にならなければならない。

【００３３】

クラスタのシンボル的な平均の計算

図５は、本発明の特徴を取り入れる平均計算ルーチン５００の例示的な実現例を示すフローチャートである。上述したように、平均計算ルーチン５００は、クラスタのシンボル的な平均を計算するために、クラスタリングルーチン４００によって呼び出される。数値データについて、平均は、分散を最小にする値である。シンボル的なデータに概念を拡げるとき、クラスタの平均は、クラスタ内の分散（それゆえクラスタの半径又は範囲）を最小にするx_μの値を求めることによって規定されることができる。

【数２】

$$\mathrm{Var}(J) = \sum_{i \in J} (x_i - x_\mu)^2 \qquad (1)$$

$$\text{Cluster radius } R(J) = \sqrt{Var(J)} \qquad (2)$$

ここで、Ｊは、同じクラス（視聴クラス又は非視聴クラス）からのテレビ番組のクラスタであり、x_iは、ショー（番組）ｉについてのシンボル的なフィーチャ値であり、x_μは、Ｖａｒ（Ｊ）を最小にするようなＪにおけるテレビ番組の１つからのフィーチャ値である。

【００３４】

こうして、図５に示すように、平均計算ルーチン５００は、まずステップ５１０の間に、所与のクラスタＪにおける現在の番組を識別する。ステップ５２０の間、考慮中の現在のシンボル的な属性について、クラスタＪの分散が、それぞれの可能なシンボル的な値x_μについて方程式（１）を使用して計算される。ステップ５３０の間、分散を最小にするシンボル的な値x_μが平均値として選択される。

【００３５】

ステップ５４０の間、考慮されるべき他のシンボル的な属性があるかを決定するためのテストが実施される。考慮されるべき他のシンボル的な属性があるとステップ５４０の間に決定される場合、プログラム制御はステップ５２０に戻り、上述した態様で続く。しかしながら、考慮されるべき他のシンボル的な属性がないとステップ５４０の間に決定される場合、プログラム制御はクラスタリングルーチン４００に戻る。

【００３６】

計算的に、Ｊにおけるそれぞれのシンボル的なフィーチャ値がx_μとして試され、分散を最小にするシンボル的な値がクラスタＪにおける考慮中のシンボル的な属性の平均になる。可能である平均計算には２つのタイプがあり、すなわちショー（番組）に基づく平均とフィーチャに基づく平均とがある。

【００３７】

フィーチャに基づくシンボル的な平均

ここに記述される例示的な平均計算ルーチン５００はフィーチャに基づく。ここで、シンボル的な属性に関する平均は、その可能な値のうちの１つでなければならないので、結果として得られるクラスタ平均は、クラスタＪにおける標本（番組）から導き出されるフィーチャ値からなる。しかしながら、クラスタ平均は、「仮定的な」テレビ番組でありうる点に留意することが重要である。この仮定的な番組のフィーチャ値は、標本のうちの１つから引き出されるチャネル値（例えばＥＢＣ）及び標本のうちの別のものから引き出されるタイトル値（例えば実際にＥＢＣで放送されることはないＢＢＣワールドニュース）を含むことができる。このように、最小分散を呈するいかなるフィーチャ値も、当該フィーチャの平均を表すために選択される。ステップ５４０の間にすべてのフィーチャ（すなわちシンボル的な属性）が考慮されたと決定されるまで、平均計算ルーチン５００はすべてのフィーチャ位置について繰り返される。こうして結果として得られた仮定的な番組は、クラスタの平均を表すために使用される。

【００３８】

番組に基づくシンボル的な平均

他の変形例では、分散に関する方程式（１）において、x_iは、テレビ番組ｉ自体でありえ、同様に、x_μは、クラスタＪにおける番組の組にわたる分散を最小にするクラスタＪにおける１つ又は複数の番組である。この例では、個々のフィーチャ値でなく番組間の距離が、最小にされるべき関連するメトリックである。加えて、この例において結果として得られる平均は、仮定的な番組でなく、組Ｊからまさに選ばれた番組である。クラスタＪにおけるすべての番組にわたる分散を最小にする、こうして見つけられたクラスタＪにおけるいかなる番組も、クラスタの平均を表すために使用される。

【００３９】

複数の番組を使用するシンボル的な平均

上述の例示的な平均計算ルーチン５００は、（フィーチャ又は番組に基づく実現例のいずれにせよ）それぞれの可能なフィーチャについて単一のフィーチャ値を使用してクラスタの平均を特徴付ける。しかしながら、平均は、もはや当該クラスタの代表的なクラスタ中心ではないので、平均計算の間にそれぞれのフィーチャについてただ１つのフィーチャ値に依存することは、多くの場合、不適当なクラスタリングをもたらす。言い換えると、ただ１つの番組によって、クラスタを表すことは望ましくないことがあり、むしろ、１つ又は複数の平均を表す複数の番組が、クラスタを表すために使用されることができる。こうして、他の変形例において、クラスタは、それぞれの可能なフィーチャに関する複数の平均又は複数のフィーチャ値によって表現されることができる。こうして、分散を最小にするＮ個のフィーチャ（フィーチャに基づくシンボル的な平均の場合）又はＮ個の番組（番組に基づくシンボル的な平均の場合）が、ステップ５３０の間に選択される。ここで、Ｎは、クラスタの平均を表すために使用される番組の数である。

【００４０】

番組とクラスタと間の距離計算

前述のように、所与のテレビ番組と所与のクラスタの平均と間の距離に基づいて、それぞれのクラスタに対するテレビ番組の近さを評価するために、距離計算ルーチン６００がクラスタリングルーチン４００によって呼び出される。計算された距離メトリックは、クラスタの範囲を決定するためにサンプルデータセットにおけるさまざまな標本間の区別を定量化する。ユーザプロファイルをクラスタ分けすることができるようにするために、視聴履歴における任意の２つのテレビ番組間の距離が計算されなければならない。概して、互いに近いテレビ番組は、１つのクラスタに入る傾向がある。数値的に表されるベクトル間の距離を計算するために多くのかなり直接的な技法が存在し、例えばユークリッド距離、マンハッタン距離及びマハラノビス距離がある。

【００４１】

しかしながら、テレビ番組は主としてシンボル的なフィーチャ値からなるので、既存の距離計算技法はテレビ番組ベクトルの場合には使用されることができない。例えば、２００１年３月２２日午後８時にＥＢＣで放送される「Ｆｉｅｎｄｓ」の一話及び２００１年

３月２５日午後８時にＦＥＸで放送される「Ｔｈｅ　Ｓｉｍｏｎｓ」の一話のような２つのテレビ番組は、以下のフィーチャベクトルを使用して表現されることができる。

タイトル：Ｆｉｅｎｄｓ　　タイトル：Ｓｉｍｏｎｓ
チャネル：ＥＢＣ　　チャネル：ＦＥＸ
放送日：２００１－０３－２２　　放送日：２００１－０３－２５
放送時間：２０００　　放送時間：２０００

【００４２】

明らかに、既知の数値距離メトリックは、フィーチャ値「ＥＢＣ」と「ＦＥＸ」との間の距離を計算するために使用されることができない。バリューディファレンスメトリック（ＶＤＭ）は、シンボル的なフィーチャで表されるドメインにおけるフィーチャの値の間の距離を測るための既存の技法である。ＶＤＭ技法は、それぞれのフィーチャのそれぞれの取り得る値についてすべての実例の分類（classification）の全体的な類似性を考慮する。この方法を使用して、フィーチャのすべての値の間の距離を規定するマトリクスが、トレーニングセットにおける標本に基づいて統計的に導き出される。シンボル的なフィーチャ値の間の距離を計算するためのＶＤＭ技法に関するより詳細な説明のため、例えばStanfill及びWaltz、「Toward Memory-Based Reasoning」、Communications of the ACM、29:12、1213-1228、1986年、を参照されたい。この内容は、参照によってここに盛り込まれる。

【００４３】

本発明は、２つのテレビ番組間又は興味のある他のアイテム間のフィーチャ値の間の距離を計算するためにＶＤＭ技法又はその変形を用いる。本来のＶＤＭ提案は、２つのフィーチャ値の間の距離計算において重みタームを使用し、これは、距離メトリックを非対照にする。変更されたＶＤＭ（ＭＶＤＭ）は、距離マトリックスを対称にするために重みタームを省く。シンボル的なフィーチャ値の間の距離を計算するためのＭＶＤＭ技法のより詳細な説明のため、Cost及びSalzberg、「A Weighted Nearest Neighbor Algorithm For Learning With Symbolic Features」、Machine Learning、Vol. 10、57-58、ボストン、マサチューセッツ、Kluwer Publishers、1993年、を参照されたい。この内容は、参照によってここに盛り込まれる。

【００４４】

ＭＶＤＭに従って、特定のフィーチャに関する２つの値Ｖ１とＶ２と間の距離δが、下式によって与えられる。

δ（Ｖ１，Ｖ２）＝　Σ｜Ｃ１ｉ／Ｃ１－Ｃ２ｉ／Ｃ２｜^{r}　　（３）

【００４５】

本発明の番組推薦環境において、ＭＶＤＭ方程式（３）は、特に「視聴」及び「非視聴」クラスを扱うために変形される。

【数３】

$$\delta(V1,V2)=\left|\frac{C1\_watched}{C1\_total}-\frac{C2\_watched}{C2\_total}\right|+\left|\frac{C1\_not\_watched}{C1\_total}-\frac{C2\_not\_watched}{C2\_total}\right| \quad (4)$$

【００４６】

方程式（４）において、Ｖ１及びＶ２は、考慮中のフィーチャに関する２つの可能な値である。上述の例を続けると、フィーチャ「チャネル」に関して、第１の値Ｖ１は「ＥＢＣ」に等しく、第２の値Ｖ２は「ＦＥＸ」に等しい。これらの値の間の距離は、標本が分類されるすべてのクラスにわたる合計である。本発明の例示的な番組推薦器の実施例の場合の関連するクラスは、「視聴」クラス及び「非視聴」クラスである。Ｃ１ｉは、Ｖ１（

ＥＢＣ）がクラスｉに分類された回数であり（１に等しいｉは、視聴クラスを意味する）、Ｃ１（Ｃ１＿ｔｏｔａｌ）は、Ｖ１がデータセットにおいて出現した合計回数である。値「ｒ」は、定数であり、通常１にセットされる。

【００４７】

方程式（４）によって規定されるメトリックは、値がすべての分類について同じ相対頻度で生じる場合、それらの値を類似するものとして識別する。タームＣ１ｉ／Ｃ１は、当該フィーチャが値Ｖ１を有すると仮定して、その中央残余（central residue）がｉとして分類されるであろう尤度を表現する。こうして、２つの値は、それらがすべての可能な分類について類似の尤度を与える場合には類似する。方程式（４）は、すべての分類にわたるこれらの尤度の差の合計を求めることによって、２つの値の間の全体的な類似性を計算する。２つのテレビ番組間の距離は、２つのテレビ番組ベクトルの対応するフィーチャ値間の距離の合計である。

【００４８】

図７Ａは、フィーチャ「チャネル」に関連するフィーチャ値に関する距離テーブルの一部である。図７Ａは、それぞれのクラスについてそれぞれのチャネルフィーチャ値の出現の数を含む。図７Ａに示される値は、例示的な第三者視聴履歴１３０から取り出されたものである。

【００４９】

図７Ｂは、ＭＶＤＭ方程式（４）を使用して図７Ａに示される例示的なカウントから計算されるそれぞれのフィーチャ値対の間の距離を示す。直観的に、ＥＢＣ及びＡＢＳは、それらが大部分は視聴クラスに出現し、非視聴クラスに出現しない（ＡＢＳは小さい非視聴成分を有する）ので、互いに「近い」とすべきである。図７Ｂは、ＥＢＣとＡＢＳとの間の小さい（ゼロでない）距離によりこの直感的事実を確認する。他方、ＡＳＰＮは、大部分は非視聴クラスに出現し、それゆえこのデータセットについてはＥＢＣ及びＡＢＳのどちらに対しても「遠い」とすべきである。図７Ｂは、最大可能距離２．０のところ、ＥＢＣとＡＳＰＮとの間の距離１．８９５を含む。同様に、ＡＢＳとＡＳＰＮと間の距離は高く、値１．８２８をもつ。

【００５０】

こうして、図６に示すように、まずステップ６１０の間、距離計算ルーチン６００は、第三者視聴履歴１３０における番組を識別する。ステップ６２０の間、距離計算ルーチン６００は、考慮中の現在番組について、（平均計算ルーチン５００により決定された）それぞれのクラスタ平均の対応するフィーチャに対するそれぞれのシンボル的なフィーチャ値の距離を計算するために方程式（４）を使用する。

【００５１】

ステップ６３０の間、現在番組とクラスタ平均と間の距離が、対応するフィーチャ値の間の距離を総計することによって計算される。ステップ６４０の間、考慮されるべき他の番組が第三者視聴履歴１３０にあるかどうか決定するためのテストが実施される。ステップ６４０の間に考慮されるべき他の番組が第三者視聴履歴１３０にあると決定される場合、次の番組がステップ６５０の間に識別され、プログラム制御はステップ６２０に進み、上述した態様で続く。

【００５２】

しかしながら、ステップ６４０の間、考慮されるべき他の番組が第三者視聴履歴１３０にないと決定される場合、プログラム制御はクラスタリングルーチン４００へ戻る。

【００５３】

「複数の番組から導き出されるシンボル的な平均」という題目のサブセクションにおいて上述したように、クラスタの平均は、（フィーチャ又は番組に基づく実現例のいずれにせよ）それぞれの可能なフィーチャについて複数のフィーチャ値を使用して特徴付けられることができる。複数の平均からの結果は、投票によるコンセンサス決定に達するために距離計算ルーチン６００の変形によってプールされる。例えば、ステップ６２０の間、番組の所与のフィーチャ値と、さまざまな平均についての各々の対応するフィーチャ値との

間で、距離が計算される。最小距離結果はプールされ、例えばコンセンサス決定に達するために多数決又は混合エキスパート（mixture of experts）を用いることによって、投票のために使用される。このような技法に関するより詳細な説明のため、例えば、J.Kittler他、「Combing Classifiers」、Proc. of the 13th Int'l Conf. on Pattern Recognition、 Vol. II、897-901、ウィーン、オーストリア、1996年、を参照されたい。その内容は参照によりここに盛り込まれる。

【００５４】

ストップ基準

上述したように、クラスタリングルーチン４００は、クラスタを生成するためのストップ基準が満たされたときを決定するために、図８に示されるクラスタリング性能評価ルーチン８００を呼び出す。標本データの更なるクラスタリングが分類精度のいかなる改善も与えないときに安定したｋに達するという条件で、例示的なクラスタリングルーチン４００は、ｋの動的な値を用いる。加えて、クラスタサイズは、空のクラスタが記録されるところまでインクリメントされることができる。こうして、クラスタの自然なレベルに達するときに、クラスタリングがストップする。

【００５５】

例示的なクラスタリング性能評価ルーチン８００は、クラスタリングルーチン４００の分類精度をテストするために、第三者視聴履歴１３０からの番組のサブセット（テストデータセット）を使用する。テストセットにおけるそれぞれの番組について、クラスタリング性能評価ルーチン８００は、当該番組に最も近いクラスタ（そのクラスタ平均が最も近い）を決定し、そのクラスタに関するクラスラベルと考慮中の番組とを比較する。合致したクラスラベルのパーセンテージは、クラスタリングルーチン４００の精度につながる。

【００５６】

こうして、図８に示すようにクラスタリング性能評価ルーチン８００は、まずステップ８１０の間に、テストデータセットの役目を果たすための番組のサブセットを第三者視聴履歴１３０から集める。そののち、ステップ８２０の間、クラスタにおける視聴される及び視聴されない番組のパーセンテージに基づいて、クラスラベルがそれぞれのクラスタに割り当てられる。例えばクラスタにおける大部分の番組が視聴されるものである場合、そのクラスタには「視聴」ラベルが割り当てられることができる。

【００５７】

テストセットにおけるそれぞれの番組に最も近いクラスタが、ステップ８３０の間に識別され、割り当てられたクラスタについてのクラスラベルは、番組が実際に視聴されたか否かと比較される。複数の番組がクラスタの平均を表示するために使用される実現例において、（それぞれの番組に対する）平均的な距離又は投票スキームが用いられることができる。プログラム制御がクラスタリングルーチン４００へ戻る前に、合致したクラスラベルのパーセンテージが、ステップ８４０の間に決定される。分類精度があらかじめ規定された閾値に達した場合、クラスタリングルーチン４００が終了する。

【００５８】

図示されると共にここに記述される実施例及び変形例は、単に本発明の原理を説明するだけであり、さまざまな変更が、本発明の範囲及び精神から逸脱することなく当業者により実現されうることが理解されるべきである。

【図面の簡単な説明】

【００５９】

【図１】本発明によるテレビ番組推薦器の概略ブロック図。

【図２】図１の例示的な番組データベースからのサンプルテーブルを示す図。

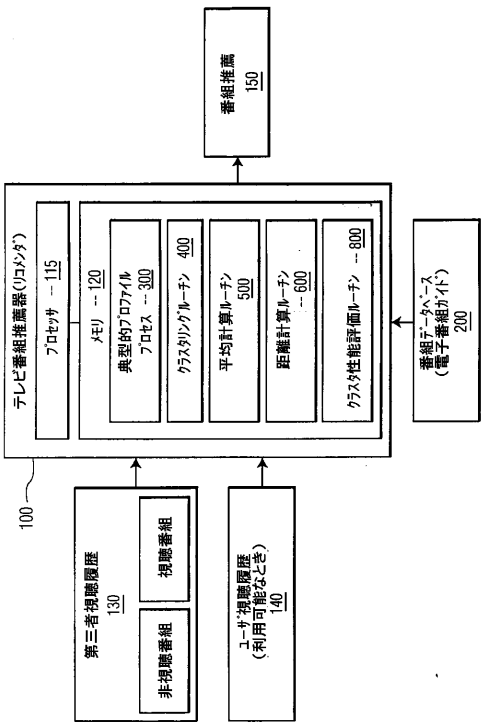
【図３】本発明の原理を具体化する図１の典型的プロファイルプロセスを示すフローチャート。

【図４】本発明の原理を具体化する図１のクラスタリングルーチンを示すフローチャート。

【図５】本発明の原理を具体化する図１の平均計算ルーチンを示すフローチャート。

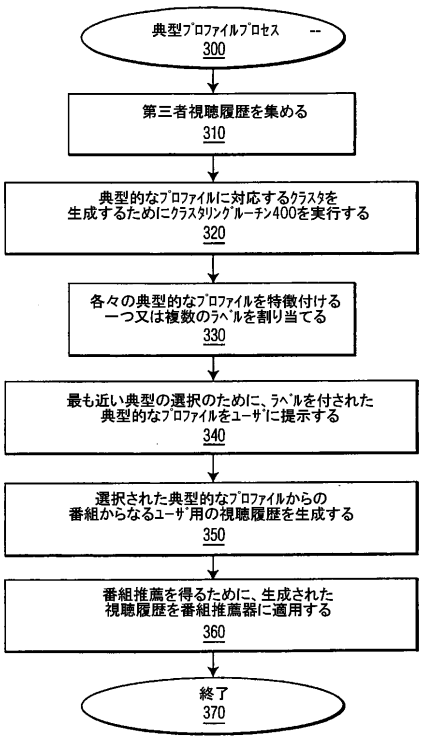
【図６】本発明の原理を具体化する図１の距離計算ルーチンを示すフローチャート。
【図７Ａ】それぞれのクラスについて、それぞれのチャネルフィーチャ値の出現の数を示す例示的なチャネルフィーチャ値出現テーブルからのサンプルテーブルを示す図。
【図７Ｂ】図７Ａに示される例示的なカウントから計算されるそれぞれのフィーチャ値対の間の距離を示す例示的なフィーチャ値対距離テーブルからのサンプルテーブルを示す図。
【図８】本発明の原理を具体化する図１のクラスタリング性能評価ルーチンを示すフローチャート。

【図１】

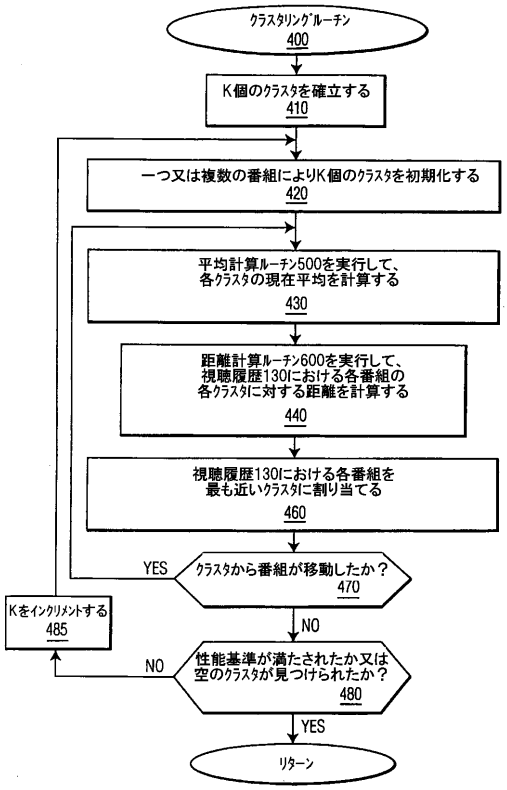


【図２】

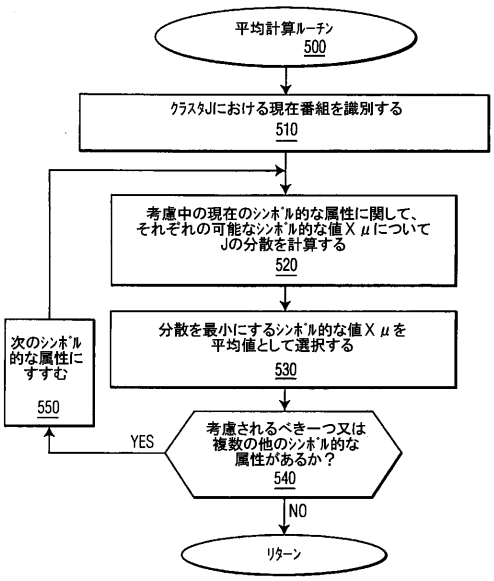
【図３】


典型ﾌﾟﾛﾌｧｲﾙﾌﾟﾛｾｽ
300
第三者視聴履歴を集める
310
典型的なﾌﾟﾛﾌｧｲﾙに対応するｸﾗｽﾀを
生成するためにｸﾗｽﾀﾘﾝｸﾞﾙｰﾁﾝ400を実行する
320
各々の典型的なﾌﾟﾛﾌｧｲﾙを特徴付ける
一つ又は複数のﾗﾍﾞﾙを割り当てる
330
最も近い典型の選択のために、ﾗﾍﾞﾙを付された
典型的なﾌﾟﾛﾌｧｲﾙをﾕｰｻﾞに提示する
340
選択された典型的なﾌﾟﾛﾌｧｲﾙからの
番組からなるﾕｰｻﾞ用の視聴履歴を生成する
350
番組推薦を得るために、生成された
視聴履歴を番組推薦器に適用する
360
終了
370

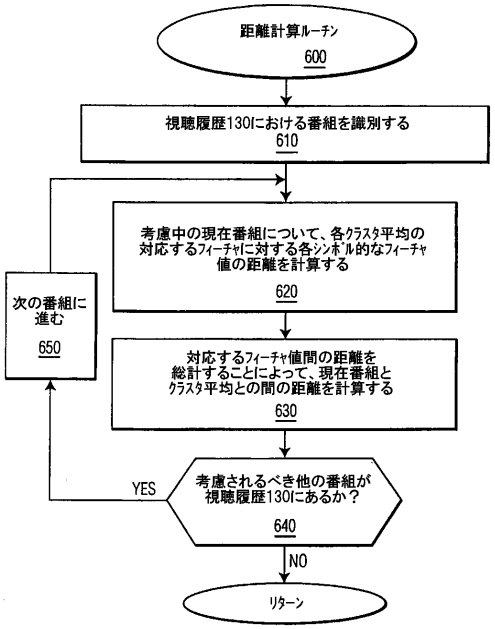
【図４】


ｸﾗｽﾀﾘﾝｸﾞﾙｰﾁﾝ
400
K個のｸﾗｽﾀを確立する
410
一つ又は複数の番組によりK個のｸﾗｽﾀを初期化する
420
平均計算ﾙｰﾁﾝ500を実行して、
各ｸﾗｽﾀの現在平均を計算する
430
距離計算ﾙｰﾁﾝ600を実行して、
視聴履歴130Iにおける各番組の
各ｸﾗｽﾀに対する距離を計算する
440
視聴履歴130Iにおける各番組を
最も近いｸﾗｽﾀに割り当てる
460
YES
ｸﾗｽﾀから番組が移動したか？
470
NO
性能基準が満たされたか又は
空のｸﾗｽﾀが見つけられたか？
480
NO
YES
Kをｲﾝｸﾘﾒﾝﾄする
485
ﾘﾀｰﾝ

【図５】


平均計算ﾙｰﾁﾝ
500
ｸﾗｽﾀJにおける現在番組を識別する
510
考慮中の現在のｼﾝﾎﾞﾙ的な属性に関して、
それぞれの可能なｼﾝﾎﾞﾙ的な値Xμについて
Jの分散を計算する
520
分散を最小にするｼﾝﾎﾞﾙ的な値Xμを
平均値として選択する
530
考慮されるべき一つ又は
複数の他のｼﾝﾎﾞﾙ的な
属性があるか？
540
YES
次のｼﾝﾎﾞﾙ
的な属性に
すすむ
550
NO
ﾘﾀｰﾝ

【図６】


距離計算ﾙｰﾁﾝ
600
視聴履歴130Iにおける番組を識別する
610
考慮中の現在番組について、各ｸﾗｽﾀ平均の
対応するﾌｨｰﾁｬに対する各ｼﾝﾎﾞﾙ的なﾌｨｰﾁｬ
値の距離を計算する
620
対応するﾌｨｰﾁｬ値間の距離を
総計することによって、現在番組と
ｸﾗｽﾀ平均との間の距離を計算する
630
考慮されるべき他の番組が
視聴履歴130Iにあるか？
640
YES
次の番組に
進む
650
NO
ﾘﾀｰﾝ

【図７Ａ】

チャネルフィーチャ値出現テーブル -- 700

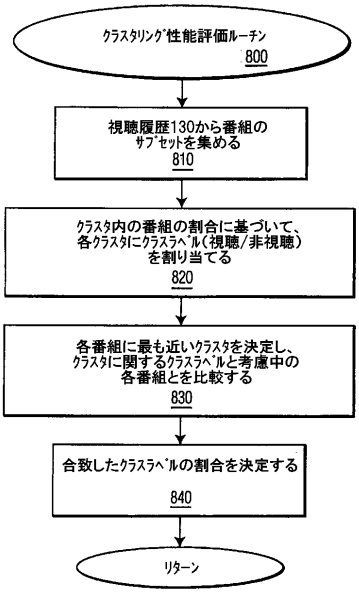
	クラス	
フィーチャ値	視聴	非視聴
EBC	353	0
ASPN	1	18
ABS	145	5

【図７Ｂ】

フィーチャ値対の距離のテーブル -- 750

	EBC	ASPN	ABS
EBC	0	1.895	0.066
ASPN	1.895	0	1.828
ABS	0.066	1.828	0

【図８】



フロントページの続き

(72)発明者 クラパティ カウシャル
オランダ国 ５６５６ アーアー アインドーフェン プロフ ホルストラーン ６

合議体
審判長 山崎 達也
審判官 原 秀人
審判官 田中 秀人

(56)参考文献 特開２００１－３０６６１２（ＪＰ，Ａ）
特開２００１－０７６００２（ＪＰ，Ａ）
特開２００１－０４３２３３（ＪＰ，Ａ）
特開２０００－２９３５３１（ＪＰ，Ａ）
河村晃好，外３名，「グループ嗜好モデルと視聴履歴を利用したコンテンツ検索サーバの試作」，電子情報通信学会技術研究報告，日本，社団法人電子情報通信学会，平成１３年7月１１日，第１０１巻 第１９２号，ｐ．９－１６
渡部勇，「緩い協調：協調情報フィルタリングシステム」，情報処理学会研究報告，日本，社団法人情報処理学会，平成3年3月8日発行，第９１巻 第１８号，ｐ．１７９－１８６
Ｓｃｏｔｔ Ｃｏｓｔ，Ｓｔｅｖｅｎ Ｓａｌｚｂｅｒｇ，“Ａ Ｗｅｉｇｈｔｅｄ Ｎｅａｒｅｓｔ Ｎｅｉｇｈｂｏｒ Ａｌｇｏｒｉｔｈｍ ｆｏｒ Ｌｅａｒｎｉｎｇ ｗｉｔｈ Ｓｙｍｂｏｌｉｃ Ｆｅａｔｕｒｅｓ”，Ｍａｃｈｉｎｅ Ｌｅａｒｎｉｎｇ，米国，Ｋｌｕｗｅｒ Ａｃａｄｅｍｉｃ Ｐｕｂｌｉｓｈｅｒｓ，１９９３年1月，Ｖｏｌｕｍｅ１０ ，Ｉｓｓｕｅ １，Ｐａｇｅｓ：５７－７８

(58)調査した分野(Int.Cl.，ＤＢ名)
G06F 17/30