

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G06F 15/18 (2006.01)



[12] 发明专利说明书

专利号 ZL 200410056761.5

[45] 授权公告日 2010 年 2 月 10 日

[11] 授权公告号 CN 100589095C

[22] 申请日 2004.8.16

[21] 申请号 200410056761.5

[30] 优先权

[32] 2003.10.10 [33] US [31] 10/683,807

[73] 专利权人 微软公司

地址 美国华盛顿州

[72] 发明人 A·拉特纳帕基 B·格罗德尼茨基
F·L·纳兰爵 R·J·拉格诺

[56] 参考文献

CN1438604A 2003.8.27

US2002107842A1 2002.8.8

审查员 林亮亮

[74] 专利代理机构 上海专利商标事务所有限公司
代理人 陈 斌

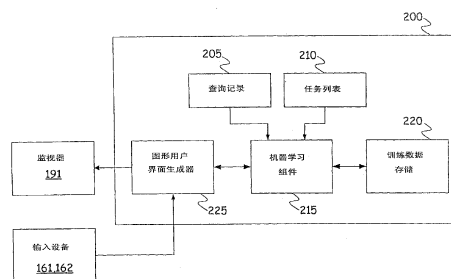
权利要求书 5 页 说明书 10 页 附图 16 页

[54] 发明名称

计算机辅助的查询与任务之间的映射

[57] 摘要

一种注释系统辅助用户将大量查询映射到任务来获取用于训练搜索组件的训练数据。该注释系统包括包含先前向搜索引擎提交的大量查询的查询记录。储存了包含多个可能任务的任务列表。机器学习组件处理查询记录数据和任务列表数据。对对应于查询记录的多个查询条目的每一个，机器学习组件将用于潜在查询与任务之间的映射的最佳猜测任务建议为训练数据的函数。图形用户界面生成组件被配置成以将每一显示的多个查询条目与其对应的建议最佳猜测任务关联的方式显示查询记录中的多个查询条目。



1. 一种用于辅助用户执行查询与任务之间的大量映射来获取用于训练搜索组件的训练数据的方法，其特征在于，所述方法包括：

获取一查询记录，所述查询记录包含先前向搜索引擎提交的查询；

获取一任务列表，所述任务列表包含多个可能的任务；

一机器学习组件将最佳猜测查询与任务之间的映射建议为所述训练数据的函数；以及

一图形用户界面生成组件以关联所述建议的最佳猜测查询与任务之间的映射的方式同时显示所述查询记录中多个查询条目的至少一部分和所述任务列表中多个任务的至少一部分；

其中当用户想要将特定的查询条目映射到其对应的建议最佳猜测任务时，所述图形用户界面生成组件进一步从所述用户接收第一类型输入，并且其中，在通过所述图形用户界面接收所述第一类型输入之后，所述机器学习组件通过将所述特定的查询条目映射到所述建议最佳猜测任务来更新所述训练数据。

2. 如权利要求1所述的方法，其特征在于，所述机器学习组件对对应于所述查询记录的多个查询条目的每一个，将用于潜在查询与任务之间的映射的最佳猜测任务建议为所述训练数据的函数。

3. 如权利要求2所述的方法，其特征在于，所述图形用户界面生成组件以将每一所显示的多个查询条目与其对应的建议最佳猜测任务关联的方式显示所述查询记录中的多个查询条目。

4. 如权利要求1所述的方法，其特征在于，当用户想要察看对潜在映射特定查询条目的下一最佳猜测任务的列表时，所述图形用户界面生成组件从所述用户接收第二类型输入。

5. 如权利要求1所述的方法，其特征在于，对应于所述查询记录的多个查询条目的每一个是一查询束，并且其中每一查询束表示所述查询记录中被捆绑在一起的多个查询。

6. 如权利要求5所述的方法，其特征在于，在通过所述图形用户界面接收所述第一类型输入之后，所述机器学习组件通过将由所述查询束表示的所述多个查询的每一个映射到所述建议最佳猜测任务来更新所述训练数据。

7. 如权利要求 5 所述的方法, 其特征在于, 对每一查询束的所述建议最佳猜测是对由所述查询束表示的多个查询的每一个的最佳猜测的加权平均值。

8. 如权利要求 5 所述的方法, 其特征在于, 当用户想要察看由特定查询束表示的查询的列表时, 所述图形用户界面生成组件从所述用户接收第二类型输入, 并且响应于接收所述第二类型输入, 所述图形用户界面生成组件显示由所述查询束表示的查询的列表。

9. 如权利要求 1 所述的方法, 其特征在于, 所述图形用户界面生成组件显示所述任务列表。

10. 如权利要求 9 所述的方法, 其特征在于, 当用户想要将所述特定的查询条目映射到所述显示的任务列表中不同于所述建议最佳猜测任务的任务时, 所述图形用户生成组件从所述用户接收第二类型输入, 并且其中在通过所述图形用户界面接收所述第二类型输入之后, 所述机器学习组件通过将所述特定的查询条目映射到所显示的任务列表中的所述任务来更新所述训练数据。

11. 如权利要求 1 所述的方法, 其特征在于, 在通过将所述特定的查询条目映射到任务来更新所述训练数据之后, 所述机器学习组件对所述多个查询条目的剩余部分的每一个, 自动更新所述最佳猜测任务作为所述更新的训练数据的函数。

12. 如权利要求 11 所述的方法, 其特征在于, 所述机器学习组件是单纯贝叶斯分类器。

13. 如权利要求 1 所述的方法, 其特征在于, 所述机器学习组件对所述任务列表中所述多个可能任务的每一个, 建议一个来自所述查询记录中可能用于查询与任务之间的映射的查询的列表作为所述训练数据的函数。

14. 一种辅助用户执行查询与任务之间的大量映射来获取用于训练搜索组件的训练数据的方法, 其特征在于, 所述方法包括:

获取包含先前向搜索引擎提交的查询的查询记录;

获取包含多个可能任务的任务列表;

对对应于所述查询记录的多个查询条目的每一个, 确定用于潜在查询与任务之间的映射的最佳猜测任务, 其中, 所述最佳猜测任务使用机器学习组件被确定为所述训练数据的函数;

以将所显示的多个查询条目的每一个与同时显示的所述任务列表中其对应的建议最佳猜测任务关联的方式显示所述查询记录中的所述多个查询条目;

当用户想要将特定的查询条目映射到其对应的建议最佳猜测任务时，从所述用户接收第一类型输入；以及

在接收所述第一类型输入之后，使用所述机器学习组件通过将所述特定的查询条目映射到所述建议最佳猜测任务来更新所述训练数据。

15. 如权利要求 14 所述的方法，其特征在于，还包括：

当用户想要察看针对所述特定查询条目的用于潜在映射的下一最佳猜测任务列表时，从所述用户接收第二类型输入；

显示所述下一最佳猜测任务的列表，它使用所述机器学习组件对所述特定的查询条目被确定为所述训练数据的函数。

16. 如权利要求 14 所述的方法，其特征在于，对应于所述查询记录的所述多个查询条目的每一个是一查询束，并且其中，每一查询束表示所述查询记录中被捆绑在一起的多个查询。

17. 如权利要求 16 所述的方法，其特征在于，在接收所述第一类型输入之后，所述方法还包括：

使用所述机器学习组件，通过将由所述查询束表示的所述多个查询的每一个映射到所述建议最佳猜测任务来更新所述训练数据。

18. 如权利要求 16 所述的方法，其特征在于，还包括：

当用户想要察看由特定的查询束表示的查询的列表时，从所述用户接收第二类型输入；以及

响应于接收所述第二类型输入，显示由所述查询束表示的查询列表。

19. 如权利要求 14 所述的方法，其特征在于，还包括显示所述任务列表。

20. 如权利要求 19 所述的方法，其特征在于，还包括：

当用户想要将所述特定的查询条目映射到所显示的任务列表中不同于所述建议最佳猜测任务的任务时，从所述用户接收第二类型输入；以及

在接收所述第二类型输入之后，使用所述机器学习组件，通过将所述特定查询条目映射到所显示的任务列表中的所述任务来更新所述训练数据。

21. 如权利要求 14 所述的方法，其特征在于，在通过将所述特定的查询条目映射到任务来更新所述训练数据之后，还包括使用所述机器学习组件，对所述多个查询条目的剩余部分的每一个，自动将所述最佳猜测任务更新为所述更新的训练数据的函数。

22. 如权利要求 21 所述的方法, 其特征在于, 所述机器学习组件是单纯贝叶斯分类器。

23. 如权利要求 14 所述的方法, 其特征在于, 对对应于所述查询记录的所述多个查询的每一个确定最佳猜测任务还包括, 使用所述机器学习组件, 对所述任务列表中的每一任务, 将所述查询记录中对查询与任务之间的映射可能的查询列表确定为所述训练数据的函数。

24. 一种辅助用户执行查询与任务之间的大量映射来获取用于训练搜索组件的训练数据的系统, 其特征在于, 所述系统包括:

用于获取包含先前向搜索引擎提交的查询的查询记录的装置;

用于获取包含多个可能任务的任务列表的装置;

用于对对应于所述查询记录的多个查询条目的每一个, 确定用于潜在查询与任务之间的映射的最佳猜测任务的装置, 其中, 所述最佳猜测任务使用机器学习组件被确定为所述训练数据的函数;

用于以将所显示的多个查询条目的每一个与同时显示的所述任务列表中其对应的建议最佳猜测任务关联的方式显示所述查询记录中的所述多个查询条目的装置;

用于当用户想要将特定的查询条目映射到其对应的建议最佳猜测任务时, 从所述用户接收第一类型输入的装置; 以及

用于在接收所述第一类型输入之后, 使用所述机器学习组件通过将所述特定的查询条目映射到所述建议最佳猜测任务来更新所述训练数据的装置。

25. 一种通过使用服务器和本地计算机来执行查询与任务之间的大量映射来获取用于训练搜索组件的训练数据的方法, 其特征在于, 所述方法包括:

从服务器将分类器模型的副本下载到本地计算机;

基于使用所述下载的分类器模型生成的猜测, 在所述本地计算机上执行查询与任务之间的映射;

基于所述本地计算机上执行的查询与任务之间的映射, 更新本地训练数据存储; 以及

使用所述本地训练数据存储更新用于创建所述分类器模型的服务器训练数据存储。

26. 如权利要求 25 所述的方法, 其特征在于, 它还包括在所述服务器上使用

所述更新的服务器训练数据存储来更新所述分类器模型。

27. 如权利要求 25 所述的方法, 其特征在于, 它还包括对多个本地计算机的每一个重复所述方法的步骤, 以使用来自所述多个计算机的每一个的本地训练数据存储来更新所述服务器训练数据存储。

28. 一种通过使用服务器和本地计算机来执行查询与任务之间的大量映射来获取用于训练搜索组件的训练数据的系统, 其特征在于, 所述系统包括:

用于从服务器将分类器模型的副本下载到本地计算机的装置;

用于基于使用所述下载的分类器模型生成的猜测, 在所述本地计算机上执行查询与任务之间的映射的装置;

用于基于所述本地计算机上执行的查询与任务之间的映射, 更新本地训练数据存储的装置; 以及

用于使用所述本地训练数据存储更新用于创建所述分类器模型的服务器训练数据存储的装置。

计算机辅助的查询与任务之间的映射

技术领域

本发明涉及训练用户接口的机器学习组件，尤其涉及通过向任务映射查询来获取训练数据。

背景技术

自然用户接口接受自然语言查询，并且作为响应，返回最可能对应于期望的查询主题的结果的列表。结果通常包括有希望地回答查询的任务、文档、文件、电子邮件或其它项目（所有这部分此处都共同地称为任务）。用于对查询生成结果的期望的技术是机器学习技术。机器学习算法使用统计数据来对特定的查询预测期望的结果。使用机器学习算法，可以在初始训练之后基于来自用户的更多查询的结果经常或频繁地更新统计数据。

在可以提供任一机器学习算法用于自然用户接口，期望它能够提供有意义的结果之前，必须使用准确的注释数据对算法进行“训练”。换言之，算法需要指示来自查询与任务之间的（query-to-task）映射的大列表的统计的训练数据。当要向顾客布置自然用户接口和对应的机器学习算法时，更需要在其布置之前使用准确的注释数据训练机器学习算法。例如，当机器学习算法的一种类型的输出是任务列表，如“安装打印机”或“打印机故障查找”时，该机器学习算法需要表示这部分任务对其为期望的结果的自然语言查询的示例的数据。

为提高自然用户接口的机器学习算法的准确度，训练数据必须表示查询与任务之间的映射的示例的相当大的列表。按照惯例，通过获取包含提交给搜索引擎的相当大量的实际查询的查询记录生成大量的查询与任务之间的映射。例如，查询记录通常包括大约 10,000 个查询或更多。用户或作者然后可以逐一地仔细检查这部分查询，并手动地注释它们（将它们与特定的任务关联）。

向任务注释查询的一种常见的方法是在电子表格数据库的第一列中表示每一查询，并在电子表格的第二列的同一行表示其对应的预期任务。因此，这一过程是相当劳动力密集型且耗时的。此外，给定要从中选择的潜在任务的相当大的列表，

选择哪一任务来与特定的查询注释变得更棘手。

因此，一种能够被用来方便更快且更准确的查询与任务之间的映射来获取训练数据的系统或方法将是本领域中重大的改进。

发明内容

一种注释系统辅助用户将大量查询映射到任务来获取用于训练搜索组件的训练数据。该注释系统包括包含先前向搜索引擎提交的大量查询的查询记录。储存了包含多个可能任务的任务列表。机器学习组件处理查询记录数据和任务列表数据。对于对应于查询记录的多个查询条目的每一个，机器学习组件将用于潜在查询与任务之间的映射的最佳猜测任务建议为训练数据的函数。图形用户界面生成组件被配置成以将所显示的多个查询条目与其对应的建议最佳猜测任务关联的方式显示查询记录中的多个查询条目。

附图说明

图 1 是可使用本发明的一个示例性环境的框图。

图 2 所示是依照本发明的注释系统的一个示例性实施例的框图。

图 3-7 所示是可以使用图 2 所示的系统生成来辅助用户有效地执行查询与任务之间的映射的图形用户界面的图解。

图 8-12 所示是本发明的方法的实施例的流程图。

图 13 和 14 所示是可以使用图 2 的系统生成来辅助用户有效地执行将查询与任务之间的映射作为以任务为中心的操作的图形用户界面的图解，它与图 3-7 所示的以查询为中心的操作相反。

图 15 所示是可使用本发明的一个示例性环境的框图。

图 16 所示是本发明的方法的一个实施例的流程图。

具体实施方式

对于用于辅助用户执行从查询记录到任务的大量映射来获取用于训练搜索组件的训练数据的注释系统来描述本发明。本发明也包括辅助用户将大量查询映射到对应的任务来获取和/或更新训练数据的方法以及配备了这部分方法的计算机可读媒质。

图 1 示出了适合在其中实现本发明的计算系统环境 100 的一个示例。计算系统环境 100 仅为合适的计算环境的一个示例,并非建议对本发明的使用或功能的范围的任何局限。也不应将计算环境 100 解释为对示例性操作环境 100 中示出的任一组件或其组合具有依赖或需求。

本发明可以使用众多其它通用或专用计算系统环境或配置来操作。适合使用本发明的众所周知的计算系统、环境和/或配置包括但不限于,个人计算机、服务器计算机、手持式或膝上设备、多处理器系统、基于微处理器的系统、机顶盒、可编程消费者电子设备、网络 PC、小型机、大型机、包括任一上述系统或设备的分布式计算环境等等。

本发明将在计算机可执行指令的一般上下文环境中描述,计算机可执行指令如程序模块,由计算机执行。一般而言,程序模块包括例程、程序、对象、组件、数据结构等等,执行特定的任务或实现特定的抽象数据类型。本发明也可以在分布式计算环境中实践,其中,任务由通过通信网络连接的远程处理设备来执行。在分布式计算环境中,程序模块可以位于本地和远程计算机存储介质中,包括存储器存储设备。

参考图 1,用于实现本发明的示例性系统包括以计算机 110 形式的通用计算装置。计算机 110 的组件可包括但不限于,处理单元 120、系统存储器 130 以及将各类系统组件包括系统存储器耦合至处理单元 120 的系统总线 121。系统总线 121 可以是若干种总线结构类型的任一种,包括存储器总线或存储器控制器、外围总线以及使用各类总线结构的本地总线。作为示例而非局限,这类结构包括工业标准体系结构 (ISA) 总线、微通道体系结构 (MCA) 总线、增强 ISA (EISA) 总线、视频电子标准协会 (VESA) 本地总线以及外围部件互连 (PCI) 总线,也称为夹层 (Mezzanine) 总线。

计算机 110 通常包括各种计算机可读介质。计算机可读介质可以是可由计算机 110 访问的任一可用介质,包括易失和非易失介质、可移动和不可移动介质。作为示例而非局限,计算机可读介质包括计算机存储介质和通信介质。计算机存储介质包括以用于储存信息的任一方法或技术实现的易失和非易失,可移动和不可移动介质,信息如计算机可读指令、数据结构、程序模块或其它数据。计算机存储介质包括但不限于, RAM、ROM、EEPROM、闪存或其它存储器技术、CD-ROM、数字多功能盘 (DVD) 或其它光盘存储、磁盒、磁带、磁盘存储或其它磁存储设备、

或可以用来储存所期望的信息并可由计算机 110 访问的任一其它媒质。通信媒质通常在诸如载波或其它传输机制的已调制数据信号中包含计算机可读指令、数据结构、程序模块或其它数据，并包括任一信息传送媒质。术语“已调制数据信号”指以对信号中的信息进行编码的方式设置或改变其一个或多个特征的信号。作为示例而非局限，通信媒质包括有线媒质，如有线网络或直接连线连接，以及无线媒质，如声学、RF、红外和其它无线媒质。上述任一的组合也应当包括在计算机可读媒质的范围之内。

系统存储器 130 包括以易失和/或非易失存储器形式的计算机存储媒质，如只读存储器（ROM）131 和随机存取存储器（RAM）132。基本输入/输出系统 133（BIOS）包括如在启动时帮助在计算机 110 内的元件之间传输信息的基本例程，通常储存在 ROM 131 中。RAM 132 通常包含处理单元 120 立即可访问或者当前正在操作的数据和/或程序模块。作为示例而非局限，图 1 示出了操作系统 134、应用程序 135、其它程序模块 136 和程序数据 137。

计算机 110 也可包括其它可移动/不可移动、易失/非易失计算机存储媒质。仅作为示例，图 1 示出了对不可移动、非易失磁媒质进行读写的硬盘驱动器 141、对可移动、非易失磁盘 152 进行读写的磁盘驱动器 151 以及对可移动、非易失光盘 156，如 CD ROM 或其它光媒质进行读写的光盘驱动器 155。可以在示例性操作环境中使用的其它可移动/不可移动、易失/非易失计算机存储媒质包括但不限于，磁带盒、闪存卡、数字多功能盘、数字视频带、固态 RAM、固态 ROM 等等。硬盘驱动器 141 通常通过不可移动存储器接口，如接口 140 连接到系统总线 121，磁盘驱动器 151 和光盘驱动器 155 通常通过可移动存储器接口，如接口 150 连接到系统总线 121。

图 1 讨论并示出的驱动器及其关联的计算机存储媒质为计算机 110 提供了计算机可读指令、数据结构、程序模块和其它数据的存储。例如，在图 1 中，示出硬盘驱动器 141 储存操作系统 144、应用程序 145、其它程序模块 146 和程序数据 147。注意，这部分组件可以与操作系统 134、应用程序 135、其它程序模块 136 和程序数据 137 相同，也可以与它们不同。这里对操作系统 144、应用程序 145、其它程序模块 146 和程序数据 147 给予不同的标号来说明至少它们是不同的副本。

用户可以通过输入设备，如键盘 162、麦克风 163 和定点设备 161，如鼠标、轨迹球或触摸板向计算机 110 输入命令和信息。其它输入设备（未示出）可包括操

纵杆、游戏垫、圆盘式卫星天线、扫描仪等等。这部分和其它输入设备通常通过耦合至系统总线的用户输入接口 160 连接至处理单元 120，但是也可以通过其它接口和总线结构连接，如并行端口、游戏端口或通用串行总线（USB）。监视器 191 或其它类型的显示设备也通过接口，如视频接口 190 连接至系统总线 121。除监视器之外，计算机也包括其它外围输出设备，如扬声器 197 和打印机 196，通过输出外围接口 195 连接。

计算机 110 可以在使用到一个或多个远程计算机，如远程计算机 180 的逻辑连接的网络化环境中操作。远程计算机 180 可以是计算机、手持式设备、服务器、路由器、网络 PC、对等设备或其它公用网络节点，并通常包括许多或所有上述与计算机 110 相关的元件，图 1 描述的逻辑连接包括局域网（LAN）171 和广域网（WAN）173，但也可以包括其它网络。这类网络环境常见于办公室、企业范围计算机网络、内联网以及因特网。

当在 LAN 网络环境中使用时，计算机 110 通过网络接口或适配器 170 连接至 LAN 171。当在 WAN 网络环境中使用时，计算机 110 通常包括调制解调器 172 或其它装置，用于通过 WAN 173，如因特网建立通信。调制解调器 172 可以是内置或外置的，通过用户输入接口 160 连接至系统总线 121。在网络化环境中，描述的与计算机 110 相关的程序模块或其部分可储存在远程存储器存储设备中。作为示例而非局限，图 1 示出了远程应用程序 185 驻留在远程计算机 180 上。可以理解，示出的网络连接是示例性的，也可以使用在计算机之间建立通信链路的其它装置。

图 2 示出了依照本发明的注释系统 200，它可以在诸如图 1 示出的处理环境中实现。注释系统 200 辅助用户执行查询—任务大量映射来获取用于训练搜索组件的训练数据。系统 200 包括，或在其它实施例中获取，包含表示先前向搜索引擎提交的查询的数据的查询记录 205。通常，查询记录包含表示大量查询，如 10,000 个或更多的数据。然而，也可以使用较小的查询记录。包括了包含表示大量可能任务的数据的任务列表 210。为获取并连续地更新训练数据 220，系统 200 辅助用户将查询记录 205 中的每一查询映射到任务列表 210 中的任务。

机器学习组件或分类器 215 从查询记录 205 中检索查询条目，并从任务列表 210 中检索任务的列表。如本发明所定义的，查询条目可以是，如，查询记录中个别的查询。可选地，查询条目可以是表示查询记录 205 被捆绑在一起的多个查询的查询束。通过查询记录群集的原始查询记录来创建查询束。查询束的一个示例是在

查询中包括各种单词的小拼写错误的一系列类似的查询。将查询束映射到任务，而不是每次将个别的查询映射到一个任务节省了可观的时间。

对于对应于查询记录的多个查询条目的每一个，机器学习组件使用训练数据 220 来建议用于潜在查询与任务之间的映射的最佳猜测任务。由机器学习组件基于先前映射的查询的统计信息生成猜测。最初，在训练数据 220 中只有少量有意义的统计数据，因此最佳猜测不会如所期望的准确。然而，当映射了更多的查询，将提高猜测的准确度。如后文所更详细描述，用户或作者总是具有选择来丢弃系统的猜测。

实际上，每次作者接受一个猜测，他或她“教”系统如何将查询映射到任务。在通过将特定的查询条目映射到任务来更新训练数据之后，机器学习组件 215 被配置成对每一剩余的查询条目自动地将最佳猜测任务更新为更新的训练数据 220 的函数。一旦注释了足够的数据，它可以用来训练机器学习算法以在自然用户接口平台的搜索组件中展开。可以以这一方式连续更新其最佳猜测的机器学习组件或分类器类型的一个示例是单纯贝叶斯分类器（Naive Bayes Classifier）。

在系统 200 中，包括了图形用户界面（GUI）生成组件 225 以在机器学习组件 215 和用户之间通过显示器（如监视器 191）和输入设备（如设备 161、162）进行接口。组件 225 被配置成以将每一显示的多个查询条目与其对应的建议最佳猜测任务关联的方式显示查询记录中的多个查询条目。组件 225 也被配置成从用户接收各种类型的输入来促进映射过程。在图 3-7 中示出了依照示例性实施例由组件 225 生成的 GUI。

现在参考图 3，示出了依照本发明的一部分实施例能够在监视器 191 上显示的 GUI 300。GUI 300 包括任务列表 302、查询条目列表（示出为查询束列表）304、指示每一查询束中的查询数量的列 306、指示查询出现的频率的列 308、由机器学习组件 215 生成的对每一查询条目的最佳猜测 310、以及指示建议最佳猜测正确的概率的列 312。任务列表 302 是如果作者不对查询与任务之间的映射选择最佳猜测任务则查询条目可以向其映射的任务的列表。同时，当查询条目 304 是查询束时，对每一查询束的建议最佳猜测任务 310 可以是对由查询束表示的多个查询的每一个的最佳猜测的加权平均值。

现在参考图 4，GUI 生成组件 225 还被配置成当用户想要将特定的查询条目映射到其对应的建议最佳猜测任务时，从用户接收第一类型输入。在图 4 所示的示例

中，用户将由查询束“password（密码）”表示的所有查询映射到由机器学习组件生成的“change password（改变密码）”的最佳猜测任务。用于指示这一映射进行的第一类型输入可以是，例如，将光标 340 定位在这一查询条目上进行鼠标类定点设备的右键点击。行 350 表示这一查询条目以及关联的统计和最佳猜测任务的选择或加亮。在通过 GUI 300 接收第一类型输入之后，机器学习组件 215 通过将特定的查询条目映射到建议最佳猜测任务来更新训练数据。如上所述，然后基于更新的训练数据更新对剩余查询条目的最佳猜测任务。

现在参考图 5，如果用户想要回顾统计地确定的下一最佳猜测任务的列表，而不是将查询条目映射到建议最佳猜测任务，则可以通过提供适当的输入来实现。例如，在一个实施例中，将光标 340 定位在建议最佳猜测“change password”任务上并点击鼠标，则显示包含下一最佳猜测的列表的下拉菜单 360。用户然后可以通过在菜单 360 中期望的任务上点击将所述的查询条目映射到下一最佳猜测任务的任一个。

现在参考图 6，如期望，用户可以察看由查询束类型的查询条目表示的个别查询，例如，在一部分实施例中，GUI 生成组件 225 被配置成当用户想要察看包含由特定的查询束表示的查询的列表的下拉菜单 370 时，从用户接收第二类型输入。响应于该输入，GUI 生成组件显示菜单 370。第二类型输入的一个示例是将光标定位在“password”查询束上进行双击。然后如需要，可以每次将菜单 370 中列出的个别查询映射到一个任务。

现在参考图 7，如果没有一个建议最佳猜测任务适合映射到查询条目，则用户能够通过提供适当的输入将该查询条目映射到任务列表 302 中的一个任务。作为示例，在一个实施例中，适当的输入包括选择一查询条目（如，如选择行 380 所示的“cd”），然后在列表 302 中的任务上点击。

现在参考图 8-12，示出了本发明的方法的一部分实施例的各种步骤的流程图 400、450、500、550 和 600。这部分流程图概括了如上文参考图 1-7 所描述的本发明的一部分方面。如图 8 的流程图 400 所示，辅助用户执行查询一任务大量映射来获取用于训练搜索组件的训练数据的方法包括获取先前向搜索引擎提交的查询的查询记录。这在块 402 中示出。如块 404 所示的，该方法也包括获取可能任务的任务列表。在块 406，示出该方法包括对对应于查询记录的多个查询条目的每一个，确定用于潜在查询与任务之间的映射的最佳猜测任务。使用机器学习组件将最佳猜

测任务确定为训练数据的函数。在块 408, 示出该方法包括以将显示的多个查询条目的每一个与其对应的建议最佳猜测任务关联的方式显示查询记录中的多个查询条目。

现在参考图 9, 该方法也可包括当用户想要将特定的查询条目映射到其对应的建议最佳猜测时, 从用户接收第一类型输入的步骤, 如块 452 所示, 以及使用机器学习组件更新训练数据的步骤, 如块 454 所示。

现在参考图 10, 在一部分实施例中, 本发明的方法还包括块 502 所示的当用户想要察看对特定查询条目的用于潜在映射的下一最佳猜测任务的列表时, 从用户接收第二类型输入的步骤。在这部分实施例中, 该方法然后包括块 504 所示的显示对特定的查询条目, 使用机器学习组件被确定为训练数据的函数的下一最佳猜测任务的列表的步骤。

在一部分实施例中, 本发明的方法包括图 11 的流程图 550 所示的另外的步骤。如块 552 所示, 在这部分实施例中, 该方法还包括当用户想要察看由特定查询束表示的查询列表时, 从用户接收第二类型输入。如块 554 所示, 然后响应于接收第二类型输入, 显示由查询束表示的查询的列表。

现在参考图 12, 示出了能够包括在一部分实施例中允许用户将特定的查询条目映射到任务列表 302 中的任务的另外的步骤。例如, 如流程图 600 的块 602 所示, 该方法还可包括当用户想要将特定的查询任务映射到显示的任务列表中不同于建议最佳猜测任务的任务时, 从用户接收第二类型输入的步骤。在如块 604 所示的步骤, 然后使用机器学习组件通过将特定的查询条目映射到显示的任务列表中的任务来更新训练数据。

图 3-7 中所示的 GUI 实施例将本发明的方法实现为“以查询为中心”的操作。换言之, 在这部分实施例中, 用户向多个查询的每一个分配最佳猜测任务。然而, 本发明也可被实现为“以任务为中心”的操作, 其中, 向特定的任务分配最佳猜测查询。

例如, 参考图 13, 示出了依照本发明的一部分实施例能够在监视器 191 上显示的 GUI 700。GUI 700 包括分层任务列表 702。执行查询与任务之间的映射的用户或作者从分层任务列表 702 中选择一个任务, 来察看用于映射到任务的建议查询的列表 704。这在图 14 中示出。

GUI 生成组件 225 被配置成当用户想要将任务列表 702 中特定的任务映射到

对应的建议最佳猜测查询的一个或多个时，从用户接收第一类型输入。例如，可以使用鼠标类定点设备在光标 740 上点击来提供这类型输入。行 750 表示这一任务的选择或加亮。响应于这一输入，系统或工具然后显示机器学习组件所建议的查询的列表 704。以匹配所选的任务的似然性的顺序显示查询。这一过程允许用户从任务映射到可能的查询，与从查询映射到可能的任务相反。在其它方面，包括随时间的推移获得猜测的改进，这一过程具有与上述的“以查询为中心”的操作相同的行为。

在本发明的实施例中，多于一个用户能够使用查询与任务之间的映射系统将查询映射到任务。例如，考虑图 15 所示的计算环境 800。计算环境 800 包括中央服务器 802 和多个用户计算机 810（示出了计算机 810-1、810-2 和 810-3）。中央服务器计算机包括储存用于建立分类器模型 806 的训练数据的训练数据存储 804。在所示的系统或处理环境 800 中，由机器学习组件用来向用户提供猜测的分类器模型 806 的创建由所有用户提交的映射的集合来训练。这部分映射储存在中央数据库训练数据存储 804 中。使用这部分映射周期性地训练分类器模型 806。尽管示出训练数据存储 804 和分类器模型 806 都储存在中央服务器 802 上，然而在其它实施例中，分类器模型 806 储存在别处，然后可以向中央服务器 802 保持，用户可以从其下载映射会话之间的更新的模型。

下载模型副本（示出为本地分类器模型 814-1 到 814-3）本地地保持在每一用户的计算机上。用户然后能够使用以从所有用户收集的映射数据至少周期性地训练的模型。由于特定的用户进行映射工作，模型的本地副本 814 适合该个别用户的映射。这由在每一用户计算机上包括本地映射数据存储 812-1 到 812-3 示出。由此，因为本地分类器模型考虑其最近的映射，用户可以从获益。这超出原始保持的模型 806 提高了猜测的准确度。此外，用户能够调节猜测被定制到最近的映射的程度。周期性地，提供了对应于本地用户机器上所执行的映射的映射数据 812-1 到 812-3，以训练中央服务器 802 上的数据存储 804，用于将来对分类器模型 806 的更新。

图 16 的流程图 850 也示出了本发明的多个用户计算机方面，描述了使用中央服务器和本地计算机执行查询与任务之间的大量映射的一般方法。如图 16 的块 852 所示，该方法包括从中央服务器将分类器模型的副本下载到用户计算机。在块 854，示出该方法包括在用户计算机上基于使用下载的分类器模型生成的猜测，作出查询与任务之间的映射的步骤。在块 856，示出该方法包括基于本地映射创建或更新本地训练数据存储的步骤。换言之，使用本地映射来更新储存在用户的计算机上的分

类器模型的本地版本。最后，如块 858 所示，该方法包括在中央服务器使用本地映射数据更新训练数据存储。由此，使用包括在每一个别的用户计算机上执行的映射的训练数据存储 804 更新储存在或下载到中央服务器上的分类器模型 806。

当有稀少的查询与任务之间的映射数据可用，可以使用其它信息源来预训练分类器模型 806。例如，可以在映射过程的最初阶段频繁地使用这一预训练。然后可以使用由所有用户从本地计算机提交的映射的集合来训练该模型。可以调节这部分信息源的每一个的相关影响。这一预训练可以在任一点上应用到模型生成过程。

标准预训练技术是从与任务关联的数据提取一组特征。然后可以在特征和任务之间的关联上训练该模型。其效果与通过将查询映射到任务所产生的效果类似。在这一预训练的最简单形式中，任务名担当对其引入到任务的映射的查询的角色（尽管未明确地保持这一查询）。也可以使用从另一源的文本映射实现预训练。无需在训练数据存储中展现或保持这部分映射。

在本地计算机 810-1 到 810-3 之一上的用户会话过程中，模型的猜测也可以在它们被生成时被影响。例如，如果一个查询已映射到任务，则它不会显示为猜测。这可防止由多个用户作出重复的映射。用户也可以明确地应用过滤器来抑制所呈现的猜测。

尽管参考具体的实施例描述了本发明，然后本领域的技术人员可以认识到，可以在不脱离本发明的精神和范围的情况下在形式和细节上作出变化。

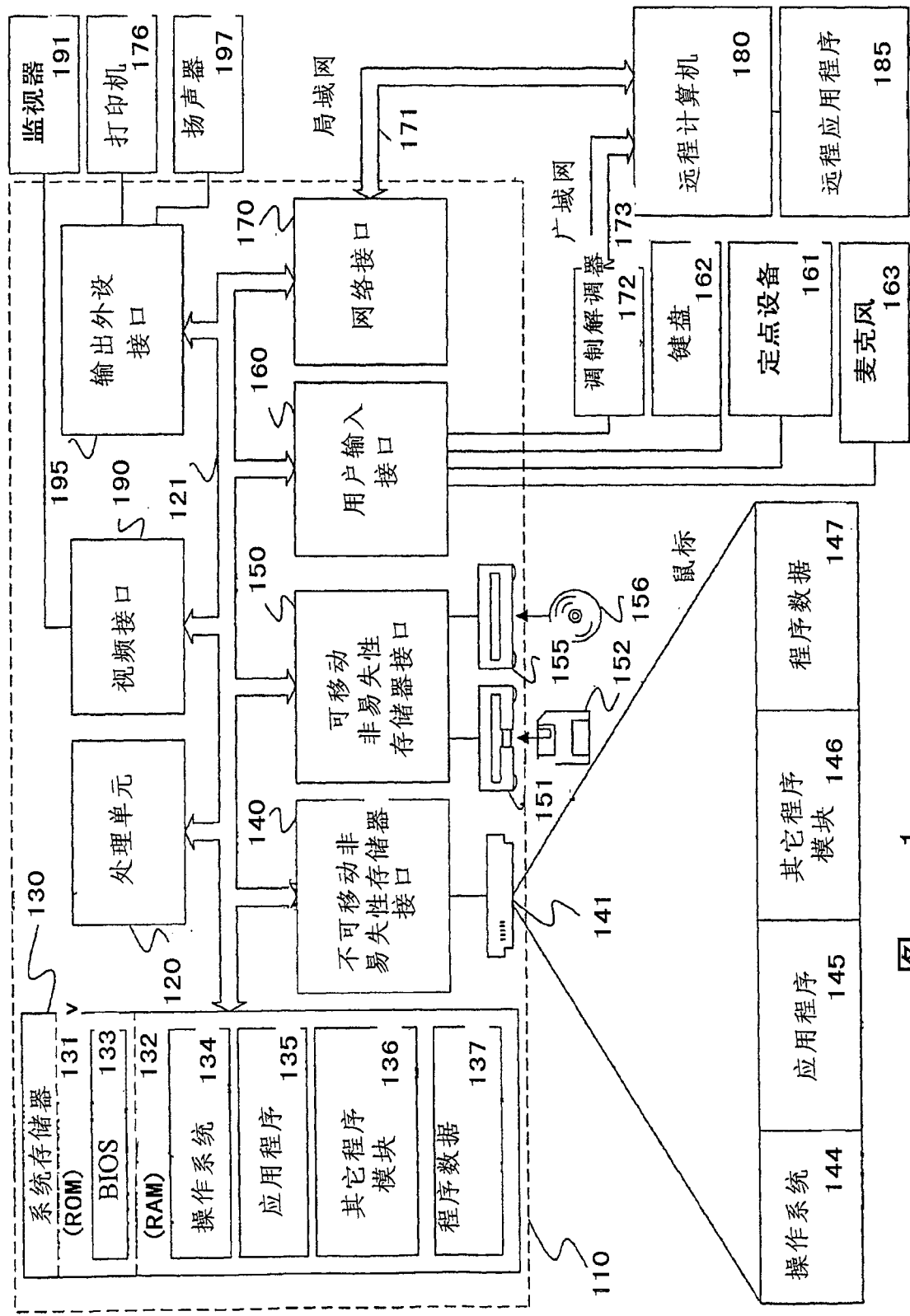


图 1

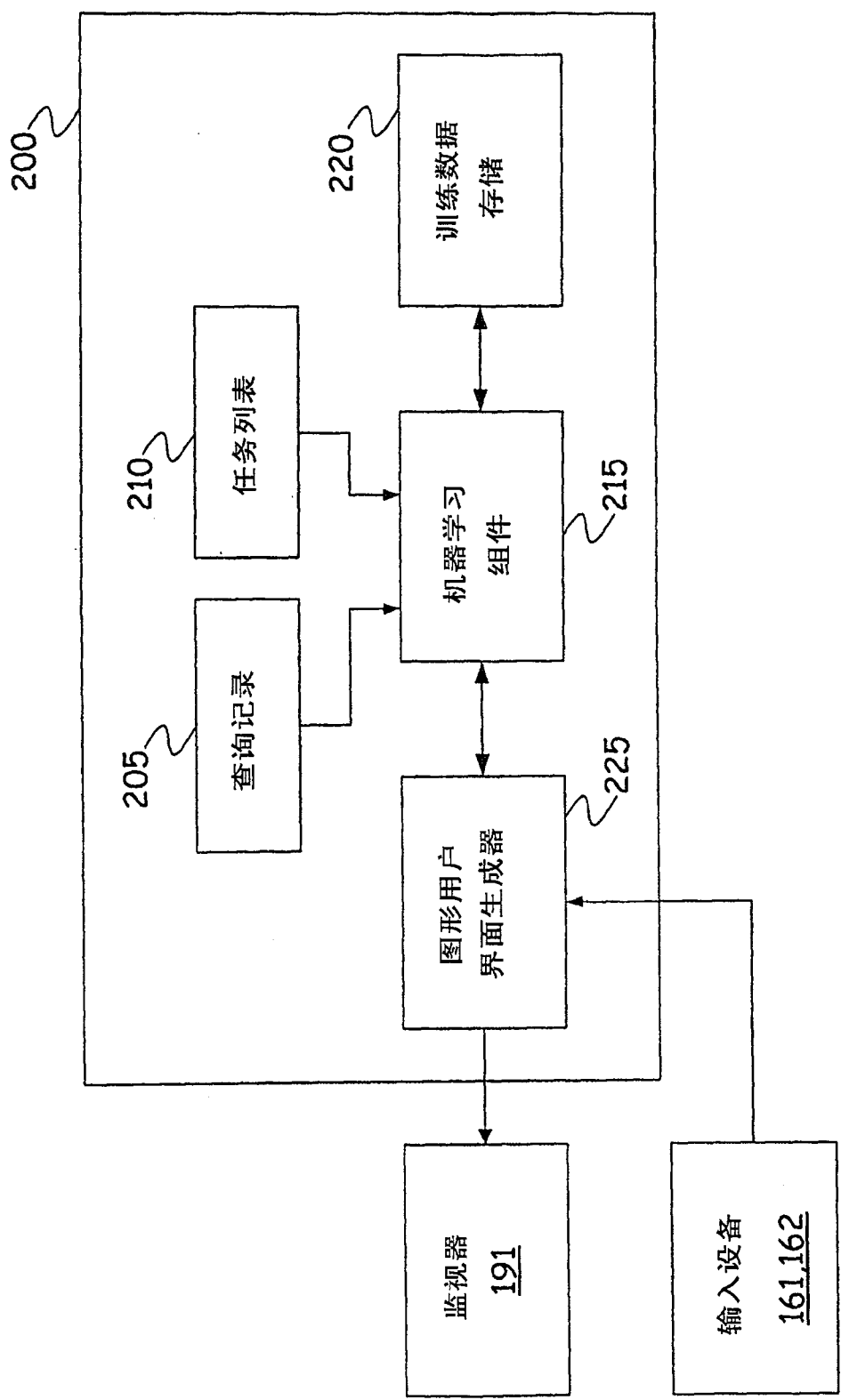


图 2

300

任务列表	查询束	查询数量	频率	最佳猜测	最佳猜测概率
access shared printer access system info add address to address add comment date and add contact add icon to desktop : backup data to tape burn file to CD burn music to CD capture screen content change font in document change password change user account change volume sound to check for e-mail messa close program all compress file and fold configure display setti configure Internet Exp configure modem configure modem for fa :	firewall register netmeeting printer boot disk format cookies password startup email dvd administrator startup disk activation backup ntfs partition cd : compatibility copy cd remote desktop	101 293 112 196 49 172 81 100 229 45 13 309 97 194 35 10 200 24 : 277 49 52	19449 11892 11647 8239 8269 7657 7652 7611 7294 7331 7361 6851 6811 6398 6350 6338 5837 5308 : 4570 4675 4670	install firewall save picture save picture configure printer copy file to floppy disk format paragraph in email save picture change password troubleshoot startup comput e-mail picture save picture administer user account troubleshoot startup comput learn about op system activat learn about backup save picture create partition save picture : save picture copy music learn about desktop	67 1 1 26 15 25 1 28 24 6 1 1 29 10 30 1 44 1 : 1 39 27



3

任务列表	查询束	查询数量	频率	最佳猜测	最佳猜测概率
access shared printer	firewall	101	19449	install firewall	67
access system info	register	293	11892	save picture	1
add address to address	netmeeting	112	11647	save picture	1
add comment date and	printer	196	8239	configure printer	26
add contact	boot disk	49	8269	copy file to floppy disk	15
add icon to desktop	format	350	7657	format paragraph in email	25
:	cookies	81	7652	save picture	1
backup data to tape	passwords	100	7611	change password	28
burn file to CD	startup	229	7294	troubleshoot startup comput	24
burn music to CD	email	45	7331	e-mail picture	6
capture screen content	dvd	13	7361	save picture	1
change font in document	administrator	309	6851	administer user account	1
change password	startup disk	97	6811	troubleshoot startup comput	29
change user account	activation	194	6398	learn about op system activat	10
change volume sound to	backup	35	6350	learn about backup	30
check for e-mail messa	ntfs	10	6338	save picture	1
close program all	partition	200	5837	create partition	44
compress file and fold	cd	24	5308	save picture	1
configure display setti	:	:	:	:	:
configure Internet Exp	compatibility	277	4570	save picture	1
configure modem	copy cd	49	4675	copy music	39
configure modem for fa	remote desktop	52	4670	learn about desktop	27

4


302	304	306	308	310	312	300
任务列表	查询束	查询数量	频率	最佳猜测	最佳猜测概率	
access shared printer	firewall	101	19449	install firewall	67	
access system info	register	293	11892	save picture	1	
add address to address	netmeeting	112	11647	save picture	1	
add comment date and	printer	196	8239	configure printer	26	
add contact	boot disk	49	8269	copy file to floppy disk	15	
add icon to desktop	format	172	7657	format paragraph in email	25	
:	cookies	81	7652	save picture	340	
backup data to tape	password	100	7611	change password	28	
burn file to CD	startup	229	7294	troubleshoot startup cannot	24	
burn music to CD	email	45	Next Best Guesses			
capture screen content	dvd	13	change password			
change font in document	administrator	309	change or delete a password			
change password	startup disk	97	passwords for files			
change user account	activation	194	password doesn't work.			
change volume sound to	backup	35	protect a document			
check for e-mail messa	ntfs	10	strategies for sharing documents			
close program all	partition	200	Require a password to modify a document			
compress file and fold	cd	24	Require a password to open a document			
configure display setti	:	:	Change OS password			
configure Internet Exp	compatibility	277	Change network password			
configure modem	copy cd	49	4675 copy music			39
configure modem for fa	remote desktop	52	4670 learn about desktop			27
:						

图 5

任务列表	查询束	查询数量	频率	最佳猜测	最佳猜测概率
access shared printer	firewall	101	19449	install firewall	67
access system info	register	293	11892	save picture	1
add address to address	netmeeting	112	11647	save picture	1
add comment date and	printer	196	8239	configure printer	26
add contact	boot disk	49	8269	copy file to floppy disk	15
add icon to desktop	format	340	7657	format paragraph in email	25
:	cookies	81	7652	save picture	1
backup data to tape	passwords	100	7611	change password	28
burn file to CD	startup	229	7294	troubleshoot startup comput	24
burn music to CD	emc	229	7294	troubleshoot startup comput	24
capture screen content	dvd	229	7294	troubleshoot startup comput	24
change font in document	adm	229	7294	troubleshoot startup comput	24
change password	stat	229	7294	troubleshoot startup comput	24
change user account	act	229	7294	troubleshoot startup comput	24
change volume sound to	bac	229	7294	troubleshoot startup comput	24
check for e-mail messa	ntf	229	7294	troubleshoot startup comput	24
close program all	par	229	7294	troubleshoot startup comput	24
compress file and fold	cd	229	7294	troubleshoot startup comput	24
configure display setti	:	229	7294	troubleshoot startup comput	24
configure Internet Exp	com	229	7294	troubleshoot startup comput	24
configure modem	cop	229	7294	troubleshoot startup comput	24
configure modem for fa	remore desktop	229	7294	troubleshoot startup comput	24
:	:	229	7294	troubleshoot startup comput	24
:	:	229	7294	troubleshoot startup comput	24

图

6

任务列表	查询束	查询数量	频率	最佳猜测	最佳猜测概率
access shared printer access system info add address to address add comment date and add contact add icon to desktop : backup data to tape 340 burn file to CD	firewall register netmeeting printer boot disk format cookies password startup email	101 293 112 196 49 172 81 100 229 45	19449 11892 11647 8239 8269 7657 7652 7611 7294 7331	install firewall save picture save picture configure printer copy file to floppy disk format paragraph in email save picture change password troubleshoot startup comput	67 1 1 26 15 25 1 28 24
burn music to CD capture screen content change font in document change password change user account change volume sound to check for e-mail messa close program all compress file and fold configure display setti configure Internet Exp configure modem configure modem for fa : compatibility copy cd remote desktop	dvd administrator startup disk activation backup ntfs partition cd : 380 : compatibility copy cd remote desktop	13 309 97 194 35 10 200 24 : 277 49 52	7361 6851 6811 6398 6350 6338 5837 5308 : 4570 4675 4670	e-mail picture save picture administer user account troubleshoot startup comput learn about op system activat learn about backup save picture create partition save picture : save picture copy music learn about desktop	6 1 1 29 10 30 1 44 1 : 1 39 27

图

7

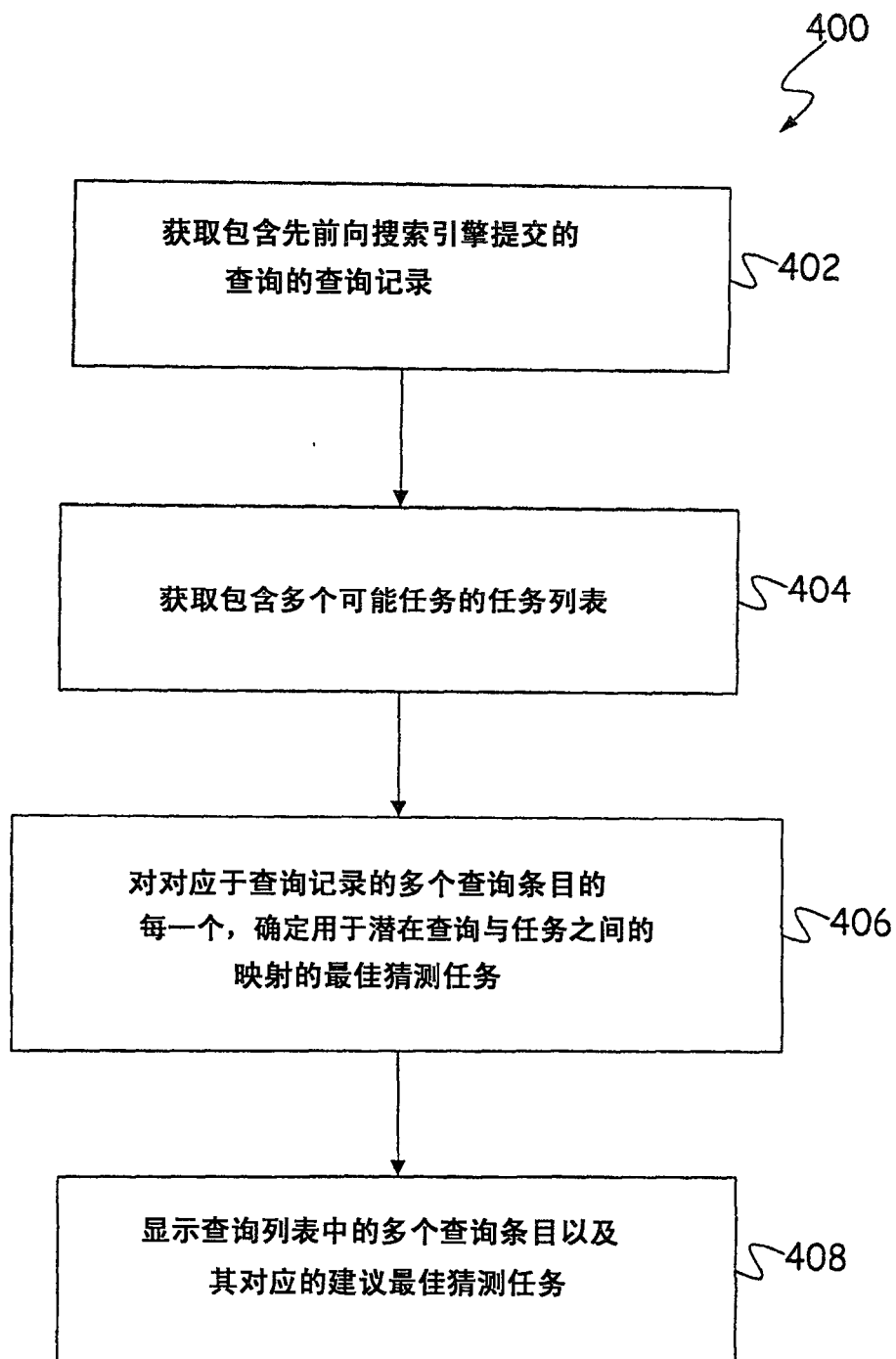


图 8

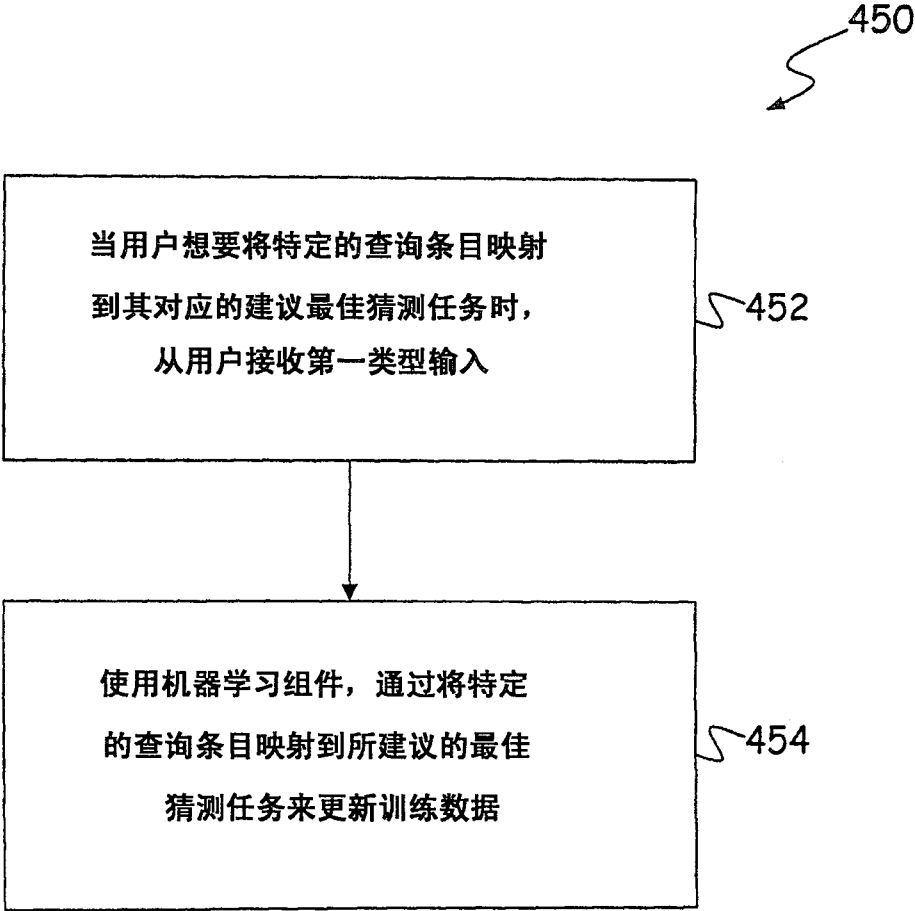


图 9

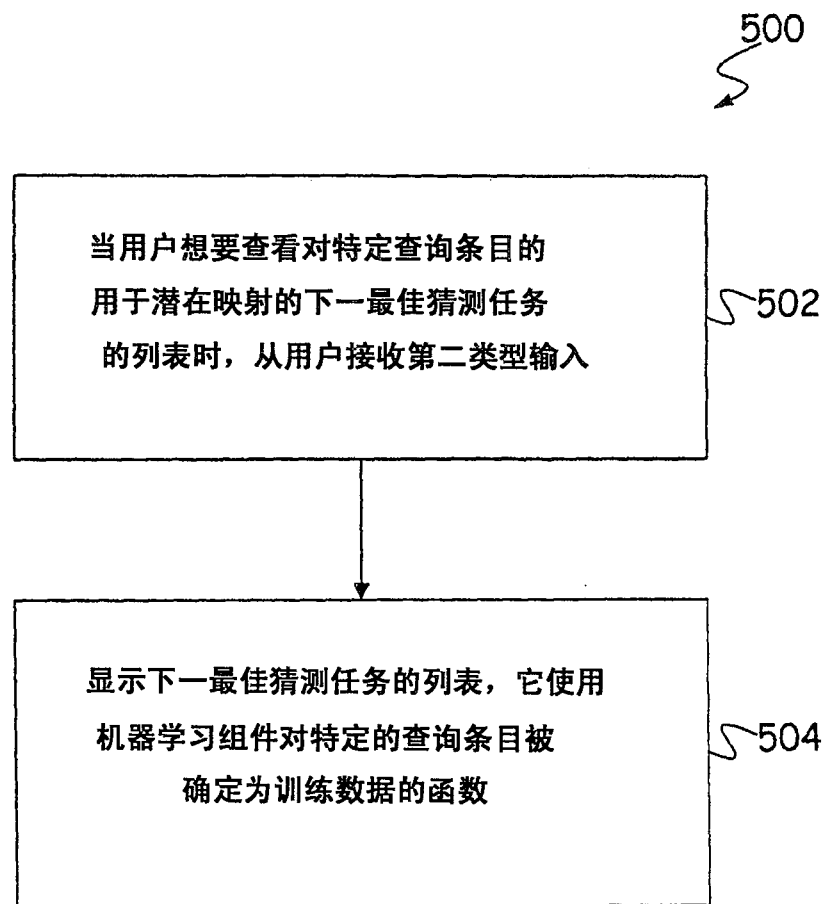


图 10

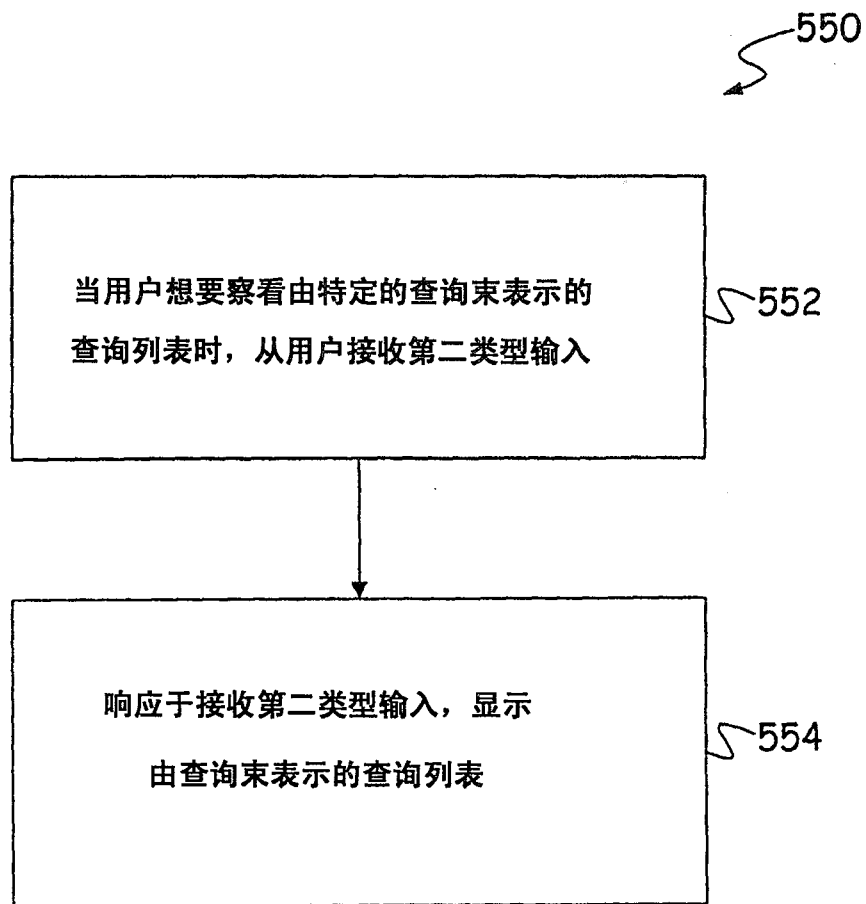


图 11

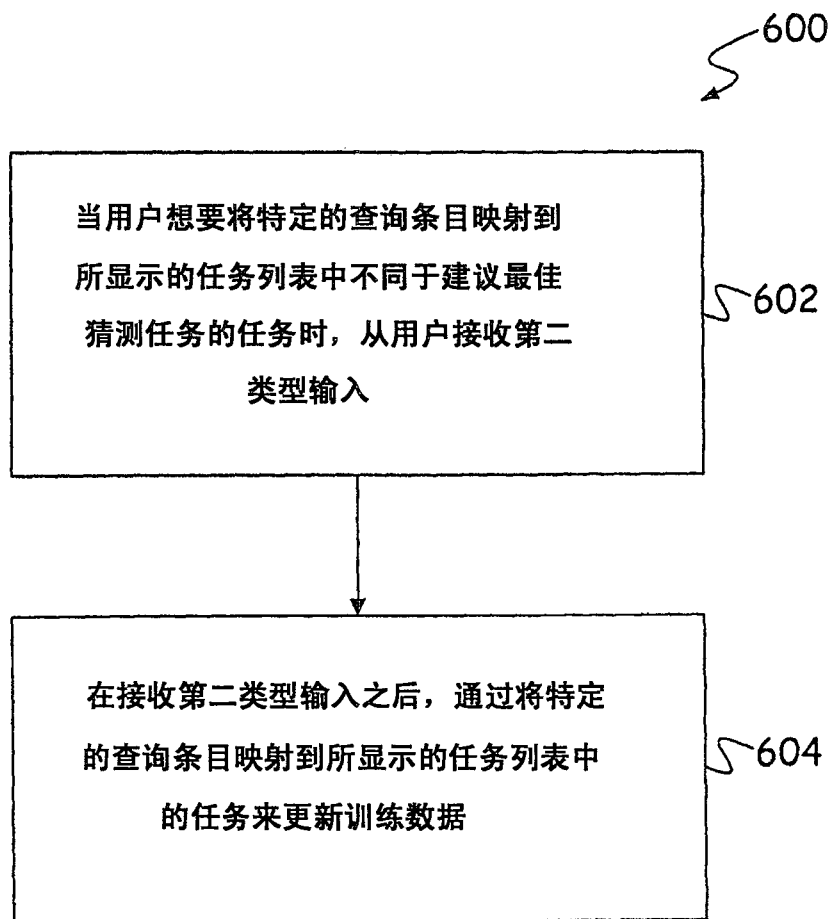


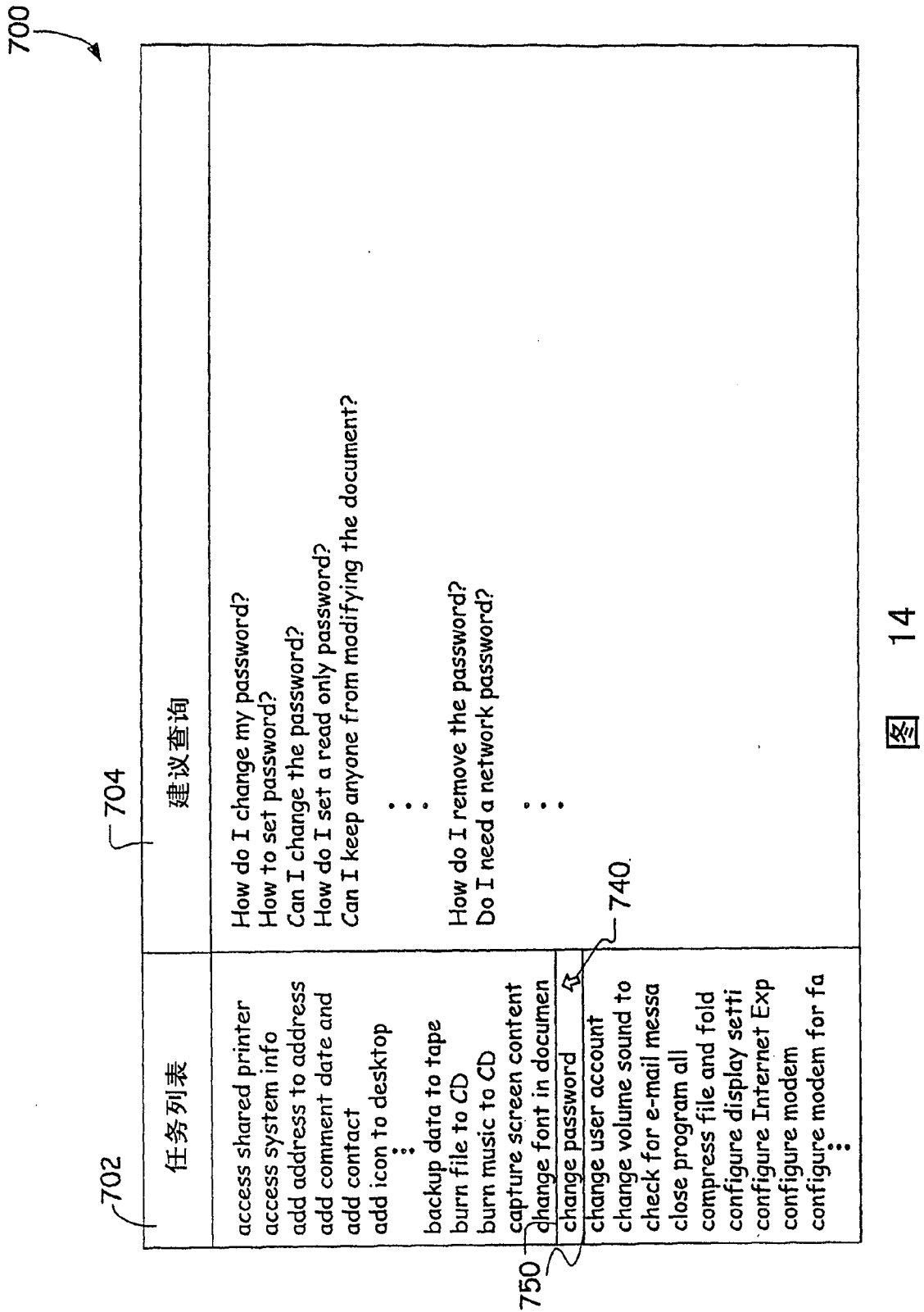
图 12

700

702

任务列表	
access shared printer access system info add address to address add comment date and add contact add icon to desktop : backup data to tape burn file to CD burn music to CD capture screen content change font in document change password change user account change volume sound to check for e-mail messa close program all compress file and fold configure display setti configure Internet Exp configure modem configure modem for fa :	

图 13



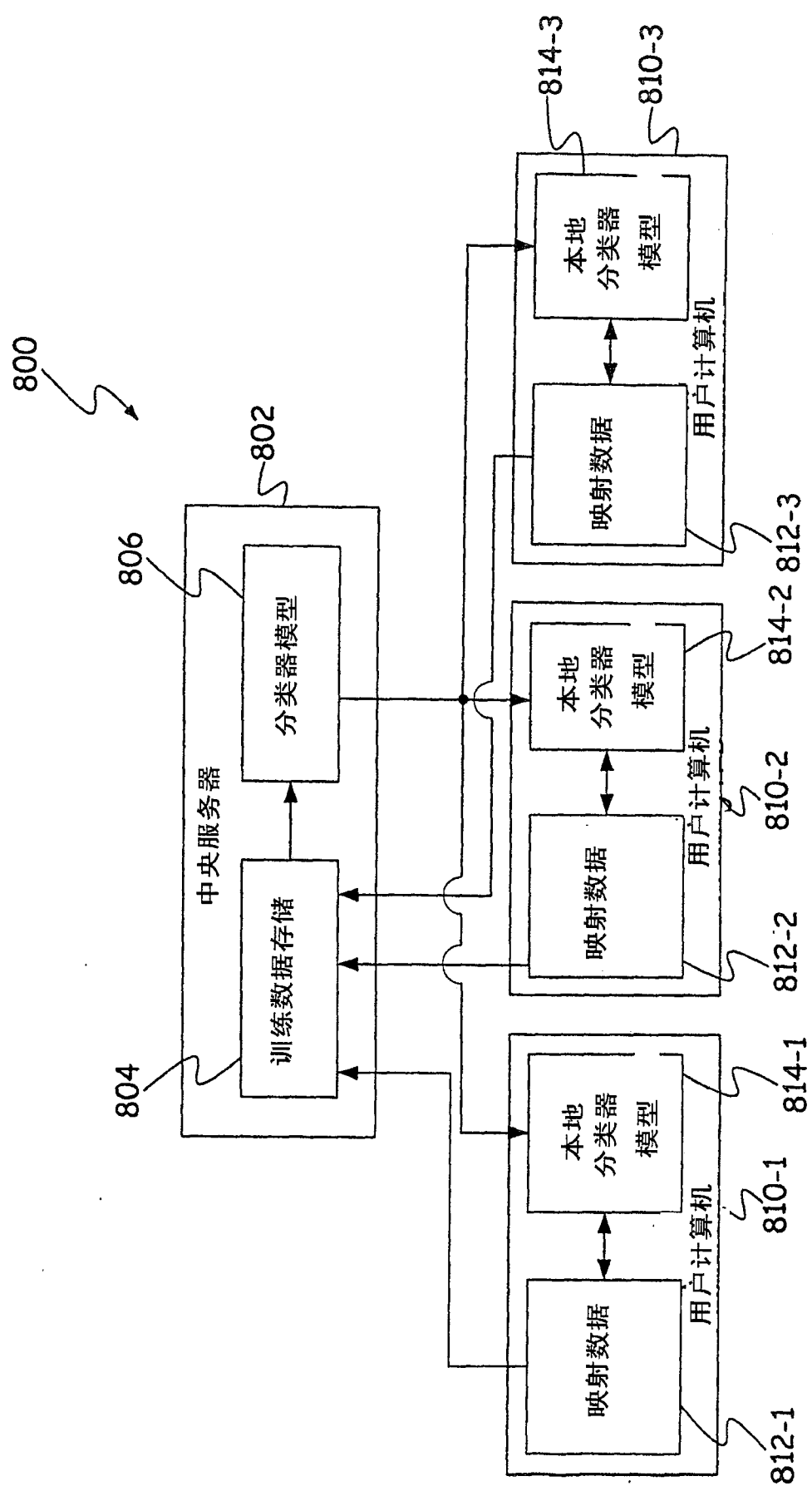


图 15

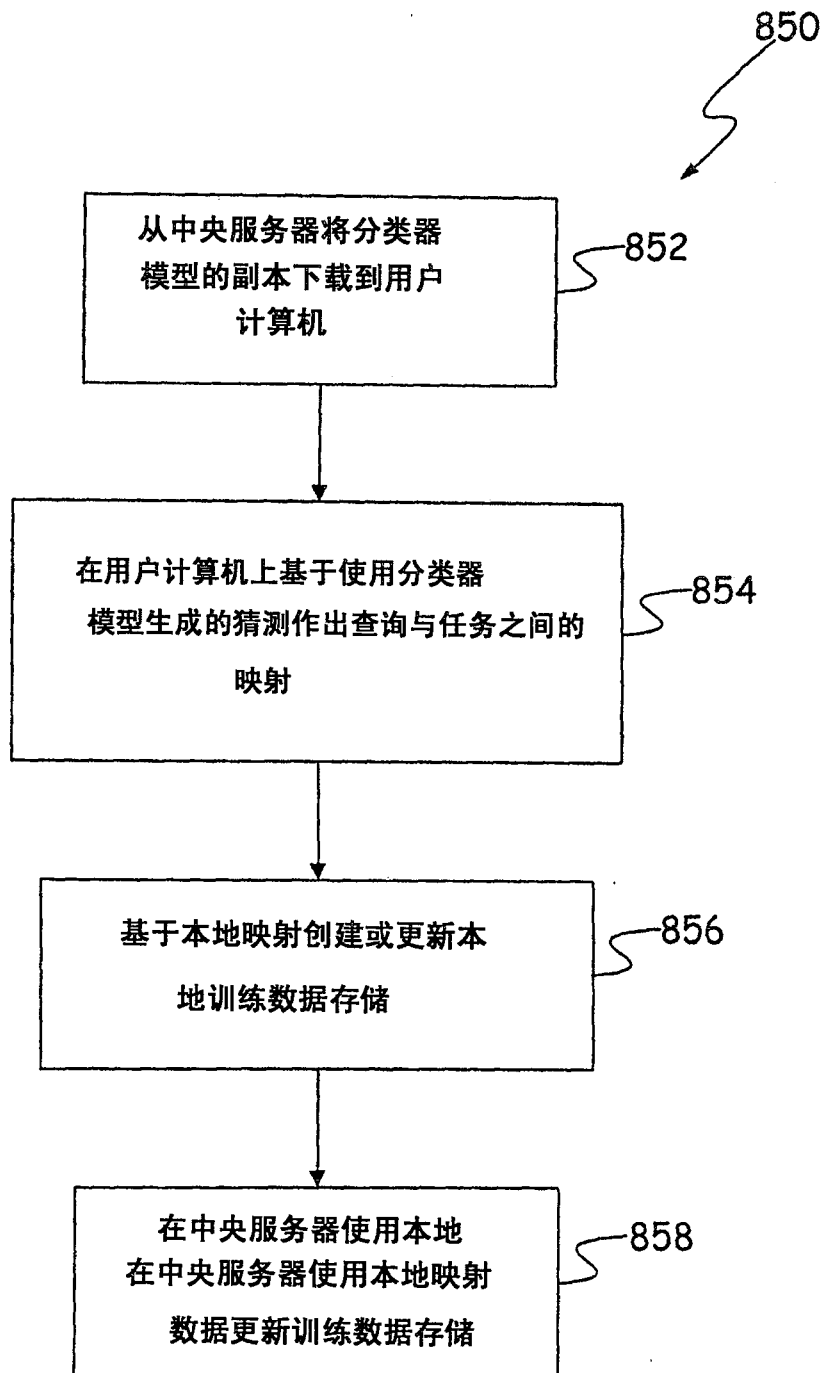


图 16