



(12)发明专利

(10)授权公告号 CN 104604200 B

(45)授权公告日 2017.12.08

(21)申请号 201380046232.4

(22)申请日 2013.07.03

(65)同一申请的已公布的文献号
申请公布号 CN 104604200 A

(43)申请公布日 2015.05.06

(30)优先权数据
12290234.9 2012.07.16 EP

(85)PCT国际申请进入国家阶段日
2015.03.04

(86)PCT国际申请的申请数据
PCT/EP2013/001958 2013.07.03

(87)PCT国际申请的公布数据
W02014/012627 EN 2014.01.23

(73)专利权人 阿尔卡特朗讯
地址 法国布洛涅-比扬古

(72)发明人 A·阿加萨延 M·布齐
M·科撒里 A·南迪

(74)专利代理机构 北京市金杜律师事务所
11256
代理人 王茂华

(51)Int.Cl.
H04L 29/06(2006.01)
H04L 29/08(2006.01)

(56)对比文件
CN 102609441 A, 2012.07.25,
WO 2007022199 A3, 2009.05.22,
US 2008205774 A1, 2008.08.28,
US 2002091736 A1, 2002.07.11,
审查员 刘婧

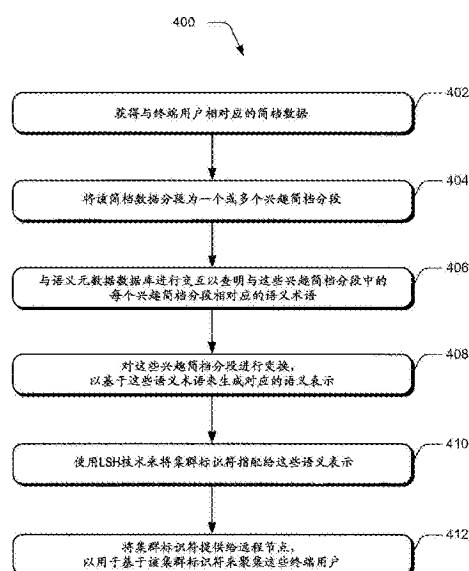
权利要求书3页 说明书15页 附图6页

(54)发明名称

用于对用户兴趣简档的隐私保护聚集的方法和装置

(57)摘要

根据本主题的一种实施方式,描述了用于对用户兴趣简档的隐私保护聚集的装置和方法。该方法包括基于终端用户的兴趣简档来生成至少一个兴趣简档分段(404)。进一步地,基于与耦合至该终端用户的用户设备的语义元数据数据库的交互,来获得与该至少一个兴趣简档分段相对应的语义术语(406),其中从该语义元数据数据库中提供的一个或多个语义术语之中获得这些语义术语。该至少一个兴趣简档分段中的每个兴趣简档分段随后被变换为至少一个语义表示(408)。进一步地,基于该至少一个语义表示,集群标识符被指配给该至少一个兴趣简档分段,其中使用局部敏感散列(LSH)技术来生成这些集群标识符(410)。



1. 一种用于对终端用户的兴趣简档的隐私保护聚集的方法,所述方法包括:
基于终端用户的兴趣简档来生成至少一个兴趣简档分段;
基于与通信地耦合至所述终端用户的用户设备的语义元数据数据库的交互,来获得与所述至少一个兴趣简档分段相对应的语义术语,其中从所述语义元数据数据库中提供的一个或多个语义术语之中获得所述语义术语;
基于所述语义术语,将所述至少一个兴趣简档分段中的每个兴趣简档分段变换为至少一个语义表示;以及
基于所述至少一个语义表示,将集群标识符指配给所述至少一个兴趣简档分段中的每个兴趣简档分段,其中使用局部敏感散列(LSH)技术来生成所述集群标识符。
2. 根据权利要求1所述的方法,其中所述方法进一步包括:将所述至少一个兴趣简档分段以及与所述至少一个兴趣简档分段中的每个兴趣简档分段相对应的所述集群标识符提供给远程节点,其中所述远程节点被配置为充当用于与所述集群标识符相关联的兴趣组的兴趣组聚合器,并且其中通过匿名通信信道来提供所述集群标识符。
3. 根据权利要求1所述的方法,其中所述方法进一步包括:
建立与远程节点的多个匿名通信信道;
将所述兴趣简档切分为一个或多个简档切片;以及
通过所述多个匿名通信信道,将所述集群标识符和所述一个或多个简档切片中的每个简档切片传输给所述远程节点,其中通过分离的匿名通信信道来传输所述一个或多个简档切片中的每个简档切片。
4. 根据权利要求3所述的方法,其中所述方法进一步包括:从所述兴趣简档中移除所述终端用户的个人可识别信息。
5. 根据权利要求1所述的方法,其中所述方法进一步包括:
通过多个匿名通信信道来传输所述集群标识符;
通过所述多个匿名通信信道来接收其他用户设备的集群标识符;
基于所述传输和所述接收,来识别具有与所述终端用户的所述集群标识符相类似的集群标识符的其他用户设备;以及
使得能够通过所述匿名通信信道在所述终端用户的所述用户设备与所述其他用户设备之间交换数据。
6. 根据权利要求1所述的方法,其中所述指配进一步包括:
获得所述至少一个兴趣简档分段中的每个兴趣简档分段与散列函数的集合之间的点积;以及
连结所述点积的符号以获得散列码作为与所述至少一个兴趣简档分段相对应的所述集群标识符。
7. 根据权利要求1-6中任一项所述的方法,其中所述变换进一步包括:基于所述语义术语来给所述至少一个兴趣简档分段中的每个兴趣简档分段加标签,以获得所述至少一个语义表示。
8. 一种与终端用户相关联的用户设备(102),包括:
处理器(202);以及
耦合至所述处理器(202)的存储器(206),所述存储器(206)包括:

简档变换模块(110),被配置为:

基于与耦合至所述用户设备(102)的语义元数据数据库(114)的交互,来获得与至少一个兴趣简档分段相对应的语义术语,其中从所述语义元数据数据库(114)中提供的一个或多个语义术语之中获得所述语义术语;并且

将所述至少一个兴趣简档分段中的每个兴趣简档分段变换为至少一个语义表示;以及

集群标识符模块(112),被配置为基于所述至少一个语义表示,将集群标识符指配给所述至少一个兴趣简档分段中的每个兴趣简档分段,其中使用局部敏感散列(LSH)技术来生成所述集群标识符。

9.根据权利要求8所述的用户设备(102),进一步包括:

兴趣简档生成模块(212),被配置为基于与所述终端用户的简档数据来生成所述终端用户的兴趣简档;以及

简档分段模块(214),被配置为将所述兴趣简档分段为所述至少一个兴趣简档分段。

10.根据权利要求8或9中任一项所述的用户设备(102),进一步包括数据传送模块(216),所述数据传送模块(216)被配置为:

建立与远程节点(104)的匿名通信信道(108),所述远程节点(104)被配置为充当用于与所述集群标识符相关联的兴趣组的兴趣组聚合器;以及

将所述至少一个兴趣简档分段以及与所述至少一个兴趣简档分段中的每个兴趣简档分段相对应的所述集群标识符提供给所述远程节点(104),其中通过所述匿名通信信道(108)来提供所述集群标识符。

11.根据权利要求8或9中任一项所述的用户设备(102),进一步包括数据传送模块(216),所述数据传送模块(216)被配置为:

建立与远程节点(104)的多个匿名通信信道(108);

从所述兴趣简档中移除所述终端用户的个人可识别信息;以及

通过所述多个匿名通信信道(108),将所述集群标识符和所述终端用户的所述兴趣简档传输给所述远程节点(104)。

12.根据权利要求8或9中任一项所述的用户设备(102),进一步包括数据传送模块(216),所述数据传送模块(216)被配置为:

将所述兴趣简档切分为一个或多个简档切片;以及

通过多个匿名通信信道(108),将所述集群标识符和所述一个或多个简档切片中的每个简档切片传输给远程节点(104),其中通过分离的匿名通信信道(108)来传输所述一个或多个简档切片中的每个简档切片。

13.根据权利要求8或9中任一项所述的用户设备(102),进一步包括数据传送模块(216),所述数据传送模块(216)被配置为:

通过多个匿名通信信道(108)来传输所述集群标识符;

通过所述多个匿名通信信道(108)来接收其他用户设备的集群标识符;

基于所述传输和所述接收,来识别具有与所述终端用户的所述集群标识符相类似的集群标识符的其他用户设备;以及

使得能够通过所述匿名通信信道(108)与所述其他用户设备交换数据。

14.根据权利要求8或9中任一项所述的用户设备(102),进一步包括数据传送模块

(216),所述数据传送模块(216)被配置为:

获得所述终端用户的兴趣组的中央聚合器节点的至少部分图像;

通过多个匿名通信信道(108)来与其他用户设备交换所述集群标识符;

基于通过所述多个匿名通信信道(108)接收的所述其他用户设备的集群标识符,来识别具有与所述终端用户的所述集群标识符相类似的集群标识符的所述其他用户设备;以及

通过所述匿名通信信道(108),将推荐服务提供给基于所述集群标识符识别的所述其他用户设备。

15.一种用于执行对终端用户的兴趣简档的隐私保护聚集的装置,所述装置包括:

用于基于终端用户的兴趣简档来生成至少一个兴趣简档分段的部件;

用于基于与耦合至所述终端用户的用户设备的语义元数据数据库的交互来获得与所述至少一个兴趣简档分段相对应的语义术语的部件,其中从所述语义元数据数据库中提供的一个或多个语义术语之中获得所述语义术语;

用于将所述至少一个兴趣简档分段中的每个兴趣简档分段变换为至少一个语义表示的部件;以及

用于基于所述至少一个语义表示而将集群标识符指配给所述至少一个兴趣简档分段中的每个兴趣简档分段的部件,其中使用局部敏感散列(LSH)技术来生成所述集群标识符。

用于对用户兴趣简档的隐私保护聚集的方法和装置

技术领域

[0001] 本主题涉及个性化服务和数据分析,并且特别地但是不排除地,涉及在将用户兴趣简档聚集用于个性化服务和数据分析时对终端用户的隐私保护。

背景技术

[0002] 由于巨大量的内容通过万维网是可用的,接入由服务提供者提供的内容的终端用户经常由服务提供商、搜索引擎、web发布商、以及广告商提供个性化的协助和数据分析,以便容易地接入相关内容。常规地,使用各种技术(诸如基于内容的推荐、协同推荐、以及数据分析)向终端用户提供个性化的服务。在基于内容的推荐中,向终端用户推荐如下的内容、服务或产品,这些内容、服务或产品类似于终端用户在过去使用或喜欢的内容、服务或产品或者匹配于终端用户的兴趣或选择。在协同推荐中,向终端用户推荐如下的内容、服务或产品,这些内容、服务或产品类似于由具有类似或相同兴趣或选择的其他终端用户所使用或喜欢的内容、服务或产品。

[0003] 在基于内容的推荐的一个示例中,电影评论网站可以监测到终端用户定期地观看某个种类的电影,例如动画电影。相应地,每次动画电影可用于观看时,可以向该终端用户提供推荐,诸如通知或警报,例如以通过进行相关支付来下载该电影。类似地,在协同推荐(也称为协同过滤)中,服务提供商可以向终端用户提供有目标的广告,其中这些广告属于由其他终端用户所偏好的产品或服务,这些其他终端用户具有与该终端用户类似和兴趣和偏好。例如,如果电视节目或者电影已经由其兴趣与终端用户的兴趣相匹配的其他终端用户所观看,则互联网协议电视(IPTV)服务提供商可以向该终端用户推荐这些电视节目或者电影。

[0004] 进一步地,用户需要将他们的人员数据提供给服务提供商用于执行数据分析应用,例如,以用于数据挖掘的目的。在用于数据分析以及由服务提供商向终端用户供应相关内容的所有这些技术中,现今的用户日益地关心他们的个人的并且可能是敏感的信息。这主要是因为,为了从个性化服务中受益并且促进数据分析,终端用户不得不透露敏感信息,但是同时,他们关心对信息的隐私保护。例如,社交联网站点的终端用户可能不反对使用他接入社交联网站点的信息来向其他终端用户并且向他自己进行与像这样的社交联网站点上的更新有关的匿名推荐,但是该终端用户可能不想要其他实体(诸如其他终端用户、服务提供商、攻击者和恶意方)知道该终端用户访问或评价的特定URL。

发明内容

[0005] 本概述被提供以介绍与在将终端用户聚集用于提供个性化服务并且执行数据分析应用时的隐私保护有关的概念。本概述不意图为识别被权利要求的主题的至关重要的特征,也不意图用于使用在确定或限制被权利要求的主题的范围中。

[0006] 在一种实施方式中,描述了一种用于对用户兴趣简档的隐私保护聚集的方法。该方法包括:基于终端用户的兴趣简档来生成至少一个兴趣简档分段。该方法进一步包括:基

于与耦合至该终端用户的用户设备的语义元数据数据库的交互,来获得与该至少一个兴趣简档分段相对应的语义术语,其中从该语义元数据数据库中提供的一个或多个语义术语之中获得这些语义术语。该至少一个兴趣简档分段中的每个兴趣简档分段随后被变换为至少一个语义表示。该方法进一步包括:基于该至少一个语义表示,将集群标识符指配给该至少一个兴趣简档分段,其中使用局部敏感散列 (LSH) 技术来生成这些集群标识符。

[0007] 在另一实施方式中,描述了一种用于对用户兴趣简档的隐私保护聚集的用户设备。在所述实施方式中,该用户设备包括处理器和耦合至该处理器的存储器。该存储器包括:简档变换模块,被配置为基于与耦合至该用户设备的语义元数据数据库的交互,来获得与至少一个兴趣简档分段相对应的语义术语。在一个实施例中,从该语义元数据数据库中提供的一个或多个语义术语之中获得这些语义术语。该简档变换模块进一步被配置为,将该至少一个兴趣简档分段中的每个兴趣简档分段变换为至少一个语义表示。该存储器进一步包括:集群标识符模块,被配置为基于该至少一个语义表示,使用局部敏感散列 (LSH) 技术来生成集群标识符,并且基于该至少一个语义表示,将该集群标识符指配给至少一个兴趣简档分段。

[0008] 在另一实施方式中,描述了一种计算机可读介质,在其上已经具体化了用于执行一种方法的计算机可读程序代码。该方法可以包括:基于终端用户的兴趣简档来生成至少一个兴趣简档分段。该方法进一步包括:基于与耦合至该终端用户的用户设备的语义元数据数据库的交互,来获得与该至少一个兴趣简档分段相对应的语义术语,其中从该语义元数据数据库中提供的一个或多个语义术语之中获得这些语义术语。该至少一个兴趣简档分段中的每个兴趣简档分段随后被变换为至少一个语义表示。该方法进一步包括:基于该至少一个语义表示,将集群标识符指配给该至少一个兴趣简档分段,其中使用局部敏感散列 (LSH) 技术来生成这些集群标识符。

附图说明

[0009] 参考附图来描述详细的描述。在这些附图中,参考标号的最左边的(多个)数字识别该参考标号首次出现在其中的附图。贯穿这些附图,使用相同的标号来参考相似的特征和组件。现在通过仅是示例的方式并且参考附图来描述根据本主题的实施例的系统和/或方法的一些实施例,在附图中:

[0010] 图1图示了根据本主题的一个实施例的用于个人信息的隐私保护的对用户简档的隐私保护聚集的示例性网络环境实施方式;

[0011] 图2图示了根据本主题的一个实施例的实施用于个人信息的隐私保护的对用户简档的隐私保护聚集的用户设备的组件;

[0012] 图3(a)、3(b)和3(c)图示了根据本主题的一个实施例的实施对用户简档的隐私保护聚集的各种应用的示例性网络环境实施方式;以及

[0013] 图4图示了根据本主题的一个实施例的用以向终端用户提供隐私保护的对用户简档的隐私保护聚集的方法。

[0014] 在本文档中,本文使用词语“示例性”来意指“用作一个示例、实例、或例证”。在本文中被描述为“示例性”的本主题的任何实施例或实施方式不必然被解释为相对其他实施例被偏好或者有优势。

[0015] 本领域的技术人员应当意识到,本文中的任何框图表示体现了本主题的原理的说明性系统的概念性视图。类似地,将意识到,任何流程图表、流程图、状态转换图、伪代码等表示各种过程,这些过程可以实质上被表示在计算机可读介质中并且如此由计算机或处理器来执行,而不论是否明确示出了这样的计算机或处理器。

具体实施方式

[0016] 描述了用于向终端用户的个人信息提供隐私性的系统和方法。这些系统和方法能够被实施在各种用户设备中,这些用户设备由终端用户用来利用服务或者查看由服务提供商提供的内容。进一步地,这些用户设备可以通过各种网络来与分布式服务器和其他用户设备进行通信。能够实施所描述的(多种)方法的用户设备和分布式服务器包括,但不限于:计算设备(诸如膝上型计算机、台式计算机、笔记本计算机)、移动电话、个人数字助理、工作站、大型计算机、机顶盒、媒体播放器、中央目录服务器、数据库服务器、文件服务器、打印服务器、web服务器、应用服务器,等等。尽管本文中的描述参考了提供推荐服务并执行数据分析应用的个人计算机和分布式服务器,但是如将由本领域的技术人员理解的,这些方法和系统可以被实施在向终端用户提供连接性以用于保护他们的隐私的其他服务器和计算机系统中。

[0017] 尽管推荐服务和个性化内容由服务提供商通过各种网络提供给不同的终端用户,但是本文所描述的方法和系统是接入独立的,并且支持多种接入类型,包括:全球移动系统(GSM)、宽带码分多址(W-CDMA)、码分多址(CDMA)、无线局域网(WLAN)、有线线路、局域网(LAN)、以及提供连接性的其他网络。

[0018] 常规地,许多服务提供商尝试基于终端用户的个人细节、偏好、以及选择,来个性化向终端用户提供的服务,诸如提供内容的服务,内容诸如视频、音频、新闻、广告等。为了这个目的,服务提供商使用各种技术,基于如个人信息(诸如终端用户的年龄、性别、以及位置)的因素,基于终端用户的过去行为或者已经被识别为具有与该终端用户的兴趣相类似的兴趣的其他终端用户的过去偏好,来推荐服务、内容、或产品并且执行数据分析应用。

[0019] 例如,如果终端用户(比如用户A)购买了由特定作者写的书,则服务提供商可以建议用户A购买由相同作者写的其他书或者与相同或相关主题有关的其他书,等等。在类似的但是另一种常规方法中,服务提供商可以向终端用户A推荐由具有与终端用户A的兴趣简档相类似的兴趣简档的其他终端用户已经偏好的那些内容或产品。为了提供个性化服务的目的并且为了数据分析,各种常规已知的方法被用来将具有类似兴趣的终端用户聚集为一个或多个兴趣组。

[0020] 由服务提供商实施的聚集的常规方法要求收集与终端用户的个人偏好、选择等有关的信息。常规地,服务提供商通过各种手段,诸如通过分析日志文件、应用历史文件、或者保存在终端用户的客户端设备上的或者由用户在向服务提供商注册时提供的其他的个人可识别信息,来监测并收集属于终端用户的信息。在另一种常规技术中,服务提供商可以保存一个文本文件,诸如超文本传送协议(http) cookie,来收集属于终端用户的信息。例如,web门户可以保存包括终端用户的web浏览器的http cookie的用户偏好数据来存储该终端用户的偏好,诸如字体大小、显示小部件的布置等。进一步地,用户偏好数据还可以存储终端用户的浏览细节,并且将这些细节发送给web门户。

[0021] 因此,在基于终端用户的个人选择来向终端用户提供推荐服务、个性化内容、服务或产品的尝试中,服务提供商经常监测并收集属于终端用户的行为的信息。在某些情形中,基于由服务提供商收集的信息来识别终端用户成为可能。这可能会导致损害终端用户的个人或机密信息,并且使终端用户暴露于潜在的隐私违反或者使他成为广告商或垃圾邮件发送者等的目标。进一步地,在极端的情况中,终端用户可能成为各种犯罪的受害者,诸如身份盗窃和信用卡诈骗。

[0022] 进一步地,现今在终端用户经由高速通信支持而通过虚拟社交网络来相互连接时,终端用户不断地向广泛类别的应用分享着私人和个人信息,诸如 **StumbleUpon**[®]、**GoogleNews**[™]、**Foursquare**[®]、**Facebook**[®]、**Yahoo!**[®]、以及 **twitter**[™]。尽管所透露的敏感个人信息的性质取决于应用的类型,诸如对于基于位置的服务 (LBS) 而言是位置并且对于统一资源定位符 (URL) 推荐器系统而言是浏览历史,但是将这些应用注入在线社交网络 (OSN) 的包罗万象的框架中 (其中 OSN 正开始充当用于供应推荐服务的“一站式商店”) 正使得用户的个人信息比以前任何时候都更加危险。

[0023] 近来,已经提议了在保留隐私的同时仍然执行数据分析应用并提供个性化服务 (诸如推荐服务)、个性化内容、个性化搜索/查询、社交网络、数据挖掘、以及其他有关服务的不同方法。

[0024] 这种已有的方法之一涉及使用基于同态加密的策略来聚集兴趣简档。在这样的方法中,使用加密技术 (诸如同态加密) 来加密用户的兴趣简档,以便向用户提供隐私性。这样的加密技术使得能够执行协议基元操作 (诸如加法和乘法), 以便使用分布式计算设定来聚集兴趣简档。使用分布式计算设定允许对终端用户的用户设备的静态集合执行多个阶段的协议基元操作。例如, Kmeans 聚集算法可以被实施在使用了同态加密技术的隐私保留方法中。尽管使用同态加密的技术促进了向终端用户提供隐私性,但是这样的技术通常是资源密集的并且执行起来是复杂的。进一步地,对终端用户的用户设备的静态集合执行多阶段的协议基元操作,限制了这样的技术在动态环境中的应用,动态环境诸如在线门户,其中终端用户的集合随着终端用户不断地加入和离开这些在线门户而动态地改变。

[0025] 进一步地,某些方法提供了集中式解决方案,其中被配置为聚集兴趣简档的集中式中介 (intermediaries) 不被终端用户所信任。在这样的情形中,每个终端用户使用基于随机投影的方法 (诸如加密) 来初始地预处理兴趣简档,以在被传达给集中式中介之前获得经投影的兴趣简档。典型地通过使用以集中式中介不可得到的随机投影函数来变换兴趣简档,以获得经投影的兴趣简档,因此确保终端用户的数据的隐私性。然而,这些随机投影函数在正意图成为共同集群的一部分的终端用户之间进行共享,以确保所有类似的兴趣简档被变换为类似的经投影的兴趣简档,以便促进高效和正确的聚集。然而,在终端用户之间共享这些随机投影函数可能使得随机投影函数的泄漏成为可能,因为恶意的终端用户可能将随机投影函数提供给集中式中介,以便从经投影的兴趣简档获得原始兴趣简档。

[0026] 用于保护私人信息的另一种技术是使用具有代理节点的流言代表 (gossip-on-behalf) 方法,该代理节点被配置为代表终端用户来传输兴趣简档。这样的技术涉及使用经加密的两跳通信而通过中介节点来将终端用户的用户简档传输给代理节点,以便对代理节点隐藏终端用户身份。然而,与代理节点共享完整的兴趣简档可能会促进关联性

(linkability) 攻击,因为这些代理节点可能通过分析一段时间上的兴趣简档来识别终端用户。

[0027] 根据本主题的一种实施方式,描述了用于在聚集终端用户的兴趣简档时确保隐私性以便向终端用户的个人信息提供隐私性的系统和方法。如之前所描述的,这些系统和方法能够被实施在能够根据被定义用于通信的各种不同标准来与网络进行通信的各种处理和通信设备中。进一步地,本文所描述的系统和方法可以通过经由不同手段提供的有线网络或者无线网络来进行连接。

[0028] 如本文所描述的系统和方法,在一方面,通过向每个终端用户和他们的私人信息提供匿名性而不限对应用的使用,向终端用户的私人信息提供了真实保护,并且在另一方面,也向终端用户提供了如下的能力:利用包括瘦客户端的任何计算设备,诸如具有有限计算能力的移动设备,来接收个性化服务或者使用要求了根据用户的私人数据的相似度的用户聚集的任何其他应用或服务。在一个实施例中,隐私保护网络包括与用于相互通信或者与分布式服务器(诸如一个或多个聚合器节点)进行通信的用户设备相关联的多个终端用户。

[0029] 根据本主题的一个实施例,为了向终端用户提供个性化内容和推荐服务,创建了具有类似兴趣简档的终端用户的集群。为了该目的,终端用户中的每个终端用户的用户设备基于简档数据来初始地生成终端用户的兴趣简档,该简档数据诸如终端用户的个人选择和偏好。例如,用户设备可以获得如下的列表:由终端用户所访问的所有URL、由终端用户所搜索或观看的所有视频、由终端用户所收听和搜索的音乐、以及与由终端用户所探寻的其他类似兴趣领域有关的数据。

[0030] 在一个实施例中,这些兴趣简档随后可以被分段为多个兴趣简档分段,并且每个分段对应于特定的兴趣领域。兴趣领域的示例包括,但不限于:情境偏好,诸如职业兴趣、休闲偏好、和宗教偏好,以及兴趣的范围,诸如音乐、舞蹈、电子购物、和社交联网。这些兴趣简档分段随后可以被处理以用于聚集为各种兴趣组。在另一实施例中,兴趣简档作为整体可以被处理以用于聚集为各种兴趣组。

[0031] 这些兴趣简档分段随后被变换为被称为语义表示的紧凑且语义密集表示。作为示例而不是限制,这些兴趣简档分段的语义表示包括:基于标签的表示、基于本体(ontology)概念的表示、以及基于主题的表示。例如,与社交联网站点“xyz.com”的URL相对应的兴趣简档可以被变换为语义表示“社交联网站点”。

[0032] 为了该目的,用户设备可以通过匿名通信信道与语义元数据数据库进行交互,语义元数据数据库具有语义术语的列表或词典,语义术语可以被用于将兴趣简档分段变换为对应的语义表示。在一个实施例中,用户设备可以将兴趣简档分段提供给语义元数据数据库,语义元数据数据库可以识别与这些兴趣简档分段中的每个兴趣简档分段相对应的语义术语,以将这些兴趣简档分段变换为对应的语义表示。在另一实施例中,用户设备可以从语义元数据数据库获得语义术语的列表,并且然后识别与兴趣简档分段相匹配的语义术语以获得对应的语义表示。在又另一实施例中,用户设备它自身可以包括元数据,该元数据具有从由终端用户所浏览的网页提取的语义术语。在所述实施例中,用户设备被配置为相互共享语义术语的词汇表,从而由不同用户设备生成的语义表示是使用相同术语来表达的并且因此是可比较的。

[0033] 可以向如此获得的语义表示指配一个集群标识符来识别终端用户能够被关联至其的兴趣组。可以使用局部敏感散列 (LSH) 的技术来指配集群标识符。LSH技术涉及,使用语义表示和由随机值定义的散列函数的集合,诸如在每个用户设备处生成的共同的随机矢量序列,来将语义表示中的每个语义表示转换为对应的散列码,即集群标识符。在用户设备处指配集群标识符促进了提供隐私保护,因为没有与远程节点共享兴趣简档。进一步地,使用LSH技术帮助增加了对用户简档的聚集中的效率,因为现在能够通过匹配散列码来容易地执行该聚集。此外,使用语义元数据数据库帮助减少了错误否定和错误肯定的发生,因此增加了聚集的效率。因此,通过LSH技术来执行聚集结合通过生成语义表示的简档变换,促进了提供具有聚集中的减少的错误的隐私保护。

[0034] 如此获得的集群标识符可以被用于将用户设备聚集为一个或多个集群,即与集群标识符相对应的兴趣组。进一步地,通过通信信道,例如匿名的通信信道,集群标识符被提供给一个或多个远程节点,例如,中央聚合器、兴趣组聚合器、以及另一用户设备。

[0035] 在一种实施方式中,可以取决于用户设备针对其而可能被使用的基于推荐的服务的应用或者类型,来选择与用户设备进行通信的远程节点。例如,在协同过滤应用的情况下,远程节点可以是中央聚合器,并且每个中央聚合器对应于至少一个集群,即兴趣组,诸如电影、舞蹈、电子购物、社交联网站点、或者它们的组合。用户设备因此可以将集群标识符中的每个集群标识符连同用户的兴趣简档一起提供给对应的中央聚合器,以使得中央聚合器能够在它们的对应兴趣组中的兴趣组的兴趣领域之后,识别最高排名(即首位排名)的兴趣组。例如,对与“音乐”有关的兴趣组负责的中央聚合器可以接收具有对应集群标识符的所有用户设备的兴趣简档。中央聚合器随后可以分析这些用户简档,来识别其他兴趣领域或者由于例如存在于多数(比如90%)终端用户的兴趣简档中而是高排名的兴趣。例如,中央聚合器可以将“舞蹈”识别为最高排名的兴趣,并且因此可以从负责兴趣领域“舞蹈”的中央聚合器获得与舞蹈相对应的推荐。中央聚合器因此可以将与音乐以及舞蹈有关的内容和推荐提供给存在于与集群“音乐”相对应的集群中的所有用户设备。中央聚合器因此可以提供不仅与用户的兴趣有关而且与兴趣组中的高排名的新兴趣或项目有关的内容和推荐。在另一实施例中,可以用充当负责一个兴趣组的中央聚合器的每个用户设备来代替中央聚合器。

[0036] 类似地,在基于内容的推荐应用的情况下,远程节点可以是兴趣组聚合器,并且每个兴趣组聚合器对应于至少一个集群,即兴趣组,诸如电影、舞蹈、电子购物、社交联网站点。用户设备在这样的情况下被配置为将集群标识符和语义表示传输给与集群标识符相对应的兴趣组聚合器,兴趣组聚合器可以基于该集群标识符来提供与该集群标识符相对应的内容。

[0037] 进一步地,在涉及端到端通信的应用中,远程节点可以是其他用户设备。在这样的应用中,用户设备可以使用集群标识符来识别具有类似集群标识符的其他用户设备,以便于使得能够使用匿名通信信道在用户设备之间进行聊天和交换数据。使用匿名通信信道帮助确保其他用户或者远程节点不能识别传输集群标识符的终端用户,因为集群标识符和其他数据是使用形成这些匿名通信信道的各种中介节点来发送的。

[0038] 本主题的系统和方法因此通过在用户设备处将集群标识符指配给兴趣简档,来促进向终端用户提供隐私性。使用LSH技术在用户设备处将集群标识符指配给兴趣简档,使得

能够在远程节点处聚集类似的用户而无需远程节点接入终端用户的兴趣简档,因此向终端用户提供隐私保护。进一步地,使用LSH技术结合生成语义描述帮助确保隐私被保护并且聚集集中的错误被减少,因为使用语义元数据数据库帮助减少了错误否定和错误肯定的发生,因此增加了聚集中的效率。因此,在用户设备处提供集群标识符促进了内容提供商执行数据分析并且更快地和高效地提供个性化服务,因为服务提供商现在不需要分析兴趣简档,并且可以使用任何内容匹配技术来将集群标识符与内容数据进行匹配。另外,使用集群标识符减少了成本并且节省了服务提供商的终端处的资源,因为服务提供商现在不需要处理或者分析兴趣简档。

[0039] 应当注意,本描述和附图仅仅说明了本主题的原理。因此将意识到,本领域的技术人员将能够设计出尽管本文没有明确描述或示出但是体现了本主题的原理并且被包括在其精神和范围内的各种布置。此外,本文所记载的所有示例主要明确地意图为仅用于教导的目的,以辅助读者理解本主题的原理和由(多位)发明人为了促进本领域而贡献的概念,并且将被解释为没有对这样具体记载的示例和条件作出限制。此外,本文中记载了本主题的原理、方面、和实施例以及它们的具体示例的所有陈述,意图为涵盖它们的等价物。

[0040] 本领域的技术人员还将意识到,如本文所使用的词语“在……期间”、“在……的同时”、“当……时”不是意味着一经发起动作就立即发生的动作的确切术语,而是在初始动作与由该初始动作发起的反应之间可能存在某种小的但是合理的延迟,诸如传播延迟。另外,自始至终地使用词语“连接的”和“耦合的”是为了本描述的清楚性并且能够包括直接连接或者间接连接。

[0041] 已经关于图1-4详细解释了聚集用户简档以用于对终端用户的隐私保护的系统和应该被实施的方式。尽管所描述的用于聚集用户简档以用于提供隐私保护的系统和方法的各方面能够被实施在任何数目的不同计算系统、传输环境、和/或配置中,但是在以下(多个)示例性系统的上下文中描述了各实施例。

[0042] 图1图示了根据本主题的一个实施例的在促进数据分析和接入个性化服务的同时对用户简档的隐私保护聚集以用于对终端用户的隐私保护的的网络环境100的实施方式。网络环境100包括通过通信网络106相互通信并且与远程节点104进行通信的一个或多个用户设备102-1、102-2、……、以及102-N,后文中统称为用户设备102并且个别地称为用户设备102。

[0043] 通信网络106可以是无线网络,或者有线网络和无线网络的组合。通信网络106能够是彼此互连并且用作单个大型网络(例如,互联网或内联网)的个体网络的集合。这样的个体网络的示例包括,但不限于:第3代合作伙伴计划(3GPP)、长期演进(LTE)、全球移动通信系统(GSM)网络、通用移动通信系统(UMTS)网络、个人通信服务(PCS)网络、时分多址(TDMA)网络、码分多址(CDMA)网络、下一代网络(NGN)、公共交换电话网络(PSTN)、综合服务数字网络(ISDN)。通信网络106可以是任何公共网络或者私人网络,包括局域网(LAN)、广域网(WAN)、因特网、内联网、端到端网络和虚拟私人网络(VPN),并且可以包括各种网络设备,诸如路由器、网桥、服务器、计算设备、存储设备,等等。

[0044] 用户设备102可以被定义为如下的用户装备(UE),这些用户装备由终端用户用来相互通信或者与其他网络实体(诸如远程节点104和分布式服务器)进行通信,以用于促进数据分析并且接入各种个性化服务(诸如推荐服务)、个性化内容、个性化搜索/查询、社交

网络、数据挖掘、以及其他有关服务。用户设备102的示例可以不带限制地包括：计算设备，诸如膝上型计算机或其他便携式计算机、台式计算机、笔记本计算机、网络计算机；移动电话；手持式设备；个人数字助理；工作站；大型计算机；机顶盒；媒体播放器；中央目录服务器；数据库服务器；文件服务器；打印服务器；web服务器；应用服务器；等等。用户设备102中的每个用户设备在由用户设备102耦合至其的通信网络106所定义的通信协议上进行工作。进一步地，用户设备102被配置为使用通信信道108-1、108-2、108-3、108-4、108-5、……、108-n(后文统称为通信信道108)中的一个或多个通信信道，通过通信网络106来彼此交互或者与其他网络实体(诸如远程节点104)进行交互。在一种实施方式中，通信信道108可以是基于例如洋葱(onion)路由框架(诸如TOR)所配置的匿名通信信道，以支持用户设备102之间以及用户设备102与其他网络实体(诸如远程节点104)之间的匿名通信。

[0045] 远程节点104能够被实施为各种计算设备，诸如膝上型计算机、台式计算机、笔记本计算机、工作站、大型计算机、服务器，等等。尽管远程节点104被示出为一个实体，但是远程节点104也可以被实施为包括分布在网络中的多个中介节点的分布式计算系统，在该网络中，每个节点能够被实施为计算设备，诸如膝上型计算机、台式计算机、笔记本计算机、工作站、大型计算机、服务器，等等。进一步地，中介节点可以通过中间网络(没有示出在附图中)而连接以用于通信和数据交换的目的。进一步地，远程节点104的示例包括，但不限于：中央聚合器、兴趣组聚合器、以及用户设备102。

[0046] 在一种实施方式中，可以取决于用户设备102针对其可以被使用的基于推荐的服务的应用或类型，来选择与用户设备102进行通信的远程节点104。例如，在协同过滤应用的情况下，远程节点104可以是中央聚合器，类似地在基于内容的推荐应用的情况下，远程节点104可以被实施为兴趣组聚合器。

[0047] 在一个实施例中，终端用户的兴趣简档是基于终端用户的行为来生成的并且由用户设备102本地保存。例如，可以基于与终端用户相对应的简档数据来生成终端用户的兴趣简档。简档数据可以指示，例如，由终端用户所访问的网站、由终端用户所播放或下载的歌曲或视频、由终端用户所使用的产品或者所利用或评论的服务，等等。基于所生成的兴趣简档，客户端设备将终端用户分类在一个或多个预定义的兴趣组中。兴趣组可以被理解为共享类似的兴趣和选择的终端用户的组。

[0048] 为了该目的，用户设备102可以包括简档变换模块110和集群标识符模块112。简档变换模块110被配置为将兴趣简档变换为紧凑且语义密集表示，称为语义表示。作为示例而不是限制，语义表示包括基于标签的表示、基于本体概念的表示、以及基于主题的表示。在一个实施例中，完整的兴趣简档可以由简档变换模块110变换为语义表示。在另一实施例中，兴趣简档的分段(后文中称为兴趣简档分段)可以被变换为语义表示。虽然本文的描述参考了兴趣简档分段，但是如将由本领域的技术人员理解的，这些方法和系统可以被实施用于变换完整的兴趣简档，尽管有些变化。

[0049] 为了变换兴趣简档，简档变换模块110可以通过通信信道108-1与诸如语义元数据库114的数据库进行交互，以获得与兴趣简档分段中的每个兴趣简档分段有关的元数据。语义元数据库114可以被理解为具有语义术语的列表的词典，这些语义术语可以被用于将兴趣简档分段变换为对应的语义表示。语义术语可以被理解为如下的术语，这些术语基于终端用户的各种兴趣领域的宽泛含义或定义来识别它们，并且因此可以被用于基于

每个兴趣领域的概念来高效和无错误地匹配各种兴趣简档。兴趣领域的示例包括,但不限于:情境偏好,诸如职业兴趣、休闲偏好、和宗教偏好,以及兴趣的范围,诸如音乐、舞蹈、电子购物、和社交联网。例如,与社交联网站点“xyz.com”的两个不同URL相对应的兴趣简档可以被匹配至语义术语“社交联网站点”,而不管URL具有不同的版本并且因此在句法上是不同的。尽管语义元数据数据库114已经被示出在用户设备102的外部,但是将理解,在其他实施例中,语义元数据数据库114可以在用户设备102的内部。进一步地,在后者的情况中,可以通过在用户设备之间交换词汇表和术语,来跨用户设备地使语义元数据数据库相一致。

[0050] 由简档变换模块110如此生成的语义表示可以被提供给集群标识符模块112,以便将集群标识符指配给语义表示中的每个语义表示。集群标识符可以被理解为如下的兴趣组标识码,这些兴趣组标识码可以被用于高效地识别与终端用户能够被关联至其的集群标识符相对应的集群,即兴趣组。集群标识符模块112可以被配置为使用LSH的技术用于指配集群标识符。LSH技术可以由集群标识符模块112用来将语义表示中的每个语义表示转换为对应的散列码,即集群标识符。为了该目的,集群标识符模块112利用由随机值定义的散列函数的集合,诸如在用户设备102中的每个用户设备处生成的共同的随机矢量序列。

[0051] 在一种实施方式中,散列函数由集群标识符模块112从中央实体获得,以确保集群识别由所有的用户设备102一致地完成,该中央实体诸如由所有的用户设备102可接入的随机种子生成器(未示出在附图中)。在另一实施方式中,集群标识符模块112可以被配置为,基于从例如中央实体获得的种子生成函数来生成散列函数。在用户设备102处指配集群标识符因此促进了提供隐私保护,因为没有与远程节点共享兴趣简档。通过LSH技术来执行聚集结合通过生成语义表示的简档变换,促进了提供具有聚集中的减少的错误的隐私保护。

[0052] 如此获得的集群标识符可以被用于将用户设备102(即终端用户)的用户简档聚集为一个或多个集群。进一步地,集群标识符模块112通过通信信道108将集群标识符提供给一个或多个远程节点104,以用于一个或多个服务,诸如聚集用于提供推荐服务或者用于端到端通信。基于用户设备102可能想要利用的基于推荐的服务的应用或类型,与用户设备102进行通信的远程节点104可以在中央聚合器和兴趣组聚合器之间变化。替换地,用户设备102可以相互通信,并且每个用户设备102充当用于其他用户设备102的远程节点。

[0053] 例如,在协同过滤应用的情况中,远程节点可以是中央聚合器,并且每个中央聚合器对应于至少一个集群,即兴趣组,诸如电影、舞蹈、电子购物、社交联网站点、以及它们的组合。远程节点104在这样的情况中被配置为向该集群的每个成员提供如下的推荐,这些推荐不仅关于与该组相对应的兴趣领域而且关于作为该集群的成员的终端用户的其他顶级评价(即高排名)的兴趣领域。在另一实施例中,可以用充当对一个兴趣组负责的中央聚合器的每个用户设备102来移除充当中央聚合器的远程节点104。

[0054] 类似地,在基于内容的推荐应用的情况中,远程节点104可以被实施为一个或多个兴趣组聚合器,并且每个兴趣组聚合器对应于至少一个集群,诸如电影、舞蹈、电子购物、社交联网站点。进一步地,在涉及端到端通信的应用中,远程节点104可以是另一其他用户设备。在这样的应用中,用户设备102可以使用集群标识符来识别具有类似集群标识符的其他用户设备102。

[0055] 本主题的系统和方法因此通过在用户设备处将集群标识符指配给兴趣简档,来促进向终端用户提供隐私性。

[0056] 图2图示了根据本主题的一个实施例的实施对用户简档的隐私保护聚集的用户设备102的组件。在所述实施例中,用户设备102包括一个或多个处理器202、(多个) I/O接口204、以及耦合至(多个)处理器202的存储器206。(多个)处理器202可以被实施为一个或多个微处理器、微计算机、微控制器、数字信号处理器、中央处理单元、状态机、逻辑电路、和/或基于操作指令来操控信号的任何设备。(多个)处理器202能够是单个处理单元或者多个单元,其中的所有单元还可以包括多个计算单元。除了其他的能力之外,(多个)处理器202还被配置为取回并执行存储器206中存储的计算机可读指令。

[0057] 该图中所示出的各种元件的功能,包括被标记为“(多个)处理器”的任何功能块,可以通过使用专属硬件以及能够执行软件的硬件联合适当软件来提供。当由处理器提供时,这些功能可以由单个专属处理器、由单个共享处理器、或者由多个个体处理器(其中的一些可以是共享的)来提供。此外,对术语“处理器”的明确使用不应当被解释为排他地指代能够执行软件的硬件,并且可以隐含地不带限制地包括:数字信号处理器(DSP)硬件、网络处理器、专用集成电路(ASIC)、现场可编程门阵列(FPGA)、用于存储软件的只读存储器(ROM)、随机存取存储器(RAM)、以及非易失性存储器。其他常规的和/或定制的硬件也可以被包括。

[0058] (多个) I/O接口204可以包括各种软件和硬件接口,例如,用于(多个)外围设备的接口,(多个)外围设备诸如键盘、鼠标、外部存储器、打印机等。进一步地,(多个) I/O接口204可以使得用户设备102能够通过通信网络106进行通信,并且可以包括用于将用户设备102与其他计算设备(诸如Web服务器和外部数据库)相连接的一个或多个端口。(多个) I/O接口204可以促进各种各样的协议和网络内的多个通信,这些网络诸如包括有线网络(例如,LAN、电缆等)和无线网络(例如,WLAN、蜂窝、卫星等)的网络。

[0059] 存储器206可以包括本领域中已知的任何计算机可读介质,包括例如,易失性存储器,诸如静态随机存取存储器(SRAM)和动态随机存取存储器(DRAM),和/或非易失性存储器,诸如只读存储器(ROM)、可擦除可编程ROM、闪存、硬盘、光盘、以及磁带。存储器206进一步包括模块208和数据210。模块208包括执行特定任务或者实施特定抽象数据类型的例程、程序、对象、组件、数据结构,等等。除了其他的事物之外,数据210还用于存储由模块208中的一个或多个模块处理、接收、以及生成的数据的储存库。

[0060] 模块208进一步包括兴趣简档生成模块212、简档分段模块214、简档变换模块110、集群标识符模块112、数据传送模块216、以及其他模块218。(多个)其他模块218可以包括补充用户设备102的应用和功能的程序或者编码的指令,例如,操作系统中的程序。

[0061] 数据210包括简档数据220、经变换的简档数据222、集群标识符数据223、以及其他数据226。其他数据226可以包括作为执行(多个)其他模块218中的一个或多个模块的结果而生成的数据。

[0062] 如先前所描述的,用户设备102被配置为通过计算用于聚集用户简档的集群标识符,来向终端用户提供隐私保护。在一个实施例中,兴趣简档生成模块212被配置为,基于简档数据220来生成用户设备102的终端用户的兴趣简档。简档数据220可以指示,例如,由终端用户所访问的网站、由终端用户所播放或下载的歌曲或视频、由终端用户所使用的产品或者所利用或评论的服务,等等。例如,兴趣简档生成模块212可以获得如下的细节:由终端用户所搜索或观看的所有视频、由终端用户所访问的所有URL、由终端用户所收听和搜索的

音乐、以及与由终端用户所探寻的其他兴趣领域有关的类似数据。如此生成的兴趣简档可以被存储在简档数据220中。简档变换模块110随后可以将兴趣简档变换为语义表示以便指配集群标识符。如先前所讨论的,简档变换模块110可以被配置为,将完整的兴趣简档或者兴趣简档分段变换为语义表示。在后者的情况中,兴趣简档在由简档变换模块110变换之前,由简档分段模块214接入并分段。

[0063] 在一个实施例中,简档变换模块110被配置为将兴趣简档分段为多个兴趣简档分段,并且每个兴趣简档分段对应于特定的兴趣领域。在一种实施方式中,简档变换模块110可以分析终端用户的兴趣简档以对兴趣简档进行分段。简档变换模块110可以基于情境偏好(诸如休闲偏好以及职业兴趣)或者兴趣的明确范围(诸如音乐、舞蹈、电子购物、以及社交联网),来将兴趣简档分段为兴趣简档分段。如此获得的兴趣简档分段可以被提供给简档变换模块110。

[0064] 简档变换模块110基于可以在用户设备102的内部或外部的语义元数据数据库114中提供的元数据,来将兴趣简档分段或者完整的兴趣简档变换为对应的语义表示。在一个实施例中,简档变换模块110可以将兴趣简档分段提供给语义元数据数据库114。语义元数据数据库114在这样的情况中可以识别与兴趣简档分段中的每个兴趣简档分段相对应的语义术语,并且将它们提供给简档变换模块110以用于将兴趣简档分段变换为对应的语义表示。在另一实施例中,简档变换模块110可以从语义元数据数据库114获得语义术语的列表。简档变换模块110在这样的情况中可以被配置为,从语义术语的列表中识别与兴趣简档分段相匹配的语义术语,以获得对应的语义表示。

[0065] 在又另一实施例中,简档变换模块110可以通过使用元数据来给兴趣简档分段或者完整的兴趣简档加标签来生成语义表示,该元数据诸如从由终端用户在用户设备102上所访问的网页提取的标签。这样本地提取的标签可以被存储作为用户设备102的其他数据226中的内部语义元数据。在所述实施例中,用户设备被配置为相互共享语义术语的词汇表,从而由不同用户设备生成的语义表示是使用相同术语来表达的并且因此是可比较的。使用语义元数据数据库帮助减少了错误否定和错误肯定的发生,因此增加了聚集的效率。如此获得的语义表示可以被存储在经变换的简档数据222中并且由集群标识符模块112接入,以便将集群标识符指配给语义表示中的每个语义表示以用于进一步的聚集。

[0066] 如图1中所描述的,集群标识符模块112被配置为将集群标识符指配给语义表示中的每个语义表示,以使得能够聚集终端用户的简档。集群标识符模块112使用LSH技术用于将集群标识符指配给语义表示。如将由本领域的技术人员理解的,在LSH技术中,如使用以下等式所表示的,两个类似的对象以高概率散列变换为相同的值,因为两个简档的散列冲突概率与这两个简档之间的相似度成比例:

[0067] $\Pr[h(x) = h(y)] \approx \text{sim}(x, y) \dots\dots\dots (1)$

[0068] 其中 x 和 y 是两个简档, $h()$ 是LSH函数,并且 $\text{sim}()$ 是两个简档之间的相似度测量。进一步地,为了实现LSH技术的功能,要确保用户设备102使用类似的散列函数用于LSH技术。

[0069] 集群标识符模块112被配置为使用由这些散列函数生成的值作为集群标识符,即具有类似兴趣的终端用户(即具有类似兴趣简档的终端用户)的兴趣组的标记或者组id。为了该目的,集群标识符模块112从中央实体(诸如随机种子生成器)初始地获得随机矢量的

集合,也称为散列函数。集群标识符模块112随后使用例如LSH技术的余弦相似度方法。在所述方法中,集群标识符模块112获得语义表示与随机矢量中的每个随机矢量之间的点积并且连结这些点积的符号,以获得散列码作为用于语义表示的集群标识符。如此获得的集群标识符可以被存储在集群标识符数据224中。尽管本文的描述参考了LSH技术的余弦相似度方法,但是如将由本领域的技术人员理解的,可以使用LSH技术的其他实施方式来实施这些方法和系统。通过LSH技术来执行聚集结合通过生成语义表示的简档变换,促进了提供具有聚集中的减少的错误的隐私保护。进一步地,在使用LSH技术来执行聚集之前生成语义表示,帮助确保远程节点104不能执行对集群成员的线性搜索。

[0070] 数据传送模块216随后通过通信信道108将集群标识符传输给远程节点104,例如,用于从服务提供商接收推荐或内容。在一种实施方式中,将集群标识符传送给远程节点104的方法可以取决于针对其来执行聚集的应用而变化。例如,在协同过滤应用的情况下,数据传送模块216可以被配置为,不仅将集群标识符而且将终端用户的兴趣简档传输给远程节点104,以使得该远程节点能够分析该用户简档以便提供组推荐。数据传送模块216在这样的情况下可以被配置为,将兴趣简档切分为更小的简档切片,并且独立地通过不同的匿名通信信道(诸如通信信道108)将更小的简档切片中的每个简档切片传输给远程节点104。不同于简档分段,简档切片被传输,而无需被转换为语义表示以便于使得该远程节点能够分析这些简档切片。通过匿名通信信道作为不同切片来传输用户简档确保了隐私保护,因为远程节点104不可能能够基于兴趣简档的切片来识别终端用户。在协同过滤应用的另一实施例中,远程节点104可以被代替并且每个用户设备(102)可以被配置为充当负责一个兴趣组的中央聚合器。

[0071] 进一步地,在基于内容的推荐应用的情况下,数据传送模块216可以被配置为将集群标识符和语义表示传输给远程节点104,以便于利用由服务提供商供应的推荐或分析服务。类似地,在端到端通信的情况下,数据传送模块216可以被配置为识别具有类似集群标识符的其他用户设备102,以便于使得能够在用户设备102之间进行聊天和交换数据。一经识别其他用户设备102,数据传送模块216就可以被配置为,传输建立其他用户设备之间的匿名消息收发信道以及匿名聊天或搜索广播信道。在描述图3(a)、3(b)和3(c)中的应用的同时,将更详细地解释这些应用中的每个应用中的数据传送模块216的工作。

[0072] 图3(a)、3(b)和3(c)分别图示了根据本主题的一个实施例的用于各种应用的对用户简档的隐私保护聚集的示例性网络环境实施方式300、302和304。

[0073] 图3(a)图示了实施对用户简档的隐私保护聚集以用于向集群的成员提供组推荐的协同过滤应用的网络环境300。在协同过滤应用的所述实施例中,远程节点104被实施为与至少一个集群(诸如电影、舞蹈、电子购物、社交联网站点)相对应的中央聚合器。用户设备102因此可以将集群标识符中的每个集群标识符连同终端用户的兴趣简档一起提供给对应的中央聚合器,以使得该中央聚合器能够在它们的对应兴趣组中的兴趣组的兴趣领域之后,识别最高排名(即首位排名)的兴趣组。

[0074] 在一个实施例中,数据传送模块216被配置为将兴趣简档切分为更小的独立切片,可以通过各个通信信道108来传达这些更小的独立切片,以便确保由远程节点104接收的各个简档分段之间的非关联性。提供该非关联性确保了远程节点104不可能能够识别不同的简档分段是属于相同的用户还是不同的用户。数据传送模块216确保没有终端用户的兴趣

简档的切片独自地包含能够被用来构建完整的兴趣简档并且推断终端用户的身份的足够的简档信息。在另一实施例中,整个兴趣简档可以作为单个切片而被发送给远程节点104。

[0075] 进一步地,由数据传送模块216通过采用确保匿名性的机制的网络(例如,实施洋葱路由的网络),经过匿名通信信道来发送兴趣简档切片中的每个兴趣简档切片和集群标识符。在一种实施方式中,洋葱路由路径被建立,其中数据传送模块216利用洋葱路由路径的出口节点的公共密钥,来加密属于终端用户的简档信息的切片和集群标识符。在到达该出口节点之前,通过一个或多个中间节点来传输属于终端用户的简档信息的各个切片和集群标识符。该出口节点解密该信息并且将该信息传输给远程节点104。

[0076] 进一步地,为了确保用户隐私性,数据传送模块216可以被配置为,从兴趣简档中移除具有可能向远程节点104透露终端用户的身份的个人可识别信息(诸如URL)的所有数据。

[0077] 一经接收到终端用户的兴趣简档和集群标识符,远程节点104就分析兴趣简档以确定与远程节点104相对应的集群中所聚集的终端用户的最高排名的兴趣或者一个或多个高排名的兴趣。例如,对兴趣组“社交联网”负责的远程节点104可以接收具有与集群“社交联网”相对应的集群标识符的所有用户设备的兴趣简档。远程节点104随后可以分析用户设备102的用户简档以识别高排名的其他兴趣。一经将例如“新闻”识别为最高排名的兴趣,远程节点104就可以从对兴趣领域“新闻”负责的中央聚合器获得与“新闻”相对应的推荐。远程节点104因此可以将与社交联网以及新闻有关的内容和推荐提供给存在于与集群“社交联网”相对应的集群中的所有用户设备102。

[0078] 在协同过滤应用的另一实施例中,可以用充当对一个兴趣组负责的中央聚合器的每个用户设备102来移除充当中央聚合器的远程节点104。为了该目的,数据传送模块216被配置为维持终端用户的兴趣组的中央聚合器节点的至少部分图像。在所述实施例中,如对本领域的技术人员将是已知的,可以基于常规的基于流言的随机对等协议来识别具有类似兴趣简档的终端用户。数据传送模块216因此可以被配置为传输和接收所有用户设备102的集群标识符,以识别具有类似集群标识符(即,类似兴趣组)的终端用户。

[0079] 图3(b)图示了实施对用户简档的隐私保护聚集以用于基于内容的推荐应用的网络环境302。在所述实施例中,远程节点104被实施为一个或多个兴趣组聚合器,并且每个兴趣组聚合器对应于至少一个集群,诸如电影、舞蹈、电子购物、社交联网站点。

[0080] 在这样的情况中,数据传送模块216被配置为通过匿名通信信道(诸如通信信道108)将集群标识符和语义表示传输给远程节点104。为了该目的,数据传送模块216建立洋葱路由路径,并且数据传送模块216如图3(a)中描述的那样利用洋葱路由路径的出口节点的公共密钥来加密属于终端用户的集群标识符。远程节点104在这样的情况中被配置为,通过使用任何公知的内容匹配技术将集群标识符与兴趣领域的列表进行匹配,来选择与由该集群标识符表示的兴趣领域有关的推荐或内容。例如,如果终端用户(比如用户A)属于具有兴趣领域“书籍”的集群,则服务提供商可以向用户A提供最新畅销书的列表。

[0081] 图3(c)图示了实施对用户简档的隐私保护聚集以用于与端到端通信有关的应用的网络环境304。在所述实施例中,远程节点104被移除,因为用户设备102使用该聚集用于在隐私保护的环境中彼此进行交互。在这样的应用中,用户设备102可以使用集群标识符来识别具有类似集群标识符的其他用户设备102。在这样的情况中,数据传送模块216被配置

为通过匿名通信信道(诸如通信信道108)将集群标识符传输给其他用户设备102。数据传送模块216使用基于流言的随机对等协议来传输集群标识符,并且识别具有类似集群标识符的用户设备102。具有共同集群标识符的用户设备102随后可以匿名地使用例如洋葱路由框架(诸如TOR)来相互连接,以便于使得能够在用户设备之间进行聊天和交换数据,而不透露它们的身份或人员信息。

[0082] 图4图示了根据本主题的一个实施例的对用户简档的隐私保护聚集以向终端用户提供隐私保护的方法。该方法被描述的顺序不意图被解释为一种限制,并且能够以任何顺序来组合任何数目的所描述的方法框,以实施方法400或者任何替换方法。另外,不偏离本文所描述的主题的精神和范围,可以从这些方法中删除个别的框。此外,能够以任何合适的硬件、软件、固件、或它们的组合来实施这些方法。

[0083] 可以在计算机可执行指令的一般性上下文中描述该(些)方法。一般而言,计算机可执行指令能够包括执行特定功能或者实施特定抽象数据类型的例程、程序、对象、组件、数据结构、过程、模块、功能,等等。该方法还可以被实行在分布式计算环境中,其中功能由通过通信网络链接的远程处理设备来执行。在分布式计算环境中,计算机可执行指令可以位于本地和远程计算机存储介质(包括存储器存储设备)两者中。

[0084] 本领域的技术人员将容易地认识到,这些方法的步骤能够由经编程的计算机来执行。在本文中,一些实施例也意图覆盖程序存储设备,例如,数字数据存储介质,这些程序存储设备是机器或计算机可读的并且对指令的机器可执行或计算机可执行的程序进行编码,其中所述指令执行所描述的方法的步骤中的一些或全部步骤。程序存储设备可以是,例如,数字存储器;磁存储介质,诸如磁盘和磁带;硬盘;或者光可读数字数据存储介质。这些实施例还意图覆盖被配置为执行示例性方法的所述步骤的通信网络和通信设备两者。

[0085] 在框402处,与终端用户相对应的简档数据由用户设备(例如,用户设备102)获得。在一种实施方式中,可以基于由终端用户所访问的网站、由终端用户所播放或下载的歌曲或视频、由终端用户所使用的产品或者所利用或评论的服务等来获得简档数据。兴趣简档生成模块(例如,兴趣简档生成模块112)可以被配置为获得如下的细节:由终端用户所搜索或观看的所有视频、由终端用户所访问的所有URL、由终端用户所进行的所有查询、以及与由终端用户所探寻的其他兴趣领域有关的类似数据。如此生成的这些兴趣简档可以被存储作为简档数据。

[0086] 在框404处,简档数据被分段为一个或多个兴趣简档分段。在一种实施方式中,简档分段模块(诸如简档分段模块214)被配置为将兴趣简档分段为多个分段,称为兴趣简档分段。这些兴趣简档分段可以被理解为兴趣简档的独立分段,并且每个分段对应于特定的兴趣领域。兴趣领域的示例包括,但不限于:情境偏好,诸如职业兴趣、休闲偏好、和宗教偏好,以及兴趣的范围,诸如音乐、舞蹈、电子购物、和社交联网。这些兴趣简档分段中的每个兴趣简档分段随后独立地被处理以获得集群标识符以用于聚集。进一步地,在另一实施例中,兴趣简档作为整体可以被提供作为单个兴趣简档以用于聚集。

[0087] 在框406处,基于与语义元数据数据库的交互来查明与这些兴趣简档分段中的每个兴趣简档分段相对应的语义术语。在一种实施方式中,简档变换模块(诸如简档变换模块110)可以被配置为,通过匿名通信信道将这些兴趣简档分段提供给语义元数据数据库。一经接收到兴趣简档分段,语义元数据数据库就从语义术语的列表中识别与这些兴趣简档分

段中的每个兴趣简档分段相对应的语义术语。替换地,简档变换模块110可以与使用如下的数据来填充的用户设备102的内部语义元数据进行交互,该数据诸如从由终端用户在用户设备102上访问的网页中提取的标签。在所述实施例中,用户设备102被配置为相互共享语义术语的词汇表,以便于维持跨所有用户设备102生成使用的语义术语的一致性。

[0088] 在框408处,兴趣简档分段被变换以基于语义术语来生成对应的语义表示。例如,简档变换模块(诸如简档变换模块110)被配置为,例如通过利用从语义元数据数据库获得的对应语义表示来给兴趣简档分段中的每个兴趣简档分段加标签,以变换这些兴趣简档分段。这些兴趣简档分段可以替换地被变换为基于本体概念或基于主题的语义表示。如此获得的语义表示可以被保存在经变换的简档数据222中。

[0089] 在框410处,使用LSH技术将集群标识符指配给语义表示。在一种实施方式中,由随机值的集合定义的散列函数(诸如共同的随机矢量序列)被用于指配这些集群标识符。例如,集群标识符模块(诸如集群标识符模块112)可以从中央实体(诸如由用户设备可接入的随机种子生成器)获得这些散列函数,以确保一致的集群识别。通过LSH技术来执行聚集结合使用共同的语义表示集合的简档变换,因此促进了提供具有聚集中的减少的错误的隐私保护。

[0090] 在框412处,与兴趣简档分段中的每个兴趣简档分段相对应的集群标识符被提供给远程节点,以便基于这些集群标识符来聚集这些用户兴趣简档。在一种实施方式中,远程节点可以是中央实体,并且可以取决于针对其来执行该聚集的应用的类型而从中央聚合器节点变化到兴趣聚合器节点。例如,在基于内容的推荐应用的情况下,远程节点可以被实施为兴趣组聚合器。用户设备在这样的情况下被配置为,通过匿名通信信道,将集群标识符中的每个集群标识符和语义表示传输给与这些集群标识符相对应的兴趣组聚合器。

[0091] 类似地,在协同过滤应用的情况下,这些远程节点可以被实施为中央聚合器,这些中央聚合器被配置为向该集群的每个成员提供如下的推荐,这些推荐不仅关于与该组相对应的兴趣领域而且关于该集群的终端用户的其他高排名的兴趣领域。用户设备在这样的情况下被配置为,通过匿名通信信道,连同终端用户的兴趣简档的切片一起传输集群标识符中的每个集群标识符,以使得这些用户设备向与这些集群标识符相对应的兴趣组聚合器。在另一实施例中,用户设备可以与其他用户设备进行交互,从而用户设备中的每个用户设备充当用于其他用户设备的远程节点或者负责一个兴趣组的中央聚合器。

[0092] 尽管已经在特定于结构特征和/或方法的语言中描述了在通信网络中用于对用户兴趣简档的隐私保护聚集的方法和系统的实施例,但是将理解,本发明不一定被限制为所描述的具体特征或方法。更确切地,这些具体特征和方法被公开作为用于提供对用户兴趣简档的隐私保护聚集的示例性实施例。

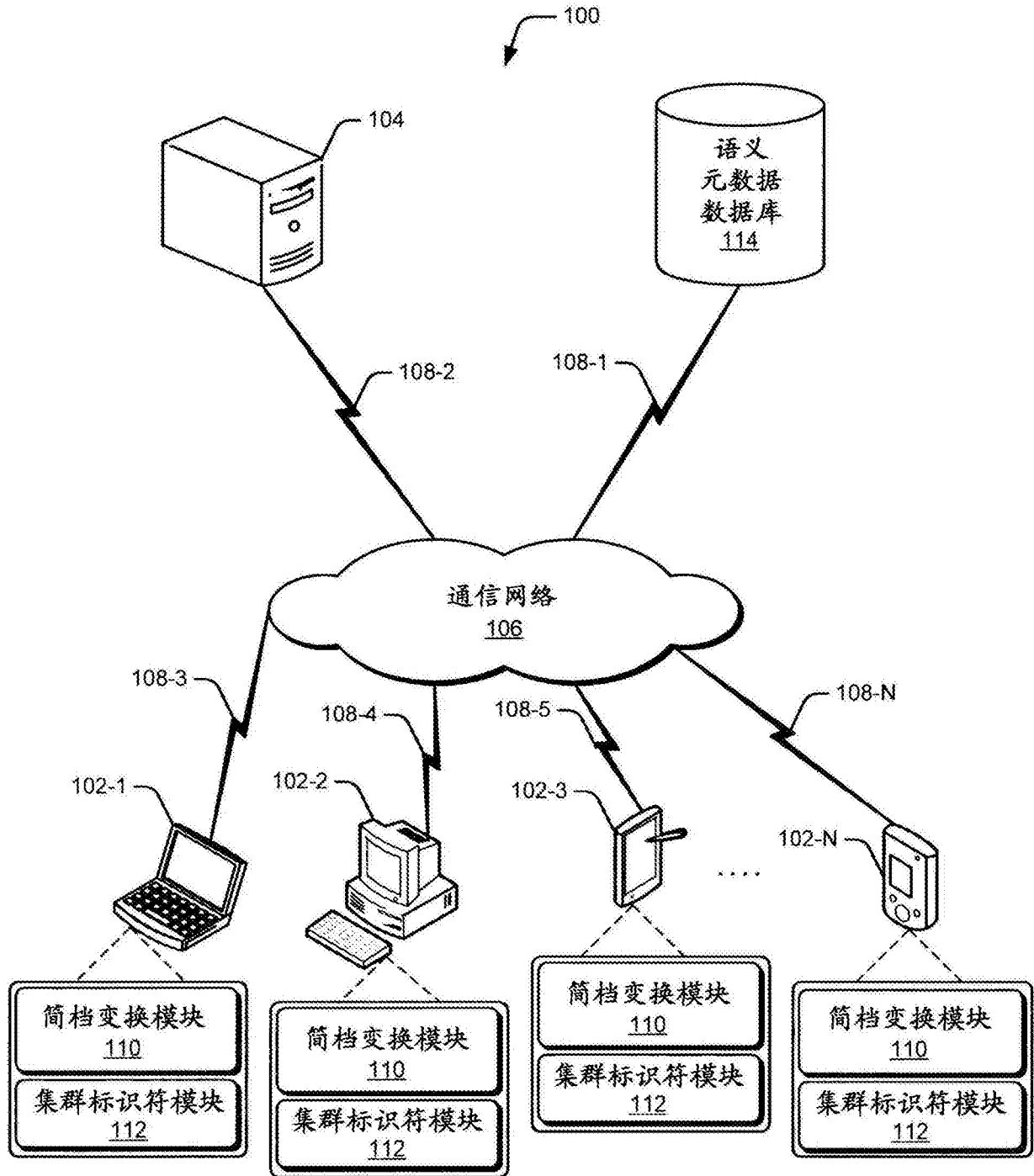


图1



图2

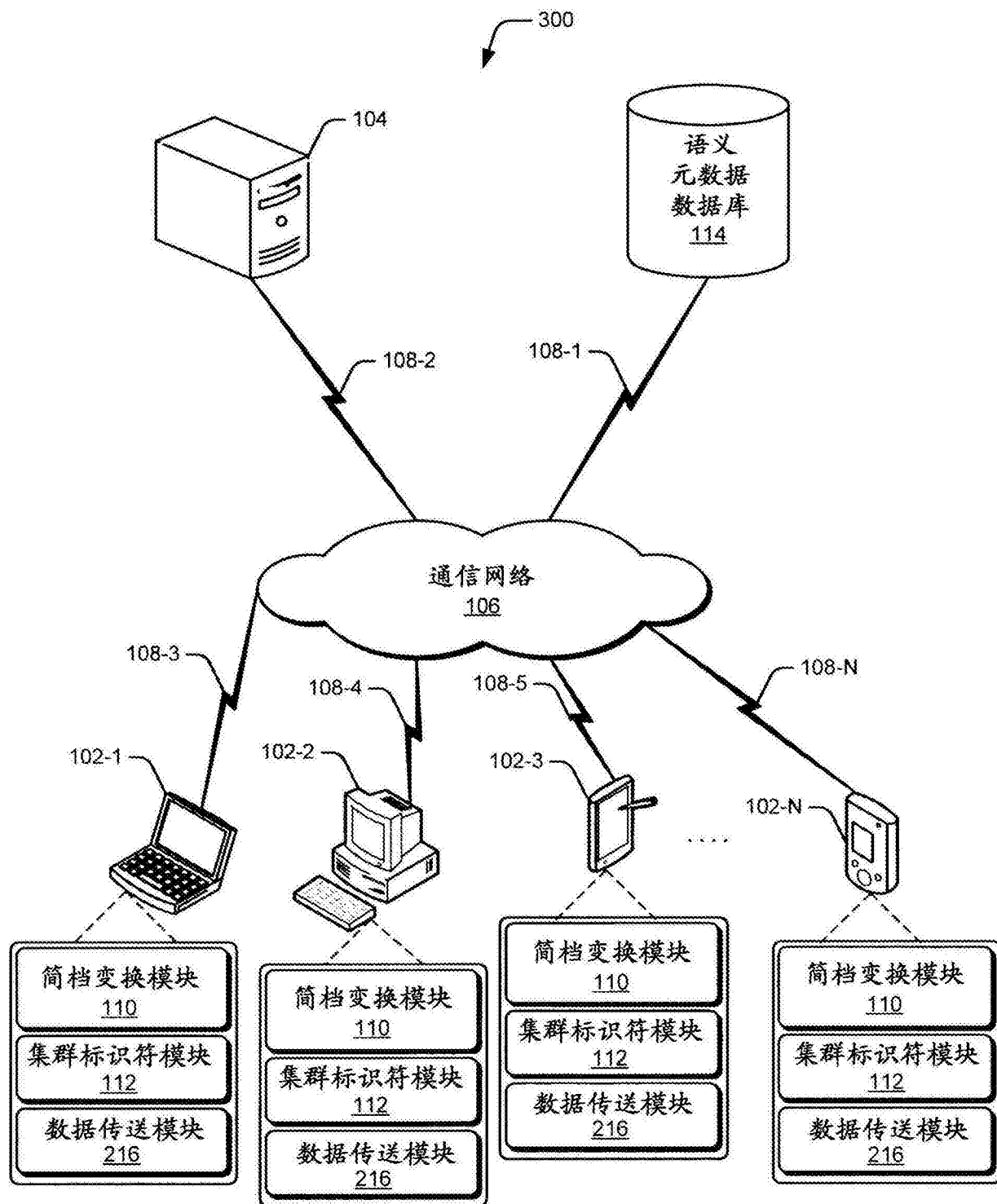


图3 (a)

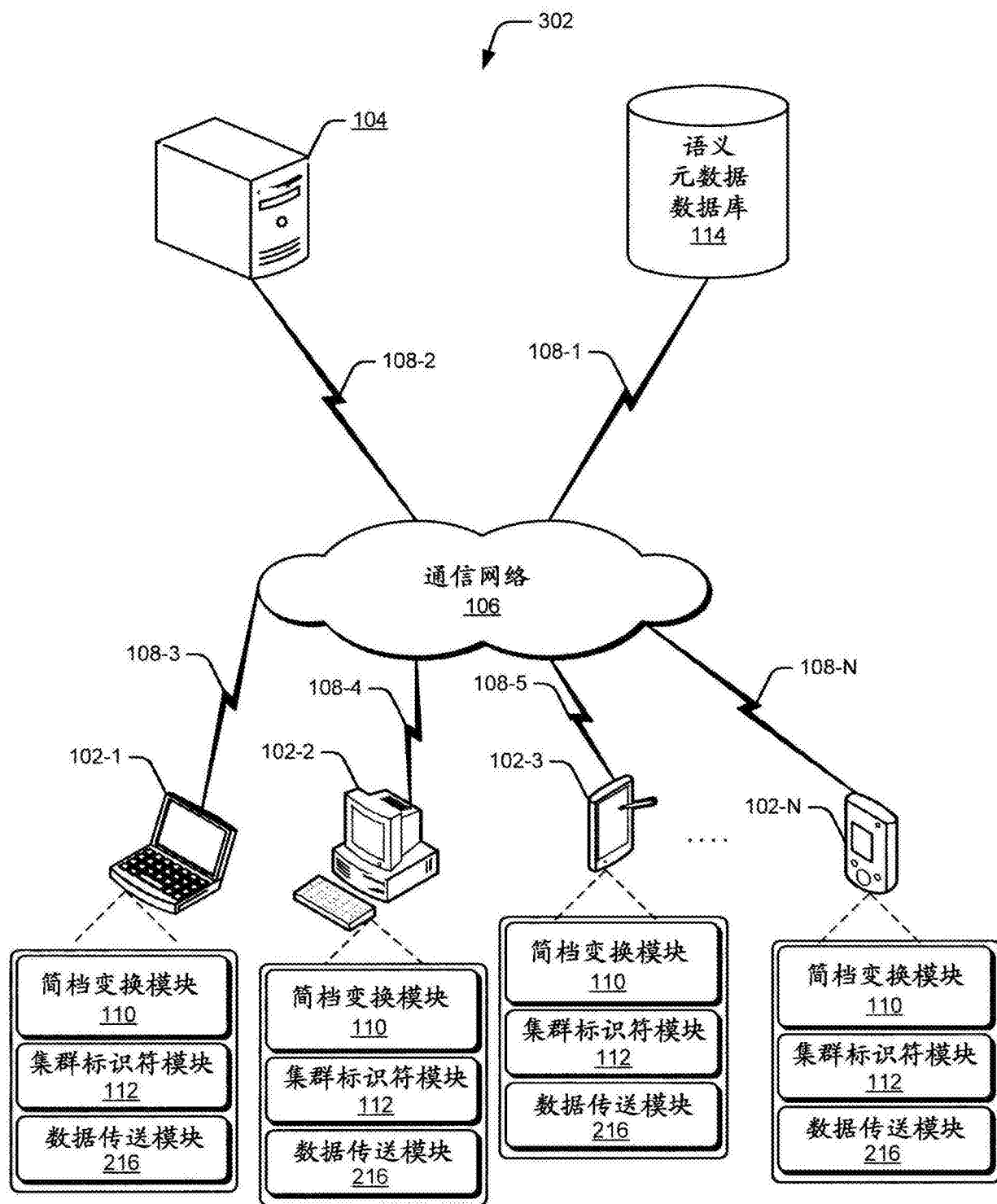


图3 (b)

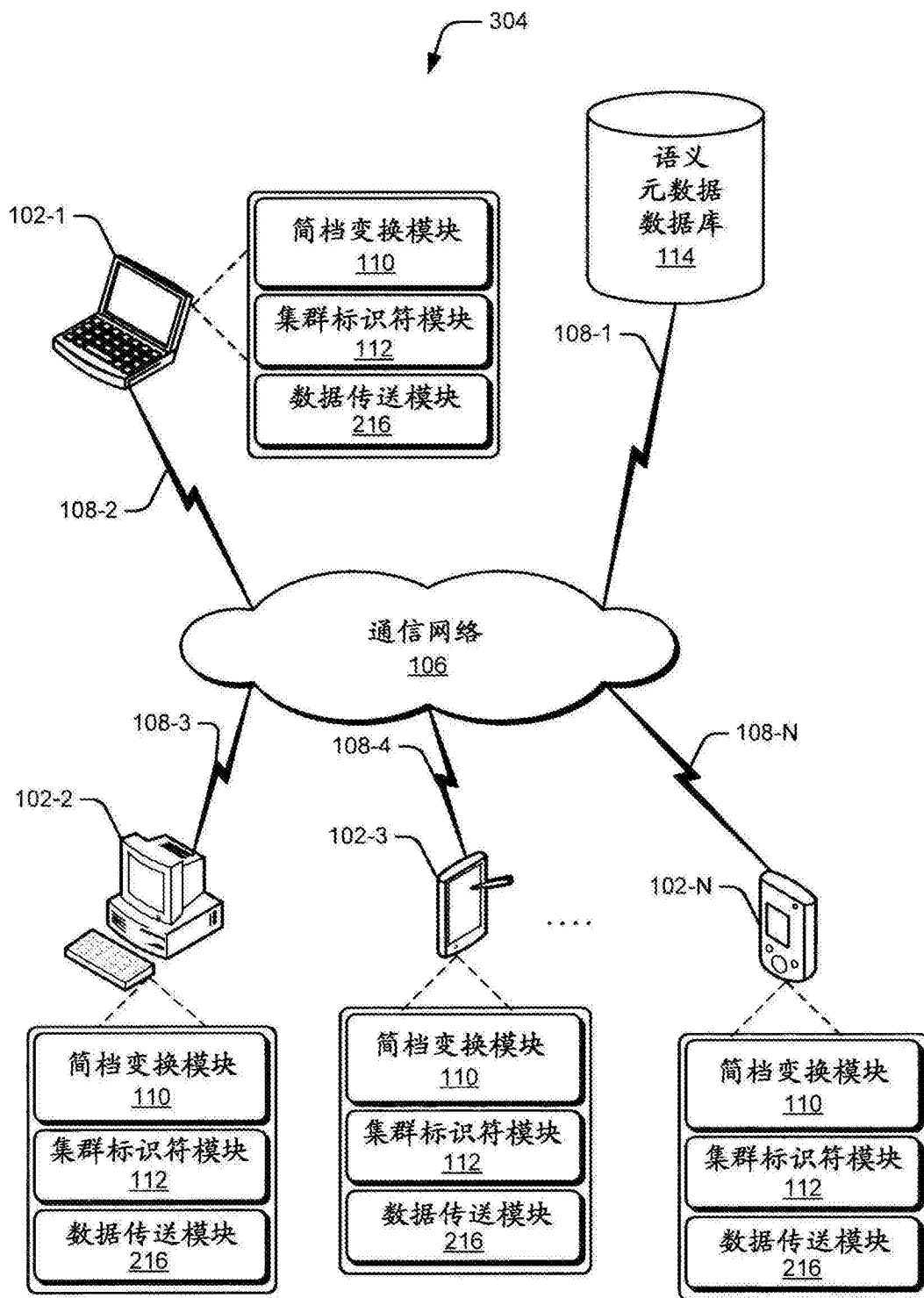


图3(c)

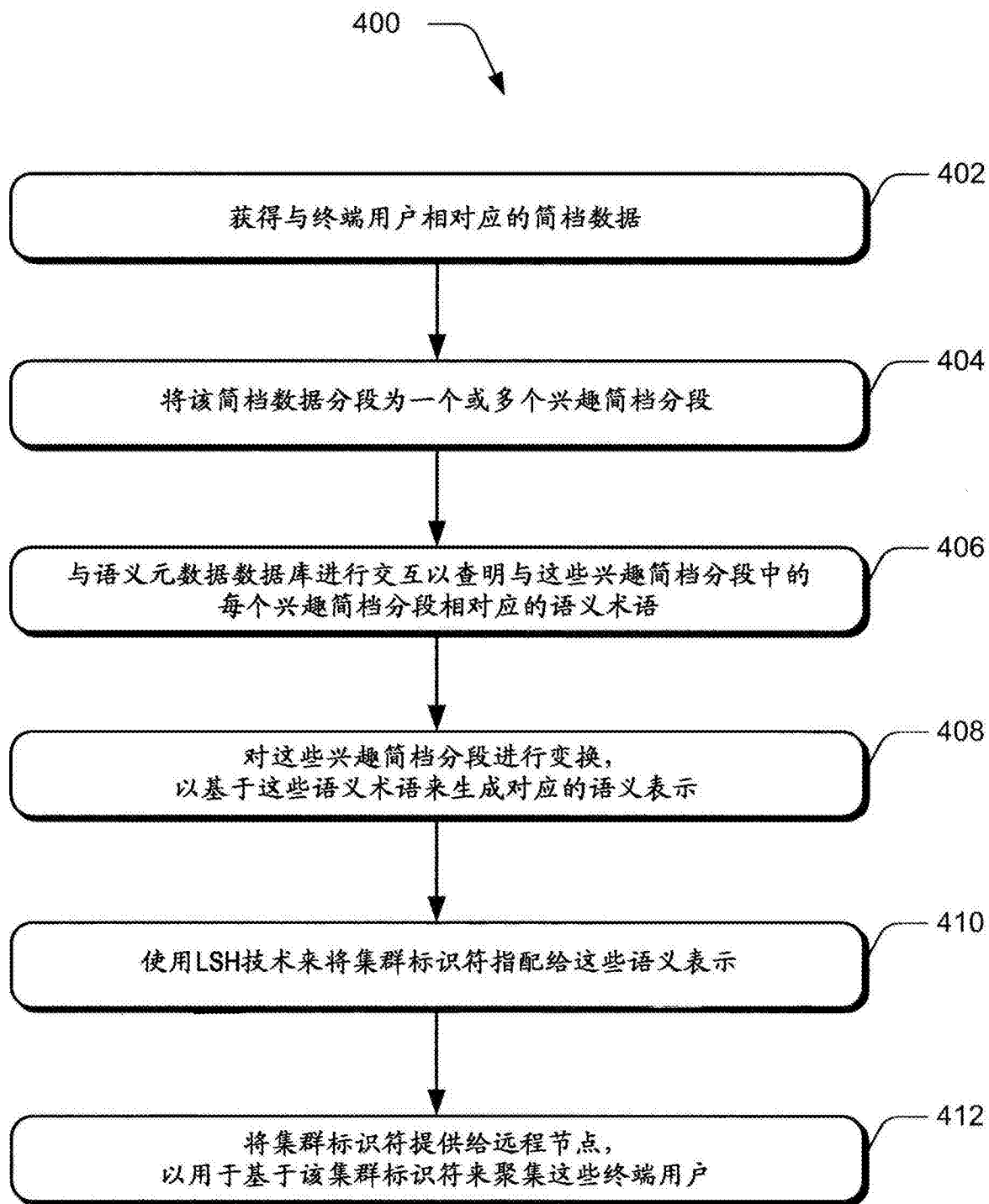


图4