

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G06F 17/00 (2006.01)



[12] 发明专利申请公布说明书

[21] 申请号 200680047941.4

[43] 公开日 2009 年 8 月 12 日

[11] 公开号 CN 101506796A

[22] 申请日 2006.12.19

[21] 申请号 200680047941.4

[30] 优先权

[32] 2005.12.19 [33] US [31] 60/752,102

[86] 国际申请 PCT/US2006/048330 2006.12.19

[87] 国际公布 WO2007/075622 英 2007.7.5

[85] 进入国家阶段日期 2008.6.19

[71] 申请人 斯专兹有限公司

地址 美国俄勒冈州

[72] 发明人 M·托伦斯 P·费雷拉

[74] 专利代理机构 北京北翔知识产权代理有限公司

代理人 杨 勇 谢 静

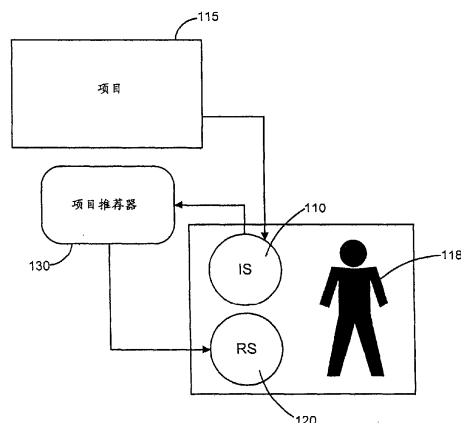
权利要求书 3 页 说明书 14 页 附图 10 页

[54] 发明名称

用户 - 用户推荐器

[57] 摘要

公开了用于将相关用户推荐给用户社区中的其它用户的系统和方法的实施方案。在此方法的一个实现中,考虑两种不同的数据集:a)用户一直在聆听(或者以其它方式使用)的音乐(或者其它项目),和 b)已经给用户的音乐(或者其它项目)推荐。在一些实施方案中,预计算方法允许系统有效地比较社区中多个用户的项目集和被推荐项目集。此比较还可以包括系统用来确定对于一个给定目标用户应当推荐哪些用户的度量。



- 1、一种用于推荐用户社区中的用户的方法，所述方法包括：
选择所述用户社区中的第一用户；
选择所述用户社区中的一个用户集；
将所述第一用户的用户简档和所述用户集中的每一用户的用户简档比较；以及
生成所述第一用户的被推荐用户集；其中所述被推荐用户集包括所述用户社区中的至少一个用户。
- 2、权利要求1的方法，其中，所述被推荐用户集包括所述用户社区中的多个用户。
- 3、权利要求1的方法，其中，选择第一用户的步骤包括接收来自所述第一用户的生成所述被推荐用户集的请求。
- 4、权利要求1的方法，其中，所述用户集包括所述用户社区中的每一个用户。
- 5、权利要求1的方法，其中，所述用户简档包括媒体项目集，所述媒体项目集包括多个媒体数据项目。
- 6、权利要求5的方法，其中，所述媒体数据项目包括识别多个媒体项目的元数据。
- 7、权利要求5的方法，其中生成一个被推荐用户集的步骤包括分析所述媒体项目集中的媒体项目的播放计数。
- 8、权利要求5的方法，其中，所述用户简档还包括被推荐媒体项目集，所述被推荐媒体项目集包括多个被推荐媒体数据项目。
- 9、权利要求8的方法，其中，所述被推荐用户集通过使用下列度量中的至少一个生成：分析所述媒体项目集和所述用户集中的每一用户的媒体数据项目的媒体项目集的相交的度量、分析所述媒体项目集和所述用户集中的每一用户的被推荐媒体数据项目的被推荐媒体项目集的相交的度量、分析所述被推荐媒体项目集和所述用户集中的每一用户的媒体数据项目的媒体项目集的相交的度量以及分析所述被推荐媒体项目集和所述用户集中的每一用户的被推荐媒体数据项目的被推荐媒体项目集的相交的度量。
- 10、权利要求9的方法，其中，通过进一步使用所述用户集中的每一用户的媒体项目集中的媒体项目的播放计数，生成所述被推荐用

户集。

11、权利要求 8 的方法，其中，通过使用下列度量生成所述被推荐用户集：分析所述媒体项目集和所述用户集中的每一用户的媒体数据项目的媒体项目集的相交的度量、分析所述媒体项目集和所述用户集中的每一用户的被推荐媒体数据项目的被推荐媒体项目集的相交的度量、分析所述被推荐媒体项目集和所述用户集中的每一用户的媒体数据项目的媒体项目集的相交的度量以及分析所述被推荐媒体项目集和所述用户集中的每一用户的被推荐媒体数据项目的被推荐媒体项目集的相交的度量。

12、一种其上存储有计算机可执行指令的计算机可读介质，所述指令用于执行用于推荐用户社区中的用户的方法，所述方法包括：

选择所述用户社区中的第一用户；

选择所述用户社区中的一个用户集；

将所述第一用户的用户简档和所述用户集中的每一用户的用户简档比较；以及

生成所述第一用户的被推荐用户集；其中所述被推荐用户集包括所述用户社区中的至少一个用户。

13、权利要求 12 的计算机可读介质，其中，选择第一用户的步骤包括接收来自所述第一用户的生成所述被推荐用户集的请求。

14、权利要求 12 的计算机可读介质，其中，所述用户简档包括媒体项目集，所述媒体项目集包括多个媒体数据项目。

15、权利要求 14 的计算机可读介质，其中，所述媒体数据项目包括识别多个媒体项目的元数据。

16、权利要求 14 的计算机可读介质，其中，生成被推荐用户集的步骤包括分析所述媒体项目集中的媒体项目的播放计数。

17、权利要求 14 的计算机可读介质，其中，所述用户简档还包括被推荐媒体数据项目的被推荐媒体项目集。

18、权利要求 17 的计算机可读介质，其中，所述被推荐用户集通过使用下列度量中的至少一个生成：分析所述媒体项目集和所述用户集中的每一用户的媒体数据项目的媒体项目集的相交的度量、分析所述媒体项目集和所述用户集中的每一用户的被推荐媒体数据项目的被推荐媒体项目集的相交的度量、分析所述被推荐媒体项目集和所

述用户集中的每一用户的媒体数据项目的媒体项目集的相交的度量以及分析所述被推荐媒体项目集和所述用户集中的每一用户的被推荐媒体数据项目的被推荐媒体项目集的相交的度量。

19、权利要求 18 的计算机可读介质，其中，通过进一步使用所述用户集中的每一用户的媒体项目集中的媒体项目的播放计数，生成所述被推荐用户集。

20、权利要求 17 的计算机可读介质，其中，通过使用下列度量生成所述被推荐用户集：分析所述媒体项目集和所述用户集中的每一用户的媒体数据项目的媒体项目集的相交的度量、分析所述媒体项目集和所述用户集中的每一用户的被推荐媒体数据项目的被推荐媒体项目集的相交的度量、分析所述被推荐媒体项目集和所述用户集中的每一用户的媒体数据项目的媒体项目集的相交的度量以及分析所述被推荐媒体项目集和所述用户集中的每一用户的被推荐媒体数据项目的被推荐媒体项目集的相交的度量。

用户-用户推荐器

附图说明

应该理解，附图仅示出本发明的某些优选实施方案，因此不能被认为是对本发明的范围进行限制，因而将通过使用以下附图，借助附加特征和细节描述和解释这些优选实施方案，附图中：

图 1 是示出根据一个实施方案的用户简档的基本组成部分和源的示意图；

图 2 示出了一个显示根据一个实施方案的 ADG 空间中的目标用户“X”的位置的图；

图 3 是示出根据一个实施方案的用户推荐器的基本结构图的示意图；

图 4 示出了指示特定项目 (item) 与用户社区中特定用户的相关性的定性范围；

图 5 是示出用户推荐系统的一个实施方案的小服务程序视图 (Servlet View) 的示意图；

图 6 是示出用户推荐系统的一个实施方案的推荐器视图的示意图；

图 7 是示出用户推荐系统的一个实施方案的管理器视图的示意图；

图 8 是用户推荐系统的一个实施方案的核心统一建模语言 (UML) 示意图。

图 9 是示出可以从用户推荐系统的一个实施方案所用的绘图机 (GraphPlotter) 工具中提取的信息的例子的示意图；

图 10 是表示根据“频率/知识”模型的听众类型之间的关系的图。

具体实施方式

在以下描述中，提供编程、软件模块、用户选择、网络交易、数据库查询、数据库结构等的某些具体细节，以便于彻底理解本发明的

特定优选实施方案。但是，本领域的普通技术人员将认识到：可以在无需一个或者多个具体细节的情况下，或者可以使用其它的方法、部件、材料等，实践各实施方案。

在一些情形下，并没有详细示出或者描述公知的结构、材料或者操作，目的是为了混淆优选实施方案的各个方面。此外，所描述的特征、结构或者特性可以在不同的替代实施方案中以任一合适方式组合。在一些实施方案中，此处所描述的方法和系统可以使用一个或者多个数字处理器实现，所述数字处理器诸如普遍用于 PC、膝上型计算机、PDA 和各种其它桌面或者便携式电子设备中的各种类型的微处理器。

公开了用于将用户推荐给用户社区中的其它用户的系统和方法的实施方案。“用户推荐器”在此处被使用时，指被结合到一个用户社区中的模块，其主要功能是将用户推荐给该社区中的其它用户。社区中可以存在一组项目供该社区中的用户交互。也可以存在一个项目推荐器将其它项目推荐给用户。可以结合此处所阐述的实施方案一起使用的推荐系统的例子在标题为“Recommender System for Identifying a New Set of Media Items Responsive to an Input Set of Media Items and Knowledge Base Metrics (用于响应输入媒体项目集和知识库度量识别新的媒体项目集的推荐系统)”的美国专利公布 NO.2006-0184558 和标题为“Dynamic Identification of a New Set of Media Items Responsive to an Input Mediaset (响应输入媒体集动态识别新的媒体项目集)”的美国专利公布 No.2006-0173910 中有描述。

术语“媒体数据项目”在此处被使用时，意图包括任意媒体项目或媒体项目的表示。“媒体项目”意图包括可以以数字媒体格式表示的任意类型的媒体文件，诸如歌曲、电影、图片、电子书、报纸、TV/无线电节目片段、游戏等等。因此，旨在术语“媒体数据项目”包括例如可播放的媒体项目文件（例如，MP3 文件）以及识别可播放媒体文件的元数据（例如，识别 MP3 文件的元数据）。因此，应该明显的是，在任何提供使用“媒体项目”的过程、步骤或者系统的实施方案

中，该过程、步骤或者系统可以代之以使用媒体项目的表示（诸如元数据），反之亦然。

用户推荐器可以为给定目标用户选择相关的用户。为了实现此，用户应该是可比较的实体。限定社区中的用户的要素可称为用户简档（profile）。因此，可以通过限定两个集合，限定用户简档，使得比较两个用户将成为将他们的用户简档集合相交（intersect）的问题。例如，参照图 1，第一集合可以是图 1 中的 110 所指的“项目集”，其可以包括特定用户 118 的最相关项目 115。第二集合可以是图 1 中的 120 所指的“推荐集”，其可以包括用户 118 的最相关被推荐项目。项目集 110 可以通过项目使用和/或某一用户和某些项目的交互推导出，而推荐集可以通过使用项目推荐器 130 推导出。在一些情形下，项目集 110 可以用作推荐器 130 的输入，从而获得推荐集作为输出。

在以下例子和实现中描述的项目将适合于音乐或者其它媒体项目。但是，应该理解，此处描述的实现并不是项目特定的，可以适合由用户社区使用/共享的任意其它类型的项目。

对于音乐或者多媒体项目（曲目（track）、艺术家、专辑等等），用户可以通过使用它们（聆听、购买等等）与其交互。这样的实施方案中的集合将被称为“音乐集”，因为其包括音乐项目。这些集合因此将被称为“音乐集”和“推荐集”。

音乐集为由用户正聆听的项目所形成的音乐集合。用户 A 的音乐集在此处将被表示为 M_a 。

推荐集是由用户正推荐的项目所形成的音乐集合。用户 A 的推荐集在此处将被表示为 R_a 。

为了比较两个用户简档，可以分析其各自的集合中的一个或者多个之间的相交（intersection）。各种不同的度量（metric）也可以应用于集合相交来提供有用的数据。一些这样的度量将描述用户之间的关系。例如，四种基本相交情况为：

$M_a \cap M_b$ 、 $M_a \cap R_b$ 、 $R_a \cap M_b$ 、和 $R_a \cap R_b$ 。分析这些情况可以得到可以被标注为或者分类为不同关系的复杂情况。例如，在一个实现中，四种相关关系可以被提取：

同等者：如果 M_a 和 M_b 充分相交，B 被认为是 A 的“同等者”。

指导者 (Guru)：如果 M_a 和 R_b 充分相交，B 被认为是 A 的“指导者”。

同等者-指导者：同等者条件加指导者条件。B 被认为是 A 的“同等者-指导者”。

追随者：如果 R_a 和 M_b 充分相交，B 被认为是 A 的“追随者”。

同等者关系可以是相关的，因为其提供给目标用户另一用户，其中该另一用户的音乐（或者其它项目）库和目标用户的音乐库在某种意义上类似。指导者关系可以是相关的，因为其提供给目标用户另一用户，其中该另一用户的音乐库包含目标用户可以享受发现的乐趣的音乐。同等者-指导者可以是相关的，因为其提供给目标用户分别对于一个或者多个被推荐用户的同等者和指导者关系的相似

(affinity) 经历和发现经历。追随者关系可以是相关的，因为其提供给用户机会来了解哪些用户可以被她或者他影响。

现在公开例如可以据其推导出前述关系的示例性具体度量。度量可以是这样一个函数，其以两个（或者更多个）用户简档作为输入并产生一个可测量结果作为输出。以下讨论的度量是不定向的，意味着参数的顺序可以改变结果。

“相似”度量回答问题：“ M_a 和 M_b 有多少相交？” 。换句话说，对于用户 B，用户 A 具有多少“相似经历”？

“发现”度量回答问题：“ R_a 和 M_b 有多少相交？” 。换句话说，对于用户 B，用户 A 具有多少“发现经历”？

“指导”度量回答问题：“ M_a 和 R_b 有多少相交？” 。换句话说，用户 A 可以指导用户 B 多少？

利用这些度量，通过求出相似度量的最大值可以发现同等者关系，通过求出发现度量的最大值可以发现指导者关系，通过求出相似和发现度量二者的最大值可以发现同等者-指导者关系，并且通过求出指导度量的最大值可以发现追随者关系。例如通过定义以这些度量中的每一个（或者多于一个）进行运算的函数，可以计算一个用户对于另一用户的总相关性。对于目标用户，社区中的所有其它用户可以

被定位为三维空间（“ADG 空间”）中的点，其中，X=相似，Y=发现，Z=指导。将这些度量定义为返回 $[0, 1]$ 内的数值，社区中的所有用户都可以被包含在上述空间中 $1 \times 1 \times 1$ 的立方体内。图 2 示出了 ADG 空间中的抽样目标用户“X”的位置。

为了在解释概念模型之后实现用户推荐器，可以建立一个基础系统。一个这样的系统可以被配置为：

- 1、存在用户社区和项目推荐器，从中可以提取社区中每一用户的用户简档。此信息可以由一个或者多个数据源馈送。
- 2、存在用户推荐器的实现，其建立该模型的数据和操作，并且从数据源收集数据。

图 3 示出了根据一种实现的用户推荐器的基本结构图。如图中所示，数据源 310 提供数据到用户社区 320（并从用户社区 320 接收数据）。数据源 310 还可以提供数据馈送（data feed）到用户推荐器 330。用户推荐器 330 与用户社区 320 交互。具体而言，可以从用户社区 320 接收请求用户推荐的请求，并且反过来将被推荐用户提供给用户社区 320。当然，可以预期，一些实现可以依赖接收来自用户的请求来生成被推荐用户集，一些实现可以生成这样的集合并将其提供给社区中的用户而不需要首先接收用户请求。

还可能期望提供可缩放的结构解决方案。对于一个请求，将目标用户和社区中的所有用户相比较可能并不可行（响应时间可能会随用户数量线性增长）。可以实现针对这个问题的多个解决方案，例如：

- 1、可以将用户数据群集，并将目标用户与合适的群（cluster）比较。
- 2、可以从用户社区中选出固定数量的用户或者用户子集。此用户子集可以被称为“可推荐用户”，并且目标用户可以与该大小固定的集合比较。可以利用允许系统推荐社区中最令人感兴趣的用户的进程来选择可推荐用户。

音乐集实体可以被建模为 N 维空间中的稀疏向量，其中 N 是我们全域（our universe）中的音乐项目的总数。每一维指不同的项目，而每一个具体的值指此项目的相关性。增加每个项目的相关值会使基

本系统或者实现可以察觉某个用户的最相关项目。

在一些实现中，项目可以是音乐曲目。但是，在这样的实施方案中，两个用户的集合中的项目之间的相交可能是概率较低的（由于稀疏问题）。此外，这些相交可能是计算成本高昂的。

在一些实施方案中通过代之以在艺术家层次上进行操作，可以解决这些问题。艺术家而不是曲目相交的可能性更高。另一方面，相关值可以取决于从中提取项目的数据源。因此可以使用归一化方法，使得所有的相关值最后都落于一个已知值范围内，诸如落于一个定性值范围内。

例如，图 4 示出了指示诸如艺术家之类的特定项目与特定用户的相关性的定性范围。对于一个给定用户的相关值低于 C1 的艺术家（或者其它项目）被认为是“低”，相关值介于 C1 和 C2 之间的艺术家被认为是“中等”，相关值介于 C2 和 C3 之间的艺术家被认为是“中等/高”，相关值高于 C3 的艺术家被认为是“高”。

稍后讨论示例性归一化过程的细节和例子以及用于获得某个项目对于某个用户的相关性的其它方法。

用户实体可以被建模为具有唯一 ID 和两个音乐集实体的实体，使得我们具有根据此处讨论的概念性模型计算相交所需的所有数据。

现在将讨论可以在本发明的一些实施方案中实现的一些操作。基本操作为比较两个用户实体所需的操作和包含音乐集之间的相交的操作。例如：

1、音乐集的大小可以表示为：

$$|M_u| = \sum_{k=1}^N M_{uk}$$

其中， M_{uk} 是集合 M_u 中的项目 k 的相关值。

2、对于 M_u 和 $M_{u'}$ 中共有的所有 M 个项目，相交的大小可以表示为：

$$|M_u \cap M_{u'}| = \sum_{k=1}^M \min(M_{uk}, M_{u'k})$$

3、相似、发现和指导度量可以表示如下。

相似度量的一个方法包括计算 M_u 和 $M_{u'}$ 的大小，并且利用 M_u 的大

小将其归一化，如下：

$$Affinity(U, U') = \frac{|M_u \cap M_{u'}|}{|M_u|}$$

作为另一种可能性，如果认为 R_u 和 $R_{u'}$ 的相交在某种意义上是相似量度 (measure)，则可以将此因素增加到整个公式中，以因子 k 将其加权然后归一化该量度：

$$Affinity(U, U') = \frac{\frac{|M_u \cap M_{u'}|}{|M_u|} + \frac{|R_u \cap R_{u'}|}{K \cdot |R_u|}}{1 + \frac{1}{K}}$$

注意， U 对 U' 的高度相似并不一定意味着 U' 对 U 的高度相似。

发现和指导的相应公式如下：

$$Discovery(U, U') = \frac{|R_u \cap M_{u'}|}{|R_u|}$$

$$Guidance(U, U') = \frac{|M_u \cap R_{u'}|}{|R_{u'}|}$$

注意， $Discovery(U, U')$ 总是等于 $Guidance(U', U)$ 。

以下模型操作是可以在用户推荐器中实现的示例性全局操作，该推荐器通过使用基本操作使得可以计算期望的结果。

-getBestUsers (User, Requirement) (获取最佳用户 (用户, 要求))：

通过计算某组度量，可以返回被推荐用户集用于目标用户。上述要求可以指定要推荐哪一种用户和要考虑哪些度量。用于此功能的总体算法可以如下：

- 1、设 TU 为目标用户，RUS 为可推荐用户集，REQ 为请求的要求。
- 2、对于 RUS 中的每一个用户 U，和 RUS 中的被比较用户 U 一起，根据 REQ 计算必要的度量 (TU, U) 并且存储结果。
- 3、按比较结果将 RUS 分类，使得在列表的开头具有根据 REQ 的最佳用户。
- 4、返回 RUS 的始于开头处的子列表。

-getRelevance (User1, User2) (获取相关性 (用户 1, 用户 2))：

通过计算用户 1 对于用户 2 的所有度量，通过计算以所有度量值执行一些计算的函数，可以返回一个浮动数 (floating number)，

其回答问题：用户 2 对于用户 1 有多相关？例如，此函数可以计算 ADG 空间中向量的长度。

用户推荐器可以被实现为 Java Web 模块。此模块例如可以被配置为 Tomcat 环境中的网络应用（webapp）。在一个实现中，用于此实现的数据源可以如下：

1、“Reach” API（到达应用程序接口）：返回每个用户的播放计数数据。一些实现可以从此数据推导出音乐集。

2、“UMA”推荐器：返回项目集的被推荐项目。一些实现可以从此数据推导出音乐集用作输入，例如，用户简档的音乐集，并且因此获得上面说明的用户简档的推荐集。

3、“KillBill” API：返回关于用户的一些额外信息，例如其别名。

一个可缩放的解决方案是通过 Reach API 获得社区中的 N 个最佳用户，并且使其成为可推荐用户。在一些实现中，可以使用 CLUTO 群集程序。CLUTO 程序为开源软件，并且可以在 <

<http://glaros.dtc.umn.edu/gkhome/cluto/cluto/download>>下载到。在一些实现中，可以使用 WEKA 群集程序。WEKA 程序也是开源程序，并且可以在 <<http://www.cs.waikato.ac.nz/ml/weka/>>下载到。在一些实现中，还可以建立和使用文件高速缓冲存储器系统。

因此，即使可以向所有用户提供推荐，也仅有最佳用户可以被推荐。如果用户不在该最佳的 N 个的集中，那么用户推荐器可以实时要求这个用户的用户简档以将其添加到数据源。

Java 实现可以由一组其逻辑可以被如下划分的类组成：

小服务程序视图：其中包括所有的小服务程序请求/响应逻辑。

推荐器视图：其中由单独的（singleton）推荐器执行主要操作，这些操作经常被授权给管理器。

管理器视图：其中实际上通过使用核心类执行所有操作。

核心视图：其中通过一组 Java 类建立模型的基础。

现在将更详细地描述前述“视图”。用户推荐器可以实现为一组 HTTP 小服务程序。推荐器可以实现为单独的实例，该实例可以由三

个不同的小服务程序使用，如图 5 所示。

1、调试小服务程序 500 管理调试命令。例如，`action=stats` 可以做一个推荐器 530 内部存储器统计图。

2、推荐小服务程序 510 根据之前描述的模型操作管理推荐请求。

3、更新小服务程序 520 对用户数据执行更新过程。

推荐器可以是主要由管理器和请求高速缓冲存储器形成的单独的类。也可以通过信号机（semaphore）控制对高速缓冲存储器的并行存取，以避免不一致的信息。例如，如果调试小服务程序发送清除高速缓冲存储器命令，如果存在推荐请求等待从高速缓冲存储器获取结果，如果在执行完清除命令之后处理该请求的话，可以接收到空响应。

在一些实现中，用于存取请求高速缓冲存储器的总体算法如下：

1、如果信号机是打开的，关闭信号机；否则等待。

2、结果在高速缓冲存储器中吗？

3、如果在，从高速缓冲存储器中取出结果。

4、打开高速缓冲存储器信号机。

可以以散列图（Hash Map）实现高速缓冲存储器，其关键是串（string）请求散列，并且其值为请求结果。可以计算串请求散列以获得用于可以产生不同结果的每一组参数的唯一串。

图 6 示出了用户-用户推荐器 600 的例子和其与系统其它元件的基本交互。推荐器 600 具有管理器 610 和请求高速缓冲存储器 620。管理器 610 具有 612 所示的一个或者多个外部连接和 614 所示的数据馈送。

管理器可以被实现为具有服务层和数据空间层两层的单独的类，如图 7 中 700 所示。

服务层 710 可以包括服务程序（service）712 的单独的实例。每一服务程序 712 可以用来执行一组特定类型的任务。例如，可以提供“更新服务程序”，其包括所有用于对数据执行更新的逻辑。还可以提供“比较服务程序”，其包括所有用于执行数据之间的比较的逻辑。这些服务程序可以直接和数据空间层 720 通信。

数据空间层 720 可以包括所有到外部服务程序（诸如 UMA 722、KillBill 724 和 Reach 726）的连接，以及永久存储可推荐用户的主存储结构 728。

用于一些实现的基本类为：用户 810、被测量的用户 820、用户相似性 830 和要求 840，如图 8 所示。在这样的实现中，用户 810 的音乐集 850 可以实现为一组对（项目 854 和相关性 856）。项目和相关性的某些实现为 ArtistItem（艺术家项目）（包括艺术家 ID）和 SimpleRelevance（简单相关性）（1 和 3 之间的一个量）。还可以存在 ItemFactory（项目工厂）和 RelevanceFactory（相关性工厂），其任务是创建项目和相关性对象，将实现的名称作为输入。在这样的实施方案中，可以在计划的任一阶段容易地改变项目和相关性的实现，而不会影响其它核心类。

当一个用户和另一用户相比较时，具有被测量的用户 820，该被测量的用户为被比较的用户加 UserSimilarity（用户相似性）830 实例。用户相似性 830 指定每一度量 860 有多和目标用户相关。客户还可以指定一个要求来求不同度量的最大值/最小值。相似、发现和指导度量分别在图 8 中示为 862、864 和 866。

如前所述，可以实现相似、发现和指导度量，但是也可以通过扩展接口度量（interface metric）来实现其它度量。

接口度量规定每一度量必须执行两个用户间的相交，其结果是可参数化的。接口也可以规定每一度量必须返回根据该度量测量两个用户之间的相关性的计算值。该值也可以是可参数化的，使得每一度量被双参数化。该计算值也可以被归一化为在 0 和 1 之间。作为例子，可以实现新的度量“年龄相似”。此度量可以返回包括两个用户年龄之间的差异的带符号的整数作为相交和/或返回表示一个用户相对另一用户的定性年龄差异（“较年轻”、“年轻得多”等等）的串。可以计算该归一化计算，使得 1 表示两个用户具有相同的年龄并且 0 表示两个用户年龄相差太远以致不能出于系统的目的而被认为是相关的。

在一个例子中，网络应用可以被配置在两个由负载平衡器控制的

专用机 (production machine) 中。可推荐用户的数量大约是 1000 个, 并且系统能够每天响应超过 500 个请求。使用可在 <

http://jakarta.apache.org/site/downloads/downloads_jmeter.cgi > 下载到的 Java 桌面应用程序 Apache Jmeter 进行压力测试。在此测试中, 发送多个请求到服务器, 并且响应时间随着同时发生的请求的数量线性增加。测试结果数量如下:

请求	每秒请求数	平均处理时间	平均响应时间
100	1	60ms	290ms
200	2	110ms	370ms
1000	10	120ms	420ms
10000	100	500ms	1800ms

在一些实现中, 通过使用绘图机工具可以绘制可视图。图 9 提供了可以从绘图机工具中提取的信息的例子。如图 9 所示, 用户社区中的不同用户之间的发现和/或相似值被表示在图中。

现在将更详细地描述特定实现的附加细节。假设具有两个用户 A 和 B, 其分别具有音乐集和推荐集 M_a 、 R_a 和 M_b 、 R_b 。如果 R 是从输入 M 生成的项目推荐器的输出, 对于一些项目推荐器 $M \cap R = \emptyset$ 总是为真。

如前所述, 四个可能相交为 $M_a \cap M_b$ 、 $M_a \cap R_b$ 、 $R_a \cap M_b$ 、 $R_a \cap R_b$ 。情况总数为 12:

同等者关系: $M_a \cap M_b$ 。A 和 B 具有共同的音乐品味。

同等者-兄弟关系: $M_a \cap M_b + R_a \cap R_b$ 。A 和 B 具有共同的音乐品味并且在未来也可以具有共同的音乐品味。

指导者-追随者关系: $M_a \cap R_b$ 。B 可以向 A 学习 (A 是 B 的指导者并且 B 是 A 追随者)。

隐-同等者关系: $R_a \cap R_b$ 。A 和 B 可以发展成具有共同的音乐品味。

同等者-指导者/同等者-追随者关系: $M_a \cap M_b + M_a \cap R_b$ 。B 可以向 A 学习, 但 B 已经从 A 学到某些东西。这种情况可以被当作同等者的特殊情况或者指导者-追随者的特殊情况。如果被当作前者, 那么可以说这是一种“较强”的同等 (第二个条件确保用户 B 的品味的下一“状

态”也是 A 和 B 之间的同等状态)，然而，如果被当作第二种，那么该情况可以被认为是“较弱”的指导者-追随者关系（追随者将在指导者的音乐中看到自己的一些音乐）。

同等者-兄弟-指导者/同等者-兄弟-追随者关系： $M_a \cap M_b + M_a \cap R_b + R_a \cap R_b$ 。与上面的情况相同，但在推荐中存在相交。

静态指导者-追随者关系： $M_a \cap R_b + R_a \cap R_b$ 。B 可以向 A 学习，并且如果 A 移向下一状态，B 将仍然向 A 学习。这是指导者-追随者的一种较强情况。

交叉列车（crossing-train）关系： $M_a \cap R_b + R_a \cap M_b$ 。B 向 A 学习，并且 A 向 B 学习。但是这些用户的下一状态将不相交，因此这是指导者-追随者的一种奇异情况（由于是双向的）。

出租车（taxi）关系： $M_a \cap R_b + R_a \cap M_b + M_a \cap M_b$ 。与上面的情况相同，但在音乐中存在相交。

相遇列车（meeting-train）关系： $M_a \cap R_b + R_a \cap M_b + R_a \cap R_b$ 。B 向 A 学习，A 向 B 学习，并且它们的下一状态将相交。如果 A 或者 B 移向下一状态，另一个仍旧可以向其学习。如果二者都移动，那么他们将成为同等者。这可能是双向指导者-追随者的最强情况。

完美连接关系： $M_a \cap R_b + R_a \cap M_b + M_a \cap M_b + R_a \cap R_b$ 。一切都相交。

还可以存在用于确定/计算一个艺术家对于一个特定用户有多相关的方法。例如，如果系统具有该用户的艺术家播放计数，通过设置绝对截止点可以归一化该数据，使得某些数量的播放计数可以被认为是“低”，某些其它数量的播放计数可以被认为是“中等”等等。

替代地，如果系统具有用户的一组播放列表，可以计算艺术家在播放列表中出现的次数。随后可以如上所述执行该方法（即，使用截止点）。

作为另一个替代方案，如果系统具有一个推荐集，可以计算艺术家的、基于其在推荐列表中占据的位置的相关性并且在分析中加以使用。当然，这假设推荐器提供被推荐艺术家的分级（ranked）列表。

在一些实现中，按照用户聆听给定艺术家的频次、其简档中有该艺术家的多少歌曲和/或别的方面，用户对给定艺术家的熟悉程度，

可以将用户进一步分类。例如，对于用户聆听的每一个艺术家，具有：

1、F：聆听频率；和

2、K：知识（用户所知的此艺术家的歌曲有多少）。

F 和 K 的值可以被分为高或者低。因此对于特定艺术家，听众可以被分为：

听众	频率	知识
A	低	低
B	低	高
C	高	低
D	高	高

通常，仅同一类型的听众会匹配，但是如果假设这些分类为正方形上的点（其中 0 为低，1 为高），A 到 B 和 C 的距离为 1，到 D 的距离为 $\sqrt{2}$ 。同样，B 到 A 和 D 的距离为 1，到 C 的距离为 $\sqrt{2}$ ，以此类推。

但是，情况也可能是，聆听频率并不是和知识一样相关。因此可以使一个维度（dimension）大于另一个维度，这使得正方形在 K 维度上变大。

使用此方法，A 高度接近 C，中等接近 B，低度接近 D。这些关系用图表示在图 10 中。因此，对于给定用户 U，艺术家 A 的相关性可以被提供为：

$$Rel(U, A) = (1 + K(U, A))^2 + f(U, A)$$

其中， $K(A) \subseteq [0, 1]$ 是测量用户 U 具有的关于艺术家 A 的知识的函数， $f(A) \subseteq [0, 1]$ 是返回用户 U 聆听该艺术家的相对频率的函数。

在一些实施方案中，K 可以从 n/N 推导出，其中，n 是用户所知的某个艺术家的歌曲的数量，N 是此艺术家的歌曲总数。同样，F 可以从 n/P 推导出，其中 n 是用户聆听的某个艺术家的播放计数次数，P 是此用户的播放计数总数。在一些实现中，F 可以通过（上面所述的）Reach API 计算。

以上描述充分公开了包括其优选实施方案的本发明。无需进一步

详细描述,相信本领域的普通技术人员可以使用前面的描述来在最大程度上使用本发明。因此,此处公开的例子和实施方案应当被认为仅是说明性的,而并不以任何方式限制本发明的范围。

对本领域的普通技术人员来说明显的是,在不脱离本发明的基本原理的情况下可以对上述实施方案的细节做出多种改变。因此,应该理解,本发明并不限于所公开的特定实施方案,并且旨在各种改型和其它实施方案被包括在所附权利要求书的范围内。

因此,本发明的范围应该仅由随后的权利要求书确定。

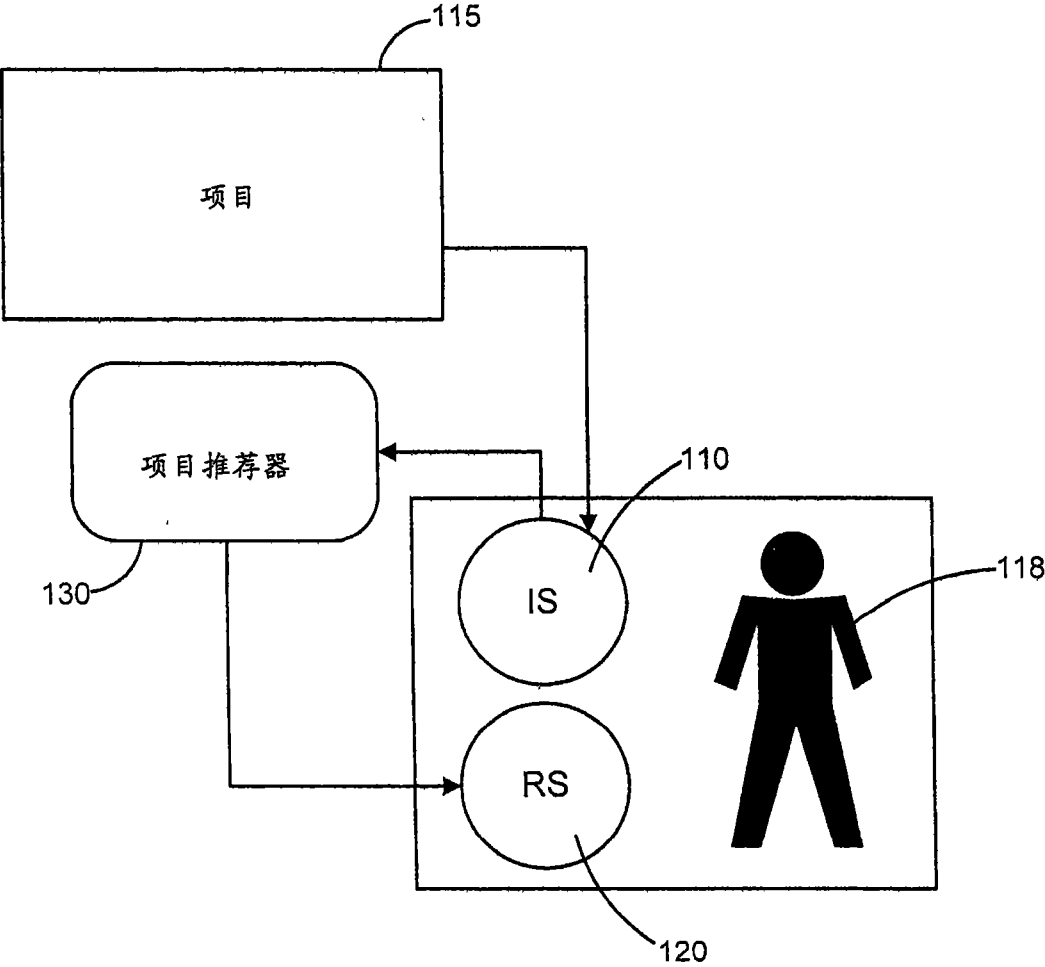


图 1

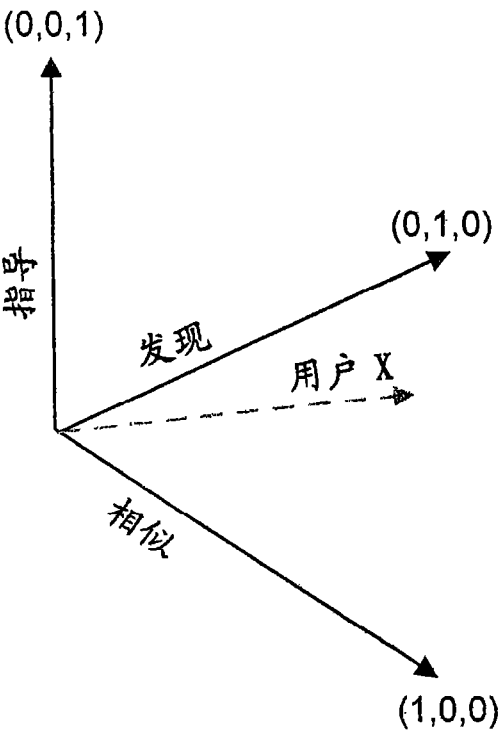


图 2

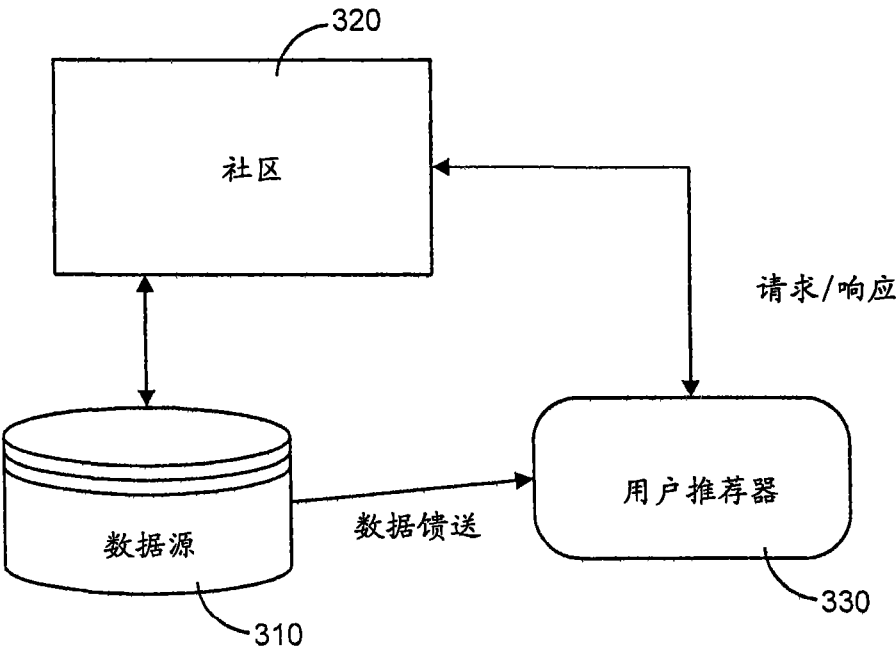


图 3

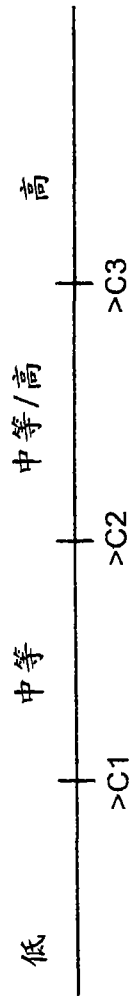


图 4

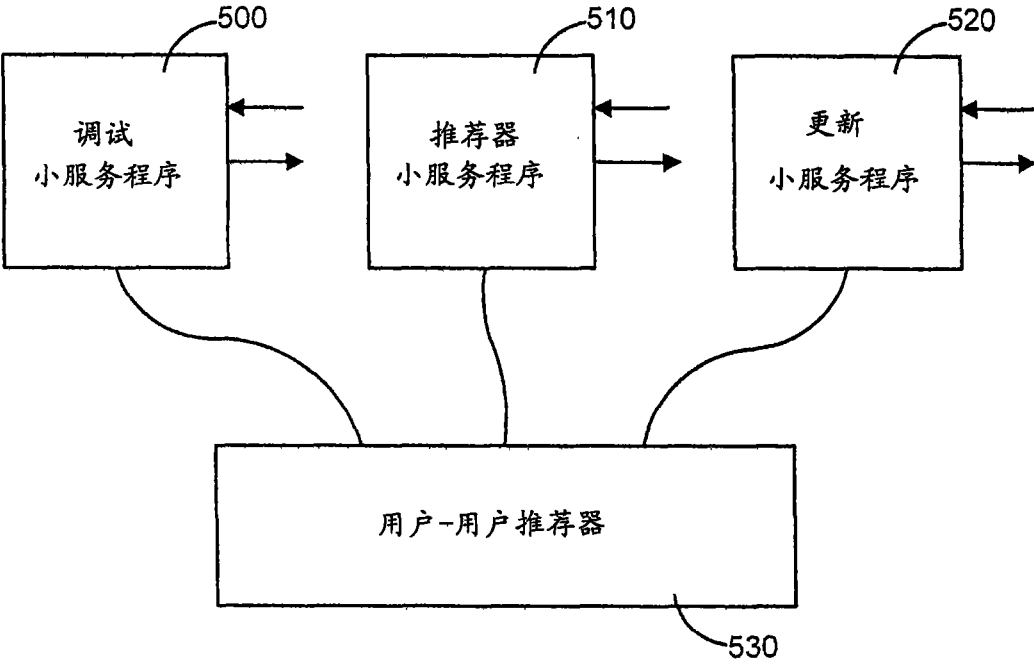


图 5

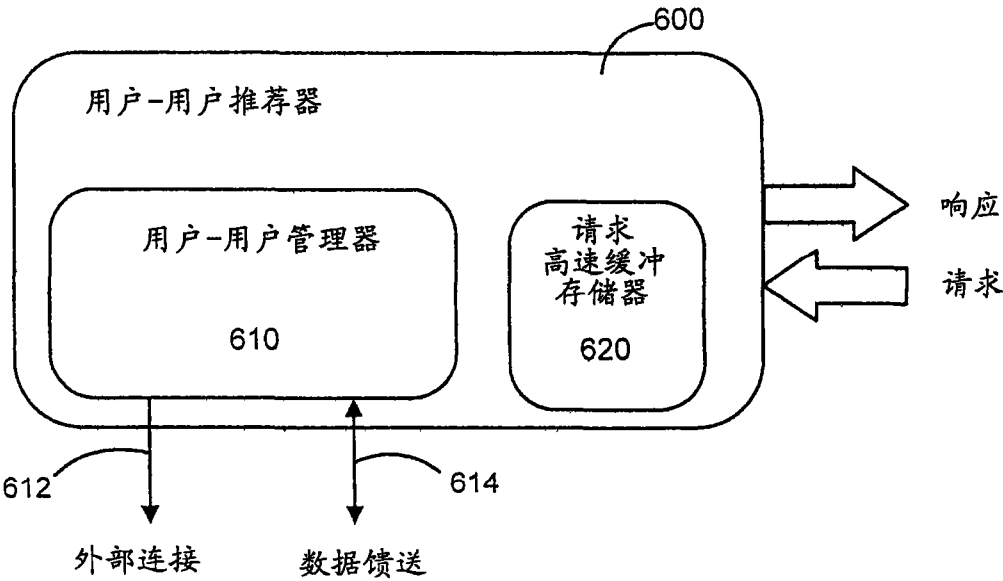


图 6

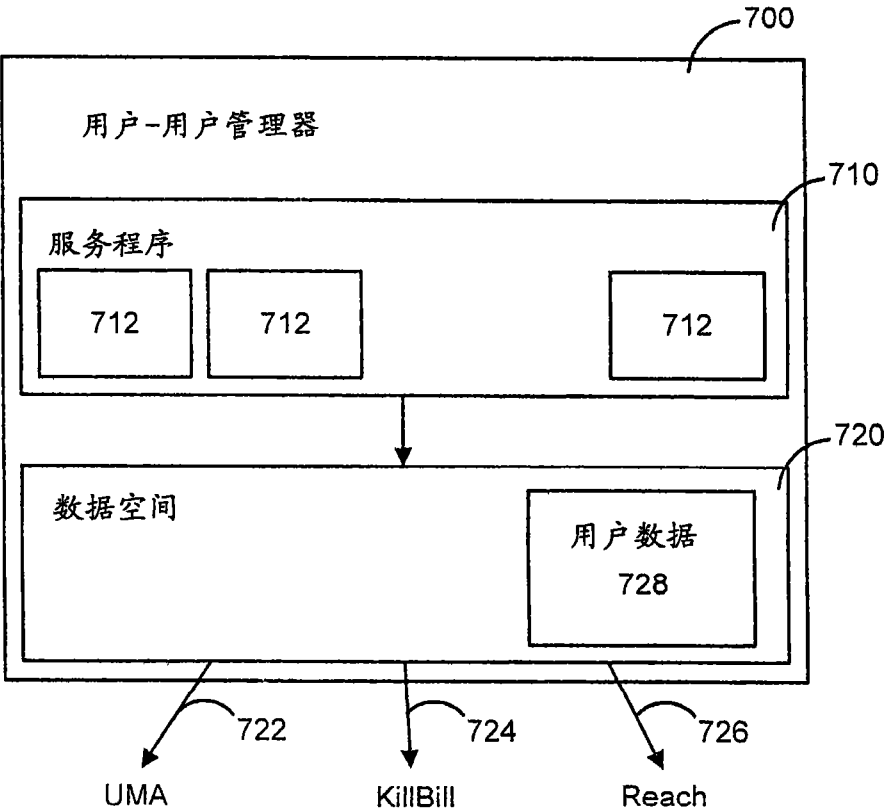


图 7

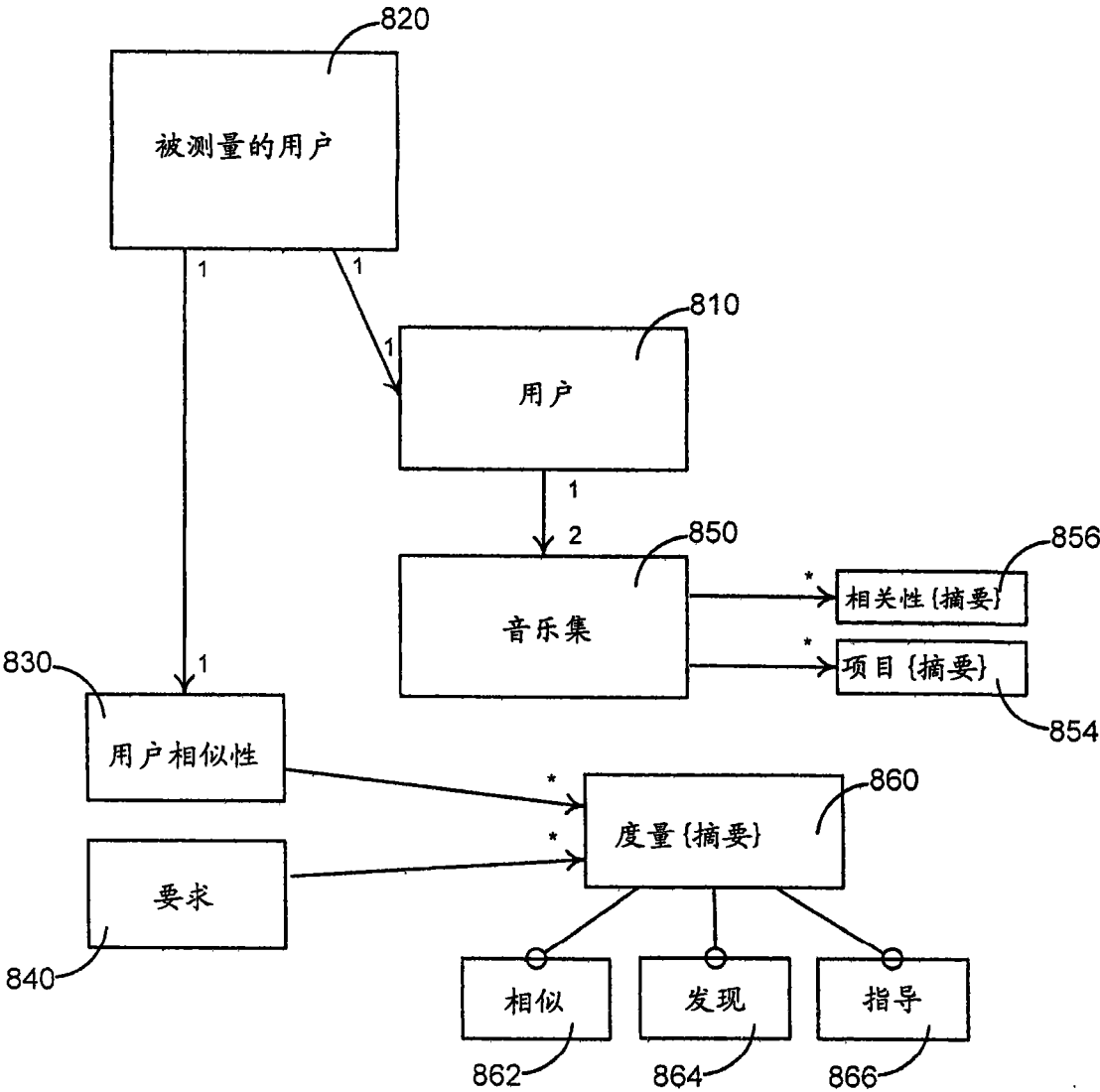


图 8

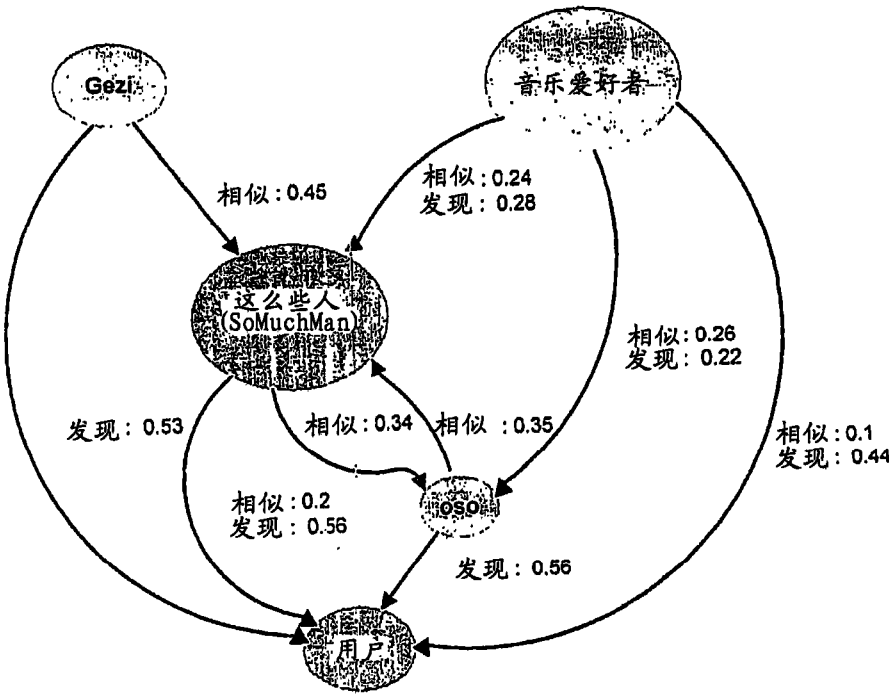


图 9

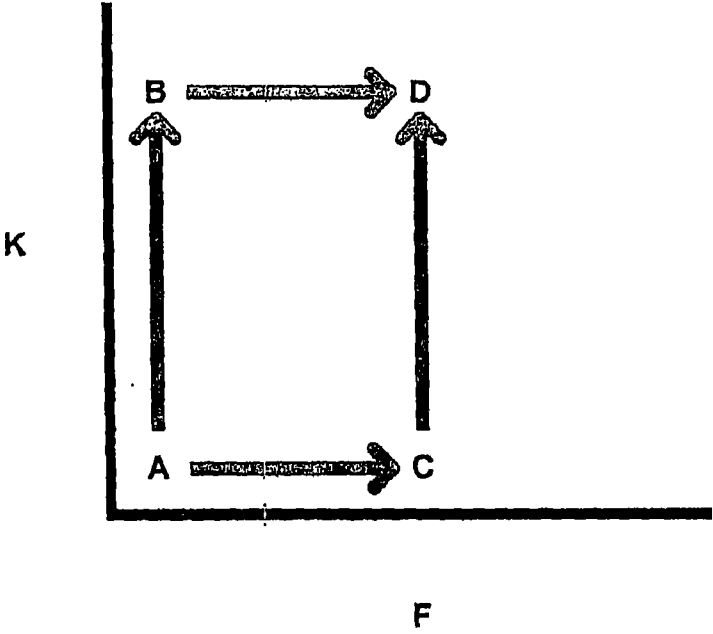


图 10