

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6745343号
(P6745343)

(45) 発行日 令和2年8月26日(2020.8.26)

(24) 登録日 令和2年8月5日(2020.8.5)

(51) Int.Cl.

G06Q 30/02 (2012.01)

F 1

G06Q 30/02 312

請求項の数 14 (全 19 頁)

(21) 出願番号 特願2018-535405 (P2018-535405)
 (86) (22) 出願日 平成28年11月15日 (2016.11.15)
 (65) 公表番号 特表2019-501464 (P2019-501464A)
 (43) 公表日 平成31年1月17日 (2019.1.17)
 (86) 國際出願番号 PCT/US2016/062032
 (87) 國際公開番号 WO2017/119952
 (87) 國際公開日 平成29年7月13日 (2017.7.13)
 審査請求日 令和1年11月13日 (2019.11.13)
 (31) 優先権主張番号 14/990,834
 (32) 優先日 平成28年1月8日 (2016.1.8)
 (33) 優先権主張国・地域又は機関
米国(US)

(73) 特許権者 502303739
オラクル・インターナショナル・コーポレーション
アメリカ合衆国カリフォルニア州94065レッドウッド・シティー、オラクル・パーカウェイ500
(74) 代理人 110001195
特許業務法人深見特許事務所
(72) 発明者 ウー, スーミン
アメリカ合衆国、02453 マサチューセツ州、ウォルサム、バーノン・ストリート、21・エイ

最終頁に続く

(54) 【発明の名称】顧客意思決定ツリー生成システム

(57) 【特許請求の範囲】

【請求項 1】

コンピュータによって実現され1つ以上のプロセッサによって実行される、顧客意思決定ツリー(CDT)を商品カテゴリーに対して生成する方法であって、前記商品カテゴリーは複数の商品属性を含み、各商品属性は対応する複数の属性値を有し、前記方法は、小売アイテム取引売上データを受け付けるステップと、

前記売上データを全ての顧客にわたってアイテム／店舗／期間レベルで集計することにより、対応する店舗識別子(ID)について、対応する商品IDと前記期間内の前記商品IDの売上数量とを含む第1の集計データを生成するステップと、

前記売上データを全ての顧客にわたって属性値／店舗／期間レベルで集計することにより、対応する店舗IDについて前記複数の属性値の各々に対する前記期間内の当該属性値の売上数量を含む第2の集計データを生成するステップと、

前記第1の集計データおよび前記第2の集計データを用いて、前記期間の第1の売上シェアおよび第2の売上シェアを判定するステップとを含み、

前記第1の売上シェアは、商品属性の他の全ての属性値に対する当該商品属性の各属性値についての売上の割合を含み、前記第2の売上シェアは、前記期間内に販売された他の全ての商品に対する各商品の売上の割合を含み、前記方法は、さらに、

前記第1の売上シェアを用いて商品属性ごとに全ての可能な属性値ペア間の相関値を計算することにより、属性値ペアについての類似度を判定するステップと、

前記第2の売上シェアを用いて全ての可能な商品ペア間の相関値を計算することにより

10

20

、商品ペアについての類似度を判定するステップと、

各商品属性ごとに、前記属性値ペアについての類似度と前記商品ペアについての類似度との相関値を計算して最も高い相関値を有する商品属性を決定することにより、最上位属性を判定するステップとを含む、方法。

【請求項 2】

前記期間は、週単位を含む、請求項 1 に記載の方法。

【請求項 3】

バイナリ属性についての類似度を判定するステップをさらに含む、請求項 1 または 2 に記載の方法。

【請求項 4】

正の値を 0 にするステップと負の値を対応する正の値に変更するステップとを含む、前記属性値ペアについての類似度および前記商品ペアについての類似度を後処理するステップをさらに含む、請求項 1 ~ 3 のいずれか 1 項に記載の方法。

【請求項 5】

前記属性値ペアについての類似度を判定するステップは、属性値ペア (X, Y) について、 X_i および Y_i は、前記属性 X および前記属性 Y の店舗 / 時間シェア値を表し、n は、X および Y の属性シェアがある店舗 / 期間の総数を表す、

【数 1】

$$SIM(X, Y) = \frac{\sum_{i=1}^n X_i Y_i - \frac{(\sum_{i=1}^n X_i)(\sum_{i=1}^n Y_i)}{n}}{\sqrt{\left(\sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n} \right) \left(\sum_{i=1}^n Y_i^2 - \frac{(\sum_{i=1}^n Y_i)^2}{n} \right)}}$$

10

20

から構成される SIM の値を判定するステップを含む、請求項 1 ~ 4 のいずれか 1 項に記載の方法。

【請求項 6】

前記バイナリ属性についての類似度を判定するステップは、

【数 2】

x_k は、期間 k におけるオーガニックシェアであり、N 個の期間があり、 \bar{x} は、前記 x_i の平均である、

$$2 \sqrt{\frac{\sum_{k=1}^N (x_k - \bar{x})^2}{N}}$$

30

を含む、請求項 3 に記載の方法

【請求項 7】

前記最上位属性を前記 CDT の第 1 レベルとして割り当てるステップと、

前記 CDT の第 2 レベルを、各々が前記最上位属性の属性値に対応する複数の下位区分に分割するステップと、

前記下位区分ごとに、前記下位区分値について、

前記小売アイテム取引売上データを受け付けるステップ、

前記売上データを前記アイテム / 店舗 / 期間レベルで集計するステップ、

前記売上データを前記属性値 / 店舗 / 期間レベルで集計するステップ、

前記期間の第 1 の売上シェアおよび第 2 の売上シェアを判定するステップ、

属性値ペア間の相関値を計算することにより、属性値ペアについての類似度を判定するステップ、

40

50

商品ペア間の相関値を計算することにより、商品ペアについての類似度を判定するステップ、および

前記属性値ペアについての類似度および前記商品ペアについての類似度に基づいて、前記最上位属性を判定するステップ、

を繰り返すステップとをさらに含む、請求項 1 ~ 6 のいずれか 1 項に記載の方法。

【請求項 8】

顧客意思決定ツリー（ C D T ）を商品カテゴリーごとに生成する C D T 生成システムであって、前記商品カテゴリーは複数の商品属性を含み、各商品属性は対応する複数の属性値を有し、

前記システムは、

集成モジュールと、

類似度モジュールとを備え、

前記集成モジュールは、小売アイテム取引売上データを受け付けることに応答して、前記売上データを全ての顧客にわたってアイテム / 店舗 / 期間レベルで集成することにより、対応する店舗識別子（ I D ）について、対応する商品 I D と前記期間内の前記商品 I D の売上数量とを含む第 1 の集成データを生成し、前記集成モジュールは、前記売上データを全ての顧客にわたって属性値 / 店舗 / 期間レベルで集成することにより、対応する店舗 I D について前記複数の属性値の各々に対する前記期間内の当該属性値の売上数量を含む第 2 の集成データを生成し、

前記類似度モジュールは、前記第 1 の集成データおよび前記第 2 の集成データを用いて前記期間の第 1 の売上シェアおよび第 2 の売上シェアを判定し、前記第 1 の売上シェアは、商品属性の他の全ての属性値に対する当該商品属性の各属性値についての売上の割合を含み、前記第 2 の売上シェアは、前記期間内に販売された他の全ての商品に対する各商品の売上の割合を含み、前記類似度モジュールは、前記第 1 の売上シェアを用いて商品属性ごとに全ての可能な属性値ペア間の相関値を計算することにより、属性値ペアについての類似度を判定し、前記類似度モジュールは、前記第 2 の売上シェアを用いて全ての可能な商品ペア間の相関値を計算することにより、商品ペアについての類似度を判定し、前記類似度モジュールは、各商品属性ごとに、前記属性値ペアについての類似度と前記商品ペアについての類似度との相関値を計算して最も高い相関値を有する商品属性を決定することにより、最上位属性を判定する、顧客意思決定ツリー（ C D T ）生成システム。

【請求項 9】

前記属性値ペアについての類似度を判定することは、属性値ペア（ X , Y ）について、 X_i および Y_i は、前記属性 X および前記属性 Y の店舗 / 時間シェア値を表し、 n は、 X および Y の属性シェアがある店舗 / 期間の総数を表す、

【数 3】

$$SIM(X, Y) = \frac{\sum_{i=1}^n X_i Y_i - \frac{(\sum_{i=1}^n X_i)(\sum_{i=1}^n Y_i)}{n}}{\sqrt{\left(\sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n} \right) \left(\sum_{i=1}^n Y_i^2 - \frac{(\sum_{i=1}^n Y_i)^2}{n} \right)}}$$

から構成される S I M の値を判定することを含む、請求項 8 に記載のシステム。

【請求項 10】

前記類似度モジュールは、さらに、

10

20

30

40

【数4】

x_k は、期間kにおけるオーガニックシェアであり、N個の期間があり、 \bar{x} は、前記 x_i の平均である、

$$2 \sqrt{\frac{\sum_{k=1}^N (x_k - \bar{x})^2}{N}}$$

から構成される、バイナリ属性についての類似度を判定する、請求項8または9に記載のシステム。

10

【請求項11】

前記期間は、週単位を含む、請求項8～10のいずれか1項に記載のシステム。

【請求項12】

前記類似度モジュールは、さらに、正の値を0にすることと負の値を対応する正の値に変更することとを含む、前記前記属性値ペアについての類似度および前記商品ペアについて類似度の後処理を行う、請求項8～11のいずれか1項に記載のシステム。

【請求項13】

レベル生成モジュールをさらに備え、

前記レベル生成モジュールは、

前記最上位属性を前記CDTの第1レベルとして割り当て、

20

前記CDTの第2レベルを、各々が前記最上位属性の属性値に対応する複数の下位区分に分割し、

前記下位区分ごとに、前記下位区分値について、

前記小売アイテム取引売上データを受け付けることと、

前記売上データを前記アイテム／店舗／期間レベルに集成することと、

前記売上データを前記属性値／店舗／期間レベルに集成することと、

前記期間の第1の売上シェアおよび第2の売上シェアを判定することと、

属性値ペア間の相関値を計算することにより、属性値ペアについての類似度を判定することと、

商品ペア間の相関値を計算することにより、商品ペアについての類似度を判定すること、

30

前記属性値ペアについての類似度および前記商品ペアについての類似度に基づいて、前記最上位属性を判定することとを繰り返す、請求項8～12のいずれか1項に記載のシステム。

【請求項14】

命令を含むコンピュータプログラムであって、前記命令は、1つ以上のプロセッサによって実行されると、前記1つ以上のプロセッサに請求項1～7のいずれか1項に記載の方法を実行させる、コンピュータプログラム。

【発明の詳細な説明】**【技術分野】**

40

【0001】

分野

一実施形態は、全体的に、コンピュータシステムを対象とし、特に、顧客意思決定ツリーを生成するコンピュータシステムを対象とする。

【背景技術】**【0002】**

背景情報

購買意思決定プロセスは、商品またはサービスの購入前、購入中、および購入後に消費者が行う潜在的な市場取引についての意思決定プロセスである。より一般的には、意思決定は、複数の選択肢の中から行動方針を選択する認知過程である。よくある例に、買い物

50

や、何を食べるかを決めることがある。

【0003】

一般に、消費者の購買意思決定を分析する3つの方法がある。（1）経済的モデル。このモデルは、非常に定量的であり、合理性の仮定およびほぼ完全な知識に基づく。消費者は、自身の効用を最重要視すると見られる。（2）心理的モデル。このモデルは、モチベーションおよびニーズの認識など、心理的および認知過程に重点を置く。このモデルは、定量的というよりむしろ定性的であり、文化的影響および家族の影響のような社会学的因素に基づく。（3）顧客行動モデル。これは、マーケターが利用する実用的なモデルである。このモデルは、通常、経済的モデルと心理的モデルとを組み合わせたモデルである。

【0004】

顧客行動モデルの1つのタイプが「顧客意思決定ツリー」（「CDT」）として知られている。CDTは、所定のカテゴリーのアイテムの購入に関する商品属性空間における、顧客の意思決定の階層を図に表したものである。CDTは、顧客が、欲しいアイテムに絞り込む前に、あるカテゴリー内の互いに異なる選択肢（属性に基づく）についてどう考えるかをモデル化し、顧客の購買意思決定を理解するのに役立つ。また、これは、「商品セグメンテーションおよびカテゴリー構造」としても一般的に知られている。CDTは、従来、市場調査のサーベイおよびその他のツールに基づいて、ブランドメーカーまたは第三者の市場調査会社によって作成される。しかしながら、これらの方法は、正確さに欠け、且つ、ブランドメーカーが提供するバイアスのかかったデータに基づくことがあるため、信憑性に欠ける可能性がある。

10

【発明の概要】

【課題を解決するための手段】

【0005】

概要

一実施形態は、顧客意思決定ツリーを生成するシステムである。システムは、小売アイテム取引売上データを受け付ける。システムは、売上データをアイテム／店舗／期間レベルに集成し、売上データを属性値／店舗／期間レベルに集成する。システムは、期間の売上シェアを判定し、属性値ペア間の相関関係に基づいて、属性値ペアについての類似度を判定する。次に、システムは、判定された類似度に基づいて、最上位属性を判定する。

【図面の簡単な説明】

30

【0006】

【図1】本発明の実施形態に係る、コンピュータサーバ／システムのブロック図である。

【図2】一実施形態に係る、小売店の取引データに基づいて自動的に生成されるヨーグルト商品カテゴリーについての例示的なCDTを示す図である。

【図3】一実施形態に係るCDTを生成する際の図1のCDT生成モジュールの機能のフロー図である。

【図4】一実施形態に係る、類似度を判定する際の図1のCDT生成モジュールの機能のフロー図である。

【図5】一実施形態に係る、類似度に基づいてCDTを生成する際の図1のCDT生成モジュールの機能のフロー図である。

40

【図6】一実施形態に係る、CDT生成モジュールが生成するCDTを示す図である。

【発明を実施するための形態】

【0007】

詳細な説明

一実施形態は、アイテムの類似度を判定するために、小売店の取引データ、具体的には、アイテム　店舗　週を集成した売上数量データを利用して、顧客意思決定ツリー（「CDT」）を自動的に生成する。そのため、CDTを生成するために、ポイントプログラムを利用していない小さな小売店も入手可能な取引データを使用することができる。さらに、実施形態は、小売店のどのアイテム同士が1つのカテゴリーにおいて同類であるかについての判定を行う。

50

【0008】

図1は、本発明の実施形態に係る、コンピュータサーバ／システム10のブロック図である。1つのシステムとして図示しているが、システム10の機能は、分散システムとして実装することができる。さらに、本明細書に開示の機能は、互いにネットワークを通じて接続され得る別々のサーバまたは装置上に実装することができる。さらに、システム10の1つ以上の構成要素が備えられなくてもよい。たとえば、サーバの機能について、システム10は、プロセッサおよびメモリを備える必要があつてもよいが、キーボードまたはディスプレイなど、図1に示すその他の構成要素のうちの1つ以上は備えなくてもよい。

【0009】

システム10は、情報を通信するためのバス12またはその他の通信機構と、情報を処理するための、バス12に接続されたプロセッサ22とを含む。プロセッサ22は、任意の種類の汎用プロセッサまたは特定用途向けプロセッサであつてもよい。システム10は、さらに、プロセッサ22によって実行される情報および命令を格納するためのメモリ14を備える。メモリ14は、Random Access Memory（「RAM」）、Read Only Memory（「ROM」）、磁気もしくは光ディスク、またはその他の種類のコンピュータ読み取り可能な媒体などの静的記憶装置、のうちの任意の組み合わせから構成できる。システム10は、さらに、ネットワークにアクセスできるようにするために、ネットワークインターフェースカードなどの通信装置20を含む。そのため、ユーザは、システム10と直接インターフェース接続する、またはネットワークもしくは任意のその他の方法を通じて間接的にインターフェース接続することができる。

【0010】

コンピュータ読み取り可能な媒体は、プロセッサ22がアクセスできる任意の入手可能な媒体であつてもよく、揮発性および不揮発性媒体の両方、リムーバブルおよびソリッドステートメディア、ならびに通信媒体を含む。通信媒体は、コンピュータ読み取り可能な命令、データ構造、プログラムモジュール、または搬送波もしくはその他の移送機構などの変調データ信号に含まれるその他のデータを含み得、任意の情報配信媒体を含む。

【0011】

プロセッサ22は、さらに、バス12を介して液晶ディスプレイ（「LCD」）などのディスプレイ24に接続される。さらに、ユーザがシステム10とインターフェース接続できるようにするために、キーボード26、およびコンピュータマウスなどのカーソル制御デバイス28がバス12に接続される。

【0012】

一実施形態において、メモリ14は、プロセッサ22によって実行されたときに機能を提供するソフトウェアモジュールを格納する。モジュールは、システム10のためのオペレーティングシステム機能を提供するオペレーティングシステム15を含む。モジュールは、さらに、小売店の消費者データからCDTを自動的に生成する顧客意思決定ツリー生成モジュール16、および本明細書に開示のすべてのその他の機能を含む。システム10は、より大きなシステムの一部であることができる。そのため、システム10は、追加機能を含むために、小売管理システム（たとえば、「オラクル社製のOracle Retail Merchandising System」または「Oracle Retail Advanced Science Engine」（「ORASE」））または（統合基幹業務「ERP」）システムなど、1つ以上の追加の機能モジュール18を含めることができる。データベース17は、モジュール16および18用の集中ストレージを提供するためにバス12に接続され、顧客データ、商品データ、取引データなどを格納する。一実施形態において、データベース17は、格納されたデータを管理するために構造化照会言語（「SQL」）を使用できるリレーションナルデータベース管理システム（「RDBMS」）である。一実施形態において、販売時点情報管理（「POS」）端末100は、CDTを生成するために使用される取引データ（たとえば、アイテム 店舗 週を集成した売上数量データ）を生成する。POS端末100自体が、一実施形態に係るCDT

10

20

30

40

50

を生成するための追加の処理機能を含めることができる。

【0013】

説明したように、CDTは、小売業界における基準の図であり、小売店が販売する商品の属性に起因すると顧客が考える重要度を示す。小売店の商品の各カテゴリーは、当該カテゴリーから商品を購入する顧客の行動を記述した独自の顧客意思決定ツリーを有してもよい。カテゴリーの属性は、ツリー状に配置され、「最も重要」な属性がツリーの根にあり、残りの属性がツリーの枝に沿って配置される。「最も重要」な属性は、カテゴリーから商品を購入する際に、カテゴリーの顧客がまず注目するカテゴリーの属性を示す。次に、枝は、カテゴリーの顧客が残りの属性を考慮する順序を示す。

【0014】

図2は、一実施形態に係る、小売店の取引データに基づいてシステム10が自動的に生成するヨーグルト商品カテゴリーの例示的なCDT200である。図2に示すように、ヨーグルト商品カテゴリーの商品属性は、大きさ、ブランド、フレーバー(flavor)、生産方法などを含む。商品属性「サイズ」の属性値には、小、中、大が含まれる。商品属性「ブランド」の属性値には、主流ブランドおよびニッチブランドが含まれる。商品属性「生産方法」の属性値には、オーガニックおよび非オーガニックが含まれる。商品属性「フレーバー」の属性値には、フレーバー無し、主流フレーバー、および特別フレーバーが含まれる。

【0015】

CDT200は、ヨーグルトを購入する際の顧客の意思決定過程への見解を小売店に提供する。たとえば、サイズはヨーグルトのカテゴリー下で第1レベル属性値であるため、CDT200は、意思決定過程の間、ヨーグルト商品202のサイズ204～206は、顧客の中で、概ね、最も重要な要因であることを示す。次に、好みのサイズによっては、ブランドまたは生産方法が2番目に重要な要因であると考えられる。たとえば、小サイズを好む顧客にとって、生産方法(たとえば、オーガニック210または非オーガニック211)は、2番目に重要な要因である。しかしながら、中サイズまたは大サイズのアイテムを好む顧客にとっては、ブランドが2番目に重要な要因であり、生産方法は、意思決定過程に何ら影響を与えない。また、フレーバーは、主流ブランドが販売する中サイズまたは大サイズのヨーグルト商品を好む顧客の中では考慮されるが、小サイズのヨーグルト商品を好む顧客の意思決定過程には何ら影響を与えない。

【0016】

これまで、CDT生成は、自動プロセスではなかった。CDT生成へのこれまでの取り組みは、顧客にインタビューする業界の専門家を雇って、店内での顧客の行動を調べる必要がある場合が多く、専門家は、その後、CDTを手作業で抽出していた。1つの既知の自動式の解決策が米国特許第8,874,499号に記載されており、これは、小売店のカテゴリーのこれまでの取引データを利用することによって、カテゴリーごとのCDTを抽出する。しかしながら、この既知の解決策は、たとえば、顧客ポイントカードを利用して、小売店がカテゴリーのこれまでの取引を顧客ごとに分けることができるることを必要とする。また、同じ顧客が比較的短期間のうちにカテゴリーにおいて複数回買い物をしていることも必要とする。取引データ上のこれらの必要条件によって、システムは、カテゴリーの顧客の「スイッチ行動」を調べることによって、属性の重要度、つまり、顧客がカテゴリーの1つの商品に常にこだわらないときに、カテゴリーの他のどの商品を顧客が購入したかを算出できるようになる。この既知の解決策は、このような「スイッチ行動」を調べるため、これまでの取引データが顧客ごとに特定でき、且つ、顧客が普段複数回買い物をするカテゴリーのCDTしか算出できない。そうでない場合、調べるスイッチ行動が存在しない。

【0017】

そのため、状況によっては、これらの既知の解決策に適していない多くのカテゴリーおよび多くの小売店がある。たとえば、多くの小売店、特に小さな小売店は、その費用の高さからポイントカードプログラムを実施していない。さらに、多くの小売店は、同じ顧客

10

20

30

40

50

による常習的な購入の可能性が極めて少ないカテゴリーを販売している。これは、たとえば、ほとんどの電子機器カテゴリーを説明している。食料雑貨店など、多くの妥当なカテゴリーを有する小売店でさえ、食料雑貨店の鍋やフライパンなど、不適当なカテゴリーを有することになる可能性がある。

【0018】

対照的に、本発明の実施形態は、顧客ポイントプログラムがなくても事実上すべての小売店によって生成されるデータである、アイテム 店舗 週を集成した売上数量データを利用する。そのため、実施形態は、費用のかかるポイントカードプログラムを実施する金銭的な余裕がない比較的小さな小売店を含むさまざまな小売店が利用することができる。さらに、実施形態は、携帯電話およびテレビなど、頻繁に購入されない商品のカテゴリーについての CDT を判定できる。10

【0019】

さらに、実施形態は、カテゴリーにおいてどのアイテム同士が同類であるかを判定できる。食料雑貨店でのヨーグルトカテゴリーなど、カテゴリーがどのアイテムから構成されているかが明確である場合が多いが、カテゴリーがあまり明確ではない多くの小売店がある。たとえば、ディズニーストアにおいて、顧客、特に子供が店舗で何かを買うとき、アイテムに特定のディズニーキャラクターが付されていれば、アイテムの機能が実際何であるかは気にしないことが多いため、カテゴリーが何であるかがはっきりしない可能性がある。そのため、たとえば、ペンが実際にマグカップの売上げを減らしてしまう可能性がある。よって、ペンとマグカップは、通常、異なるカテゴリーのアイテムであるにもかかわらず、ディズニーストアでは異なるカテゴリーでない方がよい。さらに、ペットのグルーミング商品については、異なる種類の犬のグルーミング用具は、同じ機能を果たすことができるため、用具自体は実際異なるとはいえ、互いの売上げを減らしてしまう可能性がある。20

【0020】

図3は、一実施形態に係るCDTを生成するときの、図1のCDT生成モジュール16の機能のフロー図である。一実施形態において、図3（および下記の図4および図5）のフロー図の機能は、メモリまたはその他のコンピュータ読み取り可能な媒体または有形の媒体に格納されたソフトウェアによって実装され、プロセッサによって実行される。その他の実施形態において、機能は、（たとえば、特定用途向け集積回路（「ASIC」）、プログラマブル・ゲート・アレイ（「PGA」）、フィールド・プログラマブル・ゲート・アレイ（「FPGA」）などを使用して）ハードウェアによって実行され得、またはハードウェアとソフトウェアとの任意の組み合わせによって実行され得る。30

【0021】

図3において、310において、CDT生成モジュール16は、各商品ペアと各属性値ペアとの類似度を算出する。次に、320において、CDT生成モジュール16は、310から得られた類似度に基づいて、CDTを生成する。

【0022】

図4は、一実施形態に係る、図3の310において類似度を判定するときの、図1のCDT生成モジュール16の機能のフロー図である。310において類似度を算出する際、所定のカテゴリーについての各商品ペアと属性値ペアとの類似度を判定する。一般に、実施形態は、まず、たとえば、POS端末100から売上データの形でデータ要素を受け付ける。次に、データを集成して、週単位の売上シェアを算出する。次に、属性値ペアについての類似度算出を行う。40

【0023】

データ要素については、402において、売上データを取り引レベル（つまり、取引ID / 顧客ID / 店舗 / 日付 / アイテム）レベルで受け付ける。取引とは、顧客識別情報（「ID」（customer_id））、取引ID（transaction_id）、店舗ID（store_id）、日付、および購入されたアイテムの、アイテムの売上数量、ドルでの売上高、および販売価格など、付随する情報との組み合わせによって特定され50

る売上げの発生である。この情報は、個々の小売店舗用のほとんどのPOSシステムで容易に入手可能である。下記の表1は、取引データを例示し、所定の日に所定の店舗（つまり、店舗IDが142）で同じアイテム（つまり、アイテムID(item_id)が2345）を購入した互いに異なる顧客を示している。

【0024】

【表1】

transaction_id	customer_id	store_id	item_id	日付	売上数量	売上高	販売価格
15960247	584231	142	2345	5/11/2015	34	\$305.66	\$8.99
15960248	345634	142	2345	5/11/2015	12	\$107.88	\$8.99
15960249	657856	142	2345	5/12/2015	10	\$79.90	\$7.99
15960250	123123	142	2345	5/12/2015	5	\$29.95	\$5.99
15960251	435436	142	2345	5/14/2015	50	\$449.50	\$8.99

表1

【0025】

次に、404において、データをアイテム/週レベルに集積する。その他の実施形態において、週以外の異なる期間/測定値を使用できる（たとえば、日、月など）。一実施形態において、当該所定のアイテム/店舗/週に関するすべての取引IDおよび顧客IDの取引レベルデータが、アイテム/店舗/週レベルに集積される。ここで、このレベルが売上数量およびドルで示される。ここで、販売価格が、重み付けされた平均価格：ドル売上の総額/売上総数量、として明示される。表1の上記例を使用して、2015年5月16日で終わる週について集積されたアイテム/店舗/週レベルデータは、ここで表2に示す次のようになる。

【0026】

【表2】

store_id	item_id	日付	売上数量	売上高	重み付けされた価格
142	2345	5/16/2015	111	\$972.89	\$8.76

表2

【0027】

さらに、404において、データを属性値/店舗/週レベルに集積する。その他の実施形態において、週以外の異なる期間/測定値を使用できる（たとえば、日、月など）。一実施形態において、各アイテムは、商品属性タイプと、値とを有し、このレベルで総売上が示される。属性タイプの例として、フレーバー（たとえば、「ストロベリー」または「バニラ」という値）、サイズ（たとえば、「小」、「中」、または「大」という値）、ブランド（たとえば、「コーラ」または「ペプシ」という値）などがある。下記の表3は、フレーバー属性についての売上高を示す例である。

【0028】

【表3】

store_id	フレーバー値	日付	売上数量	売上高	販売価格
2345	フレーバー1	5/16/2015	111	\$972.89	\$8.76
2345	フレーバー2	5/16/2015	23	\$184.23	\$8.01
2345	フレーバー3	5/16/2015	133	\$1,243.55	\$9.35
2345	フレーバー3	5/23/2015	78	\$692.64	\$8.88
2345	フレーバー3	5/30/2015	45	\$413.55	\$9.19

表3

【0029】

次に、集積データを使用して、実施形態は、406において、週単位の売上シェア、週単位でない場合、相応の時間測定の間の売上シェアを判定する。一実施形態において、週単位の売上シェアとは、属性値/店舗/週に属する売上の、同じ店舗/週の同じ属性タイプについての他のすべての属性値に対する割合である。所定の店舗/週について、所定の属性タイプの売上シェアの合計は、合算すると100%になる。実施形態は、データ履歴におけるすべての属性タイプ/店舗/週についての週単位の売上シェアを判定する。

【0030】

10

20

30

40

50

上記例を引き続き使用すると、下記の表4は、2015年5月16日の週についての、売上シェア = フレーバーの売上数量 / 週全体の売上数量、を示す。

【0031】

【表4】

store_id	フレーバー値	日付	売上数量	売上シェア
2345	フレーバー1	5/16/2015	111	41.6%
2345	フレーバー2	5/16/2015	23	8.6%
2345	フレーバー3	5/16/2015	133	49.8%
			合計	267
				100%

表4

10

【0032】

また、週単位の売上シェアを、1店舗 / 週全体のすべてのアイテムについて演算する。下記の表5は、一例を示す。

【0033】

【表5】

store_id	item_id	日付	売上数量	売上シェア
1001	123456	5/16/2015	22	26.5%
1001	654321	5/16/2015	44	53.0%
1001	881155	5/16/2015	5	6.0%
1001	265446	5/16/2015	12	14.5%
			合計	83
				100%

表5

20

【0034】

次に、408において、実施形態は、属性値ペアについての類似度を判定する。一実施形態において、類似度は、その売上シェア履歴を横断して属性タイプ内で演算し、フレーバーペア(X, Y)について、 X_i および Y_i は、フレーバーXおよびYの店舗 / 週のシェア値をそれぞれ表し、nは、フレーバーXおよびYのシェアがある店舗 / 週の総数を表すピアソン相関式を用いて、以下のように演算する。

30

【0035】

【数1】

$$SIM(X, Y) = \frac{\sum_{i=1}^n X_i Y_i - \frac{(\sum_{i=1}^n X_i)(\sum_{i=1}^n Y_i)}{n}}{\sqrt{\left(\sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n} \right) \left(\sum_{i=1}^n Y_i^2 - \frac{(\sum_{i=1}^n Y_i)^2}{n} \right)}} \quad (\text{式 } 1)$$

40

【0036】

実施形態は、すべてのフレーバー(X, Y)ペアについてのSIM(X, Y)を算出する。これらの類似度によって「フレーバー類似度」が構成される。SIMについての上記式は、-1と1との間の数を常に出す。属性値XおよびYについて、-1に近いSIMは、XのシェアとYのシェアとが「反相関」であることを意味し、Xのシェアが増加するとYのシェアが減少する、および、その逆の場合を意味する。したがって、顧客はXをたくさん買えば買うほどYを買わなくなる(およびその逆)ため、顧客にとって、XおよびYは互いの代用品であるという点で類似しているに違いない。-1に近づけば近づくほど、XとYとは互いの代用品になる。また、同じ方法で、実施形態は、すべてのその他の属性についての類似度も算出するため、たとえば、「ブランド類似度」、「サイズ類似度」な

50

どが得られる。

【0037】

一実施形態において、上述の相関関係は、以下の擬似コードを利用したSQLの組み込み関数「corr」を使用して算出され、

【0038】

【数2】

select

x.flavor as flavor_x, y.flavor as flavor_y,
corr(x.flavor_share, y.flavor_share) as flavor_similarity

10

from

sales_share_table x,
sales_share_table y

where

x.calendar_wk = y.calendar_wk
and x.flavor <= y.flavor

group by

flavor1, flavor2

20

【0039】

下記の表6に示す結果が得られる。

【0040】

【表6】

f l a v o r _ x	f l a v o r _ y	f l a v o r _ s i m i l a r i t y
flavor_1	flavor_1	1.00
flavor_1	flavor_2	-0.45
flavor_1	flavor_3	-0.15
flavor_2	flavor_2	1.00
flavor_2	flavor_3	0.05
flavor_3	flavor_3	1.00

30

表6

【0041】

(上記の属性値の代わりに) X と Y が 2 つの異なるアイテムを表すアイテムペアについて同様の処理を繰り返す。よって、X_i および Y_i は、特定の店舗 / 週におけるアイテム X およびアイテム Y のアイテムシェアをそれぞれ表す。そのため、ちょうど属性の属性値ペアごとに上記の SIM(X, Y) を算出したように、実施形態は、アイテムのペア(X, Y) ごとに SIM(X, Y) を算出し、下記の表7に示す以下の例示的な結果が得られる。

40

【0042】

【表7】

i t e m _ x	i t e m _ y	i t e m _ s i m i l a r i t y
2 3 4 5	2 3 4 5	1. 00
2 3 4 5	5 7 9 1	-0. 34
2 3 4 5	9 8 7 6	0. 21
5 7 9 1	5 7 9 1	1. 00
5 7 9 1	9 8 7 6	-0. 56
9 8 7 6	9 8 7 6	1. 00

表7

10

【0043】

さらに、408において、実施形態は、バイナリ属性についての類似度の算出を行う。バイナリ属性とは、2つの値しか持たない属性である。これらの値はごく一般的であり、通常、ある特性の有無を示す。下記に用いる一例は、「オーガニック」(つまり、食品アイテムがオーガニックであるかそうではないか)である。上記のSIMを求める式を単純にバイナリ属性に適用すると、結果は常にSIM = -1となる。これでは買い物客が属性をどのように扱っているかについての情報を提供できないため、バイナリ属性は特別な処理を必要とする。

【0044】

【数3】

20

代わりに、2つの値からしか選べない属性タイプ(たとえば、オーガニックおよび非オーガニック食品)については、 x_k は、週kにおけるオーガニックシェアを表し、週は、N週あり、 \bar{x} は、 x_i の平均、つまり、N週にわたっての平均オーガニックシェアである相関関係を以下のように算出する。

$$2 \sqrt{\frac{\sum_{k=1}^N (x_k - \bar{x})^2}{N}} \quad (\text{式2})$$

したがって、式2は、 x_k の標準偏差の2倍であり、平均オーガニックシェアからのオーガニックシェアの変動を計測している。一般に、変動が大きくなるほど多くの顧客がオーガニックを非オーガニック(またはその逆)に交換したことを示すので、オーガニックが非オーガニックに類似していく。非オーガニックシェアとして x_k (および、平均非オーガニックシェアとして \bar{x})が代わりに用いられた場合、同じ数字という結果になる。0から1までの計測を行うために、2の乗数を用いる(そうでない場合、 x_k がこの場合シェアであるので、 x_k が0と1との間である場合は1/2が標準偏差の最大値であるので、計測は0から1/2まで行う)。

30

【0045】

バイナリ属性についての類似度を実行するために、以下のSQL擬似コードを用いることができる。

【0046】

【数4】

```
sum(2/sqrt(n_wks)*sqrt(sum(power(abs(a.share_organic - stats.avg_share_organic),2))))
as organic_similarity,
```

```
2/sqrt(n_wks)*sqrt(sum(power(a.share_nonorganic - stats.avg_share_nonorganic,2)))
as nonorganic_similarity
```

10

from

```
(select
    avg(share_organic) as avg_share_organic,
    avg(share_nonorganic) as avg_share_nonorganic,
    count(*) as n_wks
```

from

```
sales_share_organic_values_table) stats,
```

20

```
sales_share_organic_values_table a
```

group by

```
n_wks
```

【0047】

バイナリ属性についての類似度算出の例示的な結果を下記の表8に示す。

30

【0048】

【表8】

organic_similarity	nonorganic_similarity
0.43	0.43

表8

【0049】

410において、次に、実施形態は、SIM値の後処理を行う。属性ペアおよびアイテムペアの両方のSIM値において、実施形態は、SIM値を次のように変更する。SIM値が正である場合は0に設定し、負の場合は正にする。残りの開示について、使用されるSIM値は、後処理されたSIM値である。410における後処理ステップは、バイナリ属性タイプの類似度には用いられない。バイナリ属性タイプは負ではないことが上記の式2によってすでに保証されているためである。

40

【0050】

412において、次に、実施形態は、各属性のSIM値をアイテムのSIM値と比較することによって「最上位属性」を求める。実施形態は、どの属性が顧客のアイテムレベルの購買行動を最もよく示しているのかを判定する。アイテムレベルのSIM値を各属性のSIM値と比較し、SIM値がアイテムレベルの値に最も「一致」（後述する）する属性を求める。

【0051】

50

フレーバーなど特定の属性について、実施形態は、アイテムおよび属性の S I M 値を、下記の表 9 に示すように 1 つの表に編集する。flavor_x 列は、item_x のフレーバーであり、同様に、flavor_y は、item_y のフレーバーである。flavor_similarity は、flavor_x と flavor_y との S I M 値である。なお、(item_x と item_y とが同じフレーバーであるために) flavor_x と flavor_y とが同じである場合、フレーバーが同じであるので、flavor_similarity は、1 に等しい。そうでない場合、flavor_similarity は、単に、前述したように算出された flavor_x と flavor_y との S I M 値である。

【0052】

10

【表 9】

item_x	flavor_x	item_y	flavor_y	item_similarity	flavor_similarity
4563	flavor_1	1200	flavor_3	0.58	0.45
4563	flavor_1	2345	flavor_1	0.82	1.00
4563	flavor_1	4563	flavor_1	1.00	1.00
4563	flavor_1	5665	flavor_2	0.67	0.68
4563	flavor_1	5698	flavor_4	0.65	0.21
4563	flavor_1	8758	flavor_1	0.02	1.00
4563	flavor_1	9901	flavor_2	0.10	0.68
5665	flavor_2	1200	flavor_3	0.05	0.50
5665	flavor_2	2345	flavor_1	0.98	0.68
5665	flavor_2	5665	flavor_2	1.00	1.00
5665	flavor_2	5698	flavor_4	0.68	0.29
5665	flavor_2	8758	flavor_1	0.34	0.68
5665	flavor_2	9901	flavor_2	0.58	1.00
1200	flavor_2	1200	flavor_3	1.00	0.50
1200	flavor_2	5698	flavor_4	0.12	0.29
1200	flavor_3	8758	flavor_1	0.24	0.45
...

表9

【0053】

次に、実施形態は、次の S Q L 擬似コードを利用して、アイテム類似度および属性類似度（表 9 の例において、これは、item_similarity 値および flavor_similarity 値を指す）に対して相関計算処理を実行する。これは、item_similarity 列および flavor_similarity 列に対して相関を実行することを意味し、

20

【0054】

30

【数 5】

```

select
    corr(item_similarity, flavor_similarity) as flavor_result
from
    item_flavor_similarities

```

40

【0055】

下記の表 10 に示す例示的な結果が得られる。

【0056】

【表10】

f l a v o r _ r e s u l t
0. 0804

表10

【0057】

次に、実施形態は、すべての属性について繰り返し、その結果を下記の表11の例のように編集する。 10

【0058】

【表11】

	a t t r i b u t e _ r e s u l t
b r a n d _ r e s u l t	0. 1559
o r g a n i c _ r e s u l t	0. 1235
s i z e _ r e s u l t	0. 0912
f l a v o r _ r e s u l t	0. 0804

表11

20

【0059】

最大値を有する属性は、CDTにおいて最上位を有すると考えられるため、図3の320において生成されるCDTの最上位レベル属性となる。CDTに追加するために、CDTの他のレベルおよび枝を生成するために図4の機能を繰り返す。たとえば、「ブランド」が最上位属性であると判定されると、ブランド属性におけるブランドごとに、特定ブランド内である402において受け付けられたデータ要素のサブセットのみを用いて図4の機能を実行する。

【0060】

図5は、一実施形態に係る、類似度に基づいてCDTを生成する(図3の320)際の図1のCDT生成モジュール16の機能のフロー図である。510において、同一商品カテゴリーの商品において、機能適合属性(functional-fit attribute)があるかどうかを判定する。機能適合属性とは、その値を別の値に置き換えることがまざない商品属性である。たとえば、ワイパープレードを買い物中の顧客は、車に合う対応するプレードを購入する必要がある。そのため、ワイパープレード商品カテゴリーにおいて、商品属性「サイズ」が機能適合属性として判定される。商品属性「サイズ」は、たとえば、タイヤ、エアフィルタ、集塵袋、プリンターのカートリッジなどの他の商品カテゴリーの機能適合属性にもなり得る。しかしながら、同一商品属性「サイズ」は、たとえば、フルーツ、ソフトドリンクなどの他の商品カテゴリーの機能適合属性にはならないだろう。一般に、機能適合属性は、通常、アクセサリなどの非食料品アイテムに存在する。一実施形態における機能適合属性は、生成された顧客データから直接得ることができ、通常、算出しなくてよい。たとえば、小売店は、ワイパープレードの場合、サイズが機能適合属性であるとはっきりと示すなど、通常、「機能適合」属性が何であるかをはっきりと特定する。 30 40

【0061】

すべての機能適合属性が特定されると、機能適合属性は、CDTの商品カテゴリー直下の最上位レベルに自動的に配置される。図6は、一実施形態に係る、CDT生成モジュール16が生成するCDT600を示す。CDT600は、商品カテゴリーを特定するカテゴリー・レベル610を有する。図2に示すように、ヨーグルト商品カテゴリーについては、カテゴリー・レベル610に「ヨーグルト」が表示されることになる。別の例において、「コーヒー」カテゴリーについては、カテゴリー・レベル610に「コーヒー」が表示され 50

る。次に、機能適合属性が CDT 600 の最上位レベル 620 に配置される。図 6 は、最上位レベル 620 の 2 つの機能適合属性 (FA1、FA2) 622、624 を示す。しかしながら、ヨーグルトまたはコーヒーについては、機能適合属性がないと考えられる。

【0062】

次に、図 5 の 520において、最上位属性または分割属性を特定する。最上位属性は、図 4 の機能に従って判定される。

【0063】

530において、アイテムを下位区分に分割する。下位区分は、520において特定された属性の特定の属性値に対応する。たとえば、520において、「形態」商品属性がコーヒーについての最上位属性であると判定された場合、「形態」商品属性は、コーヒーについての形態：「豆」、「挽き豆」、および「インスタント」の特定の値に各々が対応する 3 つの下位区分に分割される。下位区分は、図 6 の次のレベル 630 を形成する。下位区分は、最上位レベル 620 の下である。たとえば、図 6 は、機能適合属性 622 から枝分かれしたレベル 630 の 2 つの下位区分 (A1a、A1b) 632、634 を示す。520 および 530 を下位区分ごとに繰り返し、下位区分ごとの終端ノードに到達するまで CDT 600 を展開する (540において No)。最終的に下位区分ごとの終端ノードに到達すると (540において Yes)、プロセスは終了する。

【0064】

開示したように、ツリーは、終端ノードが特定されるまで展開される。一実施形態において、ノードが終端であると宣言するための基準は、以下の通りである。

【0065】

1. 有効属性が特定されない。

2. ノードのアイテム数 < 商品カテゴリーの総アイテムの ×%。ここで、「×」は、ツリーの大きさの上限を定める調整パラメータである。一実施形態において、×のデフォルト値は、10 である。

【0066】

3. 子ノードの平均非類似度 (「AD」) (つまり、ノードにおける商品のすべての可能性のペアの平均) が、親ノードよりも大きい。2 つの起こり得る下位のケースは、以下の通りである。

【0067】

a. すべての子ノードが親ノードよりも大きい AD 値を有する場合、親ノードは、終端ノードであると宣言される。

【0068】

b. 子ノードのうちのいくつかが親ノードよりも大きい AD 値を有する場合、これらのノードは閉じられ、その他の子ノードは、通常通り展開される。

【0069】

開示したように、実施形態は、アイテム 店舗 週を集成した売上数量データに頼って CDT を生成する。アイテム 店舗 週を集成した売上数量データは、各店舗での各アイテムの週ごとの売上げ総数量にすぎないため、このようなデータは、カテゴリーに関係なく、概して、すべての小売店から入手可能である。そのため、顧客の識別情報など、取得することがより困難または費用のかかるデータは必要ない。

【0070】

さらに、集成されたデータから CDT を生成する既知のシステムは、一般に、よりスタンダードな統計学的手法に頼るが、これらはスタンダードであるにもかかわらず CDT を算出する際に使用するには欠点がある。これらの既知の手法は、非常に大量の計算能力を必要とする可能性があり、実装することは難しいだろう。対照的に、実施形態は、standard SQL クエリを用いて実装することができ、大きな顧客データセット上であっても非常に速く動く。

【0071】

さらに、実施形態は、2 つの値 (Boolean 属性として知られる) しか有さない属

10

20

30

40

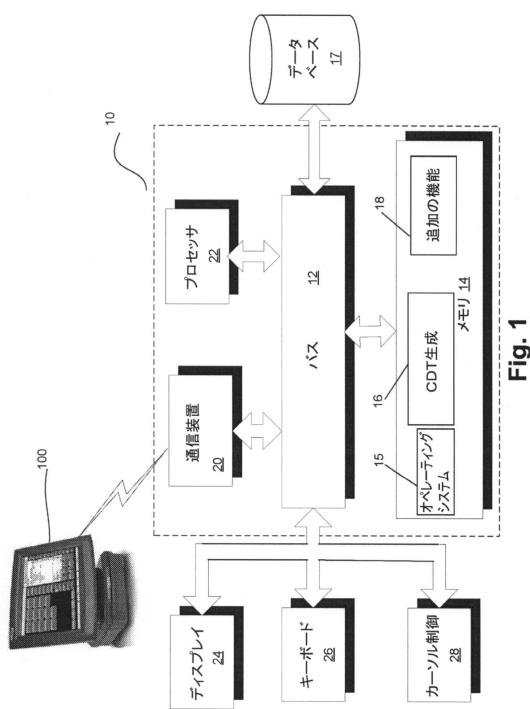
50

性を処理する。このような属性は、カテゴリーのアイテムにおける一定の特性の有無（たとえば、ヨーグルトがギリシャヨーグルトであるかどうか、またはシャンプーが低刺激性であるかどうか）を伝えるので、多くのカテゴリーにおいてとても一般的である。

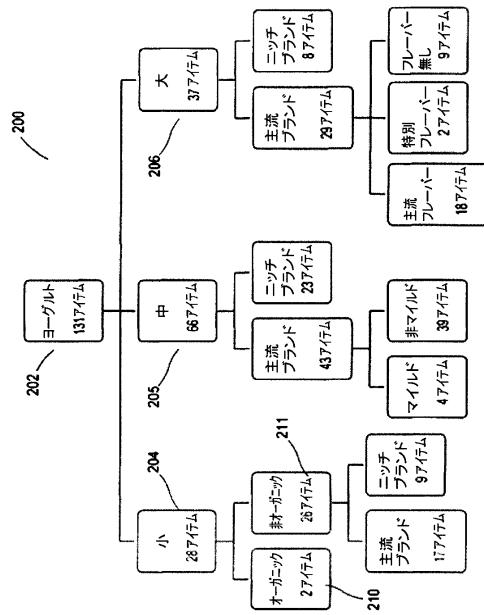
【0072】

いくつかの実施形態を本明細書において具体的に例示および／または説明した。しかしながら、開示の実施形態の変更例および変形例は、上記教示に包含され、意図する本発明の範囲の趣旨から逸脱することなく、添付の請求の範囲に含まれることがわかる。

【図1】



【図2】



【図3】

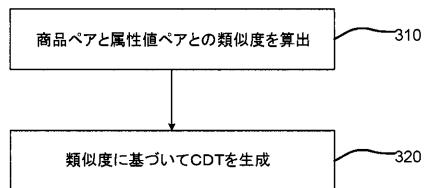


Fig. 3

【図4】

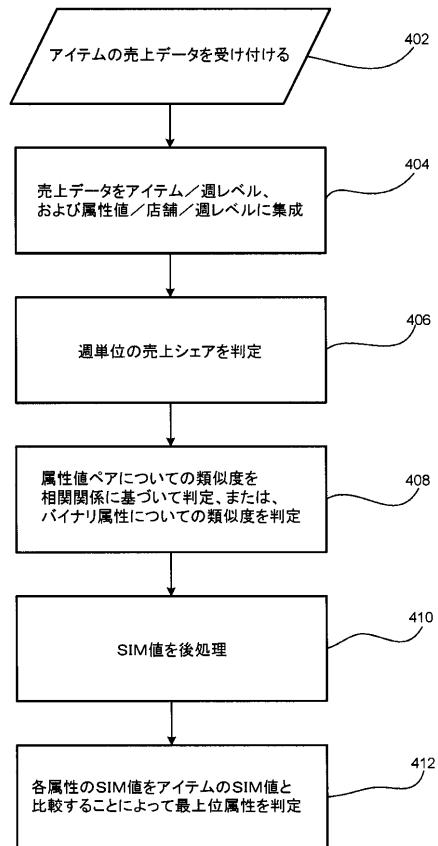


Fig. 4

【図5】

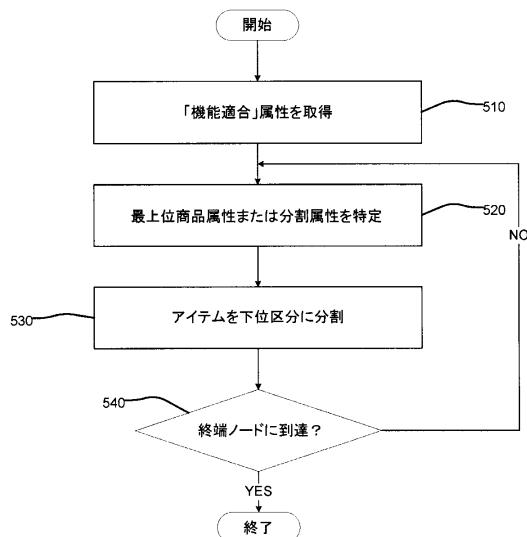


Fig. 5

【図6】

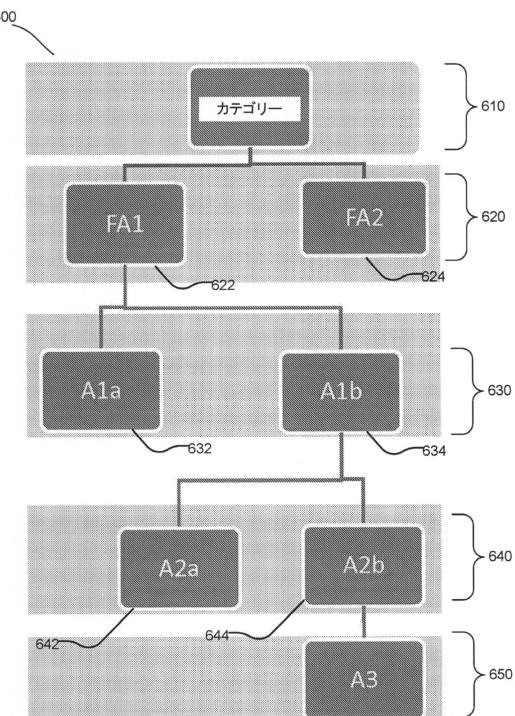


Fig. 6

フロントページの続き

(72)発明者 シン , ジヨン

アメリカ合衆国、02462 マサチューセッツ州、ニュートン、パイン・グローブ・アベニュー、
100

(72)発明者 パンチャンガム , キラン・ベンカタ

アメリカ合衆国、01730 マサチューセッツ州、ベドフォード、ケンドール・コート、78

審査官 阿部 潤

(56)参考文献 特表2015-524127 (JP, A)

米国特許出願公開第2015/0127419 (US, A1)

米国特許出願公開第2008/0300964 (US, A1)

国際公開第2010/052845 (WO, A1)

特開2015-041121 (JP, A)

(58)調査した分野(Int.Cl., DB名)

G06Q 10/00 - 99/00

G16H 10/00 - 80/00