



(12) 发明专利申请

(10) 申请公布号 CN 112384920 A

(43) 申请公布日 2021.02.19

(21) 申请号 201980046244.4

(74) 专利代理机构 上海专利商标事务所有限公司 31100

(22) 申请日 2019.07.11

代理人 徐倩 钱慰民

(30) 优先权数据

62/696,480 2018.07.11 US

(51) Int.Cl.

G06F 21/60 (2006.01)

(85) PCT国际申请进入国家阶段日

G06N 5/04 (2006.01)

2021.01.08

G06N 20/10 (2006.01)

(86) PCT国际申请的申请数据

G06Q 20/38 (2006.01)

PCT/US2019/041318 2019.07.11

(87) PCT国际申请的公布数据

W02020/014425 EN 2020.01.16

(71) 申请人 维萨国际服务协会

地址 美国加利福尼亚州

(72) 发明人 T·哈里斯 T·科罗列夫斯卡亚

李玥

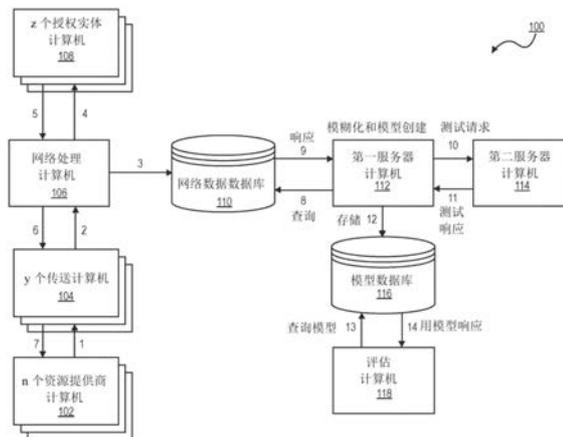
权利要求书3页 说明书24页 附图8页

(54) 发明名称

具有自动模糊变量检测功能的保护隐私的图压缩

(57) 摘要

所公开的方法包括:a)由服务器计算机接收包括多个交易的多个交易数据的网络数据。每个交易数据包括具有数据值的多个数据元素。所述多个数据元素中的至少一个包括用户的用户标识符。然后,所述服务器计算机可:b)基于所述网络数据生成包括多个社区的一个或多个图。所述服务器计算机可:c)针对所述多个交易中的每个交易确定所述数据值中的至少一些的模糊值。对于每个用户,所述服务器计算机可:d)确定所述多个社区内的社区的模糊值。然后,所述服务器计算机可:e)使用步骤c)和d)中获得的所述模糊值以及所述数据值中的至少一些来生成模型。



1. 一种方法,包括:

a) 由服务器计算机接收包括多个交易的多个交易数据的网络数据,其中每个交易数据包括具有数据值的多个数据元素,其中所述多个数据元素中的至少一个包括用户的用户标识符;

b) 由所述服务器计算机基于所述网络数据生成包括多个社区的一个或多个图;

c) 由所述服务器计算机针对所述多个交易中的每个交易确定所述数据值中的至少一些的模糊值;

d) 对于每个用户,由所述服务器计算机确定所述多个社区内的社区的模糊值;以及

e) 由所述服务器计算机使用步骤c) 和d) 中获得的所述模糊值以及所述数据值中的至少一些来生成模型。

2. 根据权利要求1所述的方法,还包括:

由所述服务器计算机接收来自评估计算机的请求,其中所述请求包括请求数据;

由所述服务器计算机确定由所述模型输出的输出数据,其中所述请求数据被输入到所述模型中;以及

由所述服务器计算机将所述输出数据提供给所述评估计算机。

3. 根据权利要求1所述的方法,其中所述数据元素包括所述用户标识符、资源提供商标识符、一年中的一天、一天中的时间、位置、IP地址和金额中的至少一个或多个。

4. 根据权利要求3所述的方法,其中所述数据元素包括所述一天中的时间,其中社区的所述模糊值是模糊社区值,并且其中对应于所述一天中的时间的所述数据值中的至少一些的所述模糊值是模糊事件值。

5. 根据权利要求4所述的方法,还包括:

由所述服务器计算机创建至少包括所述模糊事件值和所述模糊社区值的数据表。

6. 根据权利要求5所述的方法,其中创建所述数据表还包括:

由所述服务器计算机将所述多个交易数据中的每个交易数据的所述模糊事件值和所述模糊社区值存储到所述数据表中;以及

由所述服务器计算机移除所述多个交易数据中的每个交易数据的所述用户标识符。

7. 根据权利要求6所述的方法,还包括:

由所述服务器计算机将所述多个数据元素的子集存储到所述数据表中。

8. 根据权利要求1所述的方法,其中确定所述数据值中的至少一些的模糊值还包括:

由所述服务器计算机为对应于所述数据值中的所述至少一些的每个数据元素确定隶属函数集;以及

由所述服务器计算机使用所述隶属函数集确定所述数据值中的所述至少一些的所述模糊值。

9. 根据权利要求1所述的方法,其中所述模型是支持向量机、人工神经网络、决策树、贝叶斯网络和遗传算法中的一个。

10. 根据权利要求1所述的方法,其中在生成所述模型之后,所述方法还包括:

由第一服务器计算机向第二服务器计算机传输所述模型,其中所述第二服务器计算机确定外部数据和所述模型是否能用于标识至少一个用户。

11. 一种服务器计算机,包括:

处理器；

存储器；以及

计算机可读介质，其耦合到所述处理器，所述计算机可读介质包括能由所述处理器执行以实施方法的代码，所述方法包括：

a) 接收包括多个交易的多个交易数据的网络数据，其中每个交易数据包括具有数据值的多个数据元素，其中所述多个数据元素中的至少一个包括用户的用户标识符；

b) 基于所述网络数据生成包括多个社区的一个或多个图；

c) 针对所述多个交易中的每个交易确定所述数据值中的至少一些的模糊值；

d) 对于每个用户，确定所述多个社区内的社区的模糊值；以及

e) 使用步骤c) 和d) 中获得的所述模糊值以及所述数据值中的至少一些来生成模型。

12. 根据权利要求11所述的服务器计算机，其中所述方法还包括：

由所述服务器计算机接收来自评估计算机的请求，其中所述请求包括请求数据；

由所述服务器计算机确定由所述模型输出的输出数据，其中所述请求数据被输入到所述模型中；以及

由所述服务器计算机将所述输出数据提供给所述评估计算机。

13. 根据权利要求11所述的服务器计算机，其中所述数据元素包括所述用户标识符、资源提供商标识符、一年中的一天、一天中的时间、位置、IP地址和金额中的至少一个。

14. 根据权利要求13所述的服务器计算机，其中所述数据元素包括所述一天中的时间，其中社区的所述模糊值是模糊社区值，并且其中对应于所述一天中的时间的所述数据值中的至少一些的所述模糊值是模糊事件值。

15. 根据权利要求14所述的服务器计算机，其中所述方法还包括：

创建至少包括所述模糊事件值和所述模糊社区值的数据表。

16. 根据权利要求15所述的服务器计算机，其中创建所述数据表还包括：

将所述多个交易数据中的每个交易数据的所述模糊事件值和所述模糊社区值存储到所述数据表中；以及

由所述服务器计算机移除所述多个交易数据中的每个交易数据的所述用户标识符。

17. 根据权利要求16所述的服务器计算机，还包括：

将所述多个数据元素的子集存储到所述数据表中。

18. 根据权利要求11所述的服务器计算机，其中确定所述数据值中的至少一些的模糊值还包括：

由所述服务器计算机为对应于所述数据值中的所述至少一些的每个数据元素确定隶属函数集；以及

由所述服务器计算机使用所述隶属函数集确定所述数据值中的所述至少一些的所述模糊值。

19. 根据权利要求11所述的服务器计算机，其中所述模型是支持向量机、人工神经网络、决策树、贝叶斯网络和遗传算法中的一个。

20. 根据权利要求11所述的服务器计算机，其中在生成所述模型之后，所述方法还包括：

由第一服务器计算机向第二服务器计算机传输所述模型，其中所述第二服务器计算机

确定外部数据和所述模型是否能用于标识至少一个用户。

具有自动模糊变量检测功能的保护隐私的图压缩

[0001] 相关申请交叉引用

[0002] 本申请要求2018年7月11日提交的第62/696,480号美国临时申请的权益,所述美国临时申请出于所有目的以全文引用的方式并入本文中。

背景技术

[0003] 如今,收集、分析和管理大量信息的能力已经广泛普及。机器学习技术广泛用于产生用于医药、金融、推荐服务、威胁分析和认证技术的预测模型。长时间收集的大量数据支持用新的解决方案解决旧问题,深度学习的进步使得在语言、图像和文本识别方面有所突破。大型互联网公司收集用户的在线活动,以便对预测他们未来兴趣的推荐系统进行训练。来自不同医院和政府组织的健康数据可用于产生新的诊断模型,而金融公司和支付网络可以组合交易历史记录、商家数据以及账户持有人信息来训练更精确的欺诈检测引擎。但是,数据隐私是一个问题。随着许多国家对隐私的关注日益增加,需要新的改进方法来保护机器学习系统和方法中使用的数据。

[0004] 本发明的实施例单独地以及共同地解决这些问题和其它问题。

发明内容

[0005] 本发明的一些实施例涉及一种方法,包括:a)由服务器计算机接收包括多个交易的多个交易数据的网络数据,其中每个交易数据包括具有数据值的多个数据元素,其中所述多个数据元素中的至少一个包括用户的用户标识符;b)由所述服务器计算机基于所述网络数据生成包括多个社区的一个或多个图;c)由所述服务器计算机针对所述多个交易中的每个交易确定所述数据值中的至少一些的模糊值;d)对于每个用户,由所述服务器计算机确定所述多个社区内的社区的模糊值;以及e)由所述服务器计算机使用步骤c)和d)中获得的所述模糊值以及所述数据值中的至少一些来生成模型。

[0006] 本发明的另一实施例涉及一种服务器计算机,包括:处理器;存储器;以及耦合到所述处理器的计算机可读介质,所述计算机可读介质包括可由所述处理器执行以实施方法的代码,所述方法包括:a)接收包括多个交易的多个交易数据的网络数据,其中每个交易数据包括具有数据值的多个数据元素,其中所述多个数据元素中的至少一个包括用户的用户标识符;b)基于所述网络数据生成包括多个社区的一个或多个图;c)针对所述多个交易中的每个交易确定所述数据值中的至少一些的模糊值;d)对于每个用户,确定所述多个社区内的社区的模糊值;以及e)使用步骤c)和d)中获得的所述模糊值以及所述数据值中的至少一些来生成模型。

[0007] 有关本发明的实施例的更多详细信息,请参见具体实施方式和附图。

附图说明

[0008] 图1示出根据实施例的隐私保护系统的框图。

[0009] 图2示出根据实施例的服务器计算机的组件的框图。

- [0010] 图3示出根据实施例的图,所述图示出了用于确定模糊值的示例隶属函数和标签。
- [0011] 图4示出根据实施例的社区检测和模糊化的流程图。
- [0012] 图5示出根据实施例的使交互数据模糊化的流程图。
- [0013] 图6示出根据实施例的进化学习器数据模糊化方法的流程图。
- [0014] 图7示出根据实施例的隶属函数确定方法的流程图。
- [0015] 图8示出根据实施例的框图,所述框图示出了历经进化学习器的3个时期改变的隶属函数。

具体实施方式

[0016] 在论述本发明的实施例之前,可更详细地描述一些术语。

[0017] 术语“人工智能模型”或“AI模型”可以包括可用于预测结果以实现预定义目标的模型。可以使用学习算法来开发AI模型,其中基于已知或推断的模式对训练数据进行分类。AI模型也可以被称为“机器学习模型”或“预测模型”。

[0018] “对抗性AI”可以包括一种试图通过恶意输入来欺骗模型的技术。在一些实施例中,对抗性AI可以确定通过使网络数据模糊化而创建的模糊数据是否是保护隐私的。

[0019] “机器学习”可以包括人工智能过程,其中可以训练软件应用程序以通过学习做出准确预测。可以通过将输入数据应用于通过对汇总数据执行统计分析而形成的预测模型来生成预测。可以使用训练数据来训练模型,使得可以使用所述模型进行准确预测。预测可以是例如图像的分类(例如标识互联网上的猫的图像),或作为另一示例,可以是建议(例如用户可能喜欢的电影或消费者可能喜欢的餐馆)。

[0020] 在一些实施例中,模型可以是统计模型,其可以用于根据已知信息预测未知信息。例如,学习模块可以是用于根据训练数据生成回归线(监督学习)的指令集,或用于基于相似性、连接性和/或数据点之间的距离将数据分组成数据的不同分类的群集(无监督学习)的指令集。接着,回归线或数据群集可作用于根据已知信息预测未知信息的模型。一旦根据学习模块构建了模型,就可使用模型根据新请求生成预测输出。新请求可以是对与呈现的数据相关联的预测的请求。例如,新请求可以是对将图像进行分类或为用户创建推荐的请求。

[0021] “拓扑图”可以包括由边缘连接的不同顶点的平面中的图的表示。拓扑图中的不同顶点可以被称为“节点”。每个节点可以表示事件的特定信息,或者可以表示实体或对象的简档的特定信息。可以由一组边缘E使节点彼此相关。“边缘”可以被描述为由作为图 $G=(V, E)$ 的子集的两个节点构成的无序对,其中G是包括由一组边缘E连接的一组顶点(节点)V的图。例如,拓扑图可以表示交易网络,在所述交易网络中,表示交易的节点可以由边缘连接到与交易相关的一个或多个节点,例如表示装置、用户、交易类型等的信息的节点。边缘可与被称为“权重”的数值相关联,所述数值可被指派给两个节点之间的成对连接。边缘权重可以被标识为两个节点之间的连接强度和/或可以与成本或距离相关,因为它常常表示从一个节点移动到下一个节点所需的量。

[0022] “子图(subgraph或sub-graph)”可以包括由较大图的元素的子集形成的图。元素可包括顶点和连接边缘,并且子集可以是在较大图的节点和边缘的整个集合当中选择的节点和边缘的集合。例如,可以通过对图数据随机采样来形成多个子图,其中每个随机样本都

可以是子图。每个子图可以与由相同的较大图形成的另一个子图重叠。

[0023] “社区”可以包括图中的在组内密集连接的一组节点。社区可以是子图或其部分/衍生物,且子图可以是也可以不是社区和/或包括一个或多个社区。可以使用图学习算法从图标识社区,所述图学习算法例如是用于绘制蛋白质复合物的图学习算法。使用历史数据标识出的社区可以用于对新数据进行分类,以进行预测。例如,标识社区可以用作机器学习过程的一部分,在此部分中可以基于信息元素彼此的关系来对信息元素进行预测。

[0024] 术语“节点”可以包括表示指定信息的离散数据点。节点可以通过边缘在拓扑图中彼此连接,所述边缘可以被分配被称为边缘权重的值,以便描述两个节点之间的连接强度。例如,第一节点可以是表示网络中的第一装置的数据点,且第一节点可以在图中连接到表示网络中的第二装置的第二节点。连接强度可以由边缘权重定义,所述边缘权重对应于可以在两个节点之间快速且容易地传输信息的程度。边缘权重还可以用于表达从一个状态或节点移动到下一个状态或节点所需的成本或距离。例如,第一节点可以是表示机器的第一位置的数据点,且第一节点可以在图中连接到用于机器的第二位置的第二节点。边缘权重可以从第一位置移动到第二位置所需的能量。

[0025] 术语“解算器”可以包括搜索解的计算组件。例如,可以使用一个或多个解算器来计算优化问题的解。解算器可以另外被称为“代理”。一起运行以例如在蚁群优化的情况下求解给定问题的多个代理可以被称为“群体”。

[0026] 术语“时期”可以包括一段时间。例如,时期可以是训练机器学习模型的迭代的一段时间。在学习器训练学习算法的过程中,每个时期可以在定义的一组步骤已完成通过。例如,在蚁群优化中,每个时期可以在所有计算代理找到解并计算出其解的成本之后通过。在迭代算法中,时期可以包括更新模型的一个迭代或多个迭代。时期有时可能被称为“周期”。

[0027] “网络数据”可以包括与互联人员和/或事情的组和/或系统相关的数据。在一些实施例中,网络数据可以包括多个交易的多个交易数据。每个交易数据可以包括具有数据值的多个数据元素。在一些实施例中,多个数据元素中的至少一个包括用户的用户标识符。

[0028] “数据元素”可以包括具有精确含义或精确语义的数据单元。例如,数据元素可以包括邮政编码、商家标识符、用户标识符、金额、IP地址、日期、时间等。数据元素可以与数据值相关联,例如94016、商家_1234、用户_1234、\$19.99、111.111.11.111、2015年1月1日、西部时间上午11:00点分别对应于数据元素。

[0029] “交互”可以包括互惠作用或影响。“交互”可以包括各方、各装置和/或各实体之间的通信、联系或交换。示例交互包括两方之间的交易和两个装置之间的数据交换。在一些实施例中,交互可以包括用户请求访问安全数据、安全网页、安全位置等。在其它实施例中,交互可以包括支付交易,在所述支付交易中,两个装置可以交互以促进支付。

[0030] “交互数据”可以包括与交互相关的数据和/或在交互期间记录的数据。在一些实施例中,交互数据可以是网络数据的交易数据。交易数据可以包括具有数据值的多个数据元素。

[0031] “用户”可以包括个体。在一些实施例中,用户可与一个或多个个人账户和/或移动装置相关联。在一些实施例中,用户也可被称为持卡人、账户持有人或消费者。

[0032] “用户标识符”可以包括任何一条能够标识用户的数据。用户标识符可以包括任何

合适的字母数字字符串。在一些实施例中，用户标识符可以衍生自用户标识信息。在一些实施例中，用户标识符可以包括与用户关联的账户标识符。

[0033] “账户标识符”可以包括账户的任何合适的标签。账户标识符可以是任何合适的形式，并且可以包括任何合适类型的字符。账户标识符的示例包括PAN(主账号)、令牌、例如CVV的验证值(卡验证值)等。

[0034] “事件”可以包括发生的事情。在一些实施例中，事件可以在某一时间点发生，例如，事件可以包括可能在“晚上7点”发生的交互。在其它实施例中，事件可以在一段时间内发生，例如，事件可以包括可能从“晚上7点”到“晚上10点”进行的音乐会。事件可以包括交互的时间(例如，交易的时间、访问安全地点、在各方之间传送数据等)。

[0035] “资源提供商”可以是可提供资源的实体。资源可包括例如商品、服务、信息、访问等。资源提供商的示例可以包括商家、数据提供商，例如政府机构、交通机构等。“商家”通常可以是参与交易且可出售商品或服务或提供对商品或服务的取用的实体。

[0036] “访问装置”可以是用于提供对外部计算机系统的访问的任何合适的装置。访问装置可以呈任何合适的形式。访问装置的一些示例包括销售点(POS)装置、蜂窝电话、PDA、个人计算机(PC)、平板PC、手持式专用阅读器、机顶盒、电子收款机(ECR)、自动柜员机(ATM)、虚拟收款机(VCR)、查询一体机、安全系统、访问系统、网站等。访问装置可以使用任何合适的接触式或非接触式操作模式，以发送或接收来自便携式通信装置或与便携式通信装置相关联的数据。在访问装置可以包括POS终端的一些实施例中，可以使用任何合适的POS终端且其可以包括读取器、处理器和计算机可读介质。读取器可以包括任何合适的接触式或非接触式操作模式。例如，示例性读卡器可以包括与便携式通信装置、用户装置等交互的射频(RF)天线、光学扫描器、条形码读取器或磁条读取器。

[0037] “收单方”通常可以是与特定商家或其它实体具有业务关系的业务实体(例如，商业银行)。一些实体可以执行发行方功能和收单方功能两者。一些实施例可涵盖此类单个实体发行方-收单方。收单方可操作收单方计算机，其一般也可被称为“传送计算机”。

[0038] “授权请求消息”可以是请求对交互授权的电子消息。在一些实施例中，所述消息被发送到交易处理计算机和/或支付卡的发行方，以请求对交易进行授权。根据一些实施例，授权请求消息可遵守国际标准化组织(ISO) 8583，它是针对交换与用户使用支付装置或支付账户进行的支付相关联的电子交易信息的系统的标准。授权请求消息可以包括可以与支付装置或支付账户相关联的发行方账户标识符。授权请求消息还可以包括与“标识信息”对应的额外数据元素，包括(只作为示例)：服务代码、卡验证值(CVV)、动态卡验证值(dCVV)、主账户号码或“账户号码”(PAN)、支付令牌、用户名、到期日期等等。授权请求消息还可以包括“交易信息”，例如与当前交易相关联的任何信息，例如交易值、商家标识符、商家位置、收单方银行标识号(BIN)、卡片接受器ID、标识正购买的项目的信息等，以及可以用于确定是否标识和/或授权交易的任何其它信息。

[0039] “授权响应消息”可以是响应于授权请求的消息。在一些情况下，授权响应消息可以由发行金融机构或交易处理计算机生成的对授权请求消息的电子消息应答。仅借助于示例，授权响应消息可以包括以下状态指示符中的一个或多个：批准--交易被批准；拒绝--交易未被批准；或呼叫中心--响应未决的更多信息，商家必须呼叫免费授权电话号码。授权响应消息还可以包括授权代码，所述授权代码可以是信用卡发行银行响应于电子消息中的

授权请求消息(直接地或者通过交易处理计算机)返回给商家的访问装置(例如POS设备)的指示对交易的批准的代码。所述代码可充当授权的证据。

[0040] “授权实体”可以是授权请求的实体。授权实体的示例可以是发行方、政府机构、文件存储库、访问管理员等。授权实体可以操作授权实体计算机。“发行方”可以指发行并任选地维护用户账户的商业实体(例如,银行)。发行方还可以向消费者,或在一些实施例中向便携式装置发行存储在用户装置上的支付凭证,所述用户装置例如蜂窝式电话、智能卡、平板计算机或笔记本电脑。

[0041] “模糊化”可以包括域转换,其中可以将明确值转换为模糊值。明确值可以是由传感器测量的精确值,例如温度(例如80°F)、压力(例如1atm)、rpm(例如60rpm)等。在一些实施例中,每种类型的明确输入可以具有其可转换至的自身所属的一组隶属函数。这组隶属函数可存在于保留明确输入可能拥有的所有相关值的论域内。

[0042] “明确值”可以包括不同的值或输入。在一些实施例中,明确值可以包括测量参数(例如,6伏特、下午12点半等)。明确值可以包括任何合适的值。例如,明确值可以是5、10、17.26、5011等。

[0043] “模糊值”可以包括明确值与隶属函数兼容的程度。例如,模糊值可以包括0到1的值。模糊值也可以称为隶属度。在一些实施例中,明确值可以被模糊化为一个或多个模糊值。例如,可以使用与“早晨”和“中午”标签相关联的两个部分重叠的隶属函数来转换“上午11点”的明确值。明确值可以转换为两个模糊值,例如,模糊值0.4对应于“早晨”的标签,模糊值为0.6对应于“中午”标签。在一些实施例中,模糊值可以包括数据值的模糊值以及社区的模糊值。

[0044] “隶属函数”可以通过将明确值从其域映射到与隶属度相关联的集来定义模糊集。

[0045] “标签”可以包括用于标识隶属函数的描述性名称。标签可以包括任何合适的描述符。例如,标签可以包括“热”、“暖”、“凉”、“冷”、“事件1”、“事件2”,以及描述明确值可以转换到的模糊值的任何其它合适的描述符。

[0046] “范围”或“域”可以包括隶属函数的宽度。例如,范围可以包括一系列明确值(例如5-10),这些明确值的对应(例如,经由隶属函数)模糊值为非零。

[0047] “论域”可以包括明确值可能包括的一系列可能值。例如,对应于与“一天中的时间”相关的明确值的论域可以包括00:00至23:59的值。

[0048] “服务器计算机”可以包括功能强大的计算机或计算机群集。例如,服务器计算机可以是大型主机、小型计算机群集或像单元一样工作的一组服务器。在一个示例中,服务器计算机可以是耦合到网络服务器的数据库服务器。服务器计算机可以耦合到数据库,并且可包括用于服务于来自一个或多个客户端计算机的请求的任何硬件、软件、其它逻辑或前述内容的组合。服务器计算机可以包括一个或多个计算设备,且可以使用多种计算结构、布置和编译中的任一种来服务来自一个或多个客户端计算机的请求。

[0049] “处理器”可以指任何合适的一个或多个数据计算装置。处理器可以包括一起工作以实现所要功能的一个或多个微处理器。处理器可以包括CPU,所述CPU包括至少一个高速数据处理器,所述高速数据处理器足以执行用于执行用户和/或系统生成的请求的程序组件。CPU可以是微处理器,例如AMD的Athlon、Duron和/或Opteron;IBM和/或Motorola的PowerPC;IBM和Sony的Cell处理器;Intel的Celeron、Itanium、Pentium、Xeon和/或XScale;

和/或类似处理器。

[0050] “存储器”可以是可存储电子数据的任何合适的一个或多个装置。合适的存储器可包括非瞬态计算机可读介质,其存储可由处理器执行以实现所要方法的指令。存储器的示例可包括一个或多个存储器芯片、磁盘驱动器,等等。此类存储器可使用任何合适的电气、光学和/或磁性操作模式来操作。

[0051] I. 系统

[0052] 根据实施例,网络数据和/或其衍生物可以由服务器计算机转换为模糊数据,使得转换消除了在网络数据中标识特定个体(即,用户)的数据的能力,即使与外部数据耦合时也是如此。即使拥有关于个体的大量个人数据,恶意方也很难重新映射模糊数据以标识特定个体。当然,模糊数据可以保留有助于创建预测模型的预测能力。

[0053] 图1示出根据本发明的实施例的系统100的框图。系统100包括n个资源提供商计算机102、y个传送计算机104、网络处理计算机106、z个授权实体计算机108、网络数据数据库110、第一服务器计算机112、第二服务器计算机114、模型数据库116以及评估计算机118。n个资源提供商计算机102可以与y个传送计算机104中的至少一个操作性通信。y个传送计算机104可以与网络处理计算机106操作性通信。网络处理计算机106可以与z个授权实体计算机108以及网络数据数据库110操作性通信。网络数据数据库110可以与第一服务器计算机112操作性通信。第一服务器计算机112可以与网络数据数据库110以及与第二服务器计算机114和可与评估计算机118操作性通信的模型数据库116操作性通信。

[0054] 图1中的装置可以通过任何合适的通信信道或通信网络彼此操作性通信。合适的通信网络可以是下列中的任一个和/或组合:直接互连;互联网;局域网(LAN);城域网(MAN);作为互联网上节点的运行任务(OMNI);安全定制连接;广域网(WAN);无线网络(例如,使用例如但不限于无线应用协议(WAP)、I-模式等的协议);等。计算机、网络与装置之间的消息可以使用安全通信协议来传输,例如但不限于:文件传输协议(FTP);超文本传输协议(HTTP);安全超文本传输协议(HTTPS)、安全套接层(SSL)、ISO(例如,ISO 8583)等。

[0055] 出于说明简单起见,图1中示出了一定数量的组件。然而,应理解,本发明的实施例可以包括多于一个每种组件。

[0056] 在步骤1,n个资源提供商计算机102可以生成针对交互的授权请求消息。n个资源提供商计算机102可以各自包括由资源提供商运行的计算机。在一些实施例中,资源提供商计算机可以包括服务器计算机。在资源提供商与用户之间的交互期间,每个资源提供商计算机可以生成针对交互的授权请求消息。然后,资源提供商计算机可以将授权请求消息传输到y个传送计算机104中的传送计算机。n个资源提供商计算机102中的每个资源提供商计算机可以将授权请求消息传输到y个传送计算机104中的不同或类似传送计算机。

[0057] 在一些实施例中,n个资源提供商计算机102可以分别从与n个资源提供商计算机102相关联的访问装置接收授权请求消息。例如,访问装置可以包括用于向用户提供对外部计算机系统的访问的任何合适的装置。访问装置的一些示例包括销售点(POS)装置、蜂窝电话、PDA、个人计算机(PC)、平板PC、手持式专用阅读器、机顶盒、电子收款机(ECR)、自动柜员机(ATM)、虚拟收款机(VCR)、查询一体机、安全系统、访问系统、网站等。

[0058] 在步骤2,在从n个资源提供商计算机102中的一个接收到授权请求消息之后,y个传送计算机104中的传送计算机可以将授权请求消息转发到网络处理计算机106。例如,y个

传送计算机104可以包括由收单方运行的计算机和/或服务器计算机。

[0059] 在步骤3,在从y个传送计算机104中的传送计算机接收到授权请求消息之后,网络处理计算机106可以将授权请求消息和/或与其相关联的数据存储到网络数据数据库110中。例如,网络处理计算机106可以将交易的交易数据存储到网络数据数据库110中。交易数据可以包括如本文所述的与用户和发起资源提供商计算机的资源提供商之间的交易相关的任何合适的数据元素。网络数据数据库110可以包括任何合适的数据库。网络数据数据库110可以是常规的、容错的、关系式的、可扩展的、安全的数据库,例如可购自Oracle™或Sybase™的那些数据库。网络数据数据库110可以存储网络数据。

[0060] 网络处理计算机106可以包括任何合适的服务器计算机。网络处理计算机106可以包括数据处理子系统、网络,以及用来支持并递送授权服务、异常文件服务、交易评分服务以及清算与结算服务的操作。示范性网络处理计算机106可以包括VisaNet™。例如VisaNet™的处理网络能够处理信用卡交易、借记卡交易以及其它类型的商业交易。具体来说,VisaNet™可以包括处理授权请求的VIP系统(Visa集成支付系统)以及执行清算与结算服务的Base II系统。

[0061] 在步骤4,在网络数据数据库110中存储相关数据之后,网络处理计算机106可以将授权请求消息转发到z个授权实体计算机108中的一个。z个授权实体计算机108可以包括任何合适的计算机。例如,授权实体计算机可被配置成基于授权请求消息确定是否授权交互。授权实体的示例可以包括发行方、政府机构、文件存储库、访问管理员等。在接收到授权请求消息之后,授权实体计算机可以确定是否授权交互。

[0062] 在步骤5,在确定是否授权交互之后,z个授权实体计算机108中的授权实体计算机可以生成授权响应消息并将其传输到网络处理计算机106。在一些实施例中,当接收到授权响应消息时,网络处理计算机106可以将授权响应消息和/或与其相关联的数据存储到网络数据数据库110中。

[0063] 在步骤6,网络处理计算机106可以将授权响应消息转发到y个传送计算机104中的适当的传送计算机。例如,网络处理计算机106可以通过评估授权响应消息中指示适当的传送计算机的路由表和/或数据元素来确定向y个传送计算机104中的哪个传送计算机发送授权响应消息。

[0064] 在步骤7,在从网络处理计算机106接收到授权响应消息之后,y个传送计算机104中的传送计算机可以将授权响应消息传输到n个资源提供商计算机102中的适当的资源提供商计算机,如本文描述。在一些实施例中,在接收到授权响应消息之后,n个资源提供商计算机102中的资源提供商计算机可以通知用户交互的状态。例如,资源提供商计算机可以通过访问装置通知用户是否授权交互(例如,交易)。

[0065] 在任何合适的时间点,在步骤8,第一服务器计算机112可以查询网络数据数据库110中的网络数据。在步骤8之前可能已发生任何数量的交互。网络处理计算机106可以在网络数据数据库110中存储与多个交互有关的数据。例如,在第一服务器计算机112查询网络数据数据库110中的网络数据之前,网络处理计算机106可以将与10个、500个、2,000个、10,000个交互等有关的数据存储到网络数据数据库110中。

[0066] 在一些实施例中,第一服务器计算机112可以查询网络数据数据库110中与一个或多个标准相关联的网络数据。例如,第一服务器计算机112可以在查询中包括的一个标准是

时间和/或时间范围。例如,第一服务器计算机112可以查询与过去一天、过去的小时、特定日期范围(例如,2019年5月10日至2019年5月15日)相关联的网络数据。作为另一示例,第一服务器计算机112可以包括的标准是:检取的网络数据包括与在特定地理区域(例如,北美、加州等)内发生的交互相关的数据。其它示例标准可与用户人口统计资料、资源提供商人口统计资料、支出金额等相关。

[0067] 在步骤9,网络数据数据库110可以为第一服务器计算机112提供查询的网络数据。第一服务器计算机112可以接收包括多个交易(例如,5、28、500、10,000个交易等)的多个交易数据的网络数据。每个交易数据可以包括具有数据值(例如,94016、商家_1234、用户_1234、\$19.99、111.111.11.111、2015年1月1日、西部时间上午11:00点分别对应于数据元素)的多个数据元素(例如,邮政编码、商家标识符、用户标识符、金额、IP地址、日期、时间等)。在一些实施例中,多个数据元素中的至少一个可以包括用户的用户标识符。例如,网络数据数据库110可以为第一服务器计算机112提供如表1所示且在下文详细描述的交易数据。

[0068] 第一服务器计算机112然后可以基于网络数据生成包括多个社区的一个或多个图。图可以包括例如通过表示交易的边缘连接的表示用户的节点和表示资源提供商的节点。多个社区可以包括图内的任何合适的节点子集,其共享相同和/或类似的特征。第一服务器计算机112可以如本文所述的任何合适方式确定社区。

[0069] 如本文进一步详细描述,然后,第一服务器计算机112可以确定每个交易的数据值中的至少一些的模糊值。例如,第一服务器计算机112可以使对应于每个交易的时间数据元素的数据值模糊化。作为说明性示例,第一服务器计算机112可以确定“上午11点”的数据值可对应于0.5为“早晨”标签且0.5为“中午”标签的模糊值。第一服务器计算机112可以基于确定的隶属函数来确定模糊值,如本文进一步详细描述。

[0070] 对于每个用户,然后,第一服务器计算机112可以确定多个社区内的社区的模糊值。例如,第一服务器计算机112可以首先多次确定图的社区。例如,第一服务器计算机112可以确定图的社区3次。表示用户的节点可以被确定为第一社区(例如,体育社区)两次,而所述用户可以被确定为第二社区(例如,文学社区)一次。第一服务器计算机112可以确定用户可以与第一社区(例如,体育社区)的模糊值0.66、第二社区(例如,文学社区)的模糊值0.33相关联。

[0071] 然后,第一服务器计算机112可以使用数据值的模糊值和社区的模糊值以及至少一些数据值来生成模型。模型可以包括任何合适类型的模型,如本文进一步详细描述,例如,支持向量机(SVMS)、人工神经网络、决策树、贝叶斯网络、遗传算法等。在一些实施例中,模型可以包括系统或流程的数学描述,以协助计算和预测(例如,欺诈模型、异常检测模型等)。

[0072] 在一些实施例中,在步骤10,第一服务器计算机112可以将每个交易的至少一些数据值的模糊值以及社区和/或模型的模糊值以隐私保护测试请求消息传输到第二服务器计算机114。第一服务器计算机112可以请求第二服务器计算机114以确定模糊值和/或模型是否可用于确定在创建模糊值和模型期间使用的网络数据。

[0073] 在一些实施例中,第二服务器计算机114可以通过确定模糊值是否包括第一服务器计算机112尝试模糊化的网络数据来确定模糊值和/或模型是否是保护隐私的。例如,第

二服务器计算机114可以确定模糊值中的任何一个是否等于值1。例如,数据元素“时间”可以与数据值“早晨8点”相关联。第一服务器计算机112可以为数据值“早晨8点”确定“早晨”标签的模糊值为1。第二服务器计算机114可以将模糊值1与模糊值阈值(例如,0.9)进行比较。第二服务器计算机114可以确定模糊值中的任何一个是否超出模糊值阈值。以此方式,第二服务器计算机114可以捕捉在模糊值阈值的范围内不是保护隐私的任何模糊值,这些值可以预先确定或调整为表示一定量的隐私的特定值。在一些实施例中,第二服务器计算机114可以生成不满足模糊值阈值的模糊值列表(或其它合适的数据库项)。

[0074] 在其它实施例中,第二服务器计算机114可以利用外部数据来确定用户(或在一些实施例中,资源提供商)是否是可标识的。例如,第二服务器计算机114可以从任何合适的外部数据库数据库中检索包括用户就业数据(例如,工资、工作地点、职位名称、计划等)的外部数据。第二服务器计算机114可以基于用户就业数据以及模糊值来确定交易的模糊值集是否可归因于特定用户。

[0075] 例如,交易数据可以包括社区的模糊值以及时间(例如,事件)的模糊值,但可以不包括资源提供商位置(例如,街道地址)的模糊值,以及所购买资源的库存单位(SKU)。第二服务器计算机114可以确定与就业数据相关联的用户就业于X公司的电子产品职位,工作地点在远离资源提供商位置的一个城市区块,并且用户的工作日大致围绕时间的模糊值(例如,针对下午的0.8、针对晚上的0.2)结束。第二服务器计算机114还可以确定与所购买资源相关联的SKU与电子商品相对应。基于就业数据与模糊化交易数据中说明的用户之间的相似性,第二服务器计算机114可以确定重叠百分比(例如,50%、80%、90%等)。重叠百分比可以指示与外部数据相关联的用户和与模糊化交易数据相关联的用户的相似程度。如果重叠百分比大于预定重叠阈值,则第二服务器计算机114可以确定交易数据还不够模糊化,并且可以与第一服务器计算机112传达此类信息。

[0076] 在一些实施例中,第一服务器计算机112能够使用与第二服务器计算机114相同的方法,并且可以在本地确定模糊值是否是保护隐私的。

[0077] 在步骤11,在确定模糊值和/或模型是否是保护隐私的之后,第二服务器计算机114可以向第一服务器计算机112传输隐私保护测试响应消息。隐私保护测试响应消息可以包括模糊值和/或模型是否是保护隐私的指示。如果第二服务器计算机114确定模糊值和/或模型不是保护隐私的,则第二服务器计算机114可以包括(例如,模糊值和/或网络数据的值中的)哪些值尚未受到隐私保护的指示。

[0078] 在步骤12,在使网络数据模糊化并且基于其生成模型之后,第一服务器计算机112可以将模型存储在模型数据库116中。模型数据库116可以包括任何合适的数据库。模型数据库116可以是常规的、容错的、关系式的、可扩展的、安全的数据库,例如可购自Oracle™或Sybase™的那些数据库。模型数据库116可以存储一个或多个模型。

[0079] 在一些实施例中,在步骤13,评估计算机118可以查询模型数据库116中的一个或多个模型。评估计算机118可以在任何合适的时间点查询存储在模型数据库116中的任何合适数量的模型。评估计算机118可以包括外部计算机,其位于第一服务器计算机112外部。例如,评估计算机118可以查询模型数据库116中的欺诈模型和销售模型。

[0080] 在步骤14,模型数据库116可以向评估计算机118提供一个或多个模型。在接收到一个或多个模型之后,评估计算机118可以运行一个或多个模型以确定每个模型的输出数

据。例如，模型可以包括有关能够将输入数据标记为“欺诈”或“非欺诈”的模糊交易数据训练的SVM。

[0081] 在一些实施例中，评估计算机118可以向第一服务器计算机112传输请求。请求可以包括有关模型的请求数据。例如，评估计算机118可以请求第一服务器计算机112运行模型以确定加州的高端电子销售额。在接收到包括请求数据的请求之后，第一服务器计算机112可以确定输出数据。第一服务器计算机112可以将请求数据输入到模型中，以确定由模型输出的输出数据。

[0082] 图2示出根据实施例的服务器计算机200的框图。示例性服务器计算机200可以包括处理器204。处理器204可以耦合到存储器202、网络接口206、输入元件210、输出元件212和计算机可读介质208。计算机可读介质208可以包括社区确定模块208A、模糊化模块208B和模型创建模块208C。服务器计算机200可以是图1中描绘的第一服务器计算机112。

[0083] 存储器202可用于存储数据和代码。存储器202可以在内部或在外部耦合到处理器204（例如，基于云的数据存储装置），且可以包括易失性和/或非易失性存储器的任何组合，例如RAM、DRAM、ROM、闪存或任何其它合适的存储器装置。例如，存储器202可以存储密钥、网络数据、隶属函数等。

[0084] 输入元件210可以包括能够将数据输入到服务器计算机200中的任何合适的装置。输入装置的示例包括按钮、触摸屏、触摸板、麦克风、生物计量扫描器等。一个或多个输入元件210可以包括能够将数据输入到服务器计算机200中的任何合适的装置。输入元件210的示例包括按钮、触摸屏、触摸板、麦克风等。

[0085] 输出元件212可以包括可以输出数据的任何合适的装置。输出元件212的示例可以包括显示屏、扬声器和数据传输装置。一个或多个输出元件212可以包括可以输出数据的任何合适的装置。输出元件212的示例可以包括显示屏、扬声器和数据传输装置。例如，输出元件212可以包括能够向服务器计算机200的用户显示响应值的显示屏。

[0086] 计算机可读介质208可以包括代码，所述代码可由处理器204执行以执行一种方法，所述方法包括：a) 由服务器计算机接收包括多个交易的多个交易数据的网络数据，其中每个交易数据包括具有数据值的多个数据元素，其中所述多个数据元素中的至少一个包括用户的用户标识符；b) 由所述服务器计算机基于所述网络数据生成包括多个社区的一个或多个图；c) 由所述服务器计算机针对所述多个交易中的每个交易确定所述数据值中的至少一些的模糊值；d) 对于每个用户，由所述服务器计算机确定所述多个社区内的社区的模糊值；以及e) 由所述服务器计算机使用步骤c) 和d) 中获得的所述模糊值以及所述数据值中的至少一些来生成模型。

[0087] 社区确定模块208A可以包括可由处理器204执行的代码或软件，用于确定网络数据中的社区组。社区确定模块208A结合处理器204可以以任何合适的方式确定社区。例如，社区确定模块208A结合处理器204可以基于节点相对于彼此的相似性将数据项（例如，图的节点）分组为组（例如，群集）。

[0088] 在一些实施例中，社区确定模块208A结合处理器204可以执行无监督学习算法，所述无监督学习算法可以包括图学习过程，所述图学习过程可以基于距离将节点分组为密集群集。例如，学习过程可以包括以下内容：1) 使用边缘的连接性和总体计数作为权重，创建边缘的排序列表；2) 对于每个边缘，使用上述定义的权重作为排序依据，生成相邻边缘的降

序排序集合;3) 对于每个相邻边缘,生成相邻项与目标边缘之间的距离;4) 如果距离大于截止值,则将相邻边缘添加到社区;5) 重复直到所有边缘都与社区相关联为止。

[0089] 用于标识社区的合适的学习算法的示例可包括:Fastgreedy、Spinglass、Walktrap、边介数 (Edge Betweenness)、Infomap、标签传播 (Label Propagation)、最佳模块度 (Optimal Modularity) 和多层级 (Multilevel)。此外,图学习算法可以是标识彼此重叠(即,共享节点)的社区的算法。例如,通常用于基于重叠群集来标识蛋白质复合物的图学习算法还可用于对任何交互网络中的节点进行分类(例如,将交易网络的节点分组)。图学习算法可包括:基于计算出的每个边缘的权重来计算拓扑图中的每个节点的权重,以及生成包括呈权重降序的节点的队列。可以从队列顶部选择种子节点来生成社区。接着,可以使用计算出的节点之间的交互概率以迭代方式将节点添加到社区。然后,可以从队列中去除所添加的节点,并且留在队列顶部的节点可用作下一社区的种子。然后可以重复此过程,直到队列被清空为止,从而生成多个社区。有关社区团体确定的进一步详情,请参阅2017年6月30日提交的第US 2019/0005407号美国公开案,所述公开案出于所有目的以全文引用的方式并入本文。

[0090] 在其它实施例中,社区确定模块208A结合处理器204可以基于网络数据生成包括多个社区的一个或多个图。社区确定模块208A结合处理器204可以使用网络数据执行任何合适次数的社区确定。例如,社区确定模块208A结合处理器204可以在5次不同的时间确定网络数据的社区,然后比较社区确定过程的5个不同结果。每个结果可以包括多个社区。在一些实施例中,社区确定模块208A结合处理器204可以基于社区确定过程的输出而确定在不同图中的不同社区中包括特定用户或资源提供商。在一些实施例中,对于一个或多个图中的每个不同的图,社区确定模块208A结合处理器204可以执行具有参数的不同值和/或不同阈值的社区确定过程,使得每个社区确定过程的输出可以包括不同的社区组。

[0091] 模糊化模块208B可以包括可以包括由处理器204执行以用于使数据模糊化的代码或软件。模糊化模块208B结合处理器204可以针对多个交易中的每个交易确定数据值中的至少一些的模糊值。例如,模糊化模块208B结合处理器204可以通过使用一个或多个隶属函数将数据值映射到模糊值而使数据值模糊化,如本文详细描述。例如,模糊化模块208B结合处理器204可以具有如至少以下第二部分所述的功能。

[0092] 在一些实施例中,模糊化模块208B结合处理器204可以为每个用户确定多个社区内的社区的模糊值。例如,模糊化模块208B结合处理器204可以确定模糊值,所述模糊值可以表示用户与多个社区相关联。在一些实施例中,模糊化模块208B结合处理器204可以将社区确定模块208B的输出表示为模糊值。例如,第一用户可以在3个社区确定试验中的2个期间与体育社区相关联,并且在3个社区确定试验中的1个期间与文学社区相关联。模糊化模块208B结合处理器204可以将这些比率转换为模糊值。例如,模糊化模块208B结合处理器204可以确定第一用户与体育社区相关联的模糊值为0.66,并且第一用户与文学社区相关联的模糊值为0.33。

[0093] 模糊化模块208B结合处理器204能够使对应于数据元素的任何合适的数据值模糊化,例如,一天中的时间、一年中的一天、位置、IP地址、用户标识符、资源提供商标识符、金额等。

[0094] 模型创建模块208C可以包括可以包括由处理器204执行以用于创建模型的代码或

软件。模型创建模块208C结合处理器204可以使用由模糊化模块208B结合处理器204获得的模糊值以及例如邮政编码、IP地址和金额的至少一些数据值生成模型。

[0095] 模型创建模块208C结合处理器204可以创建任何合适的类型模型,例如,支持向量机(SVMS)、人工神经网络、决策树、贝叶斯网络、遗传算法等。在一些实施例中,模型可以包括系统或流程的数学描述,以协助计算和预测(例如,欺诈模型、异常检测模型等)。

[0096] 例如,模型创建模块208C结合处理器204可以创建模型,所述模型可以是统计模型,可以用来根据已知信息预测未知信息。例如,模型创建模块208C与处理器204结合可以包括用于根据训练数据生成回归线(监督学习)的指令集,或用于基于相似性、连接性和/或数据点之间的距离将数据分组为数据的不同分类的群集(无监督学习)的指令集。接着,回归线或数据群集可作用用于根据已知信息预测未知信息的模型。一旦已从模型创建模块208C构建了模型,就可以结合处理器204使用模型来根据请求生成预测的输出。例如,请求可以是对与呈现的数据相关联的预测的请求。例如,请求可以是对将交易分类为欺诈性或非欺诈性的请求,或者是对用户建议的请求。

[0097] 网络接口206可以包括可以允许服务器计算机200与外部计算机进行通信的接口。网络接口206可以使服务器计算机200能够与另一装置(例如,第二服务器计算机等)之间进行数据传送。网络接口206的一些示例可以包括调制解调器、物理网络接口(例如以太网卡或其它网络接口卡(NIC))、虚拟网络接口、通信端口、个人计算机存储卡国际协会(PCMCIA)插槽和卡等等。由网络接口206启用的无线协议可以包括Wi-Fi™。经由网络接口206传递的数据可以呈信号的形式,所述信号可以是电学、电磁、光学或能够由外部通信接口接收的任何其它信号(统称为“电子信号”或“电子消息”)。可以包括数据或指令的这些电子消息可以经由通信路径或信道提供在网络接口206与其它装置之间。如上所述,可以使用任何合适的通信路径或信道,例如电线或电缆、光纤、电话线、蜂窝链路、射频(RF)链路、WAN或LAN网络、互联网或任何其它合适的介质。

[0098] II. 模糊化概述

[0099] 模糊逻辑可以包括多值逻辑的形式,其中变量的值可以是0与1之间(包括0和1)的任何实数。相反,在布尔逻辑中,变量的值只能是整数值0或1。在模糊化过程中,可以收集明确的输入数据集,并且至少使用隶属函数将其转换为模糊的值集。然后,可以使用模糊值进行计算。

[0100] A. 值的模糊化

[0101] “模糊化”可以涉及域转换,其中可以将明确值转换为模糊值。明确值可以是测量的精确值(例如,5V的电压,50Pa的压力等)。例如,在一些实施例中,明确值可以包括来自网络数据的数据值。

[0102] 对应于数据元素(例如,用户标识符、金额、时间、位置、事件、社区等)的每种类型的明确值可以由系统使用不同组隶属函数或数据元素所转换到的集来处理。换句话说,第一数据元素(例如,时间)可对应于第一隶属函数集,而第二数据元素(例如,位置)可对应于第二隶属函数集。这组隶属函数可存在于保留明确输入可具有的一定范围的值(例如,所有相关值)的论域内。隶属函数可以通过将明确值映射到模糊值或一个或多个标签的隶属度来定义模糊集。

[0103] 图3示出根据实施例的图300,所述图示出了用于确定模糊值的示例隶属函数和标

签。图300包括四个隶属函数：第一隶属函数302、第二隶属函数304、第三隶属函数306以及第四隶属函数308。隶属函数范围跨越明确值为1到24的论域310。每个隶属函数可以跨越值的范围312，其隶属函数将明确值314映射到隶属度316的非零值。例如，范围312对应于第二隶属函数304，并且跨越明确值5到明确值19（不包括5和19）。

[0104] 每个隶属函数可以对应于标签。标签数量可以对应于论域310应被划分成的区域数量，使得每个标签描述一个行为区域。可以对每个隶属函数分配范围（例如，范围312），以数字形式标识对应于标签的一系列明确值314。

[0105] 隶属函数的形状可以代表明确值314的变量。当然，这种形状也受到可用计算资源的限制。复杂的形状需要更复杂的描述方程式或较大查找表。隶属函数的形状可以包括三角函数（例如，如图3的图300所示）、梯形函数、单模式函数、高斯函数、分段线性函数等。在一些实施例中，每个隶属函数可以包括不同的范围大小和/或可以具有不同的函数形状。

[0106] 隶属函数的数量可取决于要模糊化的明确值的类型。给定应用程序的隶属函数太少，可能导致系统产生非隐私保护的模糊化数据。然而，对于输入中的小变化，隶属函数太多可能导致输出变化较大，这会导致系统中的不稳定性。太多的隶属函数可能会对部分基于模糊数据创建的模型产生不利影响。

[0107] 为了进一步提供保护隐私的模糊化数据，隶属函数可以在范围内重叠。隶属函数之间没有任何重叠会减少到基于布尔逻辑的系统，因此，所产生的数据可能会被恶意方链接回至原始数据。论域310上的每个明确值314可以属于一个以上隶属函数的范围。

[0108] 作为说明性示例，在模糊化过程期间，服务器计算机可以映射论域310上的每个明确输入314以确定与每个隶属函数的交集。服务器计算机可以确定与特定明确值交集的每个隶属函数的隶属度316。

[0109] 例如，明确值314可以与数据元素“一天中的时间”相关。因此，论域310可涵盖当日所有时间（例如，00:00至23:59），然而，为简单说明起见，图3中描绘的明确值314包括整点1、2、3、4、...、24。每个隶属函数可以包括涵盖当日（即，论域310的）时间量的范围312。第一隶属函数302可以标记为“01:00至12:00”，第二隶属函数304可以标记为“05:00至19:00”，第三隶属函数306可以标记为“11:00至24:00”，并且第四隶属函数308可以标记为“18:00至6:00”。

[0110] 例如，输入明确值可以是09:00，其在第一隶属函数302和第二隶属函数304的范围内。明确值09:00可以映射为对于第一隶属函数302具有隶属度0.5且对于第二隶属函数304具有隶属度0.5。

[0111] 模糊值可以包括任何合适的的数据项，包括与隶属度有关的信息。例如，明确值09:00的模糊值可以是(0.5,0.5,0,0)的元组，其中元组中的元素索引对应于隶属函数和/或标签。在其它实施例中，模糊值可以包括在字典中，例如，{“01:00-12:00”：“0.5”，“05:00-19:00”：“0.5”，“11:00-24:00”：“0”，“18:00-6:00”：“0”}。

[0112] 在一些实施例中，为避免某些输入值上的边缘条件，输出可以包括具有三个或更多个隶属度的模糊值。这也可以通过跨多个隶属函数分散用户数据来降低恶意方标识特定个体的能力。同时，组合的隶属度可保持表示个体。

[0113] B. 社区模糊化

[0114] 如本文所述，装置和系统也可以被配置成使社区数据（例如，社区组）模糊化。图4

示出根据实施例的社区检测和模糊化的流程图。

[0115] 图402可以是由服务器计算机基于网络数据生成的图。图402可以包括通过边缘连接的节点。例如,图402可以包括用户节点410和资源提供商节点420。用户节点410可以基于相关联用户与资源提供商之间发生的交互(例如,交易)而连接到资源提供商节点420。

[0116] 图402可以以任何合适的方式创建。例如,图402可以是基于指示节点(例如,用户和资源提供商)之间的关系的值矩阵而创建的二分图。这种矩阵可以包括例如关联矩阵、邻接矩阵等。

[0117] 服务器计算机可以生成包括多个社区404的图402。服务器计算机可以基于聚类过程来确定社区404。例如,服务器计算机可以基于分层聚类过程、基于质心的聚类过程(例如,k均值聚类)、基于分布的聚类过程、基于密度的聚类过程和/或基于共享和/或类似特征将数据项(例如节点)分组为群集的任何合适的过程来确定社区404。

[0118] 图402的至少一些节点可以包括在多个社区404中。节点430可以包括在第一社区440和第二社区450中。例如,节点430可以表示用户。基于交易数据和有关用户的任何其它合适的的数据(例如,工资、居住地点等),服务器计算机可以确定节点430的用户包括在第一社区440(例如,体育社区)以及第二社区450(例如,文学社区)中。

[0119] 例如,节点430的用户可以执行创建用户与社区之间的关联的动作,例如在体育用品商店购物或访问图书馆。个体可以与这两个组成比例地相关联。例如,如果个体每天在体育用品商店购物,但是每月只访问图书馆一次,则可以对个体在体育社区进行更高的加权,如社区确定过程所确定。因此,服务器计算机可以确定该个体针对体育社区具有权重0.8以及针对文学社区具有权重0.2。个体可以与任何合适数量的社区相关联,例如5个社区、10个社区或任何合适数量的社区。

[0120] 在一些实施例中,服务器计算机可以确定社区组多次,其中社区组可以重叠也可以不重叠。服务器计算机可以确定社区组任何合适的次数。然后,服务器计算机可以确定每个节点包括到每个社区中的比率。例如,用户节点可以8次包括在第一社区组中,并且可以2次包括在第二社区组中。然后,用户节点可以针对第一社区组与模糊值0.8相关联,且针对第二社区组与模糊值0.2相关联。

[0121] 在确定社区404之后,服务器计算机可以生成模糊社区网络空间406。模糊社区网络空间406可以是这样的图:图中包括作为节点的社区以及表示用户和/或资源提供商的节点。例如,节点460可以表示与第一社区440和第二社区450相关联的节点430的用户。在模糊社区网络空间406中,社区节点470可以表示第一社区440,且社区节点480可以表示第二社区450。节点460可以连接到社区节点470和社区节点480两者。连接节点460的边缘可以以社区组的模糊值加权。例如,节点460与社区节点470(例如,第一社区节点)之间的边缘可以以0.8加权。类似地,节点460与社区节点480(例如,第二社区节点)之间的边缘可以以0.2加权。

[0122] III. 方法

[0123] 实施例可以使用本文描述的系统和设备来至少确定模糊值。图4-8描述这些方法的一些示例。在一些实施例中,服务器计算机可以分别包括第一服务器计算机112或图1和2的服务器计算机200。

[0124] A. 使网络数据模糊化

[0125] 图5示出根据实施例的使交互数据模糊化的流程图。如本文所述,图5的方法可由服务器计算机执行。在服务器计算机使交互数据模糊化且至少部分地基于此而创建模型的上下文中描述图5中所示的方法。然而,应理解,本发明可以适用于其它情况。例如,服务器计算机可以使任何合适的数据模糊化,例如,包括多个交易数据的网络数据。

[0126] 在步骤502,服务器计算机可以从合适的数据库中检取网络数据。例如,服务器计算机可以检取包括多个交易的多个交易数据的网络数据。多个交易如表1描绘,其中每行是多个交易数据中的一个。每个交易数据可以包括具有数据值的多个数据元素。例如,数据元素可以包括但不限于用户标识符、一年中的一天、一天中的时间、资源提供商标识符、邮政编码、IP地址和金额。每个数据元素可以与数据值相关联。例如,数据值1111可以与用户标识符的数据元素相关联。作为另一示例,数据值9可以与一天中的时间的数据元素相关联。至少表1中描绘了更多示例。在一些实施例中,多个数据元素还可以包括SKU值、地理位置(例如,地址等)以及与交易相关的其它数据元素。

[0127] 在一些实施例中,多个数据元素中的至少一个可以包括用户的用户标识符。例如,表1的第一列示出了针对多个交易数据的用户标识符的数据元素,在此示例中,每个交易可以由不同用户执行。

用户标识符	一年中的一天	一天中的时间	资源提供商标识符	邮政编码	IP 地址	金额
1111	57	9	XYZ	123		12
2222	59	12	XYZ		111.111.11.111	100
3333	63	10	XYZ	123		56
4444	57	15	XYZ	123		56
5555	88	9	XYZ		111.111.11.111	12
6666	89	8	XYZ		111.111.11.111	98
7777	90	4	XYZ	123		78
8888	110	13	XYZ		111.111.11.111	23
9999	110	10	XYZ	123		145
0000	120	12	XYZ	123		4

[0128] 表1: 多个交易的多个交易数据

[0129] 在步骤504,在检取网络数据之后,服务器计算机可以针对多个交易中的每个交易确定数据值中的至少一些的模糊值。例如,在一些实施例中,服务器计算机可以确定对应于“一天中的时间”数据元素的至少数据值的模糊值。可以将对应于“一天中的时间”或“一年中的一天”数据元素的数据值的模糊值称为模糊事件值。

[0130] 服务器计算机可以使用多个隶属函数来确定模糊事件值。如本文所述,多个隶属函数能够将明确值映射到模糊值。例如,在下表2中,事件可以包括事件1、事件2和事件3。可以将与第一交易数据相对应的第一行的“一天中的时间”数据元素的数据值模糊化为与事件1、事件2和事件3相关联。在这种情况下,模糊值可以包括与事件1相关联的0.8(即80%),与事件2相关联的0.2,以及与事件3相关联的0。

[0131] 服务器计算机可以基于多个隶属函数确定模糊事件值。例如,多个隶属函数可以包括三个隶属函数,每个隶属函数与标签和范围相关联。第一隶属函数可以具有标签“早晨”以及从0到12(即夜晚到中午)的范围。第二隶属函数可以具有标签“中午”以及从6到18的范围。第三隶属函数可以具有标签“晚上”以及从12到0的范围。这些隶属函数以说明方式描述,但是应理解,可能有更多或更少的隶属函数以及隶属函数之间有各种重叠量的更大

或更小的范围。例如,可能有第四隶属函数的标签为“夜晚”,范围为21到4。

[0133] 服务器计算机可以确定数据元素“一天中的时间”的数据值9(即,上午9点)映射到针对事件1(例如,早晨)的模糊值0.8,针对事件2(例如,中午)的模糊值0.2,以及针对事件3(例如,晚上)的模糊值0。

[0134] 下表2进一步说明了多个交易数据中的其它交易数据的模糊事件。例如,表2的第三行显示了包括模糊事件数据(即,至少一些数据值的模糊值)的第三交易数据,其中事件1的模糊事件值为0.2,事件2的模糊事件值为0.6,并且事件3的模糊事件值为0.2。这些模糊事件值可以对应于一天中的时间12(即中午),所述一天中的时间可以包括在三个上述隶属函数的范围内。第三交易数据还可以包括一年中的一天、一天中的时间、资源提供商、邮政编码、IP地址和金额的数据元素的数据值,如表1所示。

事件 1	事件 2	事件 3	一年中的一天	一天中的时间	资源提供商	邮政编码	IP 地址	金额
0.8	0.2	0	57	9	XYZ	123		12
0.2	0.6	0.2	59	12	XYZ		111.111.11.111	100
0.8	0.2	0	63	10	XYZ	123		56
0	0.8	0.2	57	15	XYZ	123		56
0	0.8	0.2	88	9	XYZ		111.111.11.111	12
0	0.8	0.2	89	8	XYZ		111.111.11.111	98
0	0.2	0.2	90	4	XYZ	123		78
0	0.2	0.2	110	13	XYZ		111.111.11.111	23
0	0.2	0.8	110	10	XYZ	123		145
0	0.2	0.8	120	12	XYZ	123		4

[0137] 表2:模糊事件

[0138] 在步骤506,在确定数据值中的至少一些的模糊值之后,服务器计算机可以基于网络数据生成包括多个社区的一个或多个图。服务器计算机可以生成如本文所述的包括多个社区的一个或多个图。例如,服务器计算机可以确定与网络数据相关联的两个社区。但是,应理解,服务器计算机可以确定超过两个社区。两个社区可以包括“高科技”社区和“户外爱好者”社区。“高科技”社区可以包括与频繁和/或大量购买新技术(例如,每6个月购买一次新的智能手机)相关联的用户,且在一些实施例中,资源提供商。“户外爱好者”社区可以包括与购买露营设备、体育用品等相关联的用户。

[0139] 在步骤508,在确定社区之后,服务器计算机可以为每个用户确定多个社区内的社区的模糊值。社区的模糊值可以是模糊社区值。下表3示出了模糊社区值。

[0140] 模糊社区值可以是与每个交易数据相关联的社区的模糊化值。例如,第一交易数据可以0.1与社区1相关联并且以0.9与社区2相关联。作为另一示例,第五数据点以0.5与社区1相关联并且以0.5与社区2相关联。

[0141] 在一些情况下,如本文所述,交易数据可以与一个或多个社区相关联,因此一个或多个社区可以被称为重叠社区组。重叠社区组可以表达为社区组之间的模糊关系。重叠社区组之间的这种关系可以表达为多重图,其中表示社区组的节点通过边缘连接到表示用户的节点(例如,如图4的模糊社区网络空间406所描绘)。

[0142] 作为说明性示例,下表3中的第一行示出了与第一用户相关联的(例如,与“1111”的用户标识符相关联的)模糊社区值。第一用户可以与第一交易数据(即,表1的第一行)相

关联。第一用户的交易数据历史记录可以指示与“高科技”社区(例如,社区1)以及“户外爱好者”社区(例如,社区2)两者的连接。第一用户可能频繁购买露营设备,但很少购买新的手机和其它技术。由于第一用户的交易数据历史记录,服务器计算机可以确定第一用户与“高科技”社区相关联的模糊值为0.1且与“户外爱好者”社区相关联的模糊值为0.9。

[0143] 在一些实施例中,在步骤504确定模糊值之前,服务器计算机可以执行步骤506至508。在其它实施例中,服务器计算机可以与步骤504并行地执行步骤506至508,如图5所描绘。

[0144]

卡号	社区1	社区2
1111	0.1	0.9
2222	0.2	0.8
3333	0.3	0.7
4444	0.4	0.6
5555	0.5	0.5
6666	0.6	0.4
7777	0.7	0.3
8888	0.8	0.2
9999	0.9	0.1
0000	0.95	0.05

[0145] 表3:使社区模糊化

[0146] 在步骤510,在针对多个交易中的每个交易确定数据值中的至少一些的模糊值以及多个社区内的社区的模糊值之后,服务器计算机可以确定数据表。数据表可以包括数据值中的至少一些的模糊值以及社区的模糊值。

[0147] 在一些实施例中,服务器计算机可以通过存储数据值中的至少一些的模糊值以及社区的模糊值来确定数据表。例如,服务器计算机可以将模糊事件数据和模糊社区数据包括到交易数据中。服务器计算机还可以移除步骤504中被模糊化的数据值。例如,在确定数据表时,服务器计算机可以移除“一天中的时间”的数据元素。数据表可以包括基于模糊事件值和模糊社区值的模糊化数据,当然,应理解,数据表可以包括任何其它合适的模糊数据,例如但不限于位置、邮编、一年中的一天、商家标识符、IP地址和金额。

[0148] 作为说明性示例,下表4中示出数据表。第一用户的模糊社区数据,包括社区1(例如“高科技”社区)的模糊值0.1和社区2(例如“户外爱好者”社区)的模糊值0.9,可以包括在表1的第一交易数据中。服务器计算机还可以将对应于与“一天中的时间”数据元素相关联的数据值的模糊值包括在交易数据中。例如,服务器计算机可以将与事件1、事件2和事件3相关联的值0.8、0.2和0分别包括在第一交易数据中。此外,服务器计算机可以从第一交易数据中移除被模糊化的数据值9。在一些实施例中,服务器计算机还可以从多个交易数据中的每个交易数据中移除用户标识符。

[0149]

社区1	社区2	事件1	事件2	事件3	资源提供商	邮政编码	IP	金额
0.1	0.9	0.8	0.2	0	XYZ	123		12

[0150]	0.2	0.8	0.2	0.6	0.2	XYZ		111.111.11.111	100
	0.3	0.7	0.8	0.2	0	XYZ	123		56
	0.4	0.6	0	0.8	0.2	XYZ	123		56
	0.5	0.5	0	0.8	0.2	XYZ		111.111.11.111	12
	0.6	0.4	0	0.8	0.2	XYZ		111.111.11.111	98
	0.7	0.3	0	0.2	0.2	XYZ	123		78
	0.8	0.2	0	0.2	0.2	XYZ		111.111.11.111	23
	0.9	0.1	0	0.2	0.8	XYZ	123		145
	0.95	0.05	0	0.2	0.8	XYZ	123		4

[0151] 表4:数据表

[0152] 在步骤512,在确定数据表之后,服务器计算机可以使用步骤504和508中获得的模糊值以及与数据元素(例如,邮政编码、IP地址、金额等)相对应的至少一些数据值来生成模型。模型可以包括统计模型,所述统计模型可以用于根据已知信息(例如,数据表)预测未知信息。例如,在训练过程中可以使用数据表来训练模型,例如提出建议(例如,建议模型)。例如,一个人可能在下雨天的中午走进商店。时间、中午和天气可以是这个人的动作(例如,进入商店)将匹配到的模糊空间域的特征。使用此数据从模型获得的高概率响应(即,输出)可能是“这个人需要雨伞和雨衣”。可以将向用户推荐雨伞和雨衣的输出发送到外部计算机。

[0153] 作为另一示例,模型可以是销售预测模型。外部计算机(例如,图1的评估计算机118)可以生成包括请求数据的请求,然后将请求传输到服务器计算机。请求可以包括请求数据,所述请求数据包括例如对中午在加利福尼亚州的预期电子产品销售的请求。从外部计算机接收到请求之后,服务器计算机可以确定由模型输出的输出数据。在一些实施例中,请求数据可以输入到模型中。例如,参考表4,服务器计算机可以确定在加利福尼亚州(例如,邮政编码)中午(例如,事件2)期间的与电子产品(例如,社区1、SKU编码等)相关联的总销售量。

[0154] 例如,服务器计算机可以使用模糊社区值和模糊事件值来对每个交易数据的金额进行加权。例如,对于第一交易,在其它数据值是表示所述请求(例如,邮政编码是加利福尼亚州的邮政编码等)的情况下,服务器计算机可以通过将高科技社区(例如,社区1)与中午事件(例如,事件2)与金额相乘来确定加权金额。服务器计算机可计算:

[0155] $0.1 * 0.2 * 12 = 0.24$

[0156] 此示例中的值0.24很小,因为相关联的交易数据与“高科技”社区略有关联,与中午事件也略有关联。服务器计算机可以确定多个交易数据中的每个交易数据的加权金额。对于表4的10行(例如10个交易数据的实例),服务器计算机可以通过将表4的每个交易数据的加权金额相加来确定总加权金额126.82。在确定输出数据之后,服务器计算机可以将输出数据提供给外部计算机。

[0157] B. 通过进化学习器进行模糊隶属函数归纳

[0158] 在一些实施例中,服务器计算机可以包括进化学习器,所述进化学习器可以确定隶属函数和模糊数据。进化学习器可以是任何合适的进化学习器。一种特别适用于在图中找到最佳和/或最不昂贵的路径的优化技术可以包括蚁群优化。蚁群优化是找到与遗传算法相似的最佳解的方法。蚁群优化和遗传算法利用模拟退火的概率技术,这两种技术都是基于生物学中的思想。遗传算法可以包括优化方法,所述优化方法迭代地修改一组单独提议的解(即总体),直到在多次连续的迭代(即,代)之后找到最佳解为止。在这种方法中,有

关解的准确性(即,适合性)的信息可以仅在所选择的图的特定域内的代理(即求解器)之间进行传达。这可能导致这些算法收敛到局部最佳或最快的合适解,而不是图的全局最佳。在某些情况下,遗传算法可以有效地定位图的全局最佳;但是,这可能会花费相当长的时间,因为可能需要多代解来深入了解整个信息空间。

[0159] 相比之下,蚁群优化技术并非基于集体蜂群的角度。在蚁群优化中,多个代理可以尝试找到最佳解,并且可以共同地互相传达反馈(即,信息素)。这些信息素可以被记录并且可以在每次迭代中继转发关于它们各自的求解路径相对于总体目标的有效性(例如,梯度或其它误差项)的信息。代理可以分布在整个信息空间中,并与整个代理(即,群体)进行沟通,从而尽管评估域内存在局部最佳,此类方法仍可以更好地达到全局最佳的解。利用蚁群优化算法的额外优势是可以在算法内配置多个学习器目标。也就是说,可以以多种方式表达学习器目标。例如,可以将蚁群算法中的代理配置为根据信噪比、最短路径、最平滑拓扑等搜索路径。

[0160] 1. 进化学习器

[0161] 图6示出根据实施例的进化学习器数据模糊化方法的流程图。如本文所述,将在确定随后可在数据模糊化期间使用的最佳隶属函数的上下文中描述图6中所示的方法。然而,应理解,实施例可以应用于其它情况。

[0162] 在步骤600之前,服务器计算机可以从网络数据数据库中检取网络数据。在步骤600,服务器计算机可以将网络数据分割成多个样本。例如,网络数据可以包括多个交易数据,服务器计算机可以将所述多个交易数据分割成2个、5个、20个样本等,每个样本包括多个交易数据的子集。在一些实施例中,每个样本可以包括相同量的交易数据(例如,交易数据为100个交易)。样本可以包括整个数据集的数据子集(例如,网络数据)。

[0163] 在步骤602,在将网络数据分割成多个样本之后,服务器计算机可以为进化学习器初始化信息素。进化学习器可以包括例如修改的蚁群过程。信息素可以是有助于代理彼此传达有关最佳解的反馈的数据项。例如,信息素可以包括算法参数(例如,S型、灵敏度等)以及隶属函数的详细信息(例如,隶属函数的提议的分割点、隶属函数类型(例如,三角形、梯形等))。

[0164] 在步骤604,服务器计算机可以确定当前时期是否小于最大时期。最大时期可以是服务器计算机可以通过修改的蚁群过程迭代的预定迭代次数。如果服务器计算机确定当前时期(例如,时期数1)小于或等于最大时期(例如,时期数7),则服务器计算机可以继续到步骤606。如果服务器计算机确定当前时期大于最后一个时期,则服务器计算机可以结束过程,并且最后确定的隶属函数可以包括最佳隶属函数。

[0165] 图8将描述为贯穿图6的方法的示例。图8示出根据实施例的框图,所述框图示出了历经进化学习器的3个时期改变的隶属函数。图8示出在图6和图7所述方法期间隶属函数的视觉描绘。图8的列表示图6的每个时期,而行表示不同的隶属函数,所述隶属函数可以如图7中所述的由不同代理在一个时期期间决定。在多个时期中,可以通过更新每个代理的信息素来改变每个隶属函数集。随着时间的流逝,每个代理可以收敛于相同/相似的隶属函数,然后可以使用这些隶属函数使网络数据模糊化。

[0166] 在图8中,针对每个时期示出四个代理。每个代理可以确定隶属函数集,所述隶属函数集可以用于例如使一天中的时间模糊化。在三个时期内,代理可以收敛于最佳隶属函

数集,所述隶属函数集可以保护明确值的隐私性并保留明确值的预测能力(例如,在模型中使用)。

[0167] 时期1期间的第一隶属函数集(第一代理集)可以包括三个隶属函数。第一隶属函数集可以对应于有关“一天中的时间”的标签。例如,第一、第二和第三隶属函数可以分别标记为早晨810、中午820和晚上830。作为另一示例,时期1期间的第四隶属函数集(第四代理集)可以包括具有标签为早晨840和晚上850的两个隶属函数。在第一时期期间,如下所述,服务器计算机可以确定第一隶属函数集表现良好(例如,提供保留与明确值类似的预测能力的模糊值,和/或为明确值提供隐私)。服务器计算机还可以确定第四隶属函数集表现不佳(例如,不能提供保留与明确值类似的预测能力的模糊值,和/或不能为明确值提供隐私)。在第二时期期间,服务器计算机可以避免在前一个时期内表现不佳的隶属函数集。例如,在第二时期期间,第四代理可以确定包括三个隶属函数而不是两个隶属函数的隶属函数集,如第一时期所确定。

[0168] 在步骤606,服务器计算机可以确定多个代理中的代理(即,蚁)的隶属函数。多个代理可以包括任何合适数量的代理。例如,多个代理可以包括3个、5个、10个、100个代理等。服务器计算机可以如下图7进一步描述地确定隶属函数。例如,在一些实施例中,在步骤606,服务器计算机可以执行图7的过程。

[0169] 在步骤608,在为多个代理中的代理(例如,第一代理)确定隶属函数之后,服务器计算机可以存储与代理相关联的隶属函数。在步骤610,在存储代理的隶属函数之后,服务器计算机可以确定是否已针对多个代理中的每个代理确定了隶属函数。如果服务器计算机确定尚未为每个代理确定隶属函数,则服务器计算机可以对多个代理中的每个代理重复步骤606至608。如果服务器计算机确定已针对多个代理中的每个代理确定了隶属函数,则服务器计算机可以继续步骤612。

[0170] 例如,在步骤606至610中,服务器计算机可以确定第一时期的隶属函数。图8的第一时期描绘了四个隶属函数集。每个隶属函数集可以与代理相关地存储。在此示例中,可能有4个代理。每个隶属函数集可以包括不同范围、形状、位置等的不同隶属函数。以下步骤611至628可以包括确定图8所描绘的在第一时期与第二时期之间的信息素趋势/变化。然后,流程可以重复任何合适的次数(例如,N个时期),以确定最佳隶属函数,例如第三时期中描述的隶属函数。

[0171] 在步骤611,在确定多个代理中的每个代理的隶属函数之后,服务器计算机可以用每个代理的隶属函数使样本的多个交易数据模糊化,如本文所述。

[0172] 在步骤612,在使用多个代理中的每个代理的隶属函数使多个交易数据模糊化之后,服务器计算机可以使用向后传播神经网络来确定多个目标变量中的目标变量的预测值。目标变量可以包括欺诈、总支出金额、交易速度等。预测值可以是预测的目标变量(例如,预测的欺诈)。在一些实施例中,向后传播神经网络可用于近似一般学习器。

[0173] 作为说明性示例,服务器计算机可以使用由代理的隶属函数模糊化的数据来预测交易数据样本的目标变量(例如,欺诈)。服务器计算机可以使用模糊交易数据来预测欺诈率。

[0174] 在步骤614,服务器计算机可以确定预测值与实际值之间的误差。实际值可以包括从数据库中检取的对应于目标变量(例如,历史欺诈)的历史值。在一些实施例中,可以由服

务器计算机使用向后传播神经网络和多个交易数据(未模糊化)来预测目标变量的实际值,从而确定实际值。服务器计算机可以以任何合适的方式确定误差。例如,在一些实施例中,服务器计算机可以确定目标变量(例如,欺诈)的均方误差(MSE)。

[0175] 例如,服务器计算机可以如下确定误差:

[0176] $MSE_l = 1/\text{长度}(S_l) (\text{实际值} - \text{预测值})^2$

[0177] 误差可以表示与未模糊化数据(例如,网络数据)相比,模糊化数据保持预测能力的程度。

[0178] 在步骤616,在确定多个目标变量中的目标变量的预测值之后,服务器计算机可以确定是否已经针对多个目标变量中的每个目标变量确定了预测值和基于预测值的误差值。例如,服务器计算机可以针对欺诈的目标变量执行步骤612至614,然后可以针对目标变量总支出金额执行步骤612至614,等。服务器计算机可以针对多个目标变量中的每个目标变量重复步骤612至614。如果服务器计算机确定已经针对代理的每个目标变量确定了预测值和误差,则服务器计算机可以执行步骤618。

[0179] 在步骤618,在针对多个代理中的代理确定多个预测值和对应多个误差之后,服务器计算机可以确定代理的平均误差。例如,代理的平均误差可以确定如下:

[0180]
$$AvgMSE_j = \frac{1}{M} * \text{sum}(MSE_l)$$

[0181] 在步骤620,在确定多个代理中的代理的平均误差之后,服务器计算机可以确定是否已针对多个代理中的每个代理确定了平均误差。服务器计算机可以为多个代理中的每个代理重复步骤612至618,直到确定多个代理中的每个代理的平均误差为止。如果服务器计算机确定已针对多个代理中的每个代理确定了平均误差,则服务器计算机可以继续步骤622。

[0182] 在步骤622,在确定包括多个代理中的每个代理的平均误差的多个平均误差之后,服务器计算机可以更新全局信息素。全局信息素可以包括关于要避免哪些隶属函数以及要移向哪些隶属函数的规则。全局信息素可以使得多个代理能够传送其个别信息素的结果(例如,平均误差)。有关代理的本地和全局信息素的更多详细信息,请参见[Bonabeau、Eric等人的“群体智能:从自然系统到人工系统(Swarm intelligence:from natural to artificial systems)”,第1期,牛津大学出版社,1999年]。

[0183] 在步骤624,在更新全局信息素之后,服务器计算机可以验证网络数据的匿名化。服务器计算机可以拒绝从网络数据中显示原始用户数据的任何隶属函数。例如,由特定代理确定的隶属函数可能不会使用户数据匿名化。在这种情况下,服务器计算机可以验证所产生的模糊数据是否是保护隐私的。在一些实施例中,服务器计算机还可以拒绝显示位置数据(例如,邮政编码、地址等)、资源提供商数据和/或隶属函数意欲匿名化的任何其它合适的原始数据的任何隶属函数。例如,如果隶属函数集不会使邮政编码模糊化,则服务器计算机可以拒绝所述隶属函数集。

[0184] 在一些实施例中,服务器计算机可以通过调整与被拒绝隶属函数相关联的代理的平均误差(例如,AvgMSE)来拒绝隶属函数。

[0185] 在一些实施例中,服务器计算机可以将模糊数据传输到第二服务器计算机,所述第二服务器计算机可以包括确定模糊数据是否是保护隐私的对抗性AI。例如,类似于服务

器计算机,第二服务器计算机可以确定模糊数据是否是保护隐私的。

[0186] 在步骤626,在拒绝非隐私保护的隶属函数(如果存在的话)之后,服务器计算机可以保存与最小平均误差相关联的隶属函数。例如,最小平均误差可以是多个代理的多个平均误差中的最小平均误差。与最小平均误差相关联的隶属函数可以是在当前时期网络数据匿名化的同时最能保持预测值确定精度的隶属函数。

[0187] 在步骤628,在以最小的平均误差保存隶属函数之后,服务器计算机可以更新每个代理的信息素(例如,本地信息素)以及增加当前时期(例如,增加1)。

[0188] 例如,在更新多个代理中的代理的信息素时,可以使用全局信息素来筛选不良的参数。例如,第一代理(i)可以将不良隶属函数(即,对应于较大平均误差和/或非隐私保护的隶属函数)确定为例如在点QUOTE处具有分割的隶属函数,而第二代理(J)可以确定在点处具有分割的隶属函数。当第二代理(J)更新其信息素时,它可以避免(即,远离)分割点。

[0189] 在更新多个代理中的每个代理的信息素之后,服务器计算机可以继续步骤604并确定当前时期是否小于或等于最大时期。如果当前时期小于或等于最大时期,则服务器计算机可以重复步骤606至628,其中更新的信息素(在步骤628更新)可以修改下一个时期的隶属函数。如果服务器计算机确定当前时期大于最大时期,则服务器计算机可以结束过程。在步骤626保存的具有最小平均误差的隶属函数可以是由服务器计算机确定的最佳隶属函数。然后,服务器计算机可以继续使用隶属函数使网络数据模糊化,然后基于模糊化数据生成模型,如本文所述。

[0190] 例如,参考图8,第二时期示出了在步骤606为每个代理确定的隶属函数。例如,在第一时期期间,第一隶属函数集和第三隶属函数集可能具有较低的误差,而第二隶属函数集和第四隶属函数集可能具有较高的误差。因此,服务器计算机可以调整全局信息素以引起向低误差隶属函数集的变化。在第二时期期间,可以部分地基于更新的全局信息素来确定新的隶属函数集。例如,第四隶属函数集在第一时期期间包括两个隶属函数,但是,在第二时期期间第四隶属函数集包括三个隶属函数,因为已针对具有三个隶属函数的代理对信息素加权。

[0191] 在第三时期期间,四个隶属函数集可能会基于从第二时期开始变化的信息素而收敛到最佳的隶属函数集。如本文所述,服务器计算机可以使用由进化学习器确定的最佳隶属函数来使交易数据模糊化。

[0192] 2. 隶属函数确定

[0193] 图7示出根据实施例的隶属函数确定方法的流程图。将在例如图6所述过程的进化学习过程期间确定代理的隶属函数的上下文中描述图7中所示的方法。然而,应理解,本发明可以适用于其它情况。有关图7和隶属函数确定的更多详细信息,请参见[Hong、Tzung-Pei和Chai-Ying Lee的“模糊规则和隶属函数的训练示例归纳(Induction of fuzzy rules and membership functions from training examples)”,模糊集和系统84.1(1996):33-47],其全文以引用方式并入本文。在此步骤中,所有训练实例的输出值通过应用以下聚类程序进行适当分组,并且产生了输出值的适当隶属函数。聚类程序将输出值相近的训练实例视为属于具有高隶属值的同一类。图7详细描述于[Hong、Tzung-Pei和Chai-Ying Lee的“模糊规则和隶属函数的训练示例归纳(Induction of fuzzy rules and membership functions from training examples)”,模糊集和系统84.1(1996):33-47]

中,在此将简要描述。

[0194] 在步骤702,服务器计算机可以对交易数据的数据值(例如,对应于“一天中的时间”、“金额”等的数据元素的数据值)进行排序。服务器计算机可以以任何合适的方式对数据值进行排序。例如,在一些实施例中,服务器计算机可以按升序或降序对数据值进行数字排序。在其它实施例中,服务器计算机可以基于与目标变量(例如,欺诈、总支出金额、交易速度等)的相关性来对数据值进行排序。

[0195] 在步骤704,在对数据值进行排序之后,服务器计算机可以确定相邻数据值之间的差值。在步骤706,在确定相邻数据值之间的差值之后,服务器计算机可以确定相邻数据值之间的相似性。相邻数据值可以在排序数据值列表中包括彼此邻近的数据值。

[0196] 在步骤708至714,在确定相邻数据值之间的相似性数据值之后,服务器计算机可以根据相似性来将数据值聚类。在步骤708,服务器计算机可以确定在聚类期间是否考虑了所有数据。如果服务器计算机确定已考虑所有数据,则服务器计算机可以继续步骤716。如果服务器计算机确定尚未考虑所有数据,则服务器计算机可以继续步骤710。

[0197] 在步骤710,在确定尚未考虑所有数据之后,服务器计算机可以确定相邻值的相似性是否小于相似性阈值。相似性阈值可以包括任何合适的阈值。如果服务器计算机确定相似性值等于或大于相似性阈值,则服务器计算机可以继续步骤712并将两个数据点纳入同一组。如果服务器计算机确定相似性值小于相似性阈值,则服务器计算机可以继续步骤714并将两个数据点纳入不同的组。在步骤712和714之后,服务器计算机可以确定步骤706后接下来两个值的相邻值之间的相似性。

[0198] 在步骤708确定已考虑(已聚类)所有数据之后,服务器计算机可以在步骤716基于聚类数据确定隶属函数,如在[Hong、Tzung-Pei和Chai-Ying Lee的“模糊规则和隶属函数的训练示例归纳(Induction of fuzzy rules and membership functions from training examples)”,模糊集和系统84.1(1996):33-47]中进一步描述。

[0199] 本发明的实施例提供若干优点。例如,可以以不可逆的方式对网络数据进行模糊化,从而在使用模糊化数据构建的任何模型中保护网络数据的隐私。例如,与交易相关联的用户可能无法通过模糊化数据标识,因此形成隐私保护数据。

[0200] 本发明的实施例提供若干额外优点。例如,网络数据的模糊化可以帮助对一个个体和位置的覆盖数据偏差进行纠正。例如,即使用户仅针对特定商品(例如天然气)使用信用卡或其它支付装置,通过模糊化关系,用户的数据仍然可以帮助为使用信用卡进行所有操作的用户改进模型。

[0201] 应理解,本发明的任何实施例都可以使用硬件(例如,专用集成电路或现场可编程门阵列)和/或使用计算机软件以控制逻辑的形式实施,其中通用可编程处理器是模块化的或集成的。如本文中所使用,处理器包括单核处理器、在同一集成芯片上的多核处理器,或在单个电路板上或网络化的多个处理单元。基于本文中所提供的公开内容和教导内容,本领域的普通技术人员将知道且了解使用硬件和硬件与软件的组合来实施本发明的实施例的其它方式和/或方法。

[0202] 本申请中描述的任何软件组件或功能可以被实施为要使用例如Java、C、C++、C#、Objective-C、Swift的任何合适计算机语言或例如Perl或Python的脚本语言,使用例如常规的或面向对象的技术由处理器执行的软件代码。软件代码可作为一系列指令或命令存储

在计算机可读介质上以供存储和/或传递,合适的介质包括随机存取存储器(RAM)、只读存储器(ROM)、例如硬盘驱动器或软盘的磁性介质,或例如光盘(CD)或数字通用盘(DVD)的光学介质、闪存存储器等等。计算机可读介质可以是此类存储或传输装置的任何组合。

[0203] 此类程序还可以使用适应于经由包括互联网的符合多种协议的有线、光学和/或无线网络进行传输的载波信号来编码和传输。因此,根据本发明的实施例的计算机可读介质可以使用以此类程序编码的数据信号来创建。以程序代码编码的计算机可读介质可与兼容装置一起封装或其它装置分开提供(例如,经由互联网下载)。任何此类计算机可读介质可以驻留于单个计算机产品(例如,硬盘驱动器、CD或整个计算机系统)之上或之内,且可存在于系统或网络内的不同计算机产品之上或之内。计算机系统可以包括监视器、打印机,或用于向用户提供本文所提及的任何结果的其它合适的显示器。

[0204] 以上描述是说明性的且不是限制性的。在本领域技术人员阅读了本公开之后,本发明的许多变体对于他们会变得显而易见。因此,本发明的范围不应该参考上面的描述来确定,而是应该参考待决的权利要求及其完整范围或等同物来确定。

[0205] 在不偏离本发明的范围的情况下,任何实施例的一个或多个特征可以与任何其它实施例的一个或多个特征组合。

[0206] 如本文中所使用,除非明确指示有相反的意思,否则使用“一个”、“一种”或“所述”旨在意指“至少一个”。

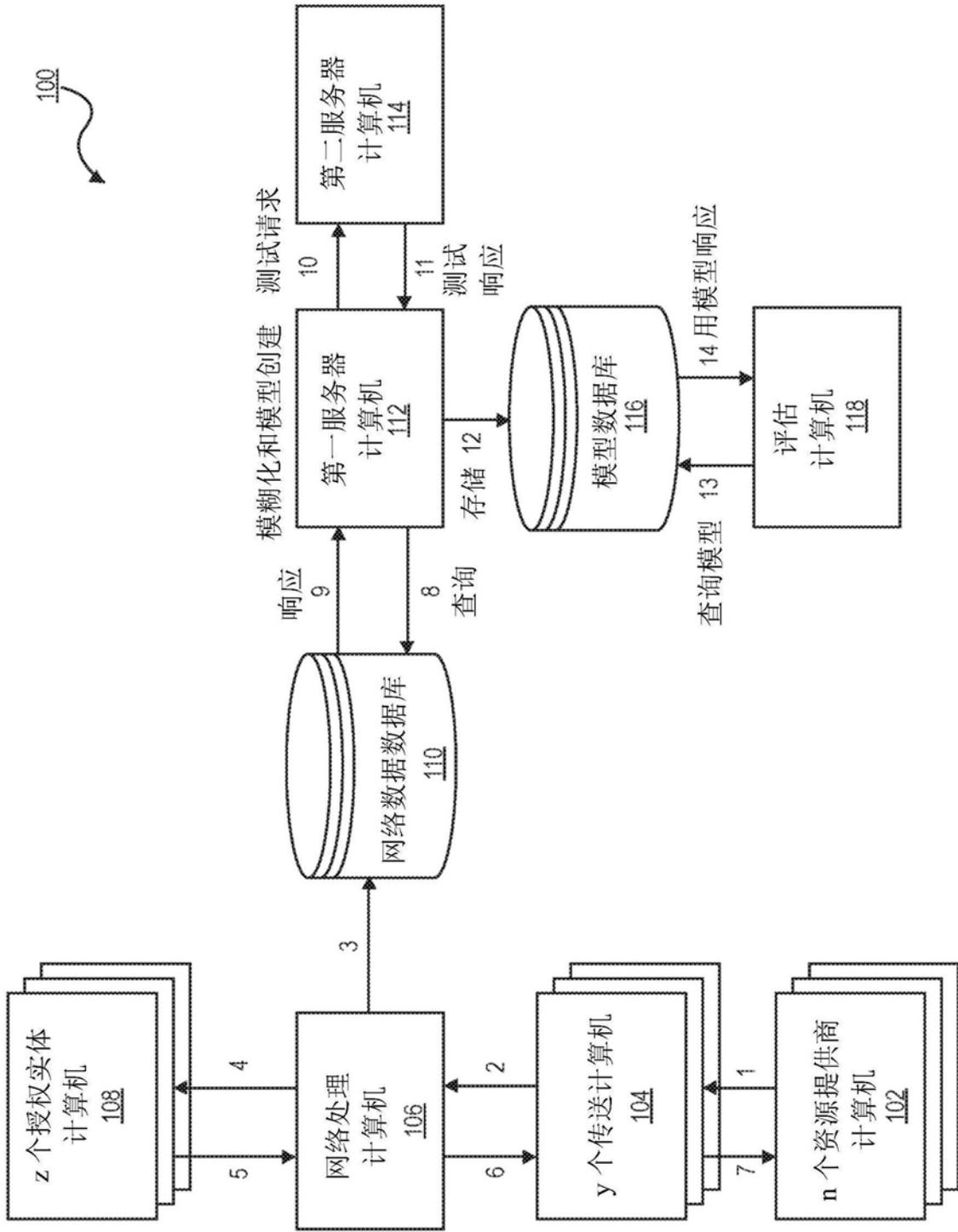


图1

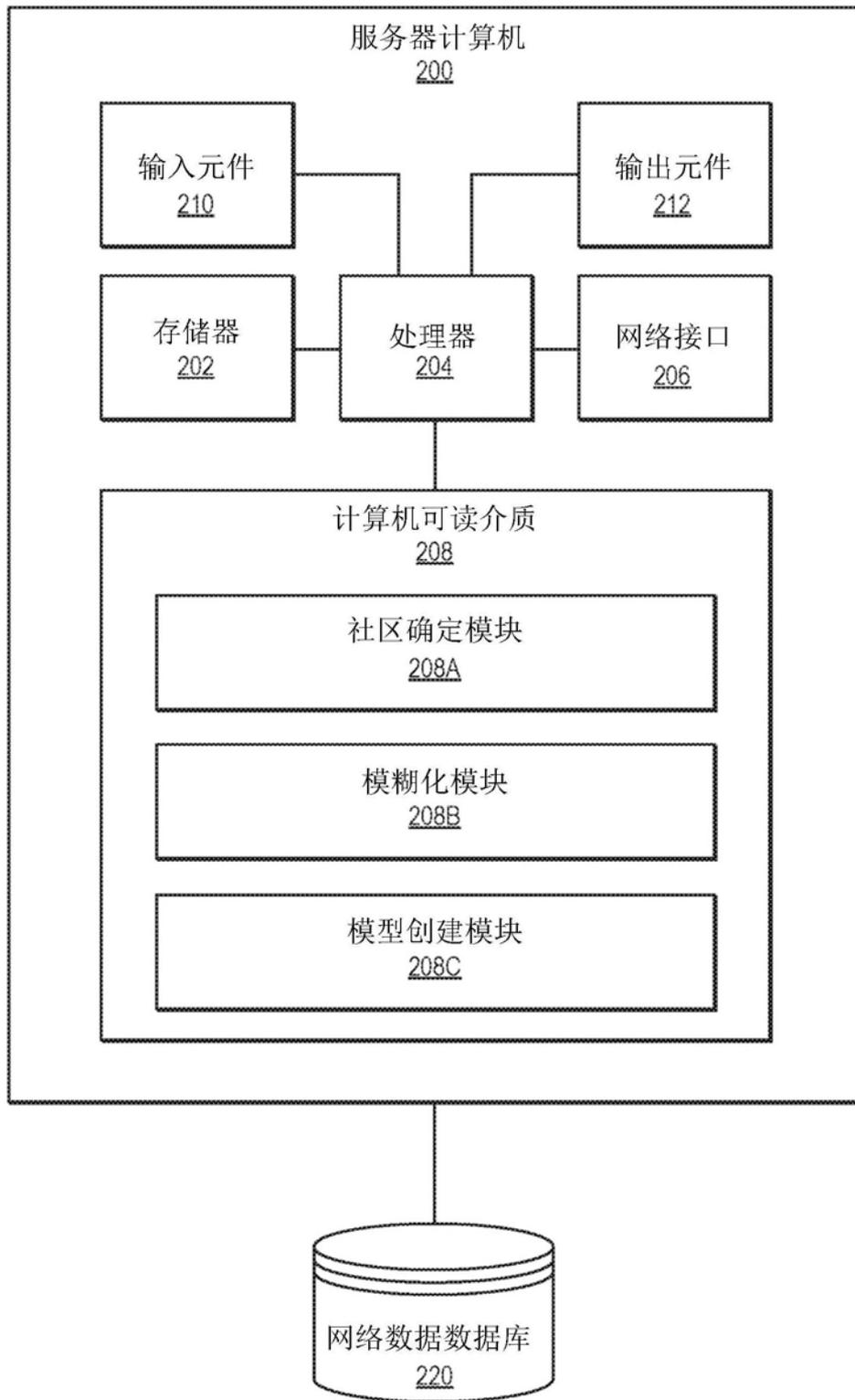


图2

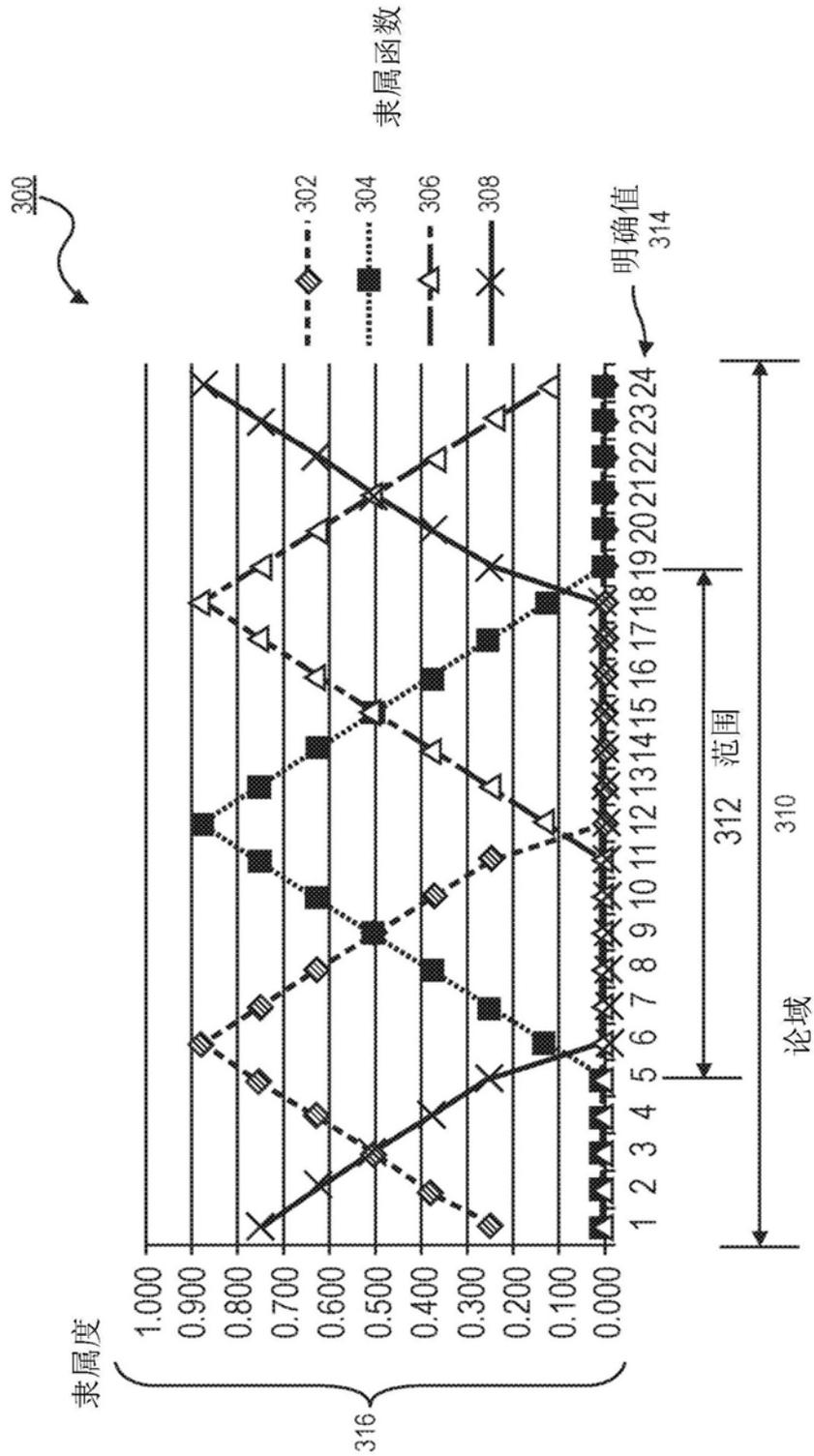


图3

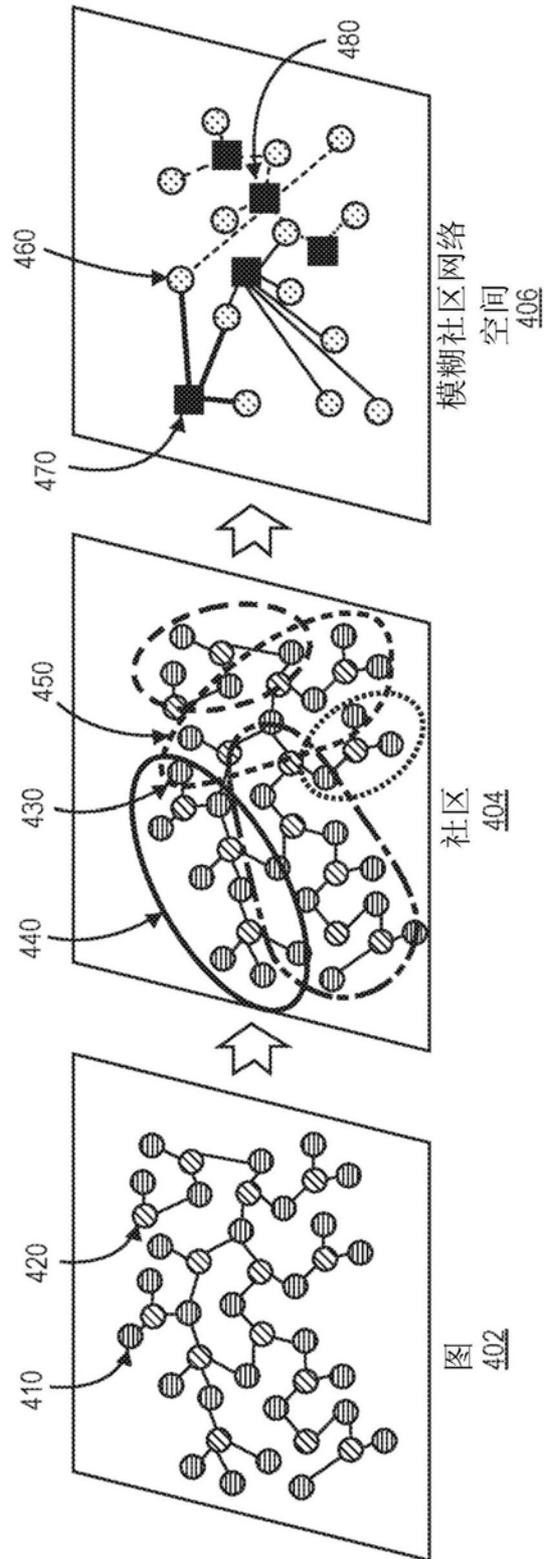


图4

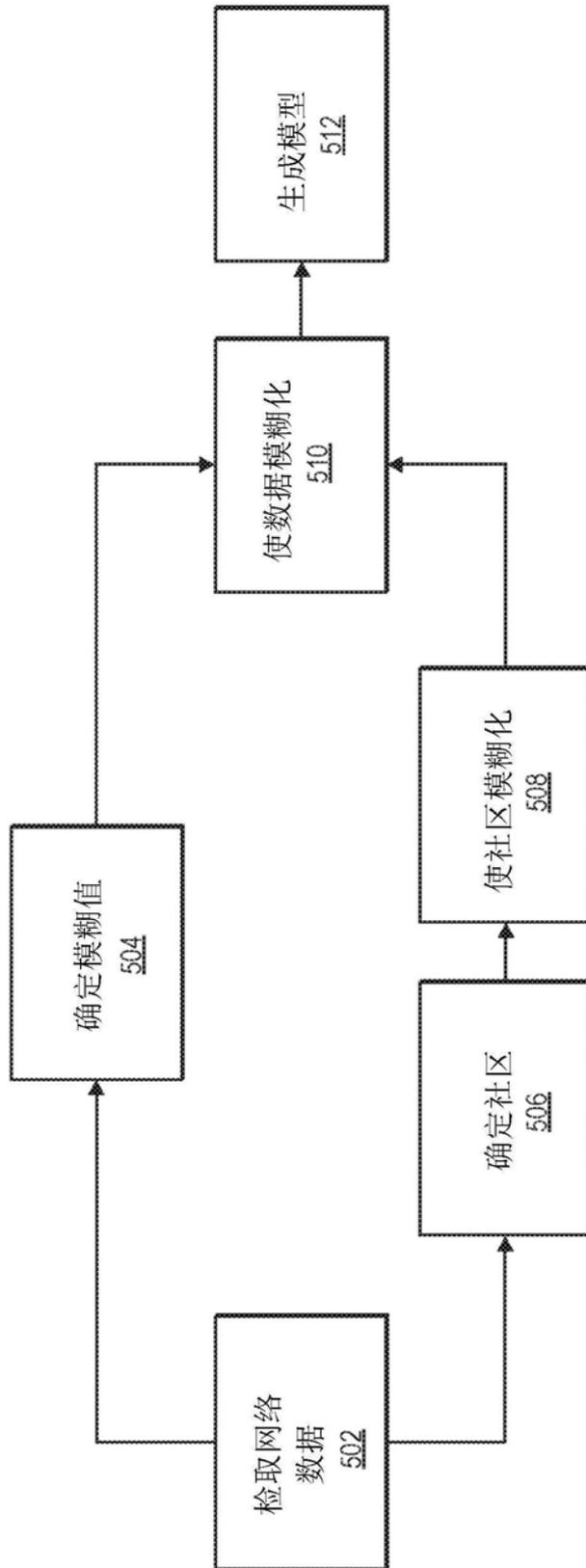


图5

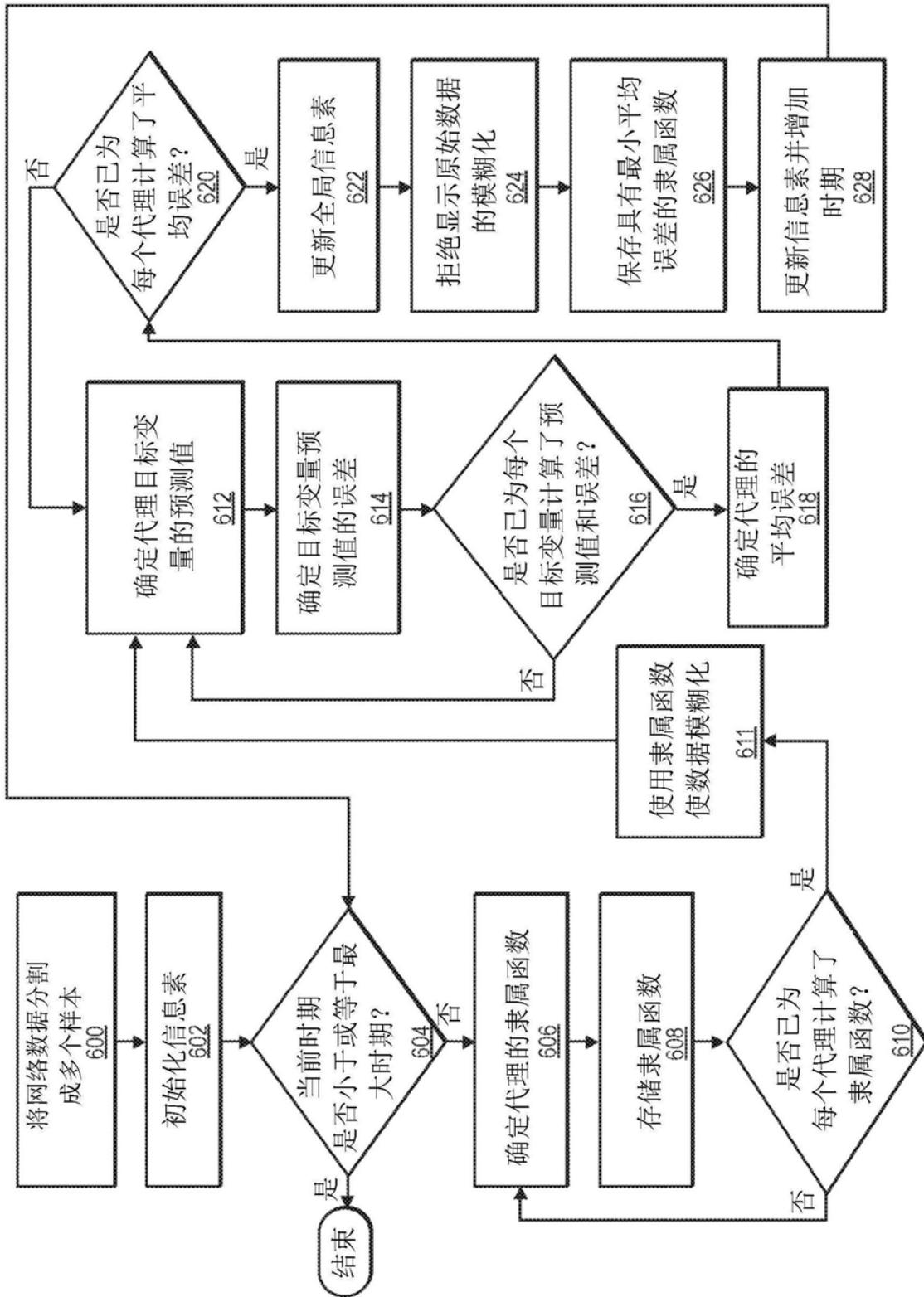


图6

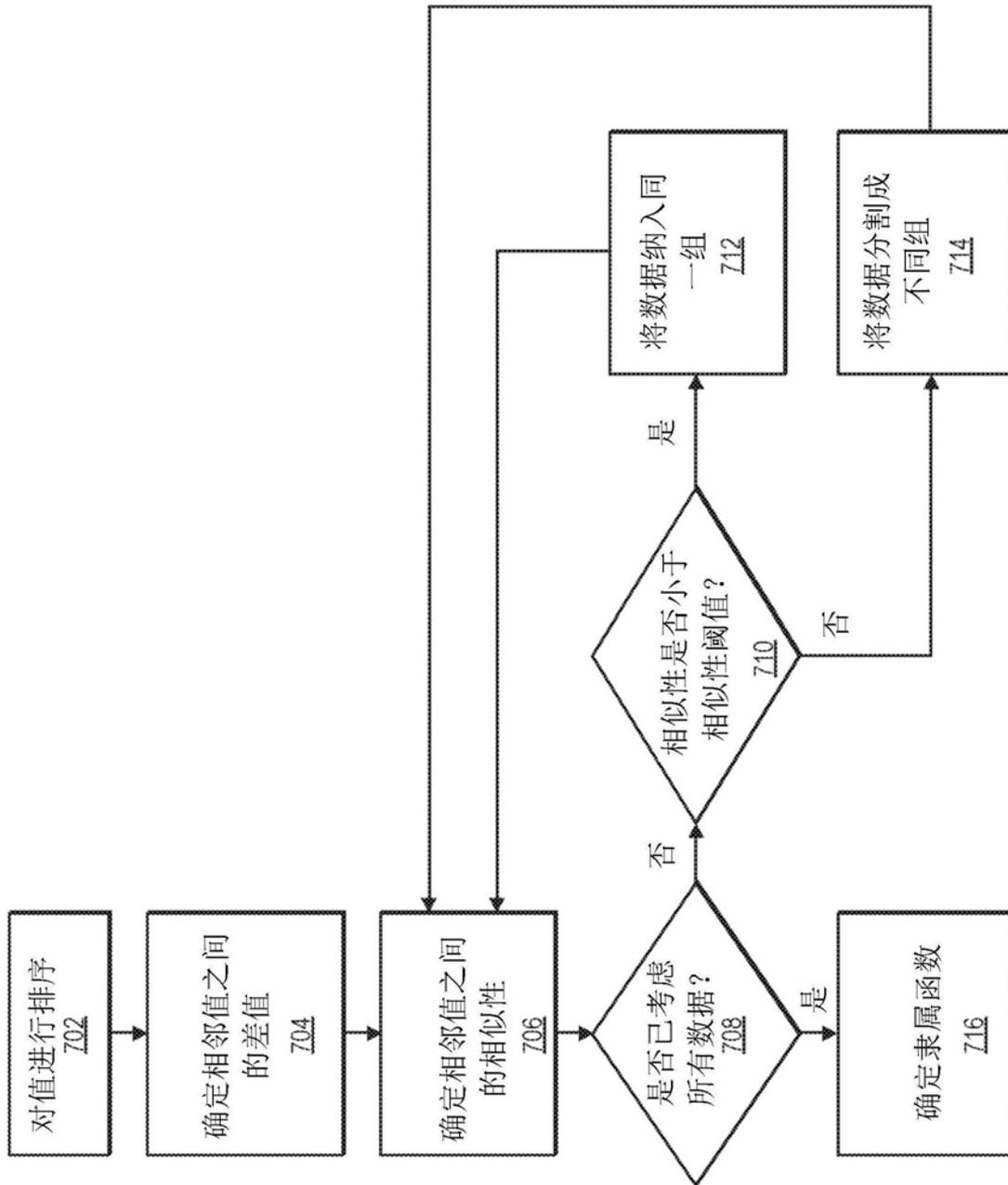


图7

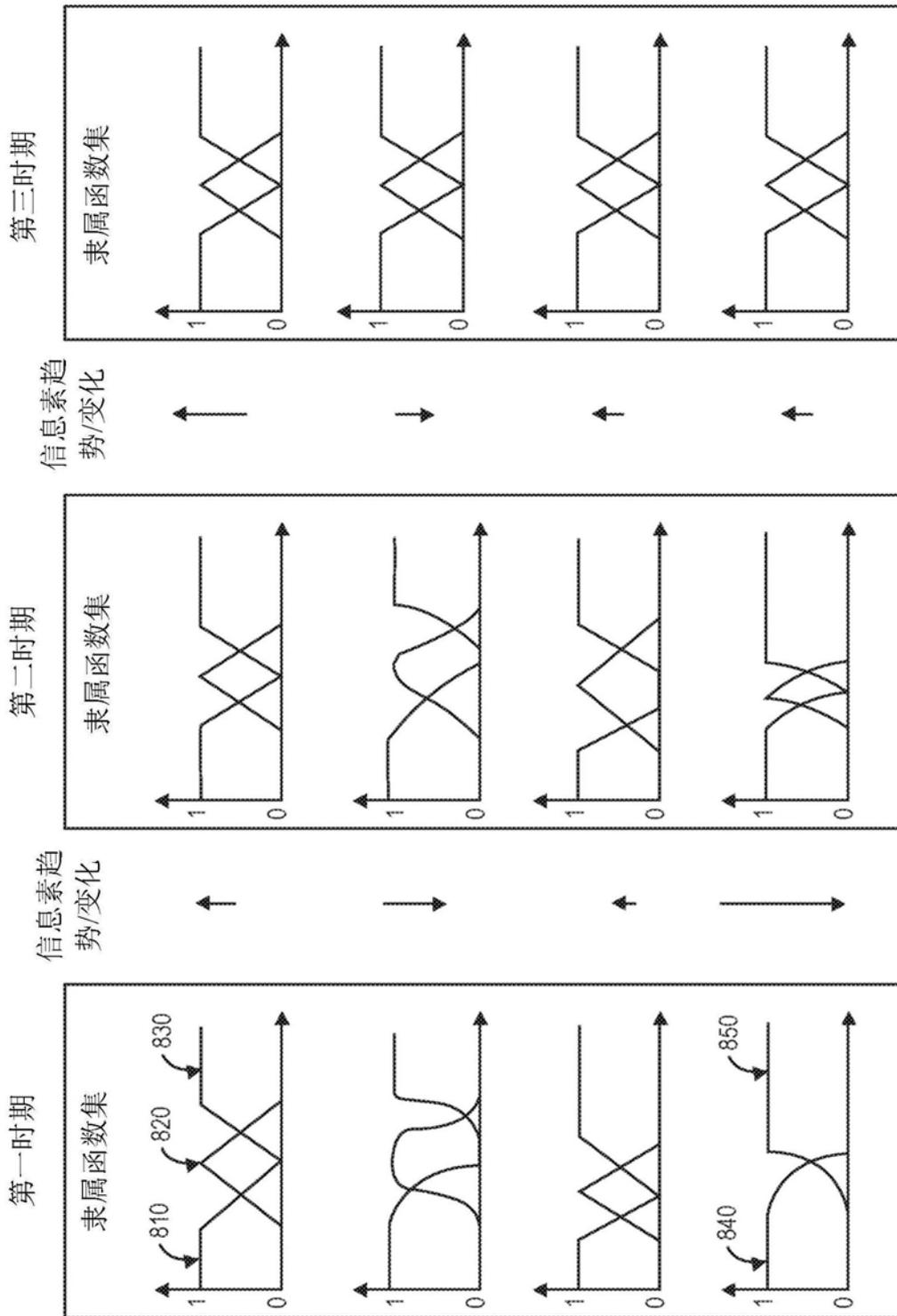


图8