(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 108292409 B
(45)授权公告日 2022.05.17

(21)申请号 201680070211.X
(22)申请日 2016.11.15
(65)同一申请的已公布的文献号
申请公布号 CN 108292409 A
(43)申请公布日 2018.07.17
(30)优先权数据
14/990,834 2016.01.08 US
(85)PCT国际申请进入国家阶段日
2018.05.31
(86)PCT国际申请的申请数据
PCT/US2016/062032 2016.11.15
(87)PCT国际申请的公布数据
WO2017/119952 EN 2017.07.13

(73)专利权人 甲骨文国际公司
地址 美国加利福尼亚
(72)发明人 吴思明 J·施恩
K·V·潘查加姆
(74)专利代理机构 中国贸促会专利商标事务所有限公司 11038
专利代理师 王希

(51)Int.Cl.
G06Q 30/02(2006.01)
G06F 17/18(2006.01)
G06Q 10/04(2006.01)
审查员 张盈盈

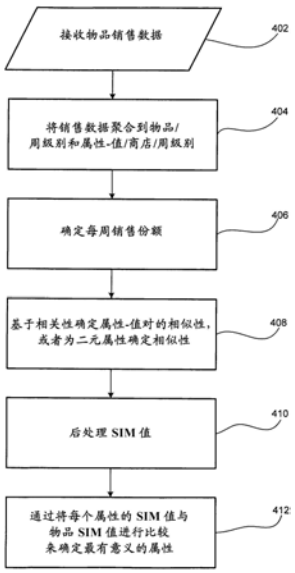
权利要求书3页 说明书11页 附图6页

(54)发明名称

消费者决策树生成系统

(57)摘要

生成消费者决策树的系统接收零售物品交易销售数据。系统将销售数据聚合到物品/商店/持续时间级别,并将销售数据聚合到属性-值/商店/持续时间级别。系统确定该持续时间的销售份额,并基于属性-值对之间的相关性确定属性-值对的相似性。系统然后基于所确定的相似性来确定最有意义的属性。



CN 108292409 B

1.一种其上存储有指令的计算机可读介质,所述指令在由处理器执行时使处理器生成一种类别的产品的消费者决策树(CDT),所述类别包括多个产品属性,每个产品属性具有对应的多个属性值,所述生成包括:

接收跨所有客户标识的零售物品交易销售数据;

将销售数据聚合到物品/商店/持续时间级别以获得第一聚合数据,第一聚合数据不包括逐个客户的数据并且包括,对于对应的商店标识符ID,对应产品ID和所述产品ID在所述持续时间期间的单元销售数目;

将销售数据聚合到属性值/商店/持续时间级别以获得第二聚合数据,第二聚合数据不包括逐个客户的数据并且包括对于对应的商店ID对于每个属性值,在所述持续时间期间的该属性值的单元销售数目;

使用第二聚合数据来确定第一销售份额,第一销售份额包括针对每个属性的每个属性值,在该持续时间期间该属性的该属性值的销售相比于该属性的全部属性值的总销售的百分比;

使用第一聚合数据来确定第二销售份额,第二销售份额包括针对每个产品,在该持续时间期间该产品的销售相比于全部产品的总销售的百分比;

使用第一销售份额来计算每个属性的全部可能的属性值对之间的相似性,所述计算包括针对每个属性值对确定属性值相关值;

使用第二销售份额来计算所有可能的产品对之间的相似性,所述计算包括为每个产品对确定产品相关值;

通过针对每个属性,将所述属性值相关值与所述产品相关值相关来确定最有意义的属性,其中具有最高相关性的属性是所述最有意义的属性,

将所述最有意义的属性指派为CDT的第一级别;

将CDT的第二级别分成多个子部分,其中每个子部分与所述最有意义的属性的属性值对应;

对于每个子部分,在移除被确定为有意义的之前的属性之后,对于子部分值重复确定第一销售份额、确定第二销售份额、计算所有可能的属性值对之间的相似性、计算所有可能的产品对之间的相似性,以及确定最有意义的属性;和

当对于每个子部分达到终端节点时,生成所述CDT,所述CDT包括客户对所述类别的产品的决策层级结构的图形表示。

2.如权利要求1所述的计算机可读介质,其中持续时间包括每周。

3.如权利要求1所述的计算机可读介质,所述生成还包括:

确定二元属性的相似性,其中二元属性是只包括两个属性值的属性。

4.如权利要求1所述的计算机可读介质,所述生成还包括对所述属性值相关值和所述产品相关值进行后处理,所述后处理包括将正值赋值为0并将负值修正为对应的正值。

5.如权利要求1所述的计算机可读介质,其中计算每个属性的全部可能的属性值对之间的相似性包括在不使用逐个客户的数据的情况下确定SIM的值,包括:

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

其中对于属性值对(X,Y)，X_i和Y_i表示属性X和Y的商店/时间份额值，并且n表示存在X和Y的属性份额的商店/持续时间的总数。

6.如权利要求3所述的计算机可读介质，其中确定二元属性的相似性包括在不使用逐个客户的数据的情况下计算：

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

其中x_k是持续时间k中的有机份额，并且存在N个持续时间，并且$\bar{x}$是x_k的平均值。

7.一种生成一种类别的产品的消费者决策树(CDT)的方法，所述类别包括多个产品属性，每个产品属性具有对应的多个属性值，该方法包括：

接收跨所有客户标识的零售物品交易销售数据；

将销售数据聚合到物品/商店/持续时间级别以获得第一聚合数据，第一聚合数据不包括逐个客户的数据并且包括，对于对应的商店标识符ID，对应产品ID和所述产品ID在所述持续时间期间的单元销售数目；

将销售数据聚合到属性值/商店/持续时间级别以获得第二聚合数据，第二聚合数据不包括逐个客户的数据并且包括对于对应的商店ID对于每个属性值，在所述持续时间期间的该属性值的单元销售数目；

使用第二聚合数据来确定第一销售份额，第一销售份额包括针对每个属性的每个属性值，在该持续时间期间该属性的该属性值的销售相比于该属性的全部属性值的总销售的百分比；

使用第一聚合数据来确定第二销售份额，第二销售份额包括针对每个产品，在该持续时间期间该产品的销售相比于全部产品的总销售的百分比；

使用第一销售份额来计算每个属性的全部可能的属性值对之间的相似性，所述计算包括针对每个属性值对确定属性值相关值；

使用第二销售份额来计算所有可能的产品对之间的相似性，所述计算包括为每个产品对确定产品相关值；

通过针对每个属性，将所述属性值相关值与所述产品相关值相关

来确定最有意义的属性，其中具有最高相关性的属性是所述最有意义的属性；

将所述最有意义的属性指派为CDT的第一级别；

将CDT的第二级别分成多个子部分，其中每个子部分与所述最有意义的属性的属性值对应；

对于每个子部分，在移除被确定为有意义的之前的属性之后，对于子部分值重复确定第一销售份额、确定第二销售份额、计算所有可能的属性值对之间的相似性、计算所有可能

的产品对之间的相似性,以及确定最有意义的属性;和

当对于每个子部分达到终端节点时,生成所述CDT,所述CDT包括客户对所述类别的产品的决策层级结构的图形表示。

8.如权利要求7所述的方法,其中持续时间包括每周。

9.如权利要求7所述的方法,还包括:

确定二元属性的相似性,其中二元属性是只包括两个属性值的属性。

10.如权利要求7所述的方法,还包括对所述属性值相关值和所述产品相关值进行后处理,所述后处理包括将正值赋值为0并将负值修正为对应的正值。

11.如权利要求7所述的方法,其中计算每个属性的全部可能的属性值对之间的相似性包括在不使用逐个客户的数据的情况下确定SIM的值,包括:

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

其中对于属性值对(X,Y),X_i和Y_i表示属性X和Y的商店/时间份额值,并且n表示存在X和Y的属性份额的商店/持续时间的总数。

12.如权利要求9所述的方法,其中确定二元属性的相似性包括在不使用逐个客户的数据的情况下计算:

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

其中x_k是持续时间k中的有机份额,并且存在N个持续时间,并且$\bar{x}$是x_k的平均值。

13.一种消费者决策树(CDT)生成系统,包括:

存储器,配置为存储计算机可执行指令,和

处理器;配置为执行所述计算机可执行指令来执行如权利要求7-12中任一项所述的方法。

消费者决策树生成系统

技术领域

[0001] 一个实施例一般而言针对计算机系统，并且特别地针对生成消费者决策树的计算机系统。

背景技术

[0002] 买方决策过程是消费者在购买产品或服务之前、期间和之后潜在的市场交易中所进行的决策制定过程。更一般而言，决策制定是从多种选择中选择行动方案的认知过程。常见的示例包括购物和决定吃什么。

[0003] 一般而言，有三种分析消费者购买决策的方法：(1)经济模型-这些模型在很大程度上是定量的，并且基于合理性和近乎完美的知识的假设。消费者被看作是最大化他们的效用；(2)心理模型-这些模型专注于心理和认知过程，诸如动机和需求识别。它们是定性的而不是定量的，并且建立在社会学因素上，如文化影响和家庭影响；(3)消费者行为模型-这些是营销人员使用的实用模型。他们通常融合经济和心理模型。

[0004] 一种类型的消费者行为模型被称为“消费者决策树”(“CDT”)。CDT是产品属性空间中消费者的决策层次的图形表示，用于购买给定类别中的物品。它建模客户在缩小到他们选择的物品之前如何考虑类别内的不同替代方案(基于属性)，并帮助理解客户的购买决策。它通常也被称为“产品细分和类别结构”。CDT按照惯例由品牌制造商或第三方市场研究公司基于调查和其它市场研究工具生成。但是，这些方法缺乏准确性，并且会缺乏真实性，因为它们可能基于品牌制造商提供的偏颇数据。

发明内容

[0005] 一个实施例是生成消费者决策树的系统。系统接收零售物品交易销售数据。系统将销售数据聚合到物品/商店/持续时间级别，并将销售数据聚合到属性-值/商店/持续时间级别。系统确定该持续时间的销售份额，并基于属性-值对之间的相关性确定属性-值对的相似性。然后系统基于所确定的相似性来确定最有意义的属性。

附图说明

[0006] 图1是根据本发明实施例的计算机服务器/系统的框图。

[0007] 图2是根据一个实施例的基于零售商的交易数据自动生成的酸奶产品类别的示例CDT。

[0008] 图3是根据一个实施例的当生成CDT时图1的CDT生成模块的功能的流程图。

[0009] 图4是根据一个实施例的当确定相似性时图1的CDT生成模块的功能的流程图。

[0010] 图5是根据一个实施例的当基于相似性生成CDT时图1的CDT生成模块的功能的流程图。

[0011] 图6图示了根据一个实施例的由CDT生成模块生成的CDT。

具体实施方式

[0012] 一个实施例使用零售商的交易数据，具体而言是物品存储周聚合销售单位数据，自动生成消费者决策树(“CDT”)，以确定物品相似性。因此，即使是不使用忠诚度计划的小零售商可用的交易数据也可以被用来生成CDT。另外，实施例提供对零售商哪些物品一起属于单个类别的确定。

[0013] 图1是根据本发明实施例的计算机服务器/系统10的框图。虽然被示为单个系统，但是系统10的功能可以被实现为分布式系统。另外，本文公开的功能可以在可以经网络耦合在一起的单独的服务器或设备上实现。另外，可以不包括系统10的一个或多个部件。例如，对于服务器的功能，系统10可以需要包括处理器和存储器，但是可以不包括图1中所示的一个或多个其它部件，诸如键盘或显示器。

[0014] 系统10包括用于传送信息的总线12或其它通信机制，以及耦合到总线12用于处理信息的处理器22。处理器22可以是任何类型的通用或专用处理器。系统10还包括用于存储要由处理器22执行的信息和指令的存储器14。存储器14可以包括随机存取存储器(“RAM”)、只读存储器(“ROM”)、诸如磁盘或光盘的静态储存器，或任何其它类型的计算机可读介质。系统10还包括通信设备20，诸如网络接口卡，以提供对网络的访问。因此，用户可以直接地或通过网络远程地或任何其它方法与系统10接口。

[0015] 计算机可读介质可以是可由处理器22访问的任何可用介质，并且包括易失性和非易失性介质、可移动和不可移动介质，以及通信介质。通信介质可以包括计算机可读指令、数据结构、程序模块或调制数据信号(诸如载波或其它传输机制)中的其它数据，并且包括任何信息输送介质。

[0016] 处理器22还经由总线12耦合到诸如液晶显示器(“LCD”)的显示器24。键盘26和诸如计算机鼠标的光标控制设备28还耦合到总线12，以使用户能够与系统10接口。

[0017] 在一个实施例中，存储器14存储当由处理器22执行时提供功能的软件模块。模块包括为系统10提供操作系统功能的操作系统15。这些模块还包括自动从零售商消费者数据生成CDT的消费者决策树生成模块16，以及本文公开的所有其它功能。系统10可以是更大系统的一部分。因此，系统10可以包括一个或多个附加的功能模块18，以包括附加功能，诸如零售管理系统(例如，来自Oracle公司的“Oracle零售销售系统”或“Oracle零售高级科学引擎”(“ORASE”))或企业资源计划(“ERP”)系统。数据库17耦合到总线12，以便为模块16和18提供集中式存储并且存储消费者数据、产品数据、交易数据等等。在一个实施例中，数据库17是关系数据库管理系统(“RDBMS”)，其可以使用结构化查询语言(“SQL”)来管理所存储的数据。在一个实施例中，专用销售点(“POS”)终端100生成用于生成CDT的交易数据(例如，物品-商店-周聚合销售单位数据)。根据一个实施例，POS终端100本身可以包括生成CDT的附加处理功能。

[0018] 如所讨论的，CDT是作为零售行业中的标准并且描绘消费者对零售商所销售产品的属性的重视程度的图。零售商的每一类产品都可以有自己的客户决策树，用于描述从那个类别购买产品的客户的行为。类别的属性布置在树中，“最重要的”属性在树的根部，然后其余属性沿着树的分支布置。“最重要的”属性指示当从该类别中购买产品时该类别的客户首先关注的类别的属性。然后分支给出该类别的客户考虑其余属性的次序。

[0019] 图2是根据一个实施例的由系统10基于零售商的交易数据自动生成的用于酸奶产

品类别的示例CDT 200。如图2中所示，酸奶产品类别的产品属性包括尺寸、品牌、风味、生产方法等等。“尺寸”产品属性的属性值包括小、中和大。“品牌”产品属性的属性值包括主流品牌和小众品牌。“生产方法”产品属性的属性值包括有机和非有机。“风味”产品属性的属性值包括无味、主流风味和特殊风味。

[0020] CDT 200为零售商提供当购买酸奶时消费者决策过程的洞察。例如，CDT 200指出，在消费者当中，酸奶产品202的尺寸204-206一般是决策过程中最重要的因素，因为尺寸是酸奶类别下面的第一级属性值。然后，取决于优选的尺寸，品牌或生产方法被视为第二重要因素。例如，对于更喜欢小尺寸的人来说，生产方法(例如，有机210或非有机211)是第二重要因素。但是，对于喜欢中型或大型物品的人来说，品牌是第二重要因素，生产方式对决策制定过程没有任何影响。而且，对于更喜欢小尺寸酸奶产品的人的决策制定过程，风味没有任何影响，但是风味也在更喜欢来自主流品牌的中等或大尺寸酸奶产品的人当中被考虑。

[0021] 历史上，CDT的生成不是自动化过程。CDT生成的历史方法经常涉及聘请行业专家访谈客户并检查店内客户行为，然后专家将手动得出CDT。一种已知的自动化解决方案在美国专利No.8,874,499中公开，该专利通过使用来自类别的零售商历史交易数据为每个类别得出CDT。但是，这种已知的解决方案要求零售商能够使用例如消费者会员卡来按消费者分离类别的历史交易。它还要求同一客户在相对较短的时间内在该类别中进行多次购买。对交易数据的这些要求允许系统通过检查类别的客户的“切换行为”来计算属性重要性，这意味着当客户不总是坚持类别中的单个产品时，他们购买了该类别中的什么其它产品。因为这种已知的解决方案检查这种“切换行为”，所以它只能为可以通过客户识别历史交易数据的类别计算CDT，其中类别是客户通常进行多次购买的类别。否则，没有切换行为可检查。

[0022] 因此，在一些情况下，存在这些已知解决方案不适用的许多类别和许多零售商。例如，许多零售商(特别是较小的零售商)由于其高成本而没有实现会员卡计划。另外，许多零售商出售同一客户极不可能经常购买的类别。例如，这描述了大多数电子产品类别。即使是拥有许多合适类别的零售商(诸如杂货商)，也可能有不适合的类别，诸如杂货店里的锅碗瓢盆。

[0023] 相反，本发明的实施例使用即使不使用消费者忠诚度计划也实际上由每个零售商生成的数据的商品-商店-周聚合销售单位数据。因此，实施例可以被广泛的零售商使用，包括没有能力实现昂贵的忠诚卡计划的相对小的零售商。另外，实施例可以确定不频繁购买的产品类别(诸如蜂窝电话和电视机)的CDT。

[0024] 另外，实施例可以确定哪些物品一起属于类别。尽管经常很清楚类别由哪些物品组成，例如杂货店的酸奶类别，但是很多零售商的类别不太清晰。例如，在迪斯尼商店里，可能并不清楚是什么类别，因为当客户(特别是儿童)在商店买东西时，他们经常不关心物品的功能实际上是什么，只要它具有特定的迪斯尼字符就可以了。因此，举例来说，钢笔实际上可以替换蚕食(cannibalize)马克杯，因此虽然钢笔和马克杯通常是分开的物品类别，但它们不应当在迪斯尼商店。另外，对于宠物美容产品，不同类型的狗美容工具可以提供相同的功能，因此即使工具本身实际上不同，也可以相互替换蚕食。

[0025] 图3是根据一个实施例的当生成CDT时图1的CDT生成模块16的功能的流程图。在一个实施例中，图3(以及下面的图4和5)的流程图的功能通过存储在存储器或其它计算机可读或有形介质中的软件来实现，并由处理器执行。在其它实施例中，功能可以由硬件(例如，

通过使用专用集成电路(“ASIC”)、可编程门阵列(“PGA”)、现场可编程门阵列(“FPGA”),等等)或硬件和软件的任意组合执行。

[0026] 在图3中,在310处,CDT生成模块16计算每个产品对与每个属性值对之间的相似性。然后,在320处,CDT生成模块16基于来自310的相似性来生成CDT。

[0027] 图4是根据一个实施例的当在图3的310处确定相似性时图1的CDT生成模块16的功能的流程图。在310处计算相似性时,确定给定类别的每个产品对与属性值对之间的相似性。一般而言,实施例首先以来自例如POS终端100的销售数据的形式接收数据元素。然后数据被聚合,然后计算每周销售份额。然后,针对属性值对执行相似性计算。

[0028] 至于数据元素,在402处在交易级别(即,交易ID/客户ID/商店/日期/物品级别)接收销售数据。交易是由消费者标识(“ID”)、交易ID、商店ID、日期和购买的物品以及附带的信息(诸如售出的单元的数量、以$为单位的销售金额以及物品的销售价格)的组合识别的销售的发生。大多数POS系统为个体零售商店容易地提供这些信息。下面的表1图示了交易数据的示例,示出了给定日期给定商店(即,商店ID为142)购买同一物品(即,物品ID为2345)的不同消费者。

[0029]

交易_ID (transaction_ID)	客户_ID (customer_ID)	商店_ID (store_ID)	物品_ID (item_ID)	日期	单元销售量	销售金额	销售价格
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

[0030] 表1

[0031] 在404处,数据然后被聚合到物品/周级别。在其它实施例中,可以使用不同于周的持续时间/测量(例如,日、月等等)。在一个实施例中,对于那个给定的物品/商店/周的所有交易ID和客户ID,交易级别的数据被聚合到物品/商店/周级别。销售单元和$现在反映这个级别。销售价格现在被定义为加权平均价格:销售$总额/售出的单元的总和。使用上述表1中的示例,对于2015年5月16日结束的周,聚合的物品/商店/周级别数据现在变成表2中所示的数据。

[0032]

商店_ID store_ID	物品_ID item_ID	日期	单元销售量	销售金额	加权价格
142	2345	5/16/2015	111	$972.89	$ 8.76

[0033] 表2

[0034] 在404处,数据被进一步聚合到属性-值/商店/周级别。在其它实施例中,可以使用不同于周的持续时间/测量(例如,日、月等等)。在一个实施例中,每个物品具有产品属性类

型和值，并且它们的集体销售被反映在这个级别。属性类型的示例是风味(例如，“草莓”或“香草”值)、尺寸(例如，“小”、“中”或“大”值)、品牌(例如，“Coke”或“Pepsi”的值)，等等。下面的表3是显示针对风味属性的销售的示例。

[0035]

商店_ID store_ID	风味值	日期	单元销售量	销售金额	销售价格
2345	风味 1	5/16/2015	111	$972.89	$8.76
2345	风味 2	5/16/2015	23	$184.23	$8.01
2345	风味 3	5/16/2015	133	$1,243.55	$9.35
2345	风味 3	5/23/2015	78	$692.64	$8.88
2345	风味 3	5/30/2015	45	$413.55	$9.19

[0036] 表3

[0037] 使用聚合数据，接下来在406处，实施例确定每周销售份额，或者如果不是每周，就确定在相关时间测量期间的销售份额。在一个实施例中，每周销售份额是属于属性值/商店/周的销售与同一商店/周中相同属性类型的所有其它属性值相比的百分比。对于给定的商店/周，用于给定属性类型的销售份额的总和最多为100％。实施例确定数据历史中用于所有属性类型/商店/周的每周销售份额。

[0038] 继续上面的示例，下面的表4示出，对于5/16/15的一周，销售份额＝一种风味的单元销售量/一周的总单元销售量。

[0039]

商店_ID store_ID	风味值	日期	单元销售量	销售份额
2345	风味 1	5/16/2015	111	41.6%
2345	风味 2	5/16/2015	23	8.6%
2345	风味 3	5/16/2015	133	49.8%
		合计	267	100%

[0040] 表4

[0041] 还跨商店/周为所有物品计算每周销售份额。下面的表5示出了示例。

[0042]

商店_ID store_ID	物品_ID item_ID	日期	单元销售量	销售份额
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%
		合计	83	100%

[0043] 表5

[0044] 在408处，实施例然后确定属性-值对的相似性。在一个实施例中，跨其销售份额历史记录中的属性类型内计算相似性，并使用Pearson相关公式如下计算：

[0045]
$$SIM(X,Y)=\frac{\sum_{i=1}^{n}X_iY_i-\frac{(\sum_{i=1}^{n}X_i)(\sum_{i=1}^{n}Y_i)}{n}}{\sqrt{\left(\sum_{i=1}^{n}X_i{}^2-\frac{\left(\sum_{i=1}^{n}X_i\right)^2}{n}\right)\left(\sum_{i=1}^{n}Y_i{}^2-\frac{\left(\sum_{i=1}^{n}Y_i\right)^2}{n}\right)}} \qquad (等式1)$$

[0046] 其中对于风味对(X,Y)，X_i和Y_i分别表示风味X和Y的商店/周份额值，并且n表示存在X和Y风味份额的商店/周的总数。

[0047] 实施例为所有风味对(X,Y)计算SIM(X,Y)。这些相似性构成“风味相似性”。上面示出的用于SIM的公式将始终产生介于-1和1之间的数字。对于属性值X和Y，SIM接近-1意味着X和Y的份额是“反相关的”，这意味着当X的份额上升时，Y的份额下降，反之亦然。因此，当客户购买更多的X时，他们购买的Y减少(反之亦然)，因此X和Y必须与客户相似，因为它们是彼此的替代。越接近-1，彼此替代的X和Y越多。以相同的方式，实施例还计算每个其它属性的相似性，并因此获得例如“品牌相似性”、“尺寸相似性”等等。

[0048] 在一个实施例中，使用以下伪代码，使用SQL中的内置函数“corr”来计算上述相关性：

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

[0049]

[0050] 结果如下表6中所示：

[0051]

风味_x	风味_y	风味_相似度
风味_1	风味_1	1.00
风味_1	风味_2	-0.45
风味_1	风味_3	-0.15
风味_2	风味_2	1.00
风味_2	风味_3	0.05
风味_3	风味_3	1.00

[0052] 表6

[0053] 对于物品对重复类似的过程，其中X和Y表示两个不同的物品(而不是如上所述的属性值)，并且因此X_I和Y_i分别表示物品X和物品Y在特定商店/周的物品份额。因此，实施例针对每个物品对(X,Y)计算SIM(X,Y)，正如实施例针对属性的每对属性值计算SIM(X,Y)一样，具有下面表7中所示的以下示例结果：

[0054]

物品_x	物品_y	物品_相似度
2345	2345	1.00
2345	5791	-0.34
2345	9876	0.21
5791	5791	1.00
5791	9876	-0.56
9876	9876	1.00

[0055] 表7

[0056] 在408处，实施例还执行用于二元属性的相似性计算。二元属性是只有两个值的属性。这些是相当普遍的，并且通常指示存在或不存在某个特性。下面使用的一个示例是“有机”(即，食物物品是有机的或不是有机的)。二元属性需要特殊处理，因为，如果简单地应用

上面给出的用于SIM的公式，那么结果将始终是SIM＝-1，这不提供关于购物者如何处理属性的信息。

[0057] 相反，对于只有两个值可供选择的属性类型(例如，有机和非有机食物)，相关性如下计算：

[0058] $$2\sqrt{\frac{\sum_{k=1}^{N}(x_k-\bar{x})^2}{N}}$$ （等式2）

[0059] 其中x_k是周k内的有机份额，并且有N周。$\bar{x}$是x_i的平均值，即，N周的平均有机份额。因此，等式2是x_k的标准偏差的2倍，并且正在测量有机份额偏离平均有机份额的波动。一般而言，波动越大，客户用有机换非有机的情况就越多(反之亦然)，因此有机与非有机越相似。如果将x_k代替地用作非有机份额(并且$\bar{x}$用作平均非有机份额)，那么将导致相同的数字。乘数2被用来使测量从0变到1(否则测量将从0变到1/2，因为如果x_k介于0和1之间(在这里就是这样，因为它们是份额)，那么1/2是标准偏差的最大值)。

[0060] 以下SQL伪代码可以被用来执行二元属性的相似性：

```
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

from
(select
```

[0061]

```
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,

sales_share_organic_values_table a

group by
n_wks
```

[0062] 下面表8中示出了用于二元属性的相似性计算的示例结果：

[0063]

有机_相似性	非有机_相似性
0.43	0.43

[0064] 表8

[0065] 在410处，实施例然后对SIM值进行后处理。在用于属性对和物品对的SIM值中，实施例如下修改SIM值：如果SIM值为正，那么将其设置为0；如果它为负，那么使其为正。对于本公开的剩余部分，所使用的SIM值是经过后处理的SIM值。由于上面的等式2确保那些已经是非负的，因此410处的后处理不用于二元属性类型的相似性。

[0066] 在412处，实施例然后通过将每个属性的SIM值与物品SIM值进行比较来找出“最有意义的属性”。实施例确定哪个属性最好地解释客户的物品级别购买行为。将物品级别的SIM值与每个属性的SIM值进行比较，并且找到其SIM值最接近“匹配”(下面公开)物品级别值的属性。

[0067] 对于诸如风味(Flavor)之类的特定属性，实施例将物品和属性SIM值编译到一个表中，如下面表9中所示。风味_x(flavor_x)列给出物品_x(item_x)的风味，同样风味_y(flavor_y)给出物品_y(item_y)的风味。风味_相似性(flavor_similarity)给出风味_x和风味_y的SIM值。要注意的是，如果风味_x和风味_y相同(因为物品_x和物品_y具有相同的风味)，那么风味_相似性等于1，因为风味相同。否则，它只是风味_x和风味_y的SIM值，如前所述计算。

[0068]

物品_x	风味_x	物品_y	风味_y	物品_相似性	风味_相似性
4563	风味_1	1200	风味_3	0.58	0.45
4563	风味_1	2345	风味_1	0.82	1.00
4563	风味_1	4563	风味_1	1.00	1.00
4563	风味_1	5665	风味_2	0.67	0.68
4563	风味_1	5698	风味_4	0.65	0.21
4563	风味_1	8758	风味_1	0.02	1.00
4563	风味_1	9901	风味_2	0.10	0.68
5665	风味_2	1200	风味_3	0.05	0.50
5665	风味_2	2345	风味_1	0.98	0.68
5665	风味_2	5665	风味_2	1.00	1.00
5665	风味_2	5698	风味_4	0.68	0.29
5665	风味_2	8758	风味_1	0.34	0.68
5665	风味_2	9901	风味_2	0.58	1.00
1200	风味_2	1200	风味_3	1.00	0.50
1200	风味_2	5698	风味_4	0.12	0.29
1200	风味_3	8758	风味_1	0.24	0.45
…	…	…	…	…	…

[0069]

[0070] 表9

[0071] 实施例然后使用以下SQL伪代码对物品和属性相似性运行相关性计算(在表9的示例中，这将参考物品_相似性(item_similarity)和风味_相似性(flavor_similarity)值)。这意味着在物品_相似性和风味_相似性列上运行关联：

[0072]
```
select
        corr(item_similarity, flavor_similarity) as flavor_result
from
        item_flavor_similarities
```

[0073] 结果如下表10中所示：

[0074]

风味_结果（flavor_result）
0.0804

[0075] 表10

[0076] 实施例然后针对所有属性重复并编译结果，如表11的以下示例中所示：

[0077]

	属性_结果
品牌_结果（brand_result）	0.1559
有机_结果（organic_result）	0.1235
尺寸_结果（size_result）	0.0912
风味_结果（flavor_result）	0.0804

[0078] 表11

[0079] 具有最大值的属性被认为在CDT中具有最大的意义，并且因此将是在图3的320处生成的CDT的顶级属性。为了添加到CDT，重复图4的功能，以产生CDT的其它级别和分支。例如，一旦确定“Brand”是最高属性，图4的功能就针对Brand属性中的每个品牌执行，但是仅使用在402处接收的在特定品牌内的数据元素的子集。

[0080] 图5是根据一个实施例的当基于相似性生成CDT时(图3的320)图1的CDT生成模块16的功能的流程图。在510处，确定相同产品类别的产品中是否存在任何适合功能的属性。适合功能的属性是产品属性，对其跨其值进行置换是极不可能的。例如，购买雨刷的客户必须购买适合对应汽车的雨刷。因此，在雨刷产品类别中，“尺寸”产品属性被确定为适合功能的属性。“尺寸”产品属性也可以是其它产品类别的适合功能的属性，例如轮胎、空气过滤器、真空袋、打印机墨盒等等。但是，相同的“尺寸”产品属性对于其它产品类别可以不是适合功能的属性，例如水果、软饮料等等。一般而言，适合功能的属性通常存在于诸如配件等等的非杂货物品中。在一个实施例中，适合功能的属性直接从生成的客户数据获得，并且通常不需要计算。例如，零售商通常将明确识别“适合功能的”属性是什么，例如，明确指出在雨刷的情况下尺寸是适合功能的属性。

[0081] 在识别了所有适合功能的属性后，适合功能的属性被自动直接放在产品类别下CDT的顶级处。图6图示了根据一个实施例的由CDT生成模块16生成的CDT 600。CDT 600具有类别级别610，用于识别产品类别。对于酸奶产品类别，将在类别级别610中显示“Yogurt”，

如图2中所示。在另一个示例中，对于“Coffee”类别，在类别级别610中显示“Coffee”。然后，适合功能的属性被放在CDT 600的顶级620处。图6示出了顶级620处的两个适合功能的属性(FA1，FA2)622、624。但是，对于Yogurt或Coffee，可能不存在任何适合功能的属性。

[0082] 在图5的520处，然后识别最有意义的属性或拆分属性。最有意义的属性根据图4的功能来确定。

[0083] 在530处，物品被分成子部分，其中每个子部分与在520处识别出的属性的特定属性值对应。例如，当在520处“形式”产品属性被确定为咖啡的最重要属性时，“形式”产品属性被分成三个子部分，每个子部分与用于咖啡的形式的特定值对应：“Bean”、“Ground”和“Instant”。子部分形成顶级620下面的下一级别630，如图6所示。例如，图6示出了级别630中的两个子部分(A1a，A1b)632、634，其从适合功能的属性622分出。对于每个子部分重复520和530，并且扩展CDT 600，直到对于每个子部分达到终端节点为止(540处的否)。如果对于每个子部分最终到达终端节点(540处的是)，那么该过程终止。

[0084] 如所公开的，树被扩展直到识别出终端节点。在一个实施例中，将节点声明为终端的标准如下：

[0085] 1.没有识别出有意义的属性。

[0086] 2.节点中物品的数量<产品类别中所有物品的x％，其中“x”是调整参数，用于限制树的尺寸。在一个实施例中，x的默认值是10。

[0087] 3.子节点的平均不相似性(“AD”)(即，节点中所有可能的产品对的平均值)大于其父节点。两种可能的子情况如下：

[0088] a.如果所有子节点的AD值都大于父节点，那么将父节点声明为终端节点。

[0089] b.如果一些子节点的AD值大于父节点，那么那些节点将被终止，并且其它子节点将以常规方式扩展。

[0090] 如所公开的，实施例在仅依赖物品-商店-周聚合销售单位数据的同时生成CDT。这些数据一般从每个零售商处可获得，而不管其类别如何，因为物品-商店-周聚合销售单元数据仅仅是在每个商店每个物品的售出单元数量的周总数。因此，不需要更困难或更昂贵地获取数据(诸如客户的身份)。

[0091] 另外，来自聚合数据的已知CDT生成系统一般依赖于更标准的统计方法，这些方法尽管是标准的，但在计算CDT时有缺点。这些已知的方法可以需要非常大量的计算能力，并且可能难以实施。相反，实施例可以用标准SQL查询来实现，并且即使在大型客户数据集上也可以非常快速地运行。

[0092] 另外，实施例处理仅具有两个值的属性(称为布尔属性)。这些属性在很多类别中相当常见，因为它们表明该类别中物品的某个属性的存在或不存在(例如，酸奶是否是希腊酸奶，或者洗发水是否是低致敏性的)。

[0093] 本文具体地图示和/或描述了几个实施例。但是，将理解的是，在不脱离本发明的精神和预期范围的情况下，所公开的实施例的修改和变化被上述教导涵盖并在所附权利要求的范围内。

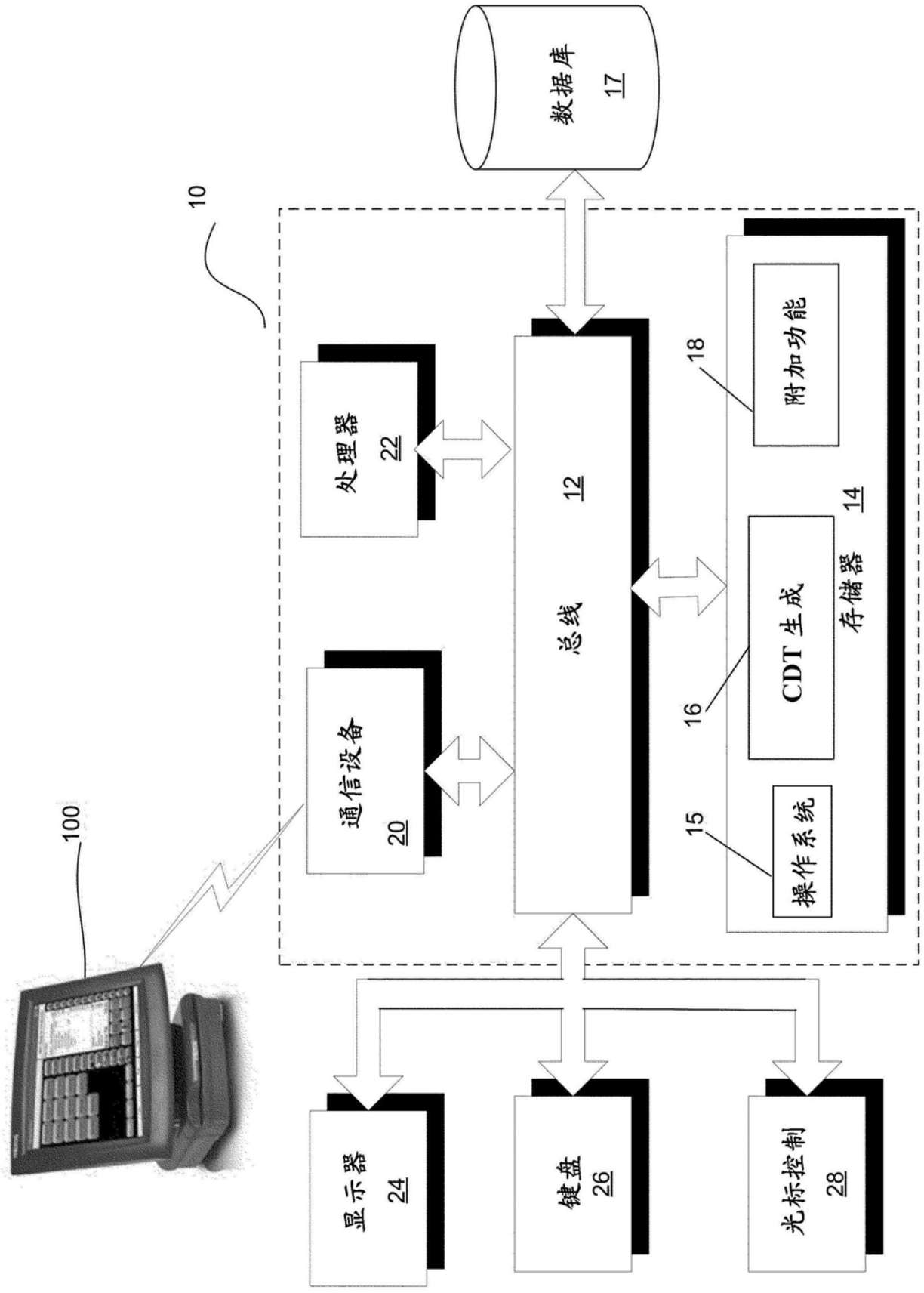

数据库
17
10
处理器
22
总线
12
附加功能
18
CDT 生成
存储器
14
16
通信设备
20
操作系统
15
100
显示器
24
键盘
26
光标控制
28

图1

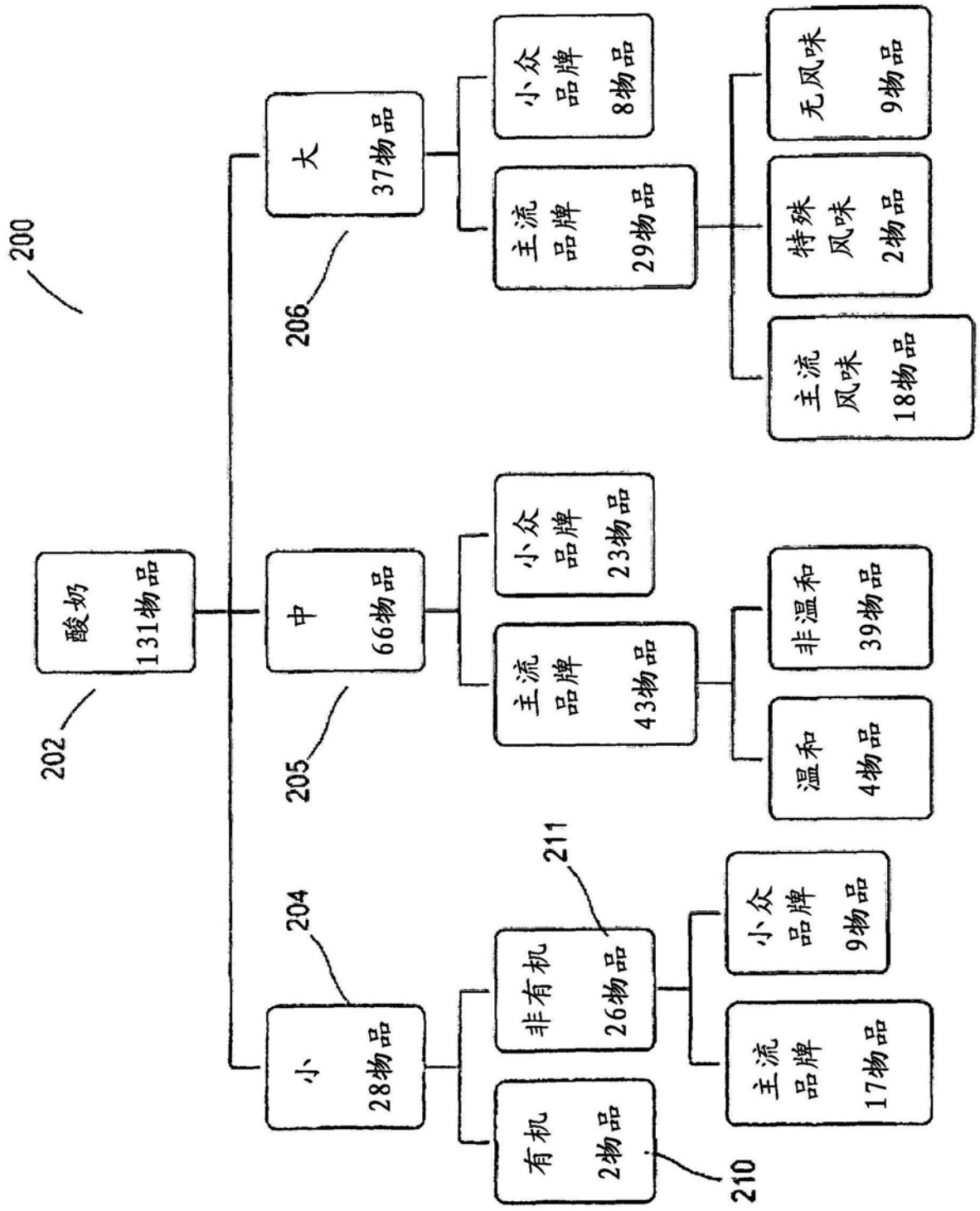

200
酸奶
131物品
202
大
37物品
206
小众
品牌
8物品
主流
品牌
29物品
无风味
9物品
特殊
风味
2物品
主流
风味
18物品
中
66物品
205
小众
品牌
23物品
主流
品牌
43物品
非温和
39物品
温和
4物品
小
28物品
204
非有机
26物品
211
小众
品牌
9物品
主流
品牌
17物品
有机
2物品
210

图2

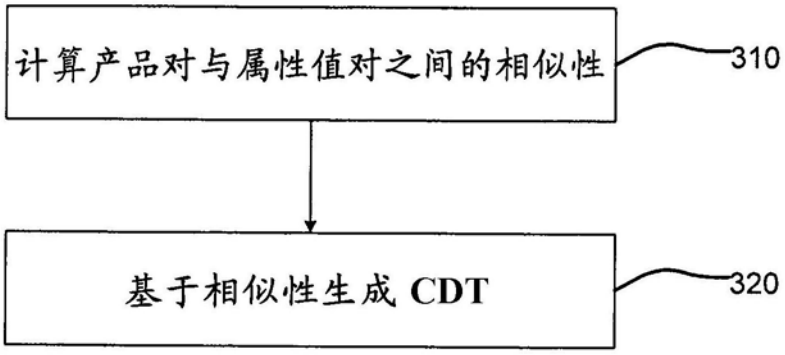


图3

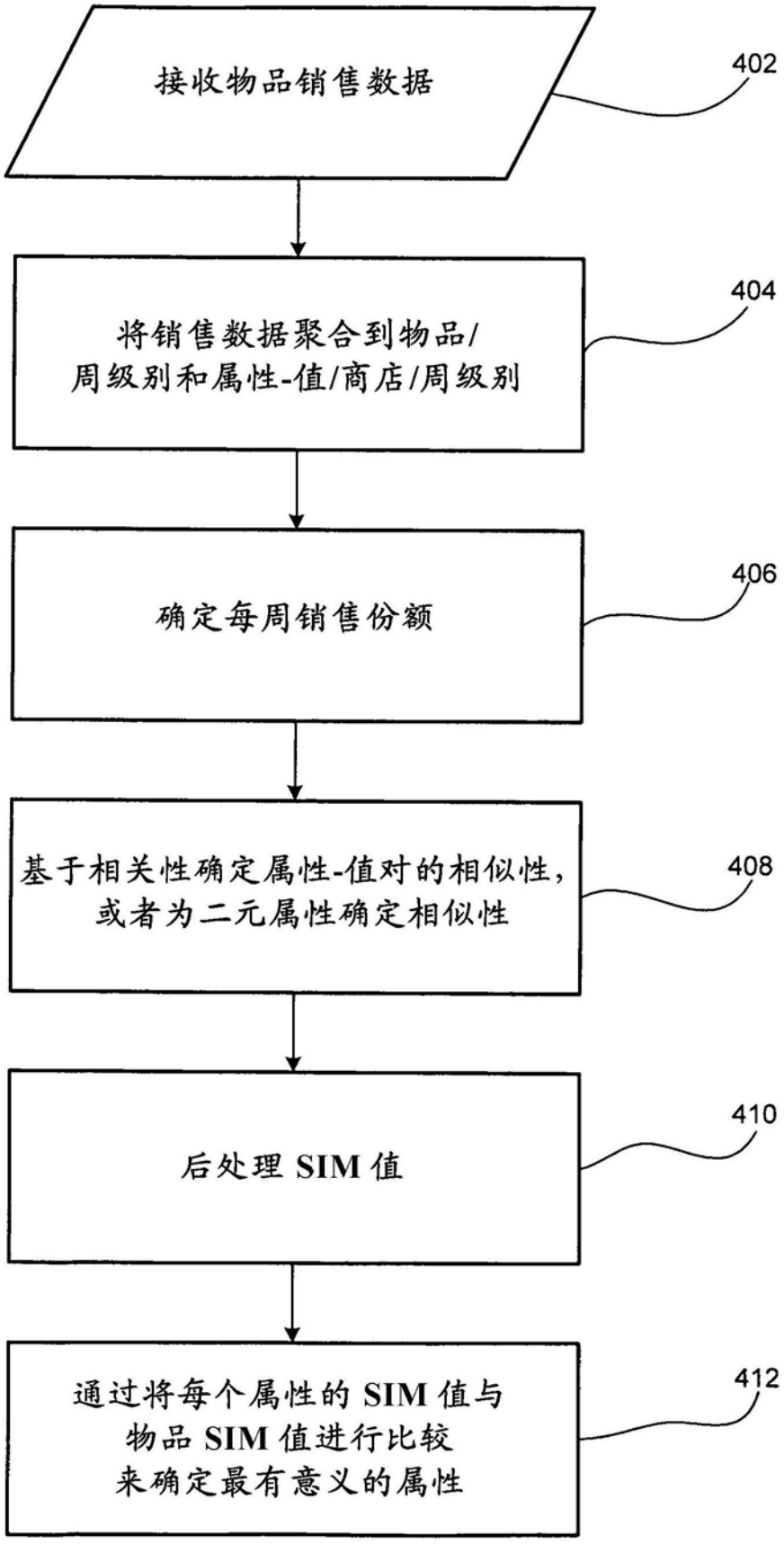

接收物品销售数据
402
将销售数据聚合到物品/
周级别和属性-值/商店/周级别
404
确定每周销售份额
406
基于相关性确定属性-值对的相似性，
或者为二元属性确定相似性
408
后处理 SIM 值
410
通过将每个属性的 SIM 值与
物品 SIM 值进行比较
来确定最有意义的属性
412

图4

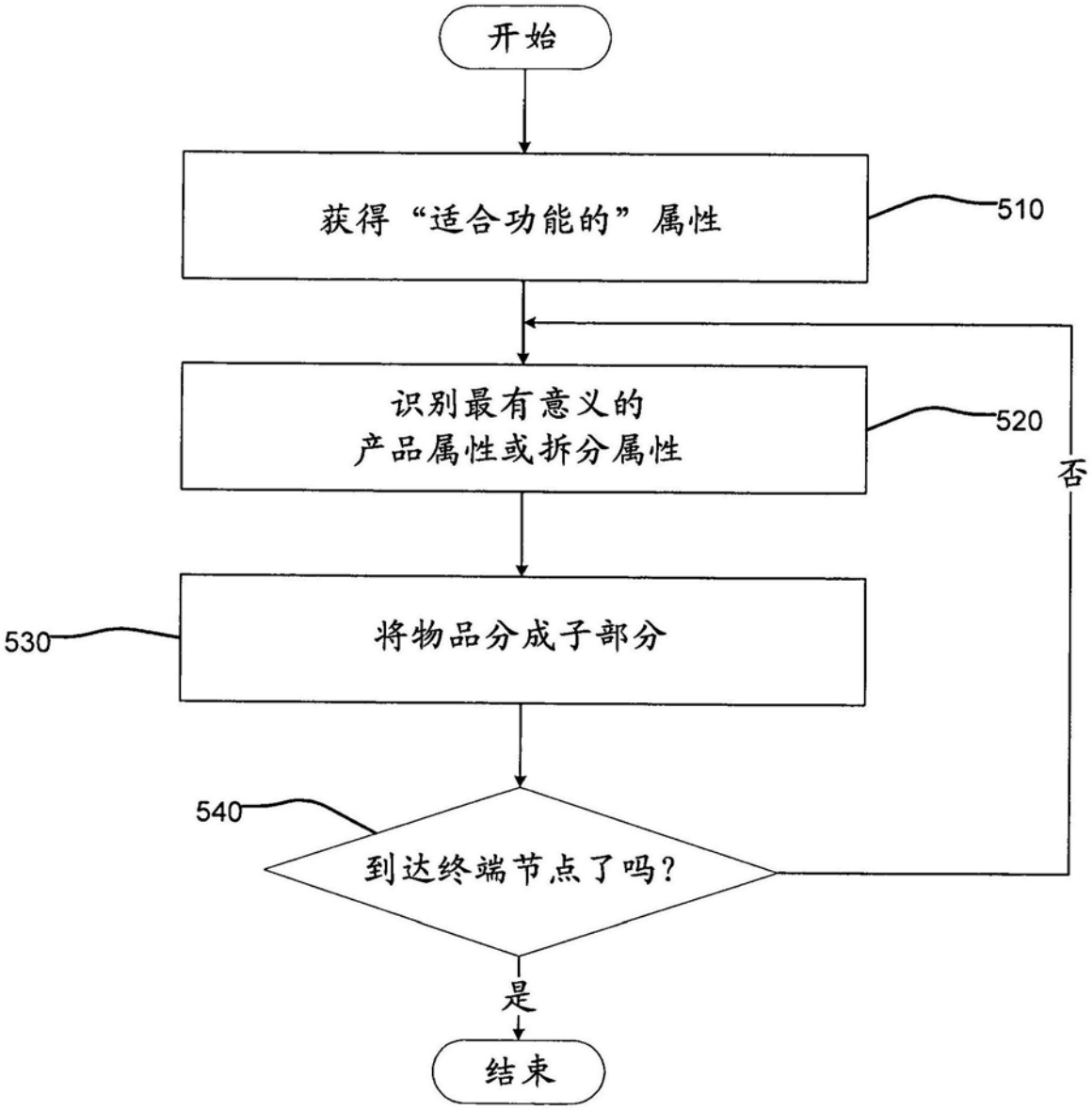

开始
获得“适合功能的”属性
510
识别最有意义的
产品属性或拆分属性
520
否
530
将物品分成子部分
540
到达终端节点了吗?
是
结束

图5

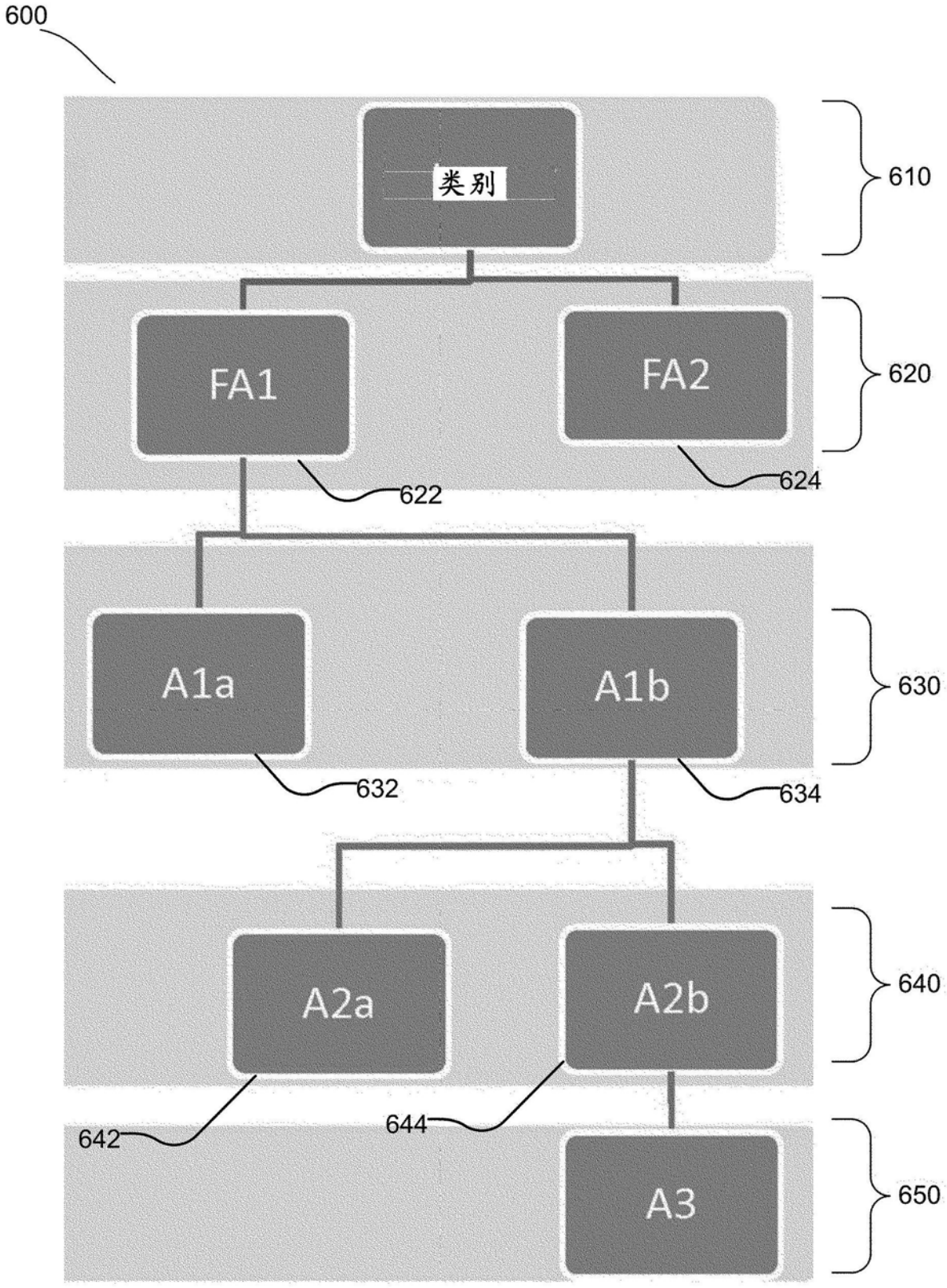

600
类别
610
FA1
FA2
620
622
624
A1a
A1b
630
632
634
A2a
A2b
640
642
644
A3
650

图6