



(12)发明专利申请

(10)申请公布号 CN 106605222 A

(43)申请公布日 2017.04.26

(21)申请号 201580047313.5

V·泽勒温斯基 T·R·斯波尔丁

(22)申请日 2015.09.22

(74)专利代理机构 中国国际贸易促进委员会专

利商标事务所 11038

(30)优先权数据

代理人 李晓芳

62/054,517 2014.09.24 US

14/678,218 2015.04.03 US

(85)PCT国际申请进入国家阶段日

(51)Int.Cl.

G06F 17/30(2006.01)

2017.03.03

(86)PCT国际申请的申请数据

PCT/US2015/051462 2015.09.22

(87)PCT国际申请的公布数据

W02016/049034 EN 2016.03.31

(71)申请人 甲骨文国际公司

地址 美国加利福尼亚

(72)发明人 U·谢费尔 A·C·波科克

B·史蒂文斯 M·艾沙奎

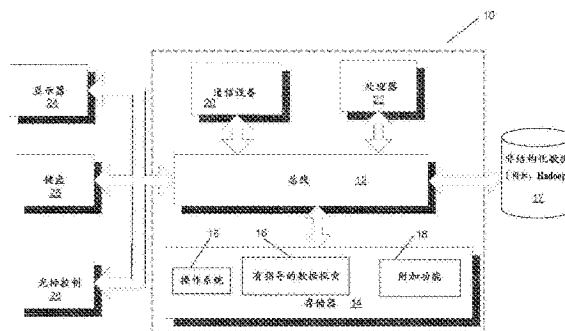
权利要求书2页 说明书6页 附图6页

(54)发明名称

有指导的数据探索

(57)摘要

用于探索数据的系统从数据库接收数据并且在服务器中索引数据。该系统显示来自被索引的数据的一个或多个可选择的数据集,每个被选择的数据集包括一个或多个属性。然后该系统按照兴趣度级别对一个或多个属性进行排序并显示经排序的属性。



1. 一种探索数据的方法,该方法包括:
从数据库接收数据;
在所述服务器中索引所述数据;
显示来自被索引的数据的一个或多个可选择的数据集,每个数据集包括一个或多个属性;
利用所述属性的熵,按照兴趣度级别对所述一个或多个属性进行排序;以及
显示经排序的属性。
2. 如权利要求1所述的方法,其中所述数据库中的至少一些数据是非结构化数据。
3. 如权利要求1所述的方法,其中排序包括:
对确定的熵范围应用低端截止并且对所述确定的熵范围应用高端截止,以及基于所述低端截止和所述高端截止对所述熵进行过滤;以及
将过滤后的熵的低值和高值映射到高兴趣度级别,并且将中间值映射到低兴趣度级别。
4. 如权利要求1所述的方法,其中显示经排序的属性包括显示用于每个经排序的属性的分块,所述分块包括所述属性的可视化。
5. 如权利要求1所述的方法,还包括:
接收对细化的选择,所述细化来自经排序的属性的显示,并且,作为响应:
将选择的细化应用到所述数据;以及
对经细化的数据的一个或多个属性进行重新排序。
6. 如权利要求1所述的方法,还包括对从所述数据库接收的数据执行元数据发现。
7. 如权利要求1所述的方法,还包括接收对所述可选择的数据集中的一个或多个数据集的选择。
8. 如权利要求1所述的方法,其中所述属性中的每一个由数据库列组成。
9. 一种其上存储有指令的计算机可读介质,当所述指令由处理器执行时,使所述处理器提供有指导的数据探索,所述提供包括:
从数据库接收数据;
在服务器中索引所述数据;
显示来自被索引的数据的一个或多个可选择的数据集,每个数据集包括一个或多个属性;
利用所述属性的熵,按照兴趣度级别对所述一个或多个属性进行排序;以及
显示经排序的属性。
10. 如权利要求9所述的计算机可读介质,其中所述数据库中的至少一些数据是非结构化数据。
11. 如权利要求9所述的计算机可读介质,其中排序包括:
对确定的熵范围应用低端截止并且对所述确定的熵范围应用高端截止,以及基于所述低端截止和所述高端截止对熵进行过滤;以及
将过滤后的熵的低值和高值映射到高兴趣度级别,并且将中间值映射到低兴趣度级别。
12. 如权利要求9所述的计算机可读介质,其中显示经排序的属性包括显示用于每个经

排序的属性的分块,所述分块包括所述属性的可视化。

13. 如权利要求9所述的计算机可读介质,所述提供还包括接收对细化的选择,所述细化来自经排序的属性的显示,并且,作为响应:

将选择的细化应用到所述数据;及

对经细化的数据的一个或多个属性进行重新排序。

14. 如权利要求9所述的计算机可读介质,所述提供还包括对从所述数据库接收的数据执行元数据发现。

15. 如权利要求9所述的计算机可读介质,所述提供还包括接收对所述可选择的数据集中的一个或多个数据集的选择。

16. 如权利要求9所述的计算机可读介质,其中所述属性中的每一个由数据库列组成。

17. 一种有指导的数据探索系统,所述有指导的数据探索系统包括:

索引模块,所述索引模块从数据库接收数据并在服务器中索引所述数据;

显示模块,所述显示模块显示来自被索引的数据的一个或多个可选择的数据集,其中每个可选择的数据集包括一个或多个属性;以及

排序模块,所述排序模块利用所述属性的熵,按照兴趣度级别对所述一个或多个属性进行排序;

其中所述显示模块还显示经排序的属性。

18. 如权利要求17所述的系统,其中排序包括:

对确定的熵范围应用低端截止并且对所述确定的熵范围应用高端截止,以及基于所述低端截止和所述高端截止对熵进行过滤;以及

将过滤后的熵的低值和高值映射到高兴趣度级别,并且将中间值映射到低兴趣度级别。

19. 如权利要求17所述的系统,其中所述排序模块还接收对细化的选择,所述细化来自经排序的属性的显示,并且,作为响应:

将选择的细化应用到所述数据;以及

对经细化的数据的一个或多个属性进行重新排序。

20. 如权利要求1所述的方法,其中所述排序模块还接收对所述可选择的数据集中的一个或多个数据集的选择。

有指导的数据探索

[0001] 对相关申请的交叉引用

[0002] 本申请要求于2014年9月24日提交的临时专利申请序列No.62/054,517的优先权，其内容通过引用被结合于此。

技术领域

[0003] 一个实施例一般而言涉及计算机系统，并且具体而言涉及通过有指导的数据探索(data exploration)提供对数据的分析的计算机系统。

背景技术

[0004] 近年来，计算机系统已经使个人和组织能够捕获和存储大量的数据。数据的激增有时被称为“大数据”，这是对于如此大或复杂以至于难以利用传统的数据处理应用来处理的数据集的任何集合的全包含术语。

[0005] 大数据的挑战包括分析、捕获、管护(curation)、搜索、共享、存储、传送、可视化和隐私问题。这种大量数据的存在导致对分析和探索数据的改进系统和方法的不断增长的需求。

发明内容

[0006] 一个实施例是用于探索数据的系统。该系统从数据库接收数据并在服务器中索引数据。该系统显示来自被索引的数据的一个或多个可选择数据集，其中所选择的数据集包括一个或多个属性。然后，该系统按照兴趣度级别对这一个或多个属性进行排序并显示经排序的属性。

附图说明

[0007] 图1是根据本发明实施例的计算机服务器/系统的框图。

[0008] 图2是根据本发明一个实施例的图1的有指导的数据探索模块以及其它元件的功能的流程图。

[0009] 图3示出了根据一个实施例的显示可用数据集的示例用户界面。

[0010] 图4示出了根据实施例在选择图3的“葡萄酒销售”数据集之后显示被示为排序的“分块(tile)”的属性的示例用户界面。

[0011] 图5示出了根据一个实施例与规格化的熵相关的兴趣度的线图。

[0012] 图6是根据本发明一个实施例的图1的有指导的数据探索模块以及其它元件的功能的流程图。

具体实施方式

[0013] 本发明的实施例提供有指导的数据探索。其中一个实施例生成显示被索引的数据集(即，数据或数据记录的集合)并允许用户选择其中一个数据集的用户界面。作为响应，该

用户界面按照兴趣度的次序显示所选择的数据集的属性。然后用户还可以选择属性的一个或多个值,这导致数据集的细化,以帮助用户选择和显示期望的数据。

[0014] 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的一个或多个组件。例如,对于用户客户端的功能,系统10可以是包括处理器、存储器和显示器的智能电话,但是可以不包括图1中所示的一个或多个其它组件。

[0015] 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可以包括以下的任何组合:随机存取存储器(“RAM”)、只读存储器(“ROM”)、诸如磁盘或光盘的静态存储器、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通过任何其它方法与系统10接口。

[0016] 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、可移除和不可移除介质以及通信介质。通信介质可以包括计算机可读指令、数据结构、程序模块或调制的数据信号(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。

[0017] 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。

[0018] 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。模块还包括用于提供有指导的数据探索功能以及本文公开的所有其它功能的有指导的数据探索模块16。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功能模块18以包括附加功能,诸如企业资源规划(“ERP”)模块,该ERP模块生成要由模块16探索的ERP数据。数据库17耦合到总线12,以提供用于模块16和18的集中式存储并且存储非结构化数据、结构化数据等。

[0019] 每个模块可以包括多个模块。在一个实施例中,模块16包括从数据库接收数据并索引数据的索引模块,显示来自被索引的数据的一个或多个可选择数据集的显示模块,以及接收对一个或多个可选择数据集的选择并通过兴趣度级别对一个或多个属性进行排序的排序模块。

[0020] 图2是根据本发明一个实施例的有指导的数据探索模块16和其它元件的功能的流程图。在一个实施例中,图2和图6的流程图的功由存储在存储器或其它计算机可读或有形介质中并由处理器执行的软件实现。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。

[0021] 在202,非结构化或部分结构化的数据存储存储在图1的数据库17中。在一个实施例中,数据存储存储在Apache Hive中,Apache Hive是在Hadoop之上构建的数据仓库基础设施,用于提供数据汇总、查询和分析。但是,在202可以接收并存储任何类型的非结构化或结构化数据。例如,在一个实施例中,数据是大量的Twitter“推文”。在一个实施例中,数据存储存储在Hadoop分布式文件系统(“HDFS”)中。

[0022] 在204,对来自202的数据执行元数据发现,然后根据元数据发现的结果来丰富数据。在一个实施例中,执行元数据发现,以从数据本身确定数据的本质/类型(例如,数据是数字、地理代码等),包括数据的基数。在一个实施例中,数据发现由来自Oracle公司的“Endeca Information Discovery Studio Provisioning Service (Endeca信息发现工作室供应服务)”执行。元数据发现生成馈送到用户界面中的元数据的列表。在一个实施例中,不执行丰富。在另一个实施例中,不执行类型发现和丰富。

[0023] 下面的表1提供了一些示例元数据:

[0024]

名称	描述
Id	唯一识别数据集的值
collectionKey	包含用于这个数据集的数据的集合的名称
sourceType	这个数据集从哪里获得其数据
sourceName	这个数据集从中获得其数据的数据库/表/文件的名称
displayName	这将作为数据集名称被显示
description	数据集的描述
timesViewed	这个数据集已经被用户查看的次数
timesFavorited	这个数据集已经被用户喜欢的次数

[0025] 表1

[0026] 在206,来自204的数据被索引到服务器中并且发布到用户界面。在一个实施例中,在206处在来自Oracle公司的“MDEX”引擎中将数据索引为Endeca索引。

[0027] 在208,生成显示所有可用数据集/数据记录或在服务器上索引的数据集的目录的用户界面。图3示出了根据一个实施例的、在301处显示可用数据集的示例用户界面300。总共有675个数据集经由接口可用(图3中并未全部示出)。每个数据集可由用户选择。在一个实施例中,仅单个数据集可用并在208处示出。还可以在208处接收对其中一个数据集的选择。在一个实施例中,在208处不需要选择。

[0028] 响应于在208处对数据集的选择,在210处按照兴趣度级别对属性进行排序,并且在212处为所选择的数据集生成“探索者”用户界面。列出数据集的属性,以供进一步探索/选择。图4示出了根据实施例在选择图3的“葡萄酒销售”数据集305之后显示被示为经排序的“分块”的属性的示例用户界面400。示例分块包括“交货日期”分块401、“邮政编码”分块402、“交货的日期”分块403等。每个分块汇总来自所选择的数据集的变量/属性,并且分块按照兴趣度级别排序(即,用户有多大可能将找到每个属性、以及用于那个感兴趣的属性的对应代表性分块的预测指标的计算),使得最能解释数据集的属性首先被显示。在一个实施例中,如下面详细描述,使用熵来确定按照兴趣度级别的排序。在另一个实施例中,按字母顺序对代表性分块进行排序。在另一个实施例中,代表性分块按照分块在数据中出现的次序列出(即,没有排序)。

[0029] 用户界面400还包括在410处在左侧示出的可用细化的列表。作为在212处生成分块的一部分,可视化的类型(例如,条形图、线图、地图等)也被确定。在一个实施例中,硬编码的决策树被用来确定可视化的类型。在图4中,每个分块表示数据集的属性(即,数据库属性的列)。用户还可以放大特定的分块/列中。因此,用户可以一目了然地看到包括元数据的

每个特定列/分块的视图。

[0030] 再次参考图2,在214,用户从可用细化的列表410中选择(或取消选择)来自一个或多个属性的一个或多个值,从而细化数据集。

[0031] 在216,基于当前数据集,再次按照兴趣度级别对属性/分块进行排序,并且在212处生成示出可选择的数据记录的修订集合的用户界面。214和216可以根据需要被重复多次,以使得从数据集中选择记录是迭代地可细化的。

[0032] 图2的实施例被认为是“单变量”,因为每列被单独考虑。

[0033] 在一个实施例中,在图2的210处,使用熵来确定分块/属性的兴趣度级别。“熵”是随机变量中的不确定性的度量。与熵一起使用的典型度量单位是“位(bit)”。随机变量的结果越不确定,就需要越多的位来表示不同的值。

[0034] 作为示例,假设需要记录可以是“1000”或“2000”的实验的结果。首先,由于只有两个可能的结果,因此不需要整个存储器(逐位(bit-wise))表示“1000”或“2000”。相反,可以使用“0”指示前者 and “1”指示后者的约定。因此,结果的真实表示只有1位,并且形成这个随机变量的熵的上限。

[0035] 第二,如果实验结果的概率是已知的,则熵值可以进一步减小,因为这个变量固有的不确定性已经减小。例如,抛一个没有偏见的硬币产生反面结果或正面结果的相等0.5机会(概率)。由于不确定性高,因此熵将反映其最高值(即,1)。但是,如果结果记录妇女是否怀孕,并且已知孕妇占妇女人口的5%,则熵将下降并且指示0.2864位的值。

[0036] 数据集(诸如图3的数据集301)的每个属性具有其自己的熵值(也被称为“自身信息”)并且表示由这个属性给出多少信息。这个信息(也被称为“信息增益”)可以基于属性的不确定性来确定。例如,如果特定属性的所有值都相同(单值结果),则该属性不携带信号,并且因此具有零信息增益。

[0037] 对于单个属性计算的熵值是非负数并且范围从0到 $\log|x|$ (表示为“x的计数的对数”),其中 $|x|$ 是这个属性的不同值的个数。对于要以位表示的熵的值,以2为底取对数。在这种情况下,例如,具有四个相等分布值的列携带 $\log_2 4 = 2$ 位的信息。作为不同的示例,对于其结果总是为零的变量,该变量仅具有一个结果值并且不携带信息,其熵值为 $\log_2 1 = 0$ 。为了提供另一个示例,如果x表示在0和65536之间均匀分布的值,每个值的不确定性高。在这个例子中,有 $65536 = 2^{16}$ 个结果,并且因此有16位的熵。

[0038] 一些实施例根据以下公式计算熵作为Shannon熵:

$$[0039] \quad H(X) = - \sum_{i=1}^n p(x_i) \log_2 p(x_i)$$

[0040] 其中 $H(X)$ 是变量X的熵,下标i在所有可能的n个结果上循环, x_i 表示可能的结果,并且 $p(x_i)$ 是结果 x_i 的概率。在一个实施例中,使用二进制(底数为2的)对数,在这种情况下,结果所得的熵按位度量。

[0041] 由于所有概率都作为在0和1之间的范围内的值被给出,因此所有对数结果都为负,并且因此在求和之外是负。此外,结果的概率越低,概率值越小,并且因此对数值越高。换句话说,不频繁出现的值事实上是熵值的最大贡献者。

[0042] 一些实施例可以通过将熵除以 $\log_2 |n|$ 来对熵进行规格化,使得结果所得的规格化的熵落在范围0-1中。

[0043] 一般来说,可以为任何离散变量评估熵。在数值变量的情况下,可以或者直接(经由计算复杂的差分方法)计算熵,或者可以对数据进行离散化或分箱(binned),以将其转换为分类变量。后一种方法可能导致精度的边际损失,但获得计算的相当大的简化并提高计算的速度。

[0044] 在离散变量的情况下,概率是在数据集中属性的频率。在一个实施例中,高基数(cardinality)数据(例如,具有与系统中的记录的数量相当的数量不同值的变量;产品评论是这种高基数数据的示例,因为每个评论可以合理地预期是不同的)可以被假设是无趣的。在包含自然语言文本的变量的情况下,词语提取可以被用来将这种高基数变量转换为较低基数(并且因此更有趣)的变量。

[0045] 在一个实施例中,熵值的使用包括计算被规格化到0-1范围的数据集的每个属性的熵。进一步的用途包括基于兴趣度函数的结果的降序对属性进行排序,以试图突出更有趣的属性,如下所述。

[0046] 一般来说,对于一个数据分析者(或科学家)可能有趣的东西对于另一个数据分析者可能不那么有趣。但是,存在由于它们的本质而根本无趣的两种退化情况。在熵范围的低端(更接近0)是保持仅单个结果的属性。直观地,这些类型不会对数据集的整体理解有贡献。类似地,在熵值的高端(更接近规格化的值1),存在高基数属性(例如,产品库存数列,其被预期对于每个产品是不同的)。这种属性也不被预期携带信息。

[0047] 在高于低端退化情况的值处,感兴趣的信号出现,表示这个变量的结果的变化。类似地,在低于高端退化情况的值处存在感兴趣的信号。

[0048] 某些实施例设想通过首先应用低端和高端截止、然后将熵的极值(低和高)值映射到感兴趣度、同时将中间值映射到低兴趣度来将精确计算出的熵转换为兴趣度值的方式。

[0049] 图5示出了兴趣度相对于规格化的熵的一个可能线图。在图5中,x轴是从0到1的规格化的熵,其中0.5是映射曲线的最小值。y轴表示属性将有多有趣。在一些实施例中,用于低级别和高级别的熵的截止是不同的并且可以独立地调整。

[0050] 由于熵的值跨度从0到 $\log |n|$,其中n是用于每个特定列或属性的不同结果(或值)的个数,因此将具有不同个数的结果的属性的熵值进行比较可能是有挑战的。一些实施例通过将用于每个特定列的熵除以 $\log |n|$ 来执行熵规格化。规格化的熵的值落在0和1之间并且因此可以直接进行比较。

[0051] 例如,考虑具有相同熵值0.5的两个列。如果第一列仅包含值“真”和“假”,则其具有 $n=2$,并且规格化的熵为 $0.5/\log(2)=0.5$ 。应用到具有相同熵值0.5但具有四个不同值的另一列的相同过程将导致规格化的熵 $0.5/\log(4)=0.25$ 。在这个示例中,在规格化之前具有相同熵值的两个列将具有不同的规格化的熵值并且因此具有不同的兴趣度。

[0052] 其它实施例可以利用熵到兴趣度的不同映射。例如,具有0-1范围中的全局最小值的抛物线曲线也将满足上述考虑。

[0053] 某些实施例可以基于每个属性类型来应用熵到兴趣度的不同映射。例如,地理编码可以被认为总是有趣的,不管它们的值的分布。

[0054] 一些实施例允许用户动态修改已经根据它们的兴趣度被排序的属性的列表。可能性包括诸如“移除”和“喜欢”按钮的用户界面元件,以对应地排除和促进所选择的属性。

[0055] 一些实施例增加了机器学习方法的使用,以基于用户示出的兴趣进一步确定用于

退化情况的特定范围/阈值。

[0056] 上述实施例独立地为每个属性计算熵。在其它实施例中，在双变量分析级别而不是在单变量级别执行熵计算。这允许在两个不同属性(相互和条件熵)之间计算熵。在这个实施例中，用户可以在基于兴趣度的排序之前选择属性。在这种场景中，计算熵并且相对于所选择的列确定兴趣度。

[0057] 在另一个实施例中，用户可以在执行最初单变量兴趣度排序之后通过某个用户手势来指示对属性或一组属性的兴趣。在这种场景中，在这个用户手势之后，考虑新信息来对数据集的属性重新排序。

[0058] 对于双变量实施例，熵计算可以基于互信息。对于两个属性X和Y，互信息I可以表示为：

$$[0059] \quad I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right),$$

[0060] 其中x和y对应地是用于属性X和Y的可能结果；p(x)和p(y)对应地是结果x和y的概率；p(x,y)是结果x和y一起出现(在同一行数据中)的联合概率，并且双重求和对所有可能的结果发生。

[0061] 作为双变量实施例的示例，假设数据具有四个属性：x1、x2、x3、x4。属性可以是离散的或数值的，在这种情况下，它们可以经由分箱离散化。对于每个属性，相对于所选择的属性计算相对熵(互信息)。如果x1表示所选择的属性，则对于每个其它属性x2、x3、x4，实施例可以计算相对于x1的互信息。然后可以根据相对于所选择的列x1的互信息的计算出的值来对这三个属性进行排序。例如，如果互信息的值为 $I(x1, x2) = 0.4$ ； $I(x1, x3) = 0.6$ ； $I(x1, x4) = 0.2$ ，则根据这个互信息对属性进行排序将导致属性的以下排序：x3、x2、x4。其它实施例可以将这个排序与其它考虑组合。例如，如果已知城市、州和邮政编码的属性彼此相关，则这些属性可以在排序过程期间保持在一起。

[0062] 在另一个实施例中，可以使用条件互信息，其中，在一个示例中，条件互信息是给定第三个随机变量的值的情况下两个随机变量的互信息的预期值。

[0063] 图6是根据本发明一个实施例的有指导的数据探索模块16和其它元件的功能的流程图。图6示出了双变量实施例。图6的实施例在202、204、206、208、210和212具有与图2的实施例类似的功能。在614，接收对分块的选择。在616，按照相对于所选择的分块的兴趣度级别来对属性进行排序。

[0064] 如所公开的，实施例提供用于数据集的探索的有指导的分析。所选择的数据集的属性按照兴趣度级别排序，并且用户通过迭代过程可以提供进一步的排序。

[0065] 本文中具体示出和/或描述了若干实施例。但是，应当认识到，在不背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由上述教导涵盖并在所附权利要求的范围内。

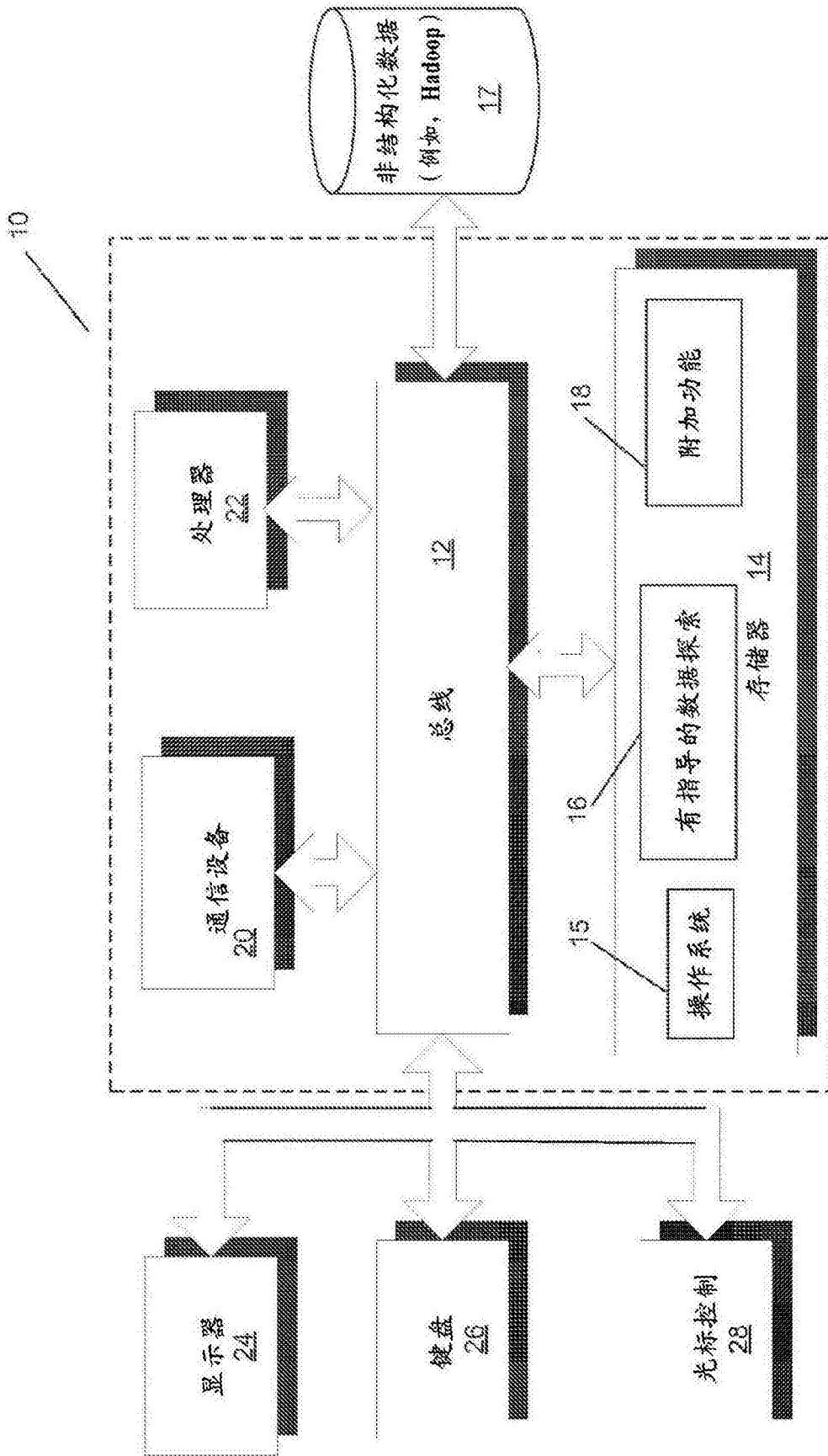


图1

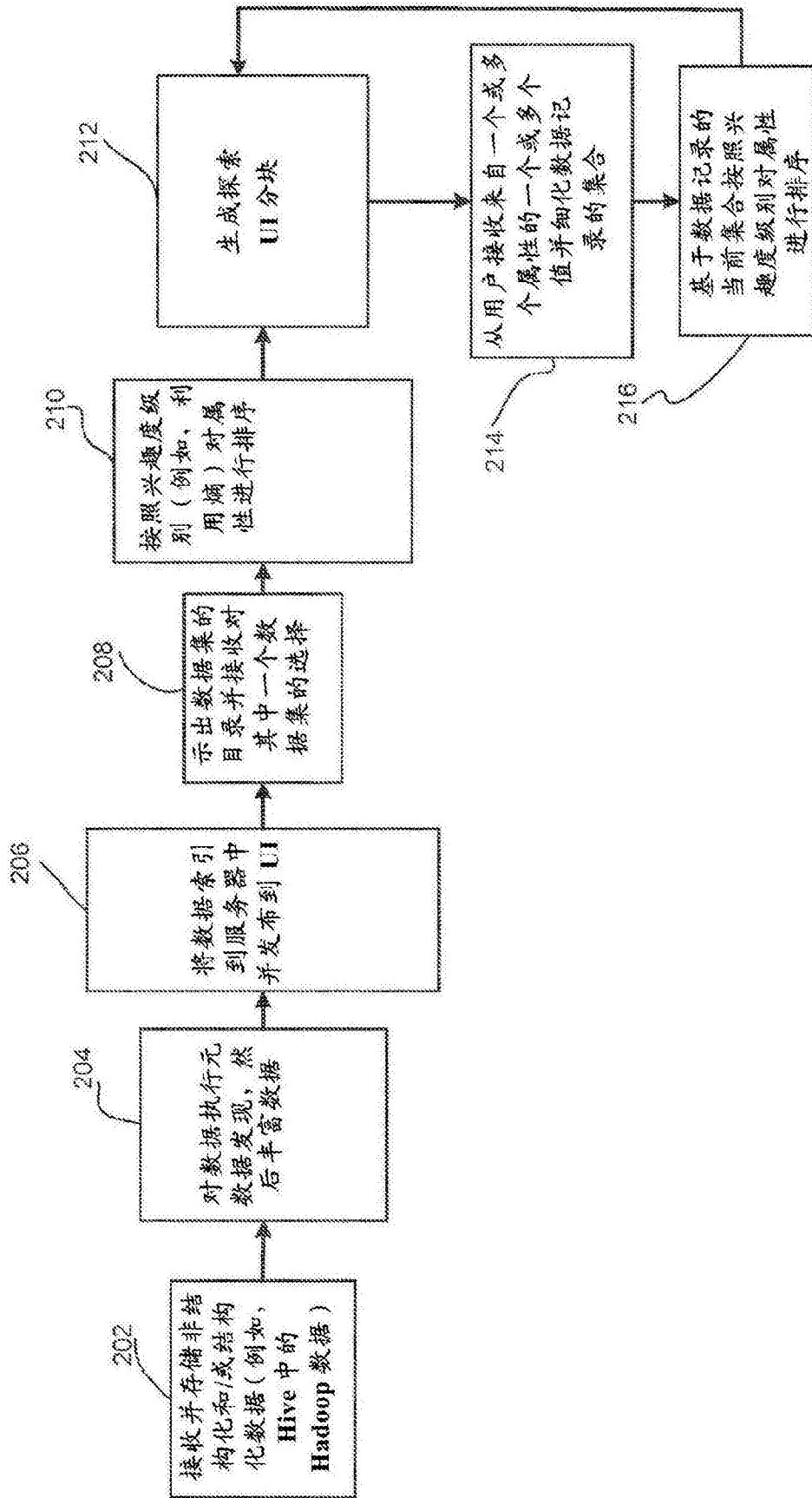


图2

ORACLE 大数据发现 | 目录

675 78

数据集 项目

创建新数据集

创建新项目

通过以下细化

- 用途
- Created by me
- My Favorite
- Certified/Gold Standard
- 内容
- Contains Dates
- Contains Locations
- Date Granularity
- Location Granularity
- Contains Language
- 元数据
- Category/Topic/Theme
- Tags
- Date Created (not available for HDFS)

DATA SETS (625)	SALES TEAM GOALS (145,208) Preview	World Wine Consumption Study (120,543)	42 ☆ 1 ●		PROJECTS (78)	
SALES ABLECHISE BY TEAM MEMBER BY PAIR	SALES SOURCE: HDFS	Created by The Wine Institute			World Wine Consumption Study (120,543)	
WINE SALES (125,482,730) Preview	305	Countries by Alcohol Consumption (79,903)	3 ☆ 1 ●		Countries by Alcohol Consumption (79,903)	
US WINE MARKET SALES TRANSACTORS	DATA SOURCE: SALES.BISERVER	Created by John Abigail			Statistical Report on world viticulture (84,816)	
WINE MARKET SHARE BY MAJOR VARIETY (124,716) Preview	305	Statistical Report on world viticulture (84,816)	2 ☆ 1 ●		Created by John Abigail	
Information on wine regions, major wineries, and the wine industry by comparing traditional wine markets with recent data on consumption.		Market Analysis of French Wine Drinkers (121,342)	14 ☆ 1 ●		Market Analysis of French Wine Drinkers (121,342)	
DATA SOURCE: WINEFILE (EXGUL)		Created by data.wineanalytics.org			The World's Wine Consumption: 2000-2012 (58,911)	
TRENDS IN THE US WINE INDUSTRY IN 2013 (1,413,206) Preview		The World's Wine Consumption: 2000-2012 (58,911)	6 ☆ 1 ●		Created by John Abigail	
San Rafael, Calif., Winey gains E. & J. Gallo and Constellation Wines dominance: the top 20 wine brands selling of grapevine through May, according to market research firm IRI, with analysis by Wines & Vines		Wine & Beer Social Media Extrad (38,474)	9 ☆ 1 ●		Created by John Abigail	
DATA SOURCE: HDFS		Created by expere.wineanal.org			National Wine Production and Consumption (120,345)	
2014 Sales Targets (114,870) Preview		National Wine Production and Consumption (120,345)	3 ☆ 1 ●		Created by John Abigail	
This dataset reflects reported 2014 from all part of a new strategy to target middle-spirit brands.		Seattle Wine Regions (63,120)	0 ☆ 1 ●		Created by John Abigail	
DATA SOURCE: SOME DATABASES (JDBC)		Created by John Abigail			Seattle Wine Regions (63,120)	
WINE MARKET SALES		Wine Industry Statistics (60,476)	4 ☆ 1 ●		Wine Industry Statistics (60,476)	
DROUGHT EFFECTS ON WINEYAC PRODUCTION (114,870) Preview		Created by John Abigail			International Wine review (321,346)	
USDA director sees drought effects on large number of wineries: Top wine grape grower issues - Drought's impact on grape vines.		International Wine review (321,346)	39 ☆ 1 ●		Created by data.pyofreywork.us	
DATA SOURCE: SOME DATABASES (JDBC)		Wine Annual Report and Statistics 2013 (100,635)	16 ☆ 1 ●		Wine Annual Report and Statistics 2013 (100,635)	
YTD Shipments (114,870) Preview		Created by John Abigail			California Sustainable Winegrowing Workbook (6,674)	
Beer, Wine, and Spirits Away from Home 2012 to 2010, June YTD including shipment volumes in the top 10 states, % of the national total, and volume change in top ten states.		California Sustainable Winegrowing Workbook (6,674)	9 ☆ 1 ●		California Sustainable Winegrowing Workbook (6,674)	
DATA SOURCE: SOME DATABASES (JDBC)						
WINE SALES, STATEWIDE						
DROUGHT EFFECTS ON WINEYAC PRODUCTION (114,870) Preview						
This dataset reflects reported incidents of wine that have occurred in the City of						

图3

300

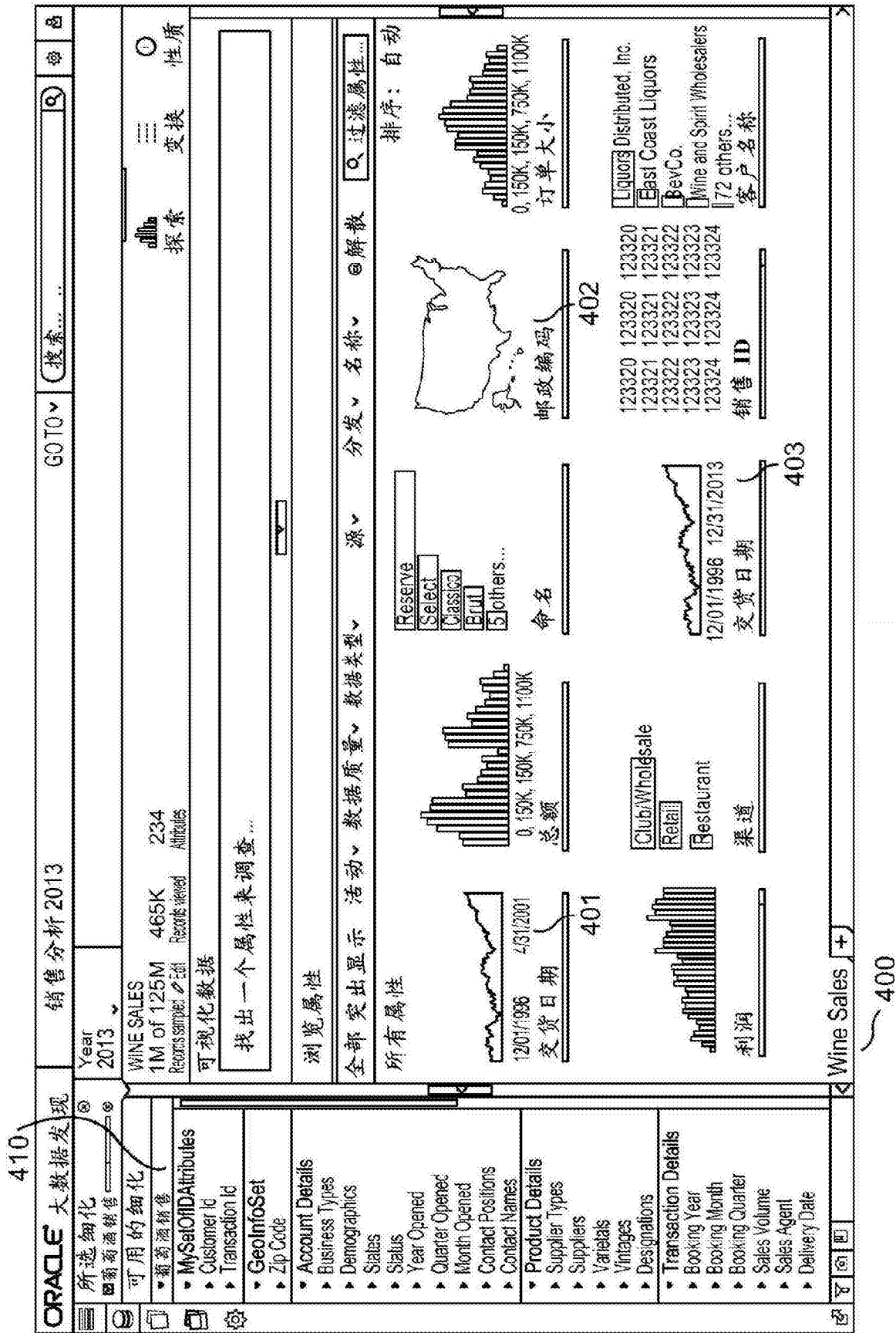


图4

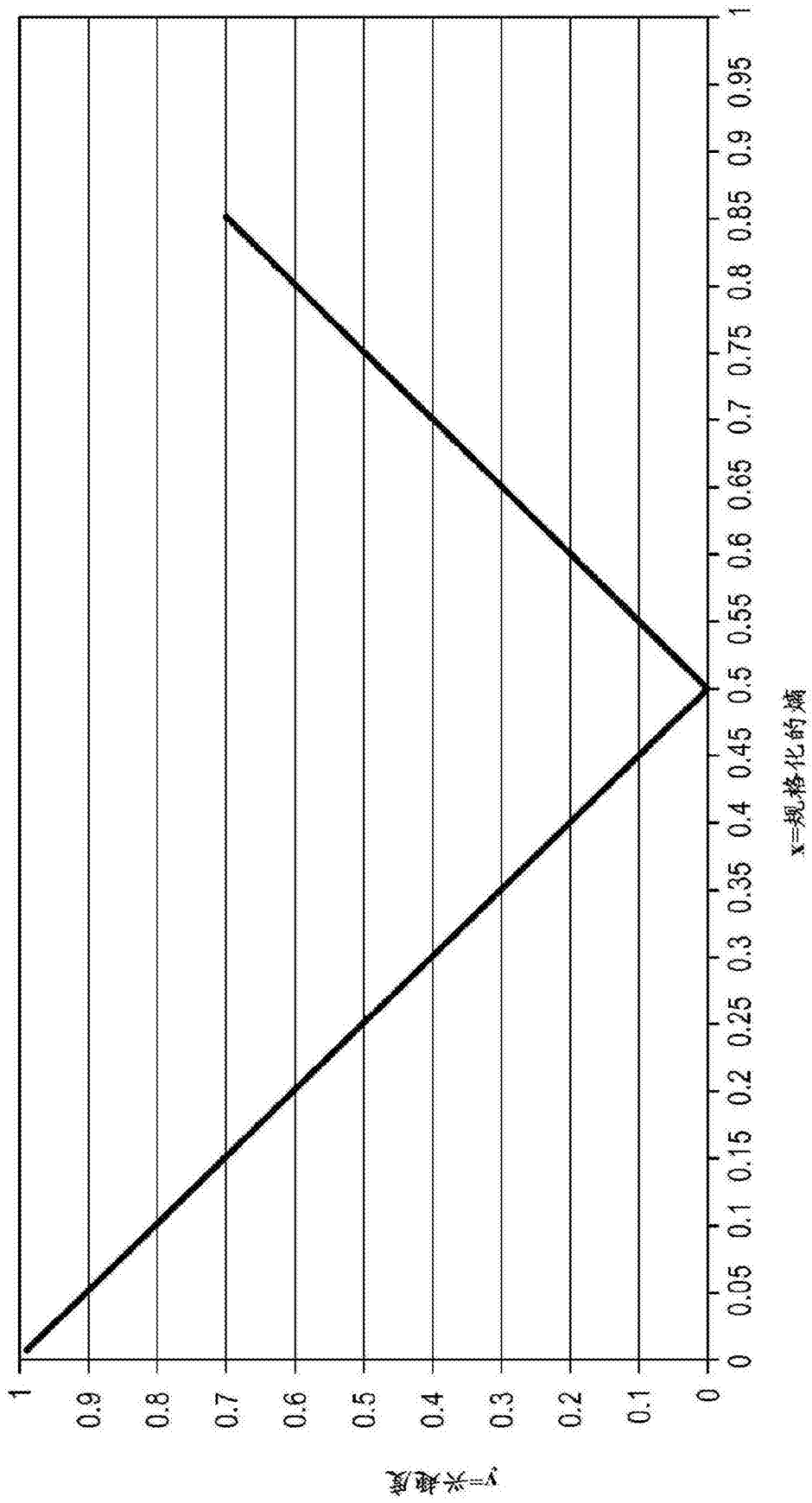


图5

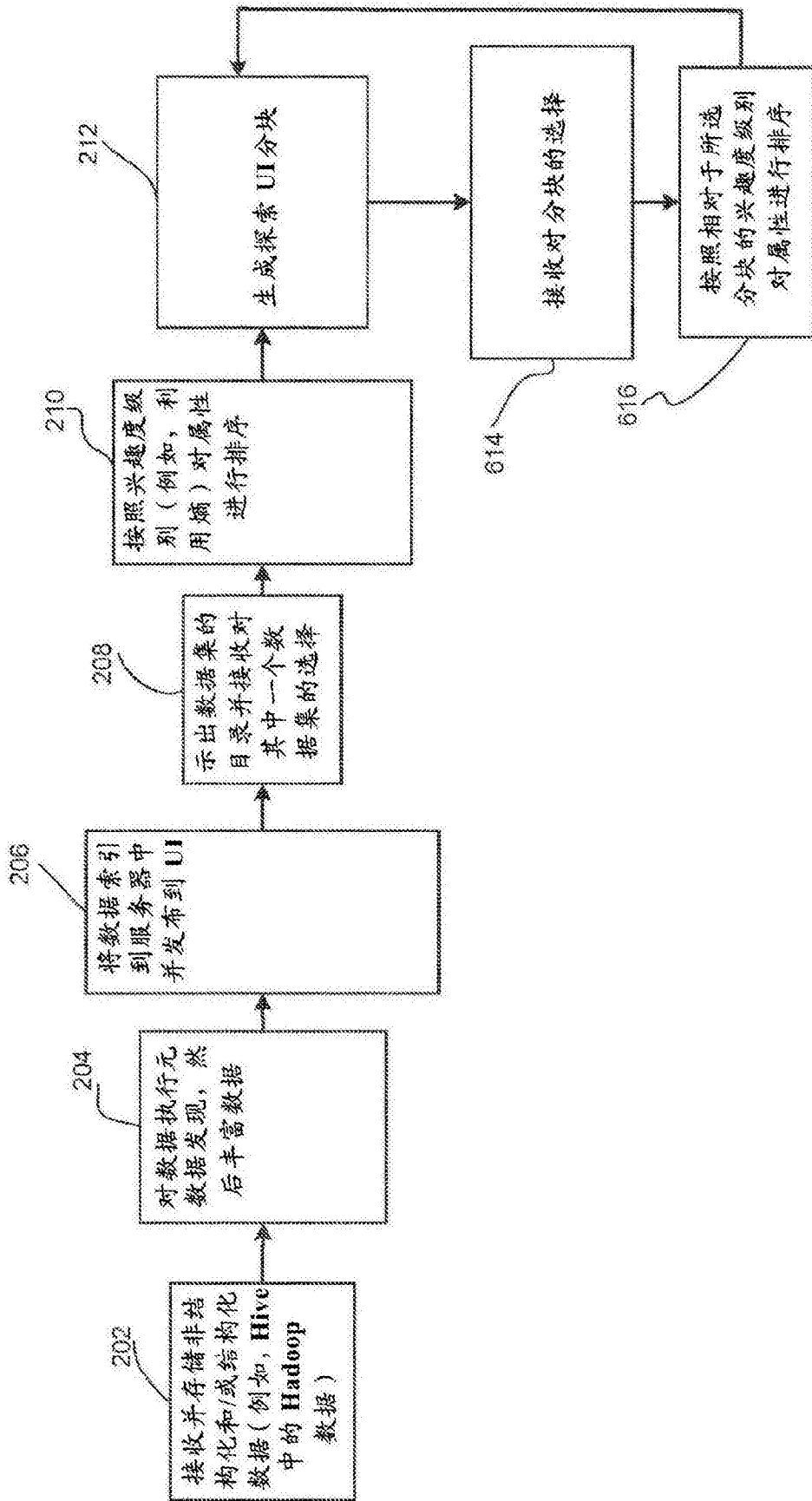


图6