



(12) 发明专利申请

(10) 申请公布号 CN 103078962 A

(43) 申请公布日 2013. 05. 01

(21) 申请号 201210545432. 1

(22) 申请日 2012. 12. 14

(30) 优先权数据

13/325, 493 2011. 12. 14 US

(71) 申请人 微软公司

地址 美国华盛顿州

(72) 发明人 铃江丰 杰米·许恩

西蒙·伯恩斯坦 马克·恩卡纳西翁

约翰逊·阿帕西夫莱

(74) 专利代理机构 北京集佳知识产权代理有限

公司 11227

代理人 唐京桥 李春晖

(51) Int. Cl.

H04L 29/12(2006. 01)

H04L 29/08(2006. 01)

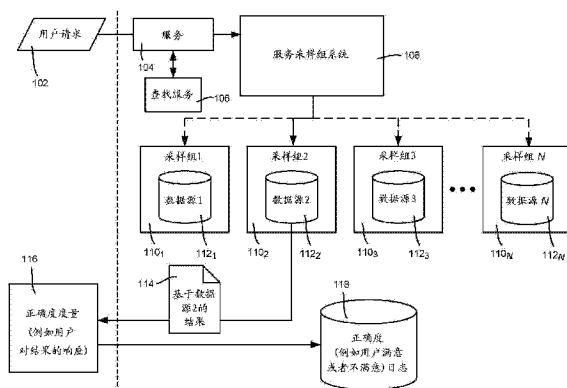
权利要求书2页 说明书13页 附图7页

(54) 发明名称

增加针对上下文信号返回的信息的准确度

(57) 摘要

本发明涉及增加针对上下文信号返回的信息的准确度。本发明针对的是这样一种技术，通过所述技术，增加了至少一个数据源针对接收的上下文数据而提供的基于上下文的信息的准确度。登录与查找的基于上下文的信息的使用相关联地接收的正确度信息。可以处理正确度信息，以通过校正数据源和 / 或通过创建混合数据源来增加总准确度，所述混合数据源包括从多个数据源之中经由正确度信息确定的最有可能准确的部分(分段)。



1. 一种计算环境中的至少部分地在至少一个处理器上执行的方法,包括:
接收(510)与至少一个数据源提供的基于上下文的信息相对应的上下文数据;
接收(514)与所述基于上下文的信息的准确度相对应的准确度信息;以及
处理(604,606)所述准确度信息,以增加与所述至少一个数据源针对其它接收的上下文信号提供的其它基于上下文的信息相对应的总准确度。
2. 根据权利要求1所述的方法,其中,接收所述上下文数据包括:
接收数据,至少一个数据源通过所述数据提供包括位置信息的基于上下文的信息,以及
 - a) 使用所述位置信息以提供包括至少一个基于位置的结果的结果集,
其中,接收所述准确度信息包括:
接收指示基于位置的结果是否可能正确的反馈;或者
 - b) 其中,接收所述准确度信息包括:
从可能更准确的数据源接收补充位置信息,并且
其中,处理所述准确度信息以增加所述总准确度包括:
如果所述位置信息不与所述补充位置信息一致,则校正所述至少一个数据源;或者
 - c) 接收所述准确度信息包括:
从可能更准确的数据源接收补充位置信息,并且
其中,处理所述准确度信息以增加所述总准确度包括:
如果所述位置信息基于所述补充位置信息已知为不正确,则动态校正所述至少一个数据源的输出;或者
a)、b)或c)的任意组合。
3. 根据权利要求1所述的方法,其中,接收所述上下文数据包括:
接收网际协议IP地址数据,所述IP地址数据用来获得包括位置信息的基于上下文的信息,并且
其中,接收所述准确度信息包括:
接收用户反馈,所述用户反馈指示所述用户对包括基于所述位置信息的至少一个结果的结果集满意还是不满意。
4. 根据权利要求1所述的方法,进一步包括:
登录从多个采样中的每个采样收集的准确度信息,
其中,所述至少一个数据源包括:
多个独立的数据源,其每个包含至少一些不准确的数据,并且
其中,处理所述准确度信息以增加所述总准确度包括:
处理所述多个采样的登录的准确度信息以从所述多个数据源之中创建混合数据源,其中包括:
将每个数据源分段成分段,以在所述数据源之中提供多个对应分段;
确定针对每个对应分段的准确度度量;以及
至少部分地基于针对每个对应分段的准确度度量,从所述对应分段之中选择用于所述混合数据源的所选分段。
5. 一种系统,包括:

采样服务(104),

所述采样服务被配置成登录用于多个采样请求中的至少一些采样请求的正确度数据(118),其中每个采样请求与上下文数据相关联,并且

对于信息被登录的每个采样请求,所述采样服务被配置成:基于与该请求相关联的上下文数据,从除了一个或多个数据源之外的至少一个实体中,获得与在所述一个或多个数据源中查找的基于上下文的信息有关的正确度度量(116),并且登录表示所述正确度度量的数据。

6. 根据权利要求 5 所述的系统,进一步包括:

被分段成多个分段的查找数据源,数据源的每个分段在每个其它数据源中具有对应分段,并且

进一步包括:

混合机构,所述混合机构被配置成处理登录的正确度数据,其包括:

确定针对每个对应分段的正确度分数;以及

基于所述对应分段的正确度分数,将所选对应分段混合成混合数据源。

7. 根据权利要求 5 所述的系统,进一步包括:

校正机构,所述校正机构被配置成:处理登录的正确度数据,以提供用于所述一个或多个数据源的更新,或者提供用于所述一个或多个数据源的动态校正数据,或者提供用于所述一个或多个数据源的更新和用于所述一个或多个数据源的动态校正数据两者。

8. 一个或多个具有计算机可执行指令的计算机可读介质,所述计算机可执行指令在被执行时执行步骤,所述步骤包括:

从每个包含至少一些不准确的数据的多个数据源(312_A-312_D)之中创建混合数据源(330),其包括:

将每个数据源分段(220,222)成分段,以在所述数据源之中提供多个对应分段;

处理(224)关于每个对应分段收集的正确度数据,以确定针对每个对应分段的正确度度量;以及

基于针对每个对应分段的正确度度量,从所述对应数据源之中选择(226)用于所述混合数据源的所选分段。

9. 根据权利要求 8 所述的一个或多个计算机可读介质,其中,所述混合数据源包括将上下文信号映射到所选分段的虚拟数据源,并且

具有进一步的计算机可执行指令,其包括:

通过访问所述虚拟数据源来查找用于上下文信号的信息,以确定哪个数据源对应于用于该上下文信号的所选分段;以及

经由该所选分段来查找所述信息。

10. 根据权利要求 8 所述的一个或多个计算机可读介质,其中,将每个数据源分段成分段包括:

使用上下文数据范围,或者

使用网际协议 IP 地址的范围。

增加针对上下文信号返回的信息的准确度

技术领域

[0001] 本发明涉及计算环境，并且具体地涉及增加针对上下文信号返回的信息的准确度。

背景技术

[0002] 用户生成上下文 (contextual) 信号，所述上下文信号经常需要在由软件系统使用之前规范化。例子包括 IP (网际协议) 地址、Wi-Fi 信号和蜂窝塔信息，一些软件系统需要将这些信息转换成 GPS 位置或者转换成城市、州、国家元组 (等)，以便由那些系统使用。另一个例子包括基于给定用户的 GPS 位置的天气应用。还有另一个例子是反向电话目录服务，其中，给定电话号码，服务返回关于该号码的所有者的信息 (例如姓名和地址)。

[0003] 在基于位置的场景和其它这样的场景中，可能存在可以提供请求信息的多个数据源。例如存在给定 IP 地址的话可以提供位置的多个数据源；针对 Wi-Fi 和蜂窝塔映射信息存在类似情形。因为组装数据的方式和收集数据的时间，有时关于输入信号和实际位置在这些源之间存在冲突映射。例如，相同的 IP 地址可能在一个数据源上映射到华盛顿特区，而在另一个数据源上映射到荷兰。

[0004] 尽管软件服务接受各种类型的用户输入，但是对这样的模糊信号进行规范化会影响在服务之下运行的应用。这不仅是因为对于每个应用而言难以实施用于减少信号模糊度的逻辑，而且还因为上下文信息需要在应用之间一致。规范化通常需要大型映射表；然而经常难以评估每个这样的映射表有多准确。例如，从 IP 地址向位置的转换需要大型查找表，以将 IP 地址范围映射到城市名、国家名等。尽管表的格式相对简单，但是表的尺寸庞大，由此确认每个 IP 范围的映射是正确的实质上不切实际。

发明内容

[0005] 提供本发明内容以便以简化的形式来介绍下文在具体实施方式中进一步描述的代表性概念的选集。本发明内容并非旨在标识要求保护的主题的关键特征或必要特征，也并非旨在以对要求保护的主题的范围进行限制的任何方式来使用。

[0006] 简言之，这里描述的主题的不同方面针对的是这样一种技术，通过所述技术，通过处理与基于上下文的信息的使用相关联地接收的正确度信息，增加了由至少一个数据源根据接收的上下文数据提供的基于上下文的信息的准确度。作为非限制性的例子，用户上下文信号如 IP 地址可以用来经由数据源查找用户的可能位置，并且另一个实体 (比如用户反馈和 / 或提供补充位置数据的可能更准确的来源) 可以提供正确度信息。

[0007] 在一个方面，一个或多个数据源提供的基于上下文的信息可以包括位置信息，该位置信息可以用来提供包括至少一个基于位置的结果的结果集。正确度信息可以包括指示基于位置的结果是否可能正确的接收反馈。

[0008] 在一个方面，可以处理正确度信息，以便如果位置信息不与补充位置信息一致，则通过校正数据源来增加总准确度。

[0009] 在一个方面,可以处理正确度信息,以便如果基于补充位置信息知道位置信息不正确,则通过动态校正数据源的输出来增加总准确度。

[0010] 在一个方面,通过将每个数据源分段成分段以在数据源之中提供多个对应分段,可以增加多个数据源的总准确度。可以处理正确度信息以确定针对每个对应分段的正确度度量,并且至少部分地基于针对每个对应分段的正确度度量,从对应分段之中选择用于混合数据源的所选分段。

[0011] 在一个实施方式中,采样服务登录用于多个采样请求中的至少一些采样请求的正确度数据,其中每个采样请求与上下文数据相关联。对于信息被登录的每个采样请求,采样服务基于与该请求相关联的上下文数据从除了一个或多个数据源之外的至少一个实体获得与在一个或多个数据源中查找的基于上下文的信息有关的正确度度量。可以处理登录的数据以增加基于查找的基于上下文的信息返回的信息的总准确度。

[0012] 其它优点可以从结合附图进行的以下具体描述中变得清楚。

附图说明

[0013] 本发明在附图中作为例子被图示而非被限制,在附图中,相同的附图标记指示相似的要素,并且在附图中:

[0014] 图 1 是表示根据一个示例实施例的示例部件的框图,所述部件被配置成登录与查找的数据的可能正确度有关的数据。

[0015] 图 2 是表示根据一个示例实施例的一个或多个示例部件的框图,所述部件被配置成处理登录的数据,以便基于登录的数据确定哪些数据源的哪些对应子集(分段)是最准确的子集(分段)。

[0016] 图 3 是根据一个示例实施例的如何将数据源的最准确对应分段合并成虚拟(或者实际)的混合数据源的表示图。

[0017] 图 4 是表示根据一个示例实施例的一个或多个示例部件的框图,所述部件被配置成处理登录的数据以便校正数据源。

[0018] 图 5 是表示根据一个示例实施例的示例步骤的流程图,所述步骤用于收集关于基于 IP 地址(或者其它)的位置查找结果而登录的用户满意度反馈。

[0019] 图 6 是表示根据一个示例实施例的示例步骤的流程图,所述步骤用于处理登录的用户反馈,以便基于反馈来确定哪些数据源分段最准确,以创建混合数据源。

[0020] 图 7 是表示这里描述的各种实施例可以实施于其中的示例非限制性联网环境的框图。

[0021] 图 8 是表示这里描述的各种实施例的一个或多个方面可以实施于其中的示例非限制性计算系统或操作环境的框图。

具体实施方式

[0022] 这里描述的本技术的不同方面一般地针对这样一种技术,通过所述技术,使用数据源来规范化大量用户输入请求的上下文信号(上下文数据)。为此目的,用户请求和关于来自数据源的返回答复的正确度度量被用来提高针对具有关联上下文信号的未来请求的返回数据的准确度。

[0023] 例如,使用多个反向 IP 地址数据库之一,基于反向 IP 地址查找(从 IP 地址到包括城市、州和国家的位置的映射),指向搜索引擎的用户查询可以导致返回基于位置的结果。用户对基于位置的结果的满意度度量可以用来确定反向 IP 地址查找有多准确,例如用户是点击返回的基于位置的结果(指示可能满意),还是提交另一个查询(指示可能不满意)。通过将每个数据源划分成称为分段的子集,并且从每个数据源选择具有最高测量正确度(例如满意度)水平的分段,可以创建“混合”数据源以返回具有更高准确度水平的未来结果。进一步,正确度信息可以用来校正数据源,以便给定未来查找请求的话得到更准确的输出。

[0024] 应当理解的是,这里的例子中的任何一个都是非限制性的。例如,尽管一些例子和描述一般地针对反向 IP 地址查找,但是其中在可用结果中存在不一致或错误可能性的任何一个或多个服务和 / 或一个或多个数据源都可以受益于这里描述的技术。同样地,本发明不限于这里描述的任何特定的实施例、方面、概念、结构、功能或例子。实际上,这里描述的任何实施例、方面、概念、结构、功能或例子都是非限制性的,并且可以用各种方式使用本发明,这些方式一般在计算和提供更准确结果方面提供益处和优点。

[0025] 图 1 是示出了一个示例实施方式中的各种部件的框图。一般而言,用户请求 102 由服务 104 处理,该服务 104 比如在用户查询被提交给搜索引擎以获得指向相关文档的链接时提供结果。如理解的那样,服务 104 有时获得或访问附加信息,以在提供结果时对服务 104 进行帮助。例如,服务 104 可以尝试比如通过向查找服务 106 提供上下文信号(例如 IP 地址数据、与 Wi-Fi 有关的数据如包括 Wi-Fi 站标识符的信号和 / 或与蜂窝塔有关的数据)来确定用户的位置,该查找服务 106 取得信号并且使用它们以查找和返回用户位置;(注意查找服务 106 可以是也可以不是与服务 104 分离的实体)。作为更具体的例子,给定用户的 IP 地址,搜索引擎可以从反向 IP 地址查找服务中获得用户的位置,由此搜索引擎能够基于返回的用户的位置返回更相关的结果。

[0026] 在一个实施方式中,查找服务 106 可以访问多个数据源中的至少一个数据源以查找基于上下文的信息。然而这些数据源可能包含错误,因此请求查找的应用或者服务有时接收到不准确的信息。数据源关于它们维护的信息经常不一致;例如,给定 IP 地址,反向 IP 地址查找服务用来获得位置的四个数据源中的三个数据源可能包含正确位置信息而第四个未包含,因此结果的准确度依赖于服务针对给定查询而使用的数据源。简单地不使用该第四数据源是欠佳的解决方案,因为对于另一个 IP 地址而言,该第四数据源可能具有相对于其它数据源而言最准确的信息。所有这样的错误在数据源之中存在于何处既不是预先已知的,也实际上不可确定。

[0027] 如图 1 中表示的那样,可以随机(或者以别的方式)采样一些接收的请求,以获得与用来向请求用户提供返回一个或多个结果的数据源的正确度有关的信息。因此,在图 1 的例子中,代替作为常规请求来处理,用户请求 102 被例示为由服务采样组系统 108 选择。

[0028] 在这个例子中,将用户请求 102 分组到 N 个采样组之一中,其中 N 表示可以从中获得信息的数据源的数目。在图 1 的例子中,用户请求被分配到 N 个采样组(110_1 - 110_N)之一,其经由实线被示出为采样组 110_2 ,对应于可用数据源 112_1 - 112_N 中的数据源 112_2 。例如,为了反向 IP 地址查找以建立用户的位置,在一个实施方式中使用四个数据源(N = 4)。

[0029] 可以用某种方式经由返回结果 114 向用户返回从所选数据源 112_2 获得的查找结

果。例如,给定搜索查询,搜索引擎可以从四个数据库之一(在这个例子中对应于所选数据源 112₂)中给定用户的 IP 地址而查找用户的位置,并且裁剪指向位置的搜索结果。作为更具体的例子,用户可以查询“披萨”,并且一些返回结果可以用于指向用户位置附近的披萨餐厅的文档链接,其中从经由数据源 112₂ 执行的反向 IP 地址查找中获得位置。

[0030] 如这里描述的那样,结果 114 的正确度的某种度量被确定(在图 1 中经由正确度度量 116 表示),并且被用来增加响应于相似用户请求而提供的查找和 / 或未来结果的准确度。作为一个例子,如果响应于查询来提供结果,则可以测量用户反馈,比如用户对那些结果的满意度。反馈可以是隐式的,由此根据用户响应或其它行为动作(包括未采取动作)来进行推断,也可以是显式的,例如经由一个或多个按钮的集合,用户通过所述按钮指示例如满意或者不满意的正确度的某种度量。

[0031] 作为推断的正确度度量的更具体的例子,如果一些结果基于位置并且用户点击了基于位置的结果之一(例如披萨餐厅链接),则这样的反馈指示有可能用户对结果满意,因此位置更有可能是正确的。如果用户代之以提交另一个查询而没有点击链接,则用户更有可能不满意,从而指示位置更有可能是不正确的。这种反馈 / 正确度信息可以与从数以千计乃至百万计的用户获得的相似“人群来源”信息一起登录于日志 118 中,并且随后被处理以确定每个数据源关于它的基于 IP 地址的位置信息有多准确。如理解的那样,本技术受益于具有充分数量的用户交互输入,以便能够观测用户的响应如满意或不满意,并且得出关于信息准确度的结论。

[0032] 注意,如果用户点击并非基于位置的链接,例如点击指向示出披萨历史的参考网站的链接,则可以不登录信息,因为用户可能不是寻找与位置有关的信息;(然而可行的是,这样的信息可以用于其它目的,因此可以被登录,但是例如具有空值或者零值位置正确度分数)。进一步要注意的是,如果用户提供更多信息如更细化的查询(例如“贝尔维尤披萨”),则该附加信息也可以用来提高准确度。更进一步,响应失败可以视为不满意的指示(例如用户去往不同的搜索引擎),然而可以对此给予比更确定的不满意的指示更少的计分加权,因为用户可能从摘录文字中获得了所需信息(比如电话号码),因此无需进一步交互。在任何情况下,给定充分数目的用户,可以识别关于基于用户满意度的数据(或其任何部分)的准确度的统计趋势。

[0033] 进一步,可以使用用于确定满意 / 不满意计分系统的训练和其它机制。例如,某一数量的经验证的已知正确(根本真实)的信息可以用来建立如下机制:对比于现有查找服务的有可能正确也有可能不正确的未验证信息(包括在实际使用场景中),用户倾向于如何对已知正确的信息做出反应。可以返回已知不良的信息(例如在训练或者其它受控场景中,以免有意地提供不良信息)以建立用户如何对不正确的信息做出响应。

[0034] 图 2 和图 3 图示了其中可以使用曾经登录的正确度度量信息的一种普通方式。在这个例子中,日志 118 被处理以创建与每个数据源 312_A-312_D 的具有最高正确度分数的部分(分段)对应的“混合”数据源 330 (图 3),其中正确度基于测量的用户满意度。一般而言,在从相对大量的用户收集正确度度量信息之后,分析用户响应的日志 118 以评估向采样的用户分配的数据源的质量。

[0035] 在这个例子中,基于被划分成称为分段的子集的每个数据源来处理日志 118;例如如图 3 中所示,将数据源 312_A 划分成段 S1A-S4A,将数据源 312_B 划分成段 S1B-S4B,

等等。因此,一个数据源的每个分段在每个其它数据源中具有对应分段。例如用于 IP 地址位置查找数据源的每个分段可以对应于 IP 地址范围。注意在图 3 中为了简化而示出每个数据源的四个分段,然而系统可以如对于给定应用而言适合的那样建立数以十计、百计、千计等量级的每个数据源的任何数目的分段。

[0036] 当针对数据源的分段计算分数时,分段分数计算机构 220 (图 2) 经由数据源分段器 222 确定正确度度量(例如用户满意或者不满意响应)属于哪个数据源和分段。注意在替换的实施方式中,比如如果固定的 IP 地址范围被预先确定,则可以预先确定用于每个响应的数据源分段,因此可以在登录时记录数据源、分段信息,而不是在处理日志时根据 IP 地址来确定。然而在处理日志时确定分段可能有利的是,分段无需是固定范围,并且可以代替地基于其它概念如均衡每个分段的响应数目等,这可能无法知道,直到登录完成为止。

[0037] 因此,分段化可以基于固定尺寸或者除了固定尺寸之外的概念。例如分段化可以基于响应数目、业务均衡等。包括在相同的数据源内,任何分段可以如希望的那样分解成子分段,或者两个或更多分段可以组合成更大分段。在这种实施方式中,无论每个数据源内的分段的相对尺寸如何(例如在图 3 中,分段 S2A 可以尺寸不同于分段 S3A),数据源每个被一致地分割,使得在数据源之中的任何对应分段(例如 S3A、S3B、S3C 和 S3D) 相互尺寸相同。

[0038] 在处理登录的条目时,分数计算机构 220 基于随响应数据登录的正确度度量调整针对该数据源和分段组合的分数。通过使系统运行足够次数,并且基于用户的响应对分数之差进行比较,系统可以判断针对每个分段的每个不同数据源的质量。一个示例计分机构是递增针对满意的响应的分段分数,并且递减针对不满意的响应的分数,并且在进行时,如果希望则以某种方式归一化针对每个分段的分数,比如归一化为百分比。以这种方式,处理日志 118 导致多个分段分数 224,一个分数针对每个数据库的每个分段,如图 3 中经由针对分段 S1A-S4D 示出的分数表示的那样。

[0039] 可以用任何方式使用分数,包括用来如这里描述的那样提高总系统准确度。一种用于提高总系统准确度的直接方式基于混合机制(算法),该混合机制(算法)选择每个数据源之中的具有最高分数的对应分段,然后使用所选分段以构建混合数据源。

[0040] 在图 2 和图 3 的例子中,基于来自数据源 312_A-312_D 中的每个数据源的反馈分数,分段混合机构 226 使用分段分数 224 以选择“最准确的”分段,所述分段在这个例子中是分段 S1C、S2A、S3D 和 S4A,每个分段相对于它在其它数据源之中的对应分段具有最高分数。结果是与图 3 的混合数据源 330 对应的混合分段数据 228。注意在平局——该平局可以是实际平局或统计平局——的情况下,冲突解决机制可以用来在分段之间进行选择(例如使用历史数据并且 / 或者使用具有总最高分数的数据源等)。

[0041] 虽然出于说明的目的,图 3 将混合数据源 330 表示为可以用于更准确的在线查找服务的单个实体,但可以理解的是,系统无需构建将所选分段组合的实际单个数据源。相反地,在在线使用期间,服务可以访问与所选分段对应的映射,以便基于混合分段数据 228 动态地选择数据源之中的哪个分段来使用。例如,给定引入的 IP 地址,该 IP 地址(经由它的 IP 地址范围)对应于分段,系统访问混合分段数据 228,以便将该 IP 地址映射到用于该范围的从多个数据源之中选择的最准确分段,并且由此从该特定映射分段获得位置结果。因此,混合数据源 330 可以是经由混合分段数据 228 等选择的虚拟数据源,而不是实际组合的物理数据源。

[0042] 注意在这个示例场景中,一旦获得了混合分段数据,一般而言,随后向用户给予虚拟的(或者可能是实际的)混合数据源的益处。然而为了再次增加准确度,用于确定正确度的采样(例如采样某一小百分比的随机选择的用户)可以继续或者在任何时间恢复。例如采样可以连续,也可以被偶尔接通或关断,等等。采样可以基于某种改变,所述改变指示需要新的正确度数据,比如无论何时显著改变数据源,例如在更新之后一周一次。采样百分比可以增加或减少,并且 / 或者可以基于一个或多个标准而随时间变化。

[0043] 尽管上面的例子一般针对的是对多个数据源进行分段以便基于用户满意或者不满意来找到每个数据源的哪个分段提供了最准确的信息,但是可以执行用于测量正确度并且提高准确度的其它方式。例如,正确度可以基于已知更准确的其它已知信息。进一步,尽管多个数据源可以受益于如上文描述的登录信息,但是甚至只有单个的数据源也可以被校正。

[0044] 作为例子,考虑数据源维护位置与 Wi-Fi 信号之间的映射,以便例如映射哪些 Wi-Fi 站 / 接入点(或者在这个例子描述中简称为接入点)在什么位置。因此可以基于使用的接入点来确定实体的位置。然而可能随时间移动接入点,由此除非更新数据源,否则任何先前映射的位置都不正确。

[0045] 如果接入点位置用来提供在提供搜索结果时使用的位置,则如上文一般地关于相似的基于 IP 地址的位置查找而描述的那样,用户对那些结果的反应是一种用于确定存储的接入点位置的正确度的可能方式。然而其它正确度信息可以是可用的。例如考虑将智能电话用于经由接入点进行无线联网的用户,由此可以查找接入点位置。取代(或者除了)任何用户满意度度量, GPS 数据和 / 或蜂窝确定的位置数据(例如基于信号强度 / 三角测量)如果可用,则可以用来获得补充位置,以便验证对比于补充位置数据,数据源的存储位置是否正确。由于数据源未必准确,所以上下文(例如位置)可能冲突;通过引用而结合于此的第 7,444,594 号美国专利一般地描述了调解计算机用户的上下文数据中的冲突。

[0046] 另一种用于获得用户反馈的方式是用户是否改变他或她的设备(例如智能电话)上的设置或其它输入以获得更准确的位置。例如,如果使用 Wi-Fi 接入点数据以获得位置,并且用户查询“披萨”,并且在获得搜索结果之后用户将查询改变成“披萨 98040”(其中“98040”代表邮政编码),则这可能是可以用来指示用户对位置不满意的信号。

[0047] 如图 4 中表示的那样,比如通过与其它这样的信息一起登录校正信息 440,并且在适当时间更新其它数据源,从一个源获得的这种校正信息 440 可以用来校正别的数据源。在图 4 中经由提供数据源更新 444 的数据源校正机构 442 来表示这一点。

[0048] 然而任何数据源校正可以更加动态。例如考虑指向纬度、经度数据源的 GPS 随着针对某一位置在给定方向上离开某一距离而被另一个信标(例如蜂窝信标)检测。动态校正数据 446 可以应用于指向纬度、经度数据源的该 GPS 的任何输出,例如作为偏移以校正输出,并且针对未来的查找请求提供更准确的位置。

[0049] 图 5 是概括各种示例步骤的流程图,所述步骤涉及收集关于例如响应于搜索查询而向用户返回的基于位置的结果的用户反馈,所述流程图始于其中接收到用户请求的步骤 502。步骤 504 确定是否采样这个请求。注意通常不采样大多数的用户请求,并且代替地在步骤 506 经由先前创建的混合数据源(或者以传统的方式,直到如这里描述的那样初始地创建混合数据源为止)处理他们的请求。

[0050] 如这里描述的那样,选择至少一些用户请求用于采样。该选择可以基于任何方案,比如每百个用户请求之中的一一个,由此一般随机地接收采样的 IP 地址。别的示例方案可以基于 IP 地址分布,以例如试图针对每个 IP 地址范围获得某种相等数目的采样。其它选择机制可以与这些方案和其它方案结合使用,比如在考虑采样之前执行过滤,以比如仅采样被认为映射到美国位置的 IP 地址。

[0051] 对于被选择用于采样的用户,步骤 508 确定用于该用户的与数据源之一对应的采样组。可以通过循环分布、通过随机分布或者通过任何其它方案来进行这种选择。对于别的方案的例子,可以针对给定的采样组 / 数据源选择更多用户,以便平衡基于位置的满意或不满意类型响应的数目,而不是平衡采样组 / 数据源之中的全部响应的数目(例如因为并非所有登录的响应都可以对应于基于位置的链接)。

[0052] 步骤 510 从数据源获得位置信息,该位置信息然后用来在步骤 512 返回一个或多个结果。对于搜索,给定查找位置信息,指向文档、广告等的返回链接中的一些链接可以是基于位置的。

[0053] 步骤 514 表示基于返回结果从用户获得反馈。在步骤 516 登录数据,例如用户的 IP 地址(或者如果范围固定则可能为分段)、用户被分配到的数据源和用户的反馈。反馈可以是分数等,例如如果用户响应指示用户满意则为一,而如果不满意则为负一,在这种情况下,步骤 514 也将确定分数等表示为获得反馈的一部分。

[0054] 基于时间窗等,对于如期望的那样多的用户请求,比如固定数目的用户请求,步骤 518 重复该过程,直到日志充满为止。注意可能存在许多类似的过程,所述过程并行地操作,向相同的日志写入,或者向以后组合的不同日志写入。步骤 520 表示关闭日志以便进一步处理,以例如执行分析,以便如参照上面的图 2 和图 3 以及下面的图 6 描述的那样对分段进行混合。

[0055] 图 6 表示用于执行示例数据源混合的流程图的示例步骤,该流程图始于其中选择每个登录的条目以便进行处理的步骤 602。步骤 604 根据登录的 IP 地址确定分段,除非如上所述用于这个条目的分段预先已知并且已经登录。

[0056] 步骤 606 表示基于反馈来调整针对这个条目的数据源、分段的分数,比如如果满意则递增累计,或者如果不满意则递减累计。另一种方案是保持针对响应计数和满意响应计数的累计。如果计分方案没有被实施为登录的一部分,则步骤 606 也表示计分方案的实施,例如将登录的用户动作(点击或者新查询或者可能的其它动作或者无动作)转换成用于调整分段分数的值。

[0057] 步骤 608 重复日志处理,直到日志被处理并且针对每个数据源的每个分段获得分数为止。步骤 610 表示如果有必要则归一化分数,以比如考虑对应分段之中的不等数目的相关响应。

[0058] 步骤 612 将分段例如虚拟地混合成混合分段数据。在这一点上,初始或者更新的混合源可用于在处置请求时使用。

[0059] 如可以看到的那样,这样以来就描述了一种可伸缩的方式以提高涉及到关联查找的返回信息的准确度。这可以包括使用推断或者明确的人群来源数据以在具有未知准确度的潜在冲突的数据源之间进行调解。

[0060] 示例联网和分布式环境

[0061] 本领域普通技术人员可以意识到的是,这里描述的各种实施例和方法可以与任何计算机或者其它客户端或服务器设备结合实施,所述计算机或者其它客户端或服务器设备可以部署为计算机网络的一部分,或者部署于分布式计算环境中,并且可以连接到任何种类的一个或多个数据存储库。就这一点而言,这里描述的各种实施例可以实施于任何如下计算机系统或环境中,所述计算机系统或环境具有任何数目的存储器或存储单元以及跨越任何数目的存储单元出现的任何数目的应用和过程。这包括但不限于具有服务器计算机和客户端计算机的环境,所述服务器计算机和客户端计算机部署在具有远程或本地存储器的网络环境或分布式计算环境中。

[0062] 通过计算设备和系统当中的通信交换,分布式计算提供了计算机资源和服务的共享。这些资源和服务包括信息的交换、用于对象如文件的高速缓冲存储器和盘存储器。这些资源和服务还包括跨越多个处理单元共享处理能力以用于负载平衡、资源扩展、处理专门化等。分布式计算利用了网络连通性,允许客户端利用它们的联合能力以有益于整个企业。就这一点而言,多种设备可以具有应用、对象或资源,所述应用、对象或资源可以参与如针对本公开的各种实施例描述的资源管理机制。

[0063] 图 7 提供了示例联网或分布式计算环境的示意图。分布式计算环境包括计算对象 710、712 等和计算对象或设备 720、722、724、726、728 等,这些对象可以包括如示例应用 730、732、734、736、738 表示的程序、方法、数据存储库、可编程逻辑等。可以意识到的是,计算对象 710、712 等和计算对象或设备 720、722、724、726、728 等可以包括诸如个人数字助理 (PDA)、音频 / 视频设备、移动电话、MP3 播放器、个人计算机、膝上型计算机等不同设备。

[0064] 每个计算对象 710、712 等和计算对象或设备 720、722、724、726、728 等可以经由通信网络 740 直接或间接与一个或多个其它计算对象 710、712 等和计算对象或设备 720、722、724、726、728 等通信。虽然在图 7 中图示为单个单元,但是通信网络 740 可以包括向图 7 的系统提供服务的其它计算对象和计算设备,并且 / 或者可以代表未示出的多个互连网络。每个计算对象 710、712 等或者计算对象或设备 720、722、724、726、728 等也可以包含诸如应用 730、732、734、736、738 之类的应用,所述应用可以利用 API 或者其它对象、软件、固件和 / 或硬件,所述 API 或者其它对象、软件、固件和 / 或硬件适合于与根据本公开的各种实施例提供的应用进行通信或者实施该应用。

[0065] 存在支持分布式计算环境的多种系统、部件和网络配置。例如计算系统可以由有线或无线系统、由本地网络或广泛分布式网络连接在一起。目前许多网络耦合到因特网,所述因特网提供了用于广泛分布式计算的基础结构并且涵盖了许多不同网络,但是任何网络基础结构都可以用于伴随如各种实施例中描述的系统进行的示例通信。

[0066] 因此,可以利用许许多多的网络拓扑和网络基础结构如客户端 / 服务器、对等或者混合体系结构。“客户端”是类或者组的成员,所述类或者组使用与其无关的另一个类或者组的服务。客户端可以是过程,例如大致为指令或者任务的集合,其请求由另一个程序或者过程提供的服务。客户端过程利用请求的服务,而无需“知道”关于另一个程序或者服务本身的工作细节。

[0067] 在客户端 / 服务器体系结构特别是联网系统中,客户端通常是这样一种计算机,所述计算机访问由另一个计算机例如服务器提供的共享网络资源。在图 7 的图示中,作为非限制性的例子,计算对象或设备 720、722、724、726、728 等可以视为客户端,并且计算对

象 710、712 等可以视为服务器, 其中充当服务器的计算对象 710、712 等提供数据服务, 比如从客户端计算对象或设备 720、722、724、726、728 等接收数据、存储数据、处理数据、向客户端计算对象或设备 720、722、724、726、728 等发送数据, 但是任何计算机可以取决于环境而视为客户端、服务器或者二者都是。

[0068] 服务器通常是通过远程或本地网络如因特网或无线网络基础结构可访问的远程计算机系统。客户端过程可以在第一计算机系统中活动, 并且服务器过程可以在第二计算机系统中活动, 通过通信介质相互通信, 由此提供分布式功能, 并且允许多个客户端利用服务器的信息收集能力。

[0069] 例如在其中通信网络 740 或总线是因特网的网络环境中, 计算对象 710、712 等可以是网络服务器, 经由若干已知协议中的任何协议如超文本传送协议 (HTTP), 其它计算对象或设备 720、722、724、726、728 与所述网络服务器通信。充当服务器的计算对象 710、712 等也可以起到客户端的作用, 例如起到计算对象或设备 720、722、724、726、728 等的作用, 如可以作为分布式计算环境的特性那样。

[0070] 示例计算设备

[0071] 如提到的那样, 这里描述的技术可以有利地应用于任何设备。因此可以理解的是, 设想所有种类的手持、便携以及其它计算设备和计算对象用于与各种实施例结合使用。因而下文在图 8 中描述的以下通用远程计算机仅为计算设备的一个例子。

[0072] 实施例可以部分地经由操作系统来实施, 以便由服务的开发者用于设备或者对象和 / 或包括在应用软件内, 所述应用软件进行操作以执行这里描述的各种实施例的一个或多个功能方面。可以在计算机可执行指令的普通上下文中描述软件, 所述计算机可执行指令比如是程序模块, 其由一个或多个计算机如客户端工作站、服务器或其它设备执行。本领域技术人员将会意识到的是, 计算机系统具有可以用来传达数据的多种配置和协议, 因此特定配置或协议未被视为限制。

[0073] 图 8 因此图示了这里描述的实施例的一个或多个方面可以实施于其中的适当计算系统环境 800 的例子, 但是如上文阐明的那样, 计算系统环境 800 仅为适当计算环境的一个例子, 而非旨在暗示关于使用或功能范围的任何限制。此外, 计算系统环境 800 并非旨在被解释为具有任何依赖性, 所述依赖性与示例计算系统环境 800 中图示的部件中的任何部件或组合有关。

[0074] 参照图 8, 用于实施一个或多个实施例的示例远程设备包括形式为计算机 810 的通用计算设备。计算机 810 的部件可以包括但不限于处理单元 820、系统存储器 830 和系统总线 822, 该系统总线 822 将包括系统存储器的各种系统部件耦合到处理单元 820。

[0075] 计算机 810 通常包括多种计算机可读介质, 所述计算机可读介质可以是计算机 810 可以访问的任何可用介质。系统存储器 830 可以包括形式为易失性和 / 或非易失性存储器的计算机存储介质如只读存储器 (ROM) 和 / 或随机存取存储器 (RAM)。作为例子而非限制, 系统存储器 830 也可以包括操作系统、应用程序、其它程序模块和程序数据。

[0076] 用户可以通过输入设备 840 向计算机 810 中输入命令和信息。监视器或其它类型的显示设备也经由接口如输出接口 850 连接到系统总线 822。除了监视器之外, 计算机也可以包括其它外围输出设备如扬声器和打印机, 它们可以通过输出接口 850 连接。

[0077] 使用指向一个或多个其它远程计算机如远程计算机 870 的逻辑连接, 计算机 810

可以在联网或分布式环境中操作。远程计算机 870 可以是个人计算机、服务器、路由器、网络 PC、对等设备或其它公共网络节点，或者是任何其它远程媒体消耗或传输设备，并且可以包括上文相对于计算机 810 描述的元件中的任何元件或所有元件。图 8 中描绘的逻辑连接包括网络 872 如局域网 (LAN) 或广域网 (WAN)，但是也可以包括其它网络 / 总线。这样的联网环境在住宅、办公室、企业范围的计算机网络、内部网和因特网中是常见的。

[0078] 如上文提到的那样，尽管已经结合各种计算设备和网络体系结构描述了示例实施例，但是基本概念可以应用于其中希望提高资源使用效率的任何网络系统和任何计算设备或系统。

[0079] 而且，存在多种方式来实施相同或相似的功能，例如适当的 API、工具箱、驱动代码、操作系统、控件、独立或可下载的软件对象等，其使得应用和服务能够利用这里提供的技术。因此从 API (或其它软件对象) 的观点以及从实施如这里描述的一个或多个实施例的软件或硬件对象的观点来设想这里的实施例。因此这里描述的各种实施例可以具有完全为硬件的方面、部分为硬件而部分为软件的方面以及为软件的方面。

[0080] 措辞“例子”这里用来指的是充当例子、实例或者示意。为了避免疑问，这里公开的主题不受这样的例子限制。此外，这里描述为“例子”的任何方面或设计不一定被解释为相对于其它方面或设计是优选或有利的，它也没有意味着排除本领域普通技术人员已知的等同示例结构和技术。进而，在使用术语“包括”、“具有”、“包含”和其它相似措辞的程度上，为了避免疑问，这样的术语旨在以与术语“包括”相似的方式具有包含意义，所述术语“包括”在应用于权利要求中时作为开放式过渡词而未排除任何附加的或其它单元。

[0081] 如提到的那样，这里描述的各种技术可以与硬件或软件或者在适合时与二者的组合结合实施。如这里使用的那样，术语“部件”、“模块”、“系统”等同样旨在指代与计算机有关的实体，所述实体为硬件、硬件与软件的组合、软件或者执行中的软件。例如部件可以是但不限于在处理器上运行的进程、处理器、对象、可执行文件、执行的线程、程序和 / 或计算机。作为示意，在计算机上运行的应用程序和计算机二者都可以是部件。一个或多个部件可以驻留于执行的进程和 / 或线程内，并且部件可以局限于一个计算机上和 / 或分布于两个或更多计算机之间。

[0082] 已经关于几个部件之间的交互描述了前述系统。可以意识到的是，这样的系统和部件可以包括那些部件或指定的子部件、指定的部件或子部件中的一些和 / 或附加部件以及根据前述各项的各种排列和组合。子部件也可以实施为通信地耦合到其它部件的部件，而不是包括在双亲部件(层级)内。此外可以注意到的是，一个或多个部件可以组合成提供聚合功能的单个部件，或者划分成几个分离的子部件，并且可以提供任何一个或多个中间层如管理层，以通信地耦合到这样的子部件，以便提供集成功能。这里描述的任何部件也可以与这里没有具体描述、但是本领域技术人员一般已知的一个或多个其它部件交互。

[0083] 鉴于这里描述的示例系统，也可以参照各种附图的流程图而意识到可以根据描述的主题实施的方法。尽管为了简化说明，方法被示出和描述为一系列的块，但是可以理解和意识到的是，各种实施例不受块的顺序限制，因为一些块可以按照与这里描绘和描述的顺序不同的顺序出现和 / 或与其它块并行出现。在经由流程图图示非序列的或者有分支的流程时可以意识到的是，可以实施各种实现相同或者相似结果的其它分支、流程路径和块的顺序。另外，一些图示的块在实施下文描述的方法时是可选的。

[0084] 结论

[0085] 尽管本发明易于具有各种修改和替代构造,但是在附图中示出并且上文已经具体描述了它的某些示意实施例。然而应当理解的是,这并非旨在使本发明限于公开的具体形式,而是恰好相反,旨在覆盖落入本发明的精神和范围内的所有修改、替代构造和等形式。

[0086] 除了这里描述的各种实施例之外,可以理解的是,可以使用其它相似实施例,或者可以对描述的(一个或多个)实施例进行修改或添加,以便在不脱离本发明的情况下执行对应(一个或多个)实施例的相同或等同的功能。更进一步,多个处理芯片或多个设备可以分担执行这里描述的一个或多个功能,并且类似地可以跨越多个设备实现存储。因而本发明不限于任何单个实施例,而是将在广度、精神和范围上根据所附权利要求来解释。

[0087] 进而,本技术还可以如下所述地进行配置。

[0088] (1)一种计算环境中的至少部分地在至少一个处理器上执行的方法,包括:接收与至少一个数据源提供的基于上下文的信息相对应的上下文数据;接收与基于上下文的信息的准确度相对应的正确度信息;以及处理所述正确度信息,以增加与所述至少一个数据源针对其它接收的上下文信号提供的其它基于上下文的信息相对应的总准确度。

[0089] (2)根据(1)所述的方法,其中,接收所述上下文数据包括:接收数据,至少一个数据源通过所述数据提供包括位置信息的基于上下文的信息,并且进一步包括:使用所述位置信息以提供包括至少一个基于位置的结果的结果集,其中,接收所述正确度信息包括:接收指示基于位置的结果是否可能正确的反馈。

[0090] (3)根据(1)所述的方法,其中,接收所述上下文数据包括:接收数据,至少一个数据源通过所述数据提供包括位置信息的基于上下文的信息,其中,接收所述正确度信息包括:从可能更准确的数据源接收补充位置信息,并且其中,处理所述正确度信息以增加所述总准确度包括:如果所述位置信息不与所述补充位置信息一致,则校正所述至少一个数据源。

[0091] (4)根据(1)所述的方法,其中,接收所述上下文数据包括:接收数据,至少一个数据源通过所述数据提供包括位置信息的基于上下文的信息,其中,接收所述正确度信息包括:从可能更准确的数据源接收补充位置信息,并且其中,处理所述正确度信息以增加所述总准确度包括:如果所述位置信息基于所述补充位置信息已知为不正确,则动态校正所述至少一个数据源的输出。

[0092] (5)根据(1)所述的方法,其中,接收所述上下文数据包括:接收网际协议(IP)地址数据,所述IP地址数据用来获得包括位置信息的基于上下文的信息,并且其中,接收所述正确度信息包括:接收用户反馈,所述用户反馈指示所述用户对包括基于所述位置信息的至少一个结果的结果集满意还是不满意。

[0093] (6)根据(1)所述的方法,进一步包括:登录从多个采样中的每个采样收集的正确度信息。

[0094] (7)根据(6)所述的方法,其中,所述至少一个数据源包括:多个独立的数据源,其每个包含至少一些不准确的数据,并且其中,处理所述正确度信息以增加所述总准确度包括:处理所述多个采样的登录的正确度信息以从所述多个数据源之中创建混合数据源,其中包括:将每个数据源分段成分段,以在所述数据源之中提供多个对应分段;确定针对每个

对应分段的正确度度量；以及至少部分地基于针对每个对应分段的正确度度量，从所述对应分段之中选择用于所述混合数据源的所选分段。

[0095] (8)一种系统，包括：采样服务，所述采样服务被配置成登录用于多个采样请求中的至少一些采样请求的正确度数据，其中每个采样请求与上下文数据相关联，并且对于信息被登录的每个采样请求，所述采样服务被配置成：基于与该请求相关联的上下文数据，从除了一个或多个数据源之外的至少一个实体中，获得与在所述一个或多个数据源中查找的基于上下文的信息有关的正确度度量，并且登录表示所述正确度度量的数据。

[0096] (9)根据(8)所述的系统，其中，所述上下文数据包括网际协议(IP)地址、Wi-Fi信号或基于蜂窝塔的数据中的至少一个。

[0097] (10)根据(8)所述的系统，其中，所述一个或多个数据源包括：多个位置查找数据源，其每个针对上下文信号返回位置信息。

[0098] (11)根据(10)所述的系统，其中，所述查找数据源被分段成多个分段，数据源的每个分段在每个其它数据源中具有对应分段，并且进一步包括：混合机构，所述混合机构被配置成处理登录的正确度数据，其包括：确定针对每个对应分段的正确度分数；以及基于所述对应分段的正确度分数，将所选对应分段混合成混合数据源。

[0099] (12)根据(8)所述的系统，其中，进一步包括：校正机构，所述校正机构被配置成：处理登录的正确度数据，以提供用于所述一个或多个数据源的更新，或者提供用于所述一个或多个数据源的动态校正数据，或者提供用于所述一个或多个数据源的更新和用于所述一个或多个数据源的动态校正数据两者。

[0100] (13)根据(8)所述的系统，其中，所述采样服务与搜索引擎相关联，其中，所述多个采样请求对应于用户查询，并且其中，所述上下文数据包括如下信息，通过所述信息可以经由所述一个或多个数据源查找与每个用户查询相关联的可能位置。

[0101] (14)根据(13)所述的系统，其中，所述可能位置用来提供响应于用户查询而返回的结果集中的至少一个基于位置的结果，并且其中，与基于上下文的信息有关的正确度度量包括关于所述结果集的用户反馈。

[0102] (15)一个或多个具有计算机可执行指令的计算机可读介质，所述计算机可执行指令在被执行时执行步骤，所述步骤包括：从每个包含至少一些不准确的数据的多个数据源之中创建混合数据源，其包括：将每个数据源分段成分段，以在所述数据源之中提供多个对应分段；处理关于每个对应分段收集的正确度数据，以确定针对每个对应分段的正确度度量；以及基于针对每个对应分段的正确度度量，从所述对应数据源之中选择用于所述混合数据源的所选分段。

[0103] (16)根据(15)所述的一个或多个计算机可读介质，其中，所述混合数据源包括将上下文信号映射到所选分段的虚拟数据源，并且具有进一步的计算机可执行指令，其包括：通过访问所述虚拟数据源来查找用于上下文信号的信息，以确定哪个数据源对应于用于该上下文信号的所选分段；以及经由该所选分段来查找所述信息。

[0104] (17)根据(15)所述的一个或多个计算机可读介质，具有进一步的计算机可执行指令，其包括：通过将采样的请求分布到与所述数据源相对应的组中来收集所述正确度数据。

[0105] (18)根据(15)所述的一个或多个计算机可读介质，其中，将每个数据源分段成分段包括使用上下文数据范围。

[0106] (19) 根据(15)所述的一个或多个计算机可读介质,其中,将每个数据源分段成分段包括使用网际协议(IP)地址的范围。

[0107] (20) 根据(15)所述的一个或多个计算机可读介质,其中,处理关于每个对应分段收集的正确度数据以确定针对每个对应分段的正确度度量包括处理登录的反馈。

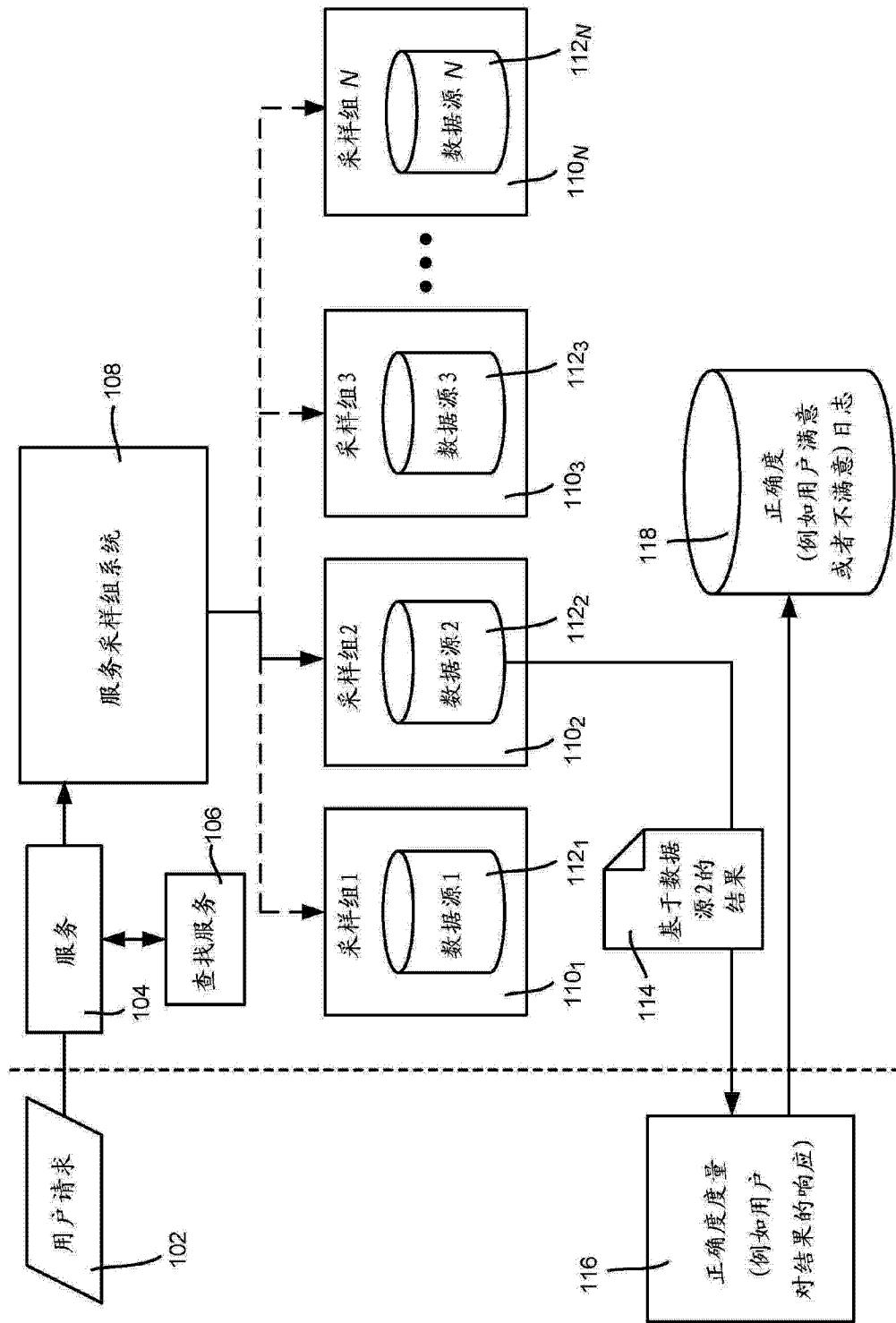


图 1

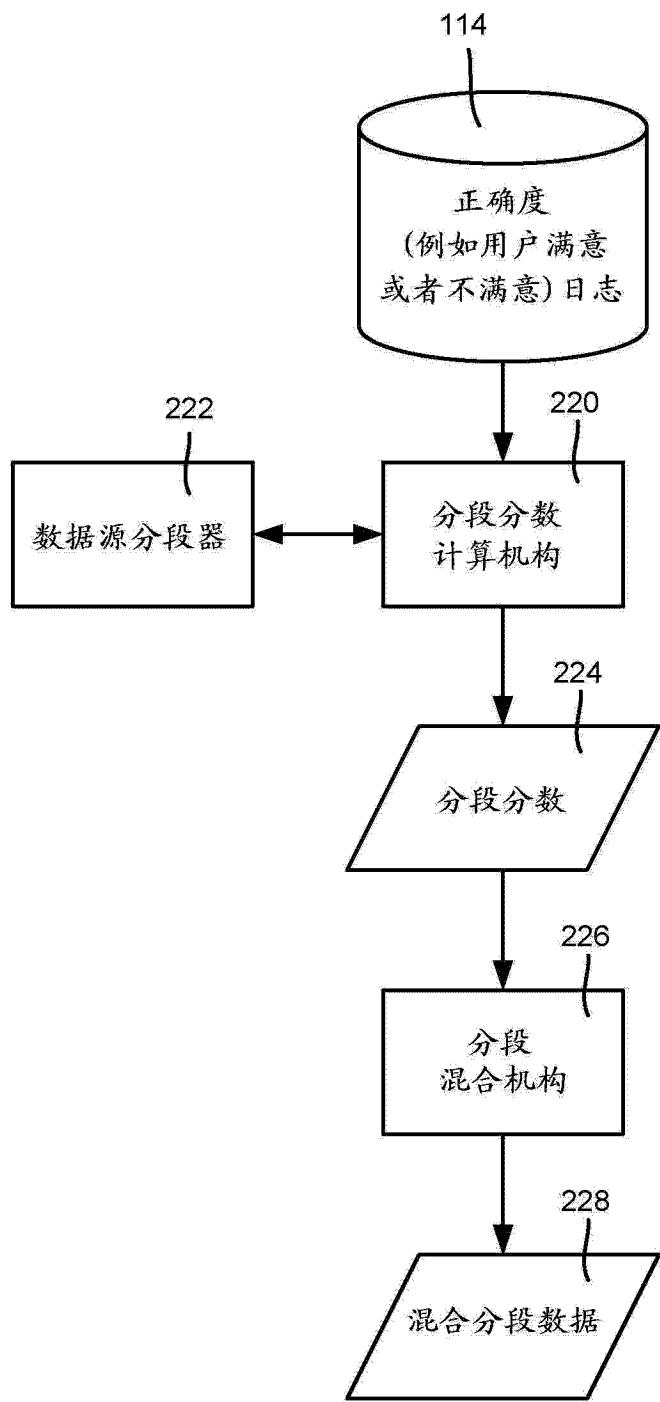


图 2

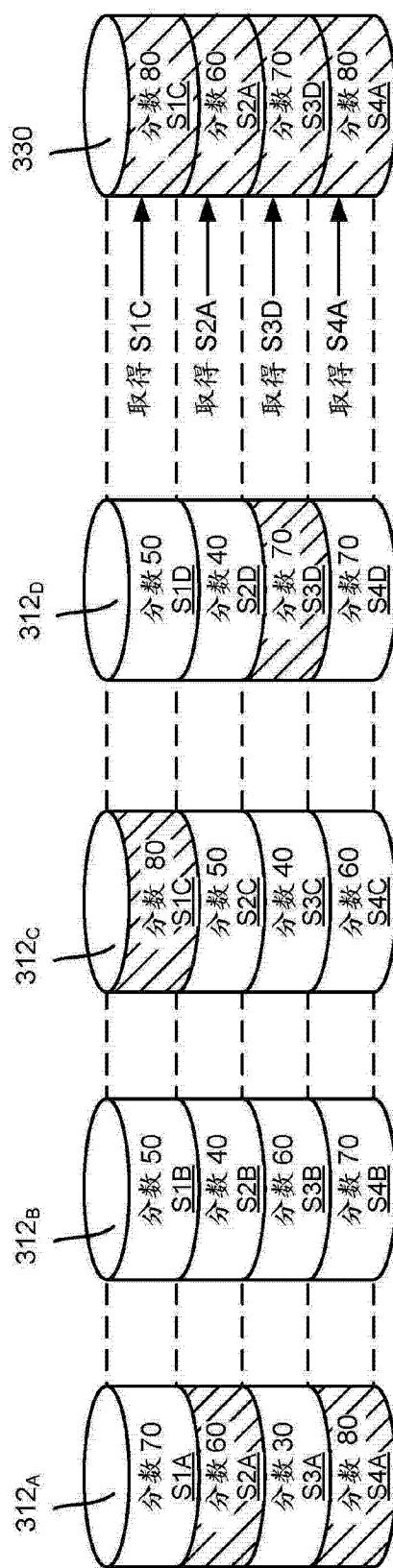


图 3

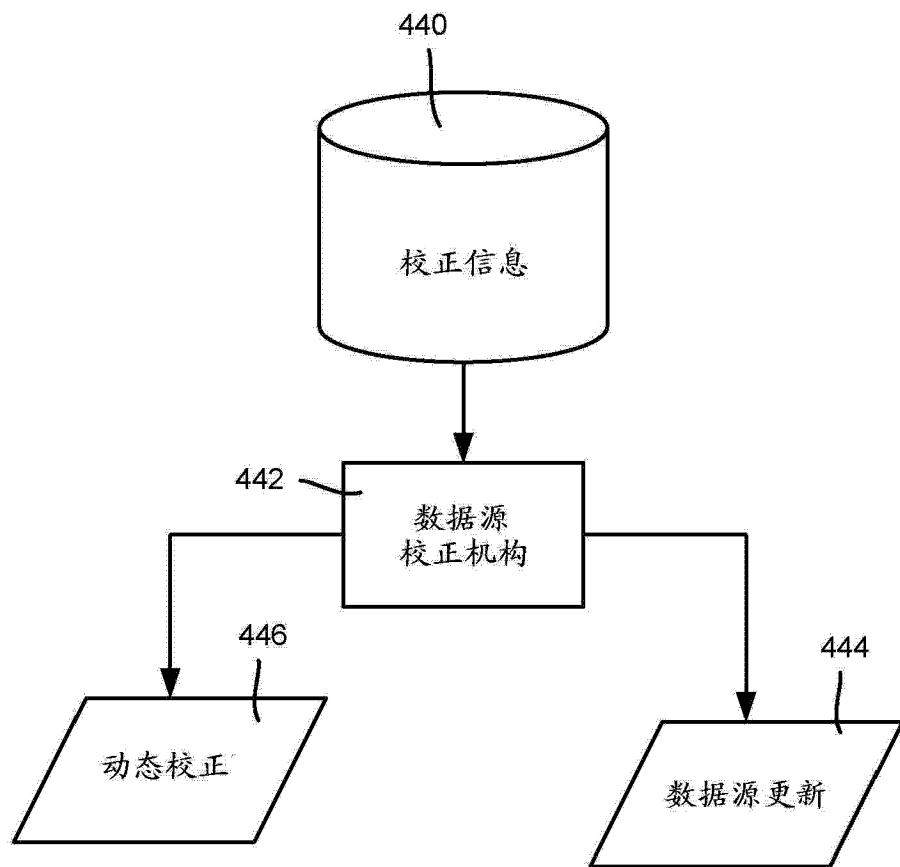


图 4

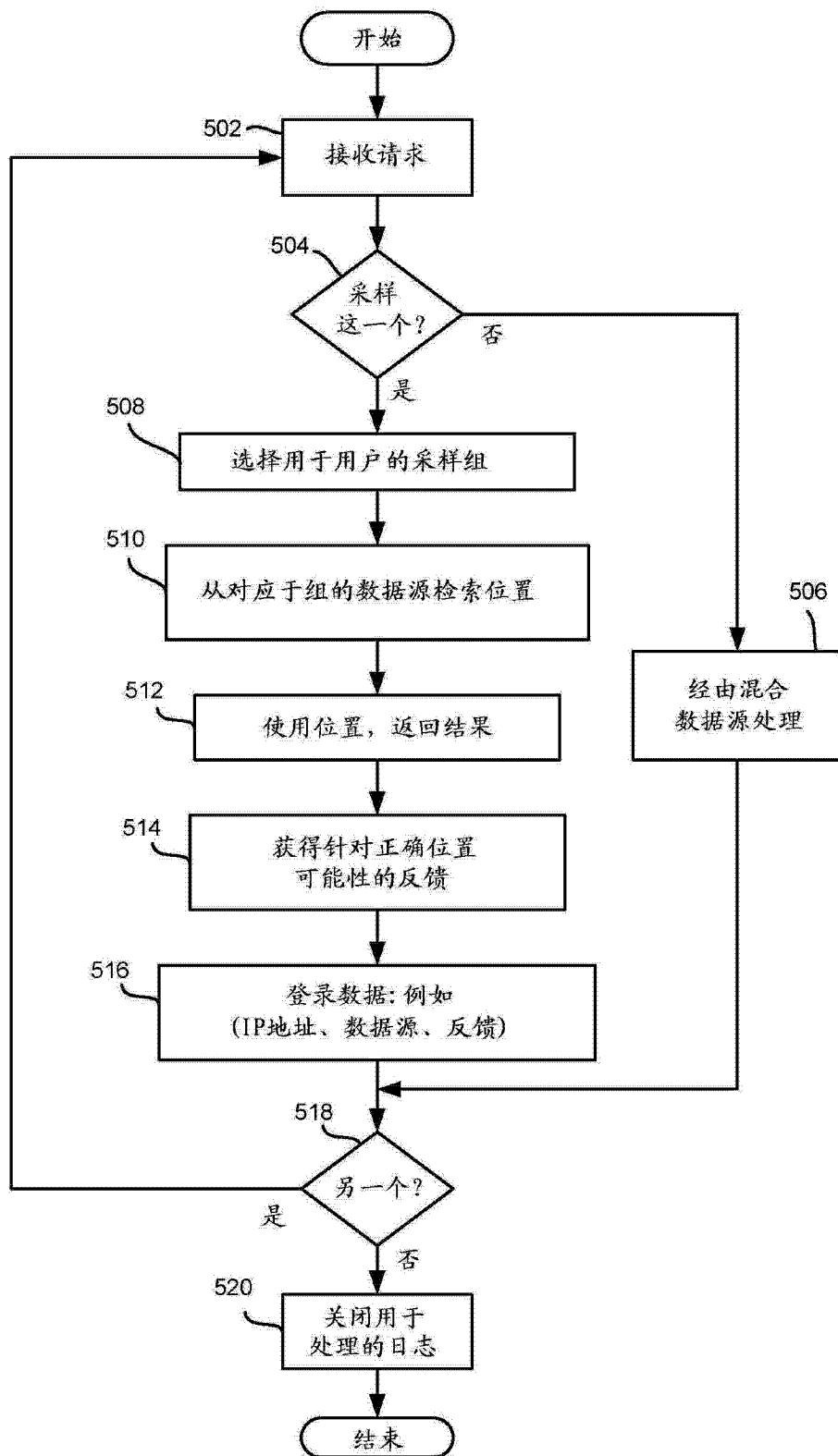


图 5

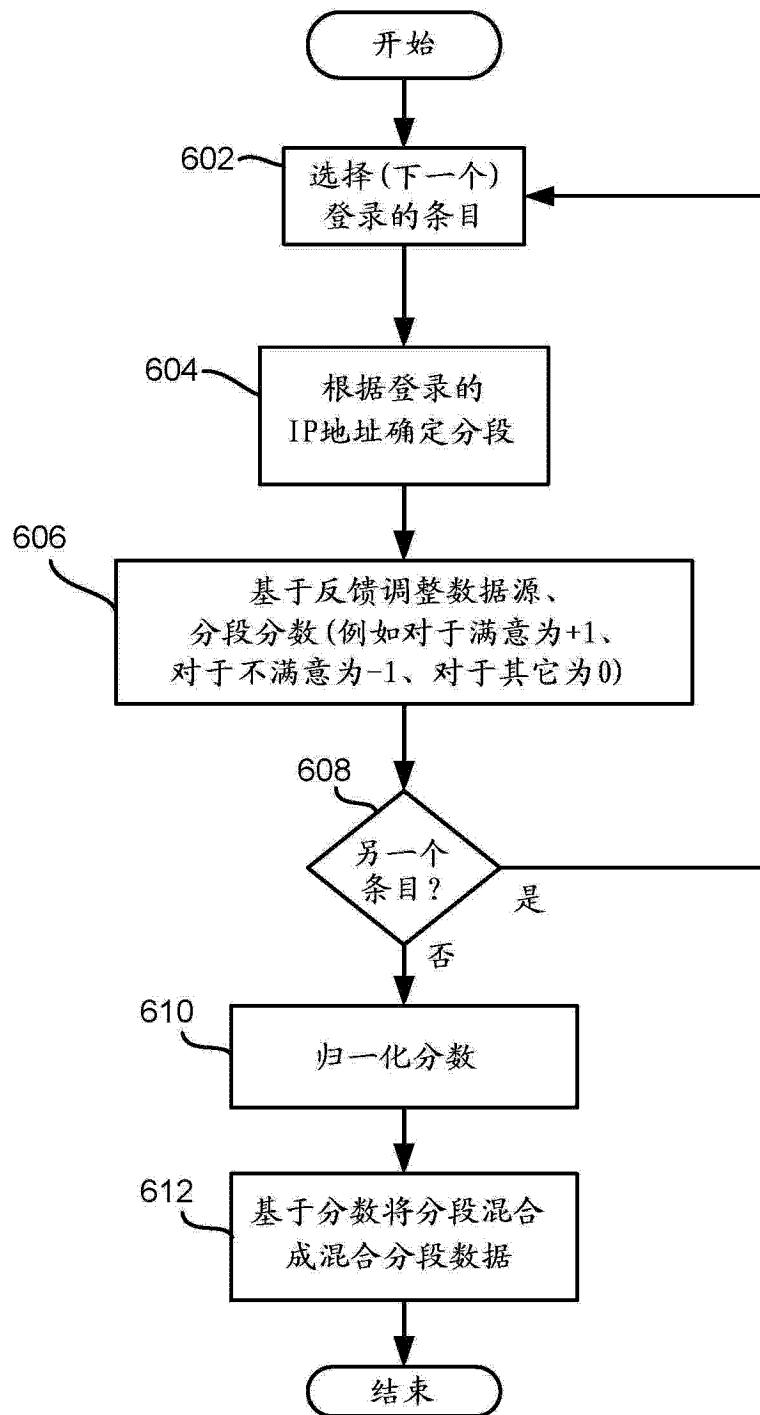


图 6

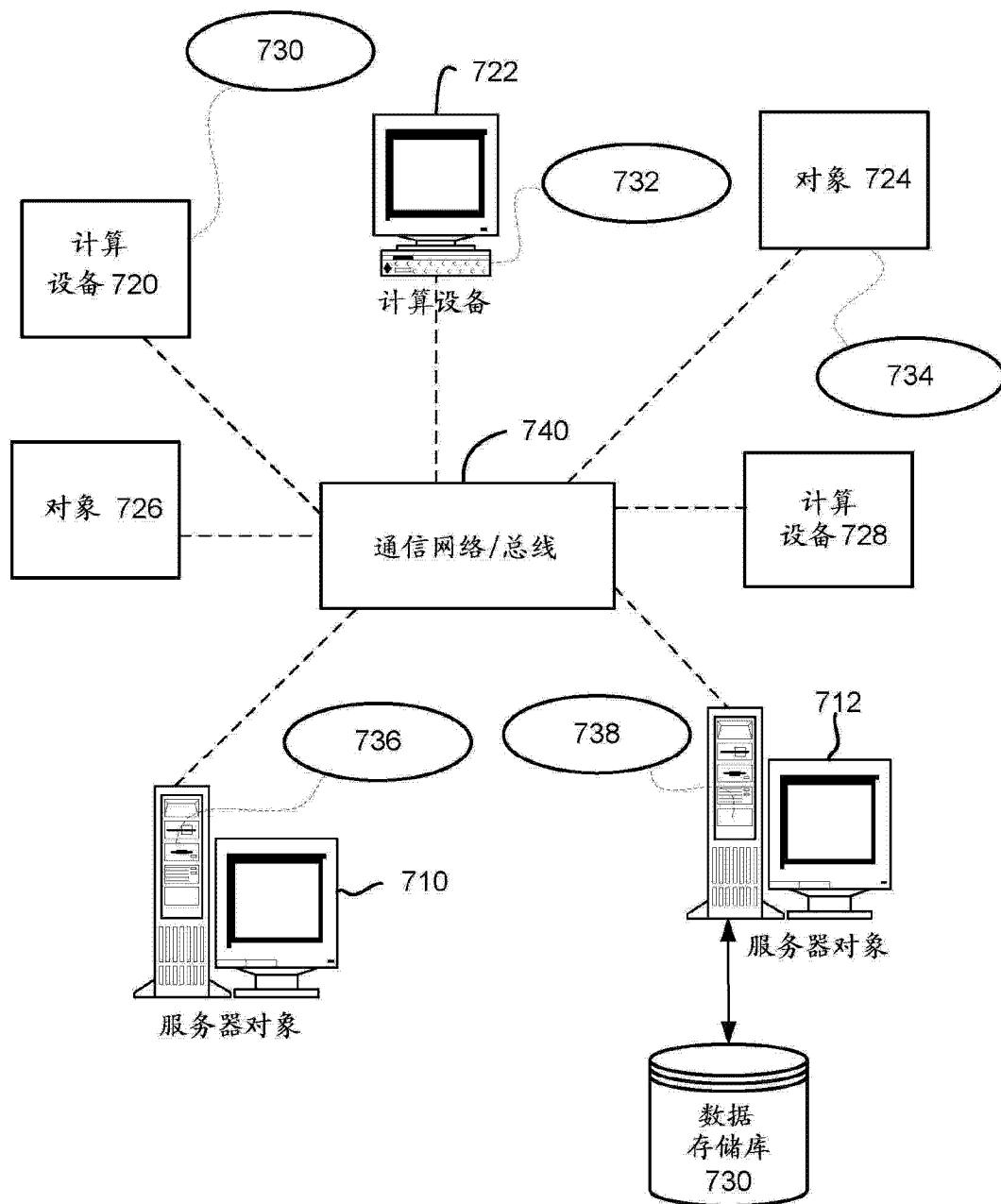


图 7

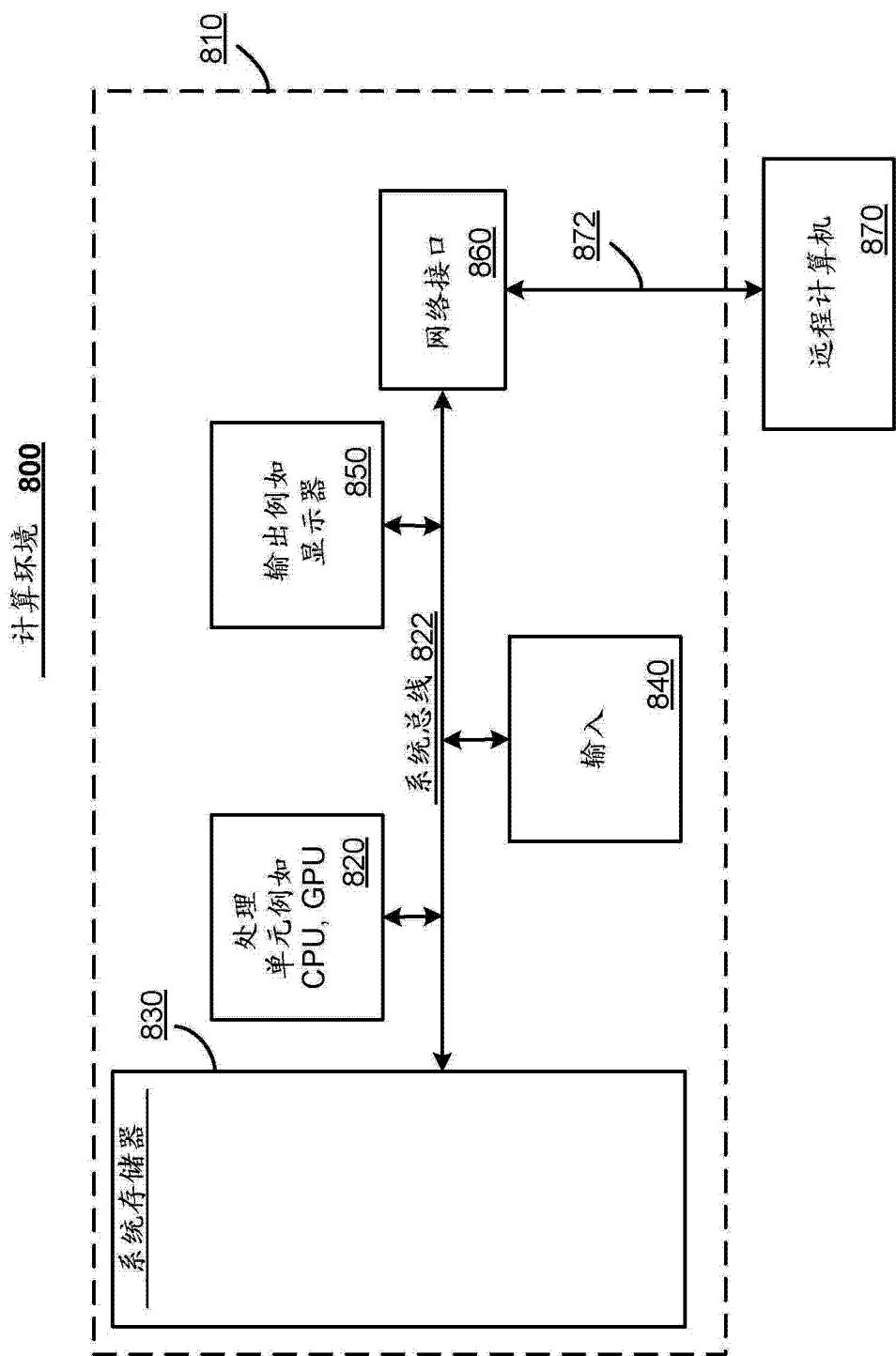


图 8