



(21) 申请号 202211492068.7

(22) 申请日 2022.11.25

(65) 同一申请的已公布的文献号

申请公布号 CN 115733763 A

(43) 申请公布日 2023.03.03

(73) 专利权人 中国银联股份有限公司

地址 201203 上海市浦东新区含笑路36号

(72) 发明人 刘红宝 何朔 高鹏飞 郑建宾

汤韬 邱震尧

(74) 专利代理机构 北京市中伦律师事务所

11410

专利代理师 操寒

(51) Int. Cl.

H04L 41/142 (2022.01)

H04L 9/40 (2022.01)

(56) 对比文件

CN 113095946 A, 2021.07.09

CN 113254717 A, 2021.08.13

审查员 李玲

权利要求书3页 说明书10页 附图6页

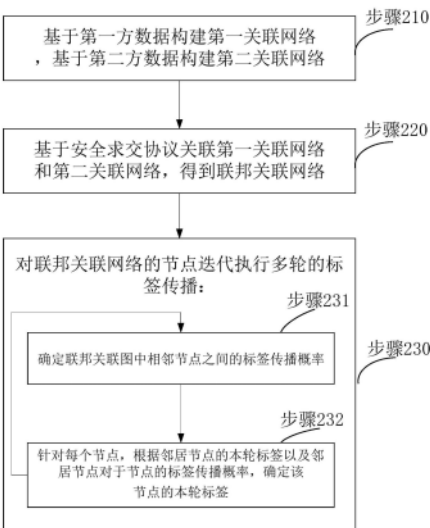
(54) 发明名称

一种关联网络的标签传播方法、装置及计算机可读存储介质

(57) 摘要

本发明提供了一种关联网络的标签传播方法、装置及计算机可读存储介质,该方法包括:基于第一方数据构建第一关联网络,基于第二方数据构建第二关联网络;基于安全求交协议关联第一关联网络和第二关联网络,得到联邦关联网络;对联邦关联网络的节点迭代执行多轮的标签传播;其中,每轮标签传播包括:确定联邦关联图中相邻节点之间的标签传播概率;针对每个节点,根据邻居节点的本轮标签以及邻居节点对于节点的标签传播概率,确定每个节点的本轮标签。利用上述方法,能够在保证隐私数据的前提下实现跨平台网络的标签传播。

200



1. 一种关联网络的标签传播方法,其特征在于,包括:
基于第一方数据构建第一关联网络,基于第二方数据构建第二关联网络;
基于安全求交协议关联所述第一关联网络和所述第二关联网络,得到联邦关联网络;
对所述联邦关联网络的节点迭代执行多轮的标签传播;
其中,每轮所述标签传播包括:确定所述联邦关联网络中节点i与其邻居节点j之间的边ij的边权重 w_{ij} ;
确定所述节点i与其全部邻居节点J之间的边权重和 $\sum_j w_{ij}$;
根据所述边权重 w_{ij} 和所述边权重和 $\sum_j w_{ij}$ 的比值,确定所述邻居节点j对于所述节点i的标签传播概率 P_{ij} ;
针对每个节点,根据邻居节点的本轮标签以及所述邻居节点对于所述节点的标签传播概率,确定每个节点的本轮标签。
2. 根据权利要求1所述的方法,其特征在于,基于安全求交协议关联所述第一关联网络和所述第二关联网络,得到联邦关联网络,还包括:
对所述第一方数据和所述第二方数据进行加密求交,确定所述第一关联网络和所述第二关联网络中的公共节点,根据所述公共节点关联所述第一关联网络和所述第二关联网络,得到联邦关联网络。
3. 根据权利要求1所述的方法,其特征在于,
若所述节点i为非共有节点,所述全部邻居节点J代表所述节点i所在图的全部邻居节点。
4. 根据权利要求1所述的方法,其特征在于,
若所述节点i为共有节点,所述全部邻居节点J代表所述节点i在所述第一关联网络的全部邻居节点a和所述节点i在所述第二关联网络的全部邻居节点b的集合。
5. 根据权利要求1所述的方法,其特征在于,
若所述节点i为共有节点,所述第一方和第三方交互所述节点i与所述第一关联网络的全部邻居节点a之间边权重和以及所述节点i与所述第二关联网络的全部邻居节点b之间的边权重和。
6. 根据权利要求1所述的方法,其特征在于,还包括:
若所述节点i为共有节点,利用以下公式确定所述边权重和 $\sum_j w_{ij}$:
$$\sum_j w_{ij} = \sum_a w_{ia} + \sum_b w_{ib};$$

其中, $\sum_a w_{ia}$ 为所述节点i与所述第一关联网络的全部邻居节点a之间边权重和,所述 $\sum_b w_{ib}$ 为所述节点i与所述第二关联网络的全部邻居节点b之间的边权重和。
7. 根据权利要求1所述的方法,其特征在于,对所述联邦关联网络的节点迭代执行多轮的标签传播,还包括:
确定所述联邦关联网络的标注节点和未标注节点;
逐轮更新所述未标注节点的标签,直至所述未标注节点的标签不再发生变化和/或超过更新轮次阈值;以及,
保持所述标注节点的标签不变。
8. 根据权利要求1所述的方法,其特征在于,根据邻居节点的本轮标签以及所述邻居节点对于所述节点的标签传播概率,确定每个节点的本轮标签,包括:

针对每个节点,确定所述节点的每个邻居节点的本轮标签,以及每个邻居节点对于所述节点的标签传播概率;

在所述节点的全部邻居节点中,计算每种标签对应的标签传播概率之和,得到每个标签对应的标签传播聚合概率;

根据具有最大标签传播聚合概率的标签更新所述节点的本轮标签。

9.根据权利要求8所述的方法,其特征在于,若所述节点为非共有节点,所述方法还包括:

所述节点所在方计算所述节点的每种邻居节点标签对应的标签传播聚合概率。

10.根据权利要求8所述的方法,其特征在于,若所述节点为共有节点,所述方法还包括:

所述第一方计算所述节点在第一关联网络中的所有邻居节点标签对应的第一标签传播聚合概率;

所述第二方计算所述节点在第二关联网络中的所有邻居节点标签对应的第二标签传播聚合概率;

所述第一方和所述第二方交互所述第一标签传播聚合概率和第二标签传播聚合概率;

所述第一方和所述第二方各自基于交互信息再次进行标签传播概率聚合,得到每种标签对应的标签传播聚合概率。

11.根据权利要求1所述的方法,其特征在于,还包括:

根据所述第一关联网络和所述第二关联网络的节点关系紧密程度,确定所述第一关联网络和所述第二关联网络的图权重;以及,

在所述第一关联网络和所述第二关联网络的交互过程中引入所述图权重。

12.根据权利要求11所述的方法,其特征在于,在所述第一关联网络和所述第二关联网络的交互过程中引入所述图权重,包括:

若所述节点*i*为共有节点,利用以下公式确定所述边权重和 $\sum_j w_{ij}$:

$$\sum_j w_{ij} = \theta_a \sum_a w_{ia} + \theta_b \sum_b w_{ib};$$

其中, $\sum_a w_{ia}$ 为所述节点*i*与所述第一关联网络的全部邻居节点*a*之间边权重和,所述 $\sum_b w_{ib}$ 为所述节点*i*与所述第二关联网络的全部邻居节点*b*之间的边权重和, θ_a 为所述第一关联网络的图权重, θ_b 为所述第二关联网络的图权重。

13.根据权利要求11所述的方法,其特征在于,在所述第一关联网络和所述第二关联网络的交互过程中引入所述图权重,包括:

所述第一方和所述第二方交互第一标签传播聚合概率和第二标签传播聚合概率之后,基于所述第一关联网络和所述第二关联网络的图权重再次进行标签传播概率聚合,得到每种标签对应的标签传播聚合概率。

14.根据权利要求1所述的方法,其特征在于,还包括:

若所述第一关联网络和所述第二关联网络为有向图网络,仅将每个节点的流入邻居节点作为所述邻居节点。

15.一种关联网络的标签传播装置,其特征在于,包括:

图构建模块,用于基于第一方数据构建第一关联网络,基于第二方数据构建第二关联网络;

联邦网络模块,用于基于安全求交协议关联所述第一关联网络和所述第二关联网络,得到联邦关联网络;

标签传播模块,用于对所述联邦关联网络的节点迭代执行多轮的标签传播;其中,每轮所述标签传播包括:

确定所述联邦关联网络中节点 i 与其邻居节点 j 之间的边 ij 的边权重 w_{ij} ;

确定所述节点 i 与其全部邻居节点 J 之间的边权重和 $\sum_j w_{ij}$;

根据所述边权重 w_{ij} 和所述边权重和 $\sum_j w_{ij}$ 的比值,确定所述邻居节点 j 对于所述节点 i 的标签传播概率 P_{ij} ;

针对每个节点,根据邻居节点的本轮标签以及所述邻居节点对于所述节点的标签传播概率,确定每个节点的本轮标签。

16.一种关联网络的标签传播装置,其特征在于,包括:

至少一个处理器;以及,与至少一个处理器通信连接的存储器;其中,存储器存储有可被至少一个处理器执行的指令,指令被至少一个处理器执行,以使至少一个处理器能够执行:如权利要求1-14中任一项所述的方法。

17.一种计算机可读存储介质,所述计算机可读存储介质存储有程序,当所述程序被多核处理器执行时,使得所述多核处理器执行如权利要求1-14中任一项所述的方法。

一种关联网络的标签传播方法、装置及计算机可读存储介质

技术领域

[0001] 本发明属于计算机领域,具体涉及一种关联网络的标签传播方法、装置及计算机可读存储介质。

背景技术

[0002] 本部分旨在为权利要求书中陈述的本发明的实施方式提供背景或上下文。此处的描述不因为包括在本部分中就承认是现有技术。

[0003] 随着隐私保护法律的趋严,机构间的数据合作越来越需要考虑数据隐私保护问题。目前隐私计算技术主要聚焦于联邦学习、安全求交、匿踪查询等场景,均为针对单点数据进行联合计算。当前关联网络中的标签传播算法其计算的数据源均在本地。无法实现隐私保护前提下的跨机构数据联合应用。标签的更新无法同时利用两方的关联网络数据,数据价值没有高效利用。

[0004] 因此,如何实现隐私保护前提下的联邦网络标签传播是一个亟待解决的问题。

发明内容

[0005] 针对上述现有技术中存在的问题,提出了一种关联网络的标签传播方法、装置及计算机可读存储介质,利用这种方法、装置及计算机可读存储介质,能够解决上述问题。

[0006] 本发明提供了以下方案。

[0007] 第一方面,提供关联网络的标签传播方法,包括:基于第一方数据构建第一关联网络,基于第二方数据构建第二关联网络;基于安全求交协议关联第一关联网络和第二关联网络,得到联邦关联网络;对联邦关联网络的节点迭代执行多轮的标签传播;其中,每轮标签传播包括:确定联邦关联图中相邻节点之间的标签传播概率;针对每个节点,根据邻居节点的本轮标签以及邻居节点对于节点的标签传播概率,确定每个节点的本轮标签。

[0008] 在一种实施方式中,基于安全求交协议关联第一关联网络和第二关联网络,得到联邦关联网络,还包括:对第一方数据和第二方数据进行加密求交,确定第一关联网络和第二关联网络中的公共节点,根据公共节点关联第一关联网络和第二关联网络,得到联邦关联网络;

[0009] 在一种实施方式中,确定联邦关联图中相邻节点之间的标签传播概率,还包括:确定联邦关联网络中节点 i 与其邻居节点 j 之间的边 ij 的边权重 w_{ij} ;确定节点 i 与其全部邻居节点 J 之间的边权重和 $\sum_j w_{ij}$;根据边权重 w_{ij} 和边权重和 $\sum_j w_{ij}$ 的比值,确定邻居节点 j 对于节点 i 的标签传播概率 P_{ij} 。

[0010] 在一种实施方式中,若节点 i 为非共有节点,全部邻居节点 J 代表节点 i 所在图的全部邻居节点。

[0011] 在一种实施方式中,若节点 i 为共有节点,全部邻居节点 J 代表节点 i 在第一关联网络的全部邻居节点 a 和节点 i 在第二关联网络的全部邻居节点 b 的集合。

[0012] 在一种实施方式中,若节点 i 为共有节点,第一方和第二方交互节点 i 与第一关联

网络的全部邻居节点a之间边权重和以及节点i与第二关联网络的全部邻居节点b之间的边权重和。

[0013] 在一种实施方式中,还包括:若节点i为共有节点,利用以下公式确定边权重和 $\Sigma_j w_{ij}$: $\Sigma_j w_{ij} = \Sigma_a w_{ia} + \Sigma_b w_{ib}$;其中, $\Sigma_a w_{ia}$ 为节点i与第一关联网络的全部邻居节点a之间边权重和, $\Sigma_b w_{ib}$ 为节点i与第二关联网络的全部邻居节点b之间的边权重和。

[0014] 在一种实施方式中,对联邦关联网络的节点迭代执行多轮的标签传播,还包括:确定联邦关联网络的标注节点和未标注节点;逐轮更新未标注节点的标签,直至未标注节点的标签不再发生变化和/或超过更新轮次阈值;以及,保持标注节点的标签不变。

[0015] 在一种实施方式中,根据邻居节点的本轮标签以及邻居节点对于节点的标签传播概率,确定每个节点的本轮标签,包括:针对每个节点,确定节点的每个邻居节点的本轮标签,以及每个邻居节点对于节点的标签传播概率;在节点的全部邻居节点中,计算每种标签对应的标签传播概率之和,得到每个标签对应的标签传播聚合概率;根据具有最大标签传播聚合概率的标签更新节点的本轮标签。

[0016] 在一种实施方式中,若节点为非共有节点,方法还包括:节点所在方计算节点的每种邻居节点标签对应的标签传播聚合概率。

[0017] 在一种实施方式中,若节点为共有节点,方法还包括:第一方计算节点在第一关联网络中的所有邻居节点标签对应的第一方标签传播聚合概率;第二方计算节点在第二关联网络中的所有邻居节点标签对应的第二标签传播聚合概率;第一方和第二方交互第一标签传播聚合概率和第二标签传播聚合概率;第一方和第二方各自基于交互信息再次进行标签传播概率聚合,得到每种标签对应的标签传播聚合概率。

[0018] 在一种实施方式中,还包括:根据第一关联网络和第二关联网络的节点关系紧密程度,确定第一关联网络和第二关联网络的图权重;以及,在第一关联网络和第二关联网络的交互过程中引入图权重。

[0019] 在一种实施方式中,在第一关联网络和第二关联网络的交互过程中引入图权重,包括:若节点i为共有节点,利用以下公式确定边权重和 $\Sigma_j w_{ij}$:

[0020] $\Sigma_j w_{ij} = \theta_a \Sigma_a w_{ia} + \theta_b \Sigma_b w_{ib}$;其中, $\Sigma_a w_{ia}$ 为节点i与第一关联网络的全部邻居节点a之间边权重和, $\Sigma_b w_{ib}$ 为节点i与第二关联网络的全部邻居节点b之间的边权重和, θ_a 为第一关联网络的图权重, θ_b 为第二关联网络的图权重。

[0021] 在一种实施方式中,在第一关联网络和第二关联网络的交互过程中引入图权重,还包括:第一方和第二方交互第一标签传播聚合概率和第二标签传播聚合概率之后,基于第一关联网络和第二关联网络的图权重再次进行标签传播概率聚合,得到每种标签对应的标签传播聚合概率。

[0022] 在一种实施方式中,还包括:若第一关联网络和第二关联网络为有向图网络,仅将每个节点的流入邻居节点作为邻居节点。

[0023] 第二方面,提供一种关联网络的标签传播装置,包括:图构建模块,用于基于第一方数据构建第一关联网络,基于第二方数据构建第二关联网络;联邦网络模块,用于基于安全求交协议关联第一关联网络和第二关联网络,得到联邦关联网络;标签传播模块,用于对联邦关联网络的节点迭代执行多轮的标签传播;其中,每轮标签传播包括:确定联邦关联图中相邻节点之间的标签传播概率;针对每个节点,根据邻居节点的本轮标签以及邻居节点

对于节点的标签传播概率,确定每个节点的本轮标签。

[0024] 第三方面,提供一种关联网络的标签传播装置,包括:至少一个处理器;以及,与至少一个处理器通信连接的存储器;其中,存储器存储有可被至少一个处理器执行的指令,指令被至少一个处理器执行,以使至少一个处理器能够执行:如第一方面的方法。

[0025] 第四方面,提供一种计算机可读存储介质,计算机可读存储介质存储有程序,当程序被多核处理器执行时,使得多核处理器执行如第一方面的方法。

[0026] 上述实施方式的优点之一,能够在保证隐私数据的前提下实现跨平台网络的标签传播。。

[0027] 本发明的其他优点将配合以下的说明和附图进行更详细的解说。

[0028] 应当理解,上述说明仅是本发明技术方案的概述,以便能够更清楚地了解本发明的技术手段,从而可依照说明书的内容予以实施。为了让本发明的上述和其它目的、特征和优点能够更明显易懂,以下特举例说明本发明的具体实施方式。

附图说明

[0029] 通过阅读下文的示例性实施方式的详细描述,本领域普通技术人员将明白本文所述的优点和益处以及其他优点和益处。附图仅用于示出示例性实施方式的目的,而并不认为是对本发明的限制。而且在整个附图中,用相同的标号表示相同的部件。在附图中:

[0030] 图1为根据本发明一实施方式的关联网络的标签传播设备的结构示意图;

[0031] 图2为根据本发明一实施方式的关联网络的标签传播方法的流程示意图;

[0032] 图3为根据本发明一实施方式的第一关联网络和第二关联网络的示意图;

[0033] 图4为根据本发明一实施方式的联邦关联网络的示意图;

[0034] 图5为根据本发明一实施方式的确定第一关联网络和第二关联网络的标签传播概率的示意图;

[0035] 图6为根据本发明一实施方式的确定联邦关联网络的标签传播概率的示意图;

[0036] 图7为根据本发明一实施方式的关联网络的标签传播的示意图;

[0037] 图8为根据本发明一实施方式的关联网络的标签传播的示意图;

[0038] 图9为根据本发明一实施方式的关联网络的标签传播装置的结构示意图。

[0039] 在附图中,相同或对应的标号表示相同或对应的部分。

具体实施方式

[0040] 下面将参照附图更详细地描述本公开的示例性实施方式。虽然附图中显示了本公开的示例性实施方式,然而应当理解,可以以各种形式实现本公开而不应被这里阐述的实施方式所限制。相反,提供这些实施方式是为了能够更透彻地理解本公开,并且能够将本公开的范围完整的传达给本领域的技术人员。

[0041] 在本申请实施方式的描述中,应理解,诸如“包括”或“具有”等术语旨在指示本说明书中所公开的特征、数字、步骤、行为、部件、部分或其组合的存在,并且不旨在排除一个或多个其他特征、数字、步骤、行为、部件、部分或其组合存在的可能性。

[0042] 除非另有说明,“/”表示或的意思,例如,A/B可以表示A或B;本文中的“和/或”仅仅是一种描述关联对象的关联关系,表示可以存在三种关系,例如,A和/或B,可以表示:单独

存在A,同时存在A和B,单独存在B这三种情况。

[0043] 术语“第一”、“第二”等仅用于描述目的,而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量。由此,限定有“第一”、“第二”等的特征可以明示或者隐含地包括一个或者更多个该特征。在本申请实施方式的描述中,除非另有说明,“多个”的含义是两个或两个以上。

[0044] 为清楚阐述本申请实施方式,首先将介绍一些后续实施方式中可能会出现概念。

[0045] 下面将参考附图并结合实施方式来详细说明本发明。

[0046] 首先参见图1,其示意性示出了其中可以使用根据本公开的示例性实现方式的环境100的示意图。

[0047] 图1示出了根据本公开的実施方式的计算设备100的示例的示意图。需要说明的是,图1即可为关联网络的标签传播方法的硬件运行环境的结构示意图。本发明实施方式基于关联网络的标签传播方法设备可以是PC,便携计算机等终端设备。

[0048] 如图1所示,该关联网络的标签传播方法设备可以包括:处理器1001,例如CPU,网络接口1004,用户接口1003,存储器1005,通信总线1002。其中,通信总线1002用于实现这些组件之间的连接通信。用户接口1003可以包括显示屏(Display)、输入单元比如键盘(Keyboard),可选用户接口1003还可以包括标准的有线接口、无线接口。网络接口1004可选的可以包括标准的有线接口、无线接口(如WI-FI接口)。存储器1005可以是高速RAM存储器,也可以是稳定的存储器(non-volatile memory),例如磁盘存储器。存储器1005可选的还可以是独立于前述处理器1001的存储装置。

[0049] 本领域技术人员可以理解,图1中示出的关联网络的标签传播设备结构并不构成对关联网络的标签传播方法设备的限定,可以包括比图示更多或更少的部件,或者组合某些部件,或者不同的部件布置。

[0050] 如图1所示,作为一种计算机存储介质的存储器1005中可以包括操作系统、网络通信模块、用户接口模块以及关联网络的标签传播方法程序。其中,操作系统是管理和控制关联网络的标签传播设备硬件和软件资源的程序,支持关联网络的标签传播程序以及其它软件或程序的运行。

[0051] 在图1所示的关联网络的标签传播设备中,用户接口1003主要用于接收第一终端、第二终端和监管终端发送的请求、数据等;网络接口1004主要用于连接后台服务器与后台服务器进行数据通信;而处理器1001可以用于调用存储器1005中存储的关联网络的标签传播程序,并执行以下操作:

[0052] 基于第一方数据构建第一关联网络,基于第二方数据构建第二关联网络;基于安全求交协议关联第一关联网络和第二关联网络,得到联邦关联网络;对联邦关联网络的节点迭代执行多轮的标签传播;其中,每轮标签传播包括:确定联邦关联图中相邻节点之间的标签传播概率;针对每个节点,根据邻居节点的本轮标签以及邻居节点对于节点的标签传播概率,确定每个节点的本轮标签。

[0053] 由此,双方仅需要交互标签传播概率等非隐私数据,即可能能够在双方原始数据不出库的前提下,进行跨机构数据关联网络联合计算与应用方案。

[0054] 图2示出了根据本公开的實施方式的用于执行一种关联网络的标签传播方法的流

程图。该方法例如可以由如图1所示的计算设备100来执行。应当理解的是,方法200还可以包括未示出的附加框和/或可以省略所示出的框,本公开的范围在此方面不受限制。

[0055] 步骤210,基于第一方数据构建第一关联网络,基于第二方数据构建第二关联网络;

[0056] 例如,参见图3,A方与B方分别基于自身的数据,形成关联网络中的节点和边。假设A方为银行,通过用户间的转账数据形成A方关联网络为转账关联网络,其中,用户手机号为关联网络中的节点,对有转账关系的节点则进行连边,将转账金额为节点间的边权重值。B方为运营商,通过用户间的通话数据形成B方关联网络为通话关联网络,用户手机号为关联网络中的节点,对有通话记录的节点则进行连边,将通话次数为节点间的边权重值。可选地,各自关联网络的边权重值可进行归一化。

[0057] 步骤220,基于安全求交协议关联第一关联网络和第二关联网络,得到联邦关联网络;

[0058] 在一种实施方式中,上述步骤220进一步包括:对第一方数据和第二方数据进行加密求交,确定第一关联网络和第二关联网络中的公共节点,根据公共节点关联第一关联网络和第二关联网络,得到联邦关联网络。

[0059] 例如,参见图4,可以利用安全求交算法(比如基于RSA+HASH的隐私求交算法),对两方节点数据进行安全求交,在不暴露原始数据的前提下发现共有节点,从而形成虚拟的联邦关联网络。如图4所示,可以将图3中的第一关联网络和第二关联网络关联起来,得到该联邦关联网络。其中, V_a 代表A方的节点, V_b 代表B方的节点, V_{ab} 代表双方共有节点, V_{ab1} 、 V_{ab2} 、 V_{ab3} 为双方的共有节点。以节点 V_{ab1} 示例,单独从A方关联网络来看, V_{ab1} 的邻居节点有2个。而从全局数据来看, V_{ab1} 的邻居节点有4个。

[0060] 步骤230,对联邦关联网络的节点迭代执行多轮的标签传播;

[0061] 具体地,联邦关联网络的节点可以包括有标注节点和未标注节点,比如,可以在第一关联网络和第二关联网络中分别存在一部分未标注节点。又比如,第一关联网络中全为有标注节点,第二关联网络中全为未标注节点。诸如此类。

[0062] 在一种实施方式中,针对上述联邦关联网络同时包括有标注节点和未标注节点的情况,上述步骤230可以进一步包括如下步骤:首先,确定联邦关联网络的标注节点和未标注节点;逐轮更新未标注节点的标签,直至未标注节点的标签不再发生变化和/或超过更新轮次阈值;以及,保持标注节点的标签不变。如此,可以保证原始样本的标签不变,保证标签传播准确度。

[0063] 可选地,标注节点的标签也可以逐轮动态更新,也即逐轮更新联邦关联网络中所有节点的标签,直至节点的标签不再发生变化和/或超过更新轮次阈值。如此可以矫正原始标签,挖掘出隐藏风险标签。

[0064] 在上述步骤230中,每轮标签传播具体包括以下步骤231-232:

[0065] 步骤231,确定联邦关联图中相邻节点之间的标签传播概率;

[0066] 在一种实施方式中,上述步骤231具体可以包括:

[0067] (1) 确定联邦关联网络中节点 i 与其邻居节点 j 之间的边 ij 的边权重 w_{ij} ;

[0068] (2) 确定节点 i 与其全部邻居节点 J 之间的边权重和 $\sum_j w_{ij}$;

[0069] 在一种实施方式中,若节点 i 为非共有节点,全部邻居节点 J 代表节点 i 所在图的全

部邻居节点。

[0070] 在一种实施方式中,若节点*i*为共有节点,全部邻居节点*J*代表节点*i*在第一关联网络的全部邻居节点*a*和节点*i*在第二关联网络的全部邻居节点*b*的集合。

[0071] 进一步地,若节点*i*为共有节点,第一方和第二方交互节点*i*与第一关联网络的全部邻居节点*a*之间边权重和 $\sum_a w_{ia}$ 以及节点*i*与第二关联网络的全部邻居节点*b*之间的边权重和 $\sum_b w_{ib}$,如此第一方和第二方各自在其方均可以基于双方交互的两方边权重和来计算节点*i*与其全部邻居节点*J*之间的边权重和 $\sum_j w_{ij}$ 。

[0072] 进一步地,在一种实施方式中,若节点*i*为共有节点,基于双方交互的两方边权重和,第一方和第二方均可以利用以下公式确定边权重和 $\sum_j w_{ij}$:

$$[0073] \quad \sum_j w_{ij} = \sum_a w_{ia} + \sum_b w_{ib};$$

[0074] 其中, $\sum_a w_{ia}$ 为节点*i*与第一关联网络的全部邻居节点*a*之间边权重和, $\sum_b w_{ib}$ 为节点*i*与第二关联网络的全部邻居节点*b*之间的边权重和。

[0075] 可选地,在另一种实施方式中,可以进一步考虑由于业务场景对节点关系紧密度的影响。比如,在金融场景中,转账关系为强关系,通话关系为弱关系,因此在计算边的标签传播概率时,可考虑双方不同业务场景下的边关系强弱进行加权聚合。

[0076] 在此情况下,第一方和第二方均可以利用以下公式确定加权边权重和 $\sum_j w_{ij}$:

$$[0077] \quad \sum_j w_{ij} = \theta_a \sum_a w_{ia} + \theta_b \sum_b w_{ib};$$

[0078] 其中, θ_a 为第一关联网络的对应的图权重, θ_b 为第二关联网络的图权重。

[0079] (3) 根据边权重 w_{ij} 与边权重和 $\sum_j w_{ij}$ 的比值,确定邻居节点*j*对于节点*i*的标签传播概率 P_{ij} 。

[0080] 具体地,联邦关联网络各边的标签传播概率 $P_{ij} = \frac{w_{ij}}{\sum_j w_{ij}}$,其中, w_{ij} 代表边*ij*的权重值。此处,对于非共有节点,*J*代表节点*i*的邻居节点;对于共有节点,*J*代表节点*i*分别在A、B两方的所有邻居节点。

[0081] $\sum_j w_{ij}$ 的计算逻辑为,A方计算本方节点*i*的邻居节点权重和为 $\sum_a w_{ia}$;B方计算本方节点*i*的邻居节点权重和为 $\sum_b w_{ib}$ 。双方交互 $\sum_a w_{ia}$ 和 $\sum_b w_{ib}$,得到最终的权重计算分母值为 $\sum_j w_{ij} = \sum_a w_{ia} + \sum_b w_{ib}$ 。

[0082] 参考图5,此处以节点Vab2为例。在A方本地网络中,Vab2的邻居节点有1个,单独计算本方的标签传播概率为 $P=0.1/0.1=1$;B方本地网络中,Vab2的邻居节点有3个,单独计算本方的标签传播概率分别为 $P=0.2/(0.2+0.4+0.8)=1/7$, $P=0.4/(0.2+0.4+0.8)=2/7$, $P=0.8/(0.2+0.4+0.8)=4/7$;进一步地,双方交换目标节点Vab2与邻居节点的权重值,A方为0.1,B方为0.2+0.4+0.8=1.4。结合联邦关联网络,更新目标节点的标签传播概率。

[0083] 参考图6,经过上述计算,在A方,其邻居节点对于该节点*i*的标签传播概率 $P_{ij} = \frac{w_{ij}}{\sum_j w_{ij}} = \frac{0.1}{0.1+0.2+0.8+0.4} = 1/15$;在B方,类似地,其邻居节点对于该节点*i*的标签传播概率分别为2/15,4/15,8/15。

[0084] 步骤232,针对每个节点,根据邻居节点的本轮标签以及邻居节点对于该节点的标签传播概率,确定每个节点的本轮标签。

[0085] 参考图7,此处继续以Vab2所形成的联邦关联网络为例,如下所示。节点5为风险节点,显示为白色节点,标签设置为“1”;其余节点为未知节点,显示为灰色节点,标签设置为

“0”。标签传播过程中,节点5的标签始终为“1”,其余节点的标签则逐轮更新,直到所有节点的标签不再发生变化或超过更新轮次阈值为止。

[0086] 在一种实施方式中,上述步骤232中,进一步包括以下步骤:

[0087] 步骤2321,针对节点 i ,确定节点 i 的每个邻居节点的本轮标签,以及每个邻居节点对于节点 i 的标签传播概率;

[0088] 步骤2322,在节点 i 的全部邻居节点中,计算每种标签对应的标签传播概率之和,得到每个标签对应的标签传播聚合概率;

[0089] 具体地,若节点 i 为非共有节点,则仅由节点 i 所在方计算节点 i 的每种邻居节点标签对应的标签传播聚合概率。

[0090] 具体地,若节点 i 为共有节点,则执行以下步骤:首先,第一方计算节点 i 在第一关联网络中的所有邻居节点标签对应的第一方标签传播聚合概率;第二方计算节点 i 在第二关联网络中的所有邻居节点标签对应的第二标签传播聚合概率;其次,第一方和第二方交互第一标签传播聚合概率和第二标签传播聚合概率;最后,第一方和第二方各自基于交互信息再次进行标签传播概率聚合,得到每种标签对应的标签传播聚合概率。

[0091] 步骤2323,根据具有最大标签传播聚合概率的标签更新节点 i 的本轮标签。

[0092] 上述步骤2321-步骤2323所示出的节点标签更新规则,可以包括如下具体步骤:

[0093] 首先,对于节点 i 的第 T 轮更新,设其邻居节点集合为 $J((J_1, L_1, P_{i1}), (J_2, L_2, P_{i2}), (J_j, L_j, P_{ij}), \dots, (J_n, L_n, P_{in}))$,其中 J_j 为邻居节点 j 的标识, L_j 为邻居节点 j 的标签, P_{ij} 为边 $\langle i, j \rangle$ 的传播概率。

[0094] 其次,计算邻居节点集合的所有标签的聚合传播概率。具体为 $P(L_j) = \sum P_{ij}$,其中, P_{ij} 为标签为 L_j 的邻居节点对于目标节点 i 的标签传播概率。其中,若节点 i 为非共有节点,则只需计算本方邻居节点标签的传播概率。若节点 i 为共有节点,则A方计算本方所有邻居节点标签对应的概率 $P(L_{aj}) = \sum P_{iaj}$,B方计算本方邻居节点标签对应的概率 $P(L_{bj}) = \sum P_{ibj}$,A方和B方交互 $P(L_{aj})$ 与 $P(L_{bj})$,并各自在本方进行再次标签传播概率聚合,得到最终结合双方关联网络信息的标签传播聚合概率 $P(L_j) = P(L_{aj}) + P(L_{bj})$ 。

[0095] 最后,选择最大的 $P(L_j)$ 所对应的标签 L_j 为节点 i 的本轮标签。重复以上步骤,直到所有节点的标签不再发生变化。

[0096] 一个具体示例中,参考图7和图8给出标签更新的具体计算示例。

[0097] 参考图7,针对第一轮传播,对于节点Vab2,执行以下计算:

[0098] (1) 计算A方邻居节点的标签传播聚合概率为 $\langle \text{“0”}, 1/15 \rangle$,其中,“0”代表无风险标签,1/15代表标签“0”对应的标签传播聚合概率。可以理解,由于A方Vab2仅有一个邻居节点1,且其初始标签值为“0”,在上文中已经计算出节点1对节点Vab2的标签传播概率为1/15,因此对于A方Vab2节点来说,其仅存在一个可传播标签“0”,且该可传播标签“0”对应的标签传播聚合概率为1/15。

[0099] (2)、计算B方邻居节点的标签传播聚合概率为 $\langle \text{“0”}, 6/15 \rangle$ 、 $\langle \text{“1”}, 8/15 \rangle$,其中,“0”代表无风险标签,6/15代表标签“0”对应的标签传播聚合概率。“1”代表有风险标签,8/15代表标签“1”对应的标签传播聚合概率。由于B方Vab2有三个邻居节点(3、4、5),且节点3、4的初始标签值为“0”,且节点5的初始标签值为“1”。在上文中已经计算出节点3对节点Vab2的标签传播概率为2/15,节点4对节点Vab2的标签传播概率为4/15,节点5对节点Vab2的标签

传播概率为8/15,因此对于B方Vab2节点来说,其存在2个可传播标签“0”和“1”,且该可传播标签“0”对应的标签传播聚合概率为 $6/15=2/15+4/15$,该可传播标签“1”对应的标签传播聚合概率为8/15。

[0100] (3)、双方交换标签传播聚合概率,并将相同标签对应的标签传播聚合概率累加,可以各自计算得到节点Vab2的标签传播聚合概率为 $\langle\text{“0”}, 7/15\rangle, \langle\text{“1”}, 8/15\rangle$,也即该可传播标签“0”对应的标签传播聚合概率为7/15,该可传播标签“1”对应的标签传播聚合概率为8/15。

[0101] (4)、选择具有最大标签传播聚合概率 $\langle\text{“1”}, 8/15\rangle$ 对应的标签“1”为节点Vab2本轮的标签。其他节点类似上述步骤。

[0102] 经过第一轮传播后,该联邦关联网络更新后的节点标签分布图如图8所示,其中节点Vab1和Vab2均更新为标签“1”。继续下一轮标签传播,直到节点的标签不再发生变化或传播轮次大于一定阈值为止。

[0103] 在一种实施方式中,根据第一关联网络和第二关联网络的节点关系紧密程度,确定第一关联网络和第二关联网络的图权重;以及,在第一关联网络和第二关联网络的交互过程中引入所述图权重。

[0104] 比如,可以根据业务场景确定第一关联网络和第二关联网络为强关联或弱关联,进而可以在计算边的标签传播概率时,引入上述第一关联网络和第二关联网络的图权重。当然,也可以在计算每种标签的标签传播聚合概率时,引入上述第一关联网络和第二关联网络的图权重,本申请对此不作具体限制。

[0105] 在一种实施方式中,在第一关联网络和第二关联网络的交互过程中引入图权重,至少包括如下两种引入方式:

[0106] (1)在上述步骤231中,若节点i为共有节点,利用以下公式确定边权重和 $\sum_j w_{ij}$:

[0107] $\sum_j w_{ij} = \theta_a \sum_a w_{ia} + \theta_b \sum_b w_{ib}$;其中, $\sum_a w_{ia}$ 为节点i与第一关联网络的全部邻居节点a之间边权重和, $\sum_b w_{ib}$ 为节点i与第二关联网络的全部邻居节点b之间的边权重和, θ_a 为第一关联网络的图权重, θ_b 为第二关联网络的图权重。

[0108] (2)在上述步骤232中,第一方和第三方交互第一标签传播聚合概率和第二标签传播聚合概率之后,基于第一关联网络和第二关联网络的图权重再次进行标签传播概率聚合,得到每种标签对应的标签传播聚合概率。例如,若节点i为共有节点,则A方计算本方所有邻居节点标签对应的概率 $P(L_{aj}) = \sum P_{iaj}$,B方计算本方邻居节点标签对应的概率 $P(L_{bj}) = \sum P_{ibj}$,A方和B方交互 $P(L_{aj})$ 与 $P(L_{bj})$,并各自在本方进行再次标签传播概率聚合,得到最终结合双方关联网络信息的标签传播聚合概率 $P(L_j) = \theta_a P(L_{aj}) + \theta_b P(L_{bj})$ 。

[0109] 在一种实施方式中,还包括:若第一关联网络和第二关联网络为有向图网络,仅将每个节点的流入邻居节点作为邻居节点。例如,对于有向图,计算节点的传播概率时,可以只考虑目标节点的流入邻居节点。具体可结合业务场景进行判断。

[0110] 在本说明书的描述中,参考术语“一些可能的实施方式”、“一些实施方式”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施方式或示例描述的具体特征、结构、材料或者特点包含于本发明的至少一个实施方式或示例中。在本说明书中,对上述术语的示意性表述不必针对的是相同的实施方式或示例。而且,描述的具体特征、结构、材料或者特点可以在任一个或多个实施方式或示例中以合适的方式结合。此外,在不相互矛盾

的情况下,本领域的技术人员可以将本说明书中描述的不同实施方式或示例以及不同实施方式或示例的特征进行结合和组合。

[0111] 关于本申请实施方式的方法流程图,将某些操作描述为以一定顺序执行的不同的步骤。这样的流程图属于说明性的而非限制性的。可以将本文中所描述的某些步骤分组在一起并且在单个操作中执行、可以将某些步骤分割成多个子步骤、并且可以以不同于在本文中所示出的顺序来执行某些步骤。可以由任何电路结构和/或有形机制(例如,由在计算机设备上运行的软件、硬件(例如,处理器或芯片实现的逻辑功能)等、和/或其任何组合)以任何方式来实现流程图所示出的各个步骤。

[0112] 基于相同的技术构思,本发明实施方式还提供一种关联网络的标签传播装置,用于执行上述任一实施方式所提供的关联网络的标签传播方法。图9为本发明实施方式提供的一种关联网络的标签传播装置结构示意图。

[0113] 如图9所示,装置900包括:

[0114] 图构建模块910,用于基于第一方数据构建第一关联网络,基于第二方数据构建第二关联网络;

[0115] 联邦网络模块920,用于基于安全求交协议关联所述第一关联网络和所述第二关联网络,得到联邦关联网络;

[0116] 标签传播模块930,用于对所述联邦关联网络的节点迭代执行多轮的标签传播;其中,每轮所述标签传播包括:确定所述联邦关联图中相邻节点之间的标签传播概率;针对每个节点,根据邻居节点的本轮标签以及所述邻居节点对于所述节点的标签传播概率,确定每个节点的本轮标签。

[0117] 需要说明的是,本申请实施方式中的装置可以实现前述方法的实施方式的各个过程,并达到相同的效果和功能,这里不再赘述。

[0118] 根据本申请的一些实施方式,提供了关联网络的标签传播方法的非易失性计算机存储介质,其上存储有计算机可执行指令,该计算机可执行指令设置为在由处理器运行时执行:上述实施方式所述的方法。

[0119] 本申请中的各个实施方式均采用递进的方式描述,各个实施方式之间相同相似的部分互相参见即可,每个实施方式重点说明的都是与其他实施方式的不同之处。尤其,对于装置、设备和计算机可读存储介质实施方式而言,由于其基本相似于方法实施方式,所以其描述进行了简化,相关之处可参见方法实施方式的部分说明即可。

[0120] 本申请实施方式提供的装置、设备和计算机可读存储介质与方法是一一对应的,因此,装置、设备和计算机可读存储介质也具有与其对应的方法类似的有益技术效果,由于上面已经对方法的有益技术效果进行了详细说明,因此,这里不再赘述装置、设备和计算机可读存储介质的有益技术效果。

[0121] 本领域内的技术人员应明白,本发明的实施方式可提供为方法、装置(设备或系统)、或计算机可读存储介质。因此,本发明可采用完全硬件实施方式、完全软件实施方式、或结合软件和硬件方面的实施方式的形式。而且,本发明可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机可读存储介质的形式。

[0122] 本发明是参照根据本发明实施方式的方法、装置(设备或系统)、和计算机可读存

储介质的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0123] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0124] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0125] 在一个典型的配置中,计算设备包括一个或多个处理器(CPU)、输入/输出接口、网络接口和内存。

[0126] 内存可能包括计算机可读介质中的非永久性存储器,随机存取存储器(RAM)和/或非易失性内存等形式,如只读存储器(ROM)或闪存(flash RAM)。内存是计算机可读介质的示例。

[0127] 计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括,但不限于相变内存(PRAM)、静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、其他类型的随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带,磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质,可用于存储可以被计算设备访问的信息。此外,尽管在附图中以特定顺序描述了本发明方法的操作,但是,这并非要求或者暗示必须按照该特定顺序来执行这些操作,或是必须执行全部所示的操作才能实现期望的结果。附加地或备选地,可以省略某些步骤,将多个步骤合并为一个步骤执行,和/或将一个步骤分解为多个步骤执行。

[0128] 虽然已经参考若干具体实施方式描述了本发明的精神和原理,但是应该理解,本发明并不限于所公开的具体实施方式,对各方面的划分也不意味着这些方面中的特征不能组合以进行受益,这种划分仅是为了表述的方便。本发明旨在涵盖所附权利要求的精神和范围内所包括的各种修改和等同布置。

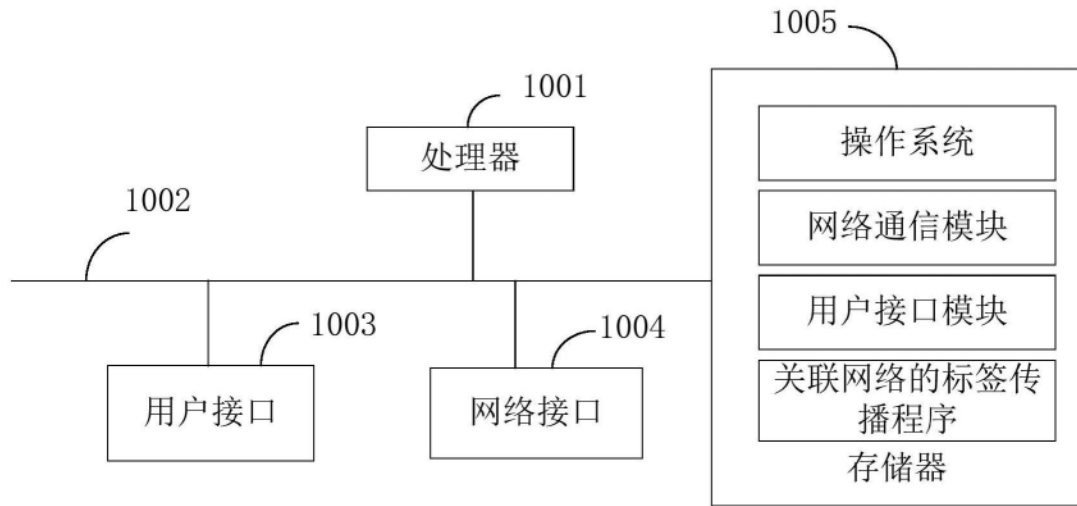


图1

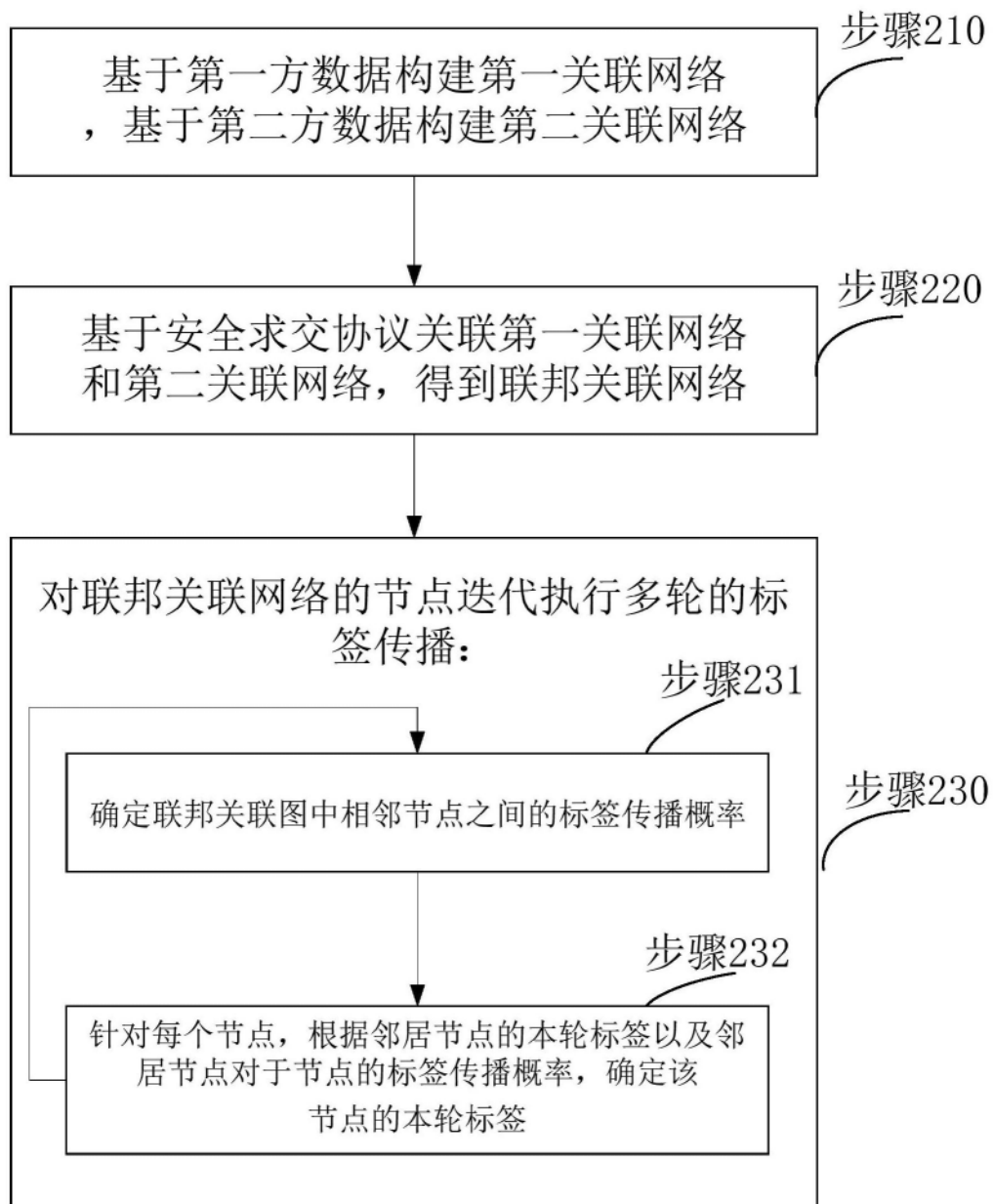
200

图2

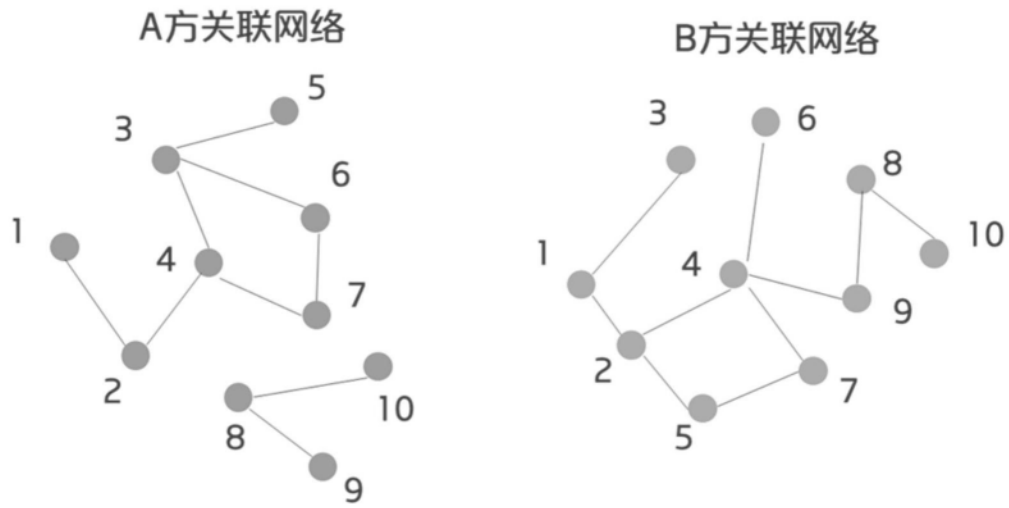


图3

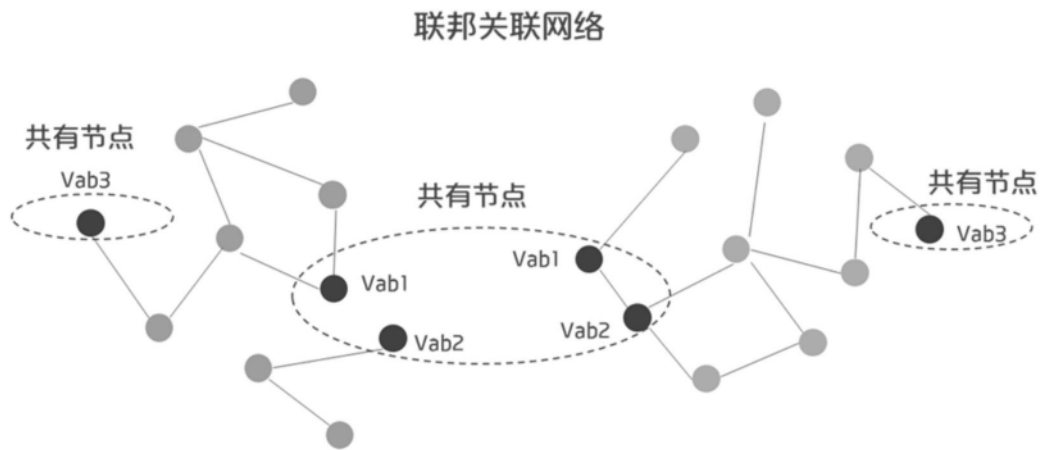


图4

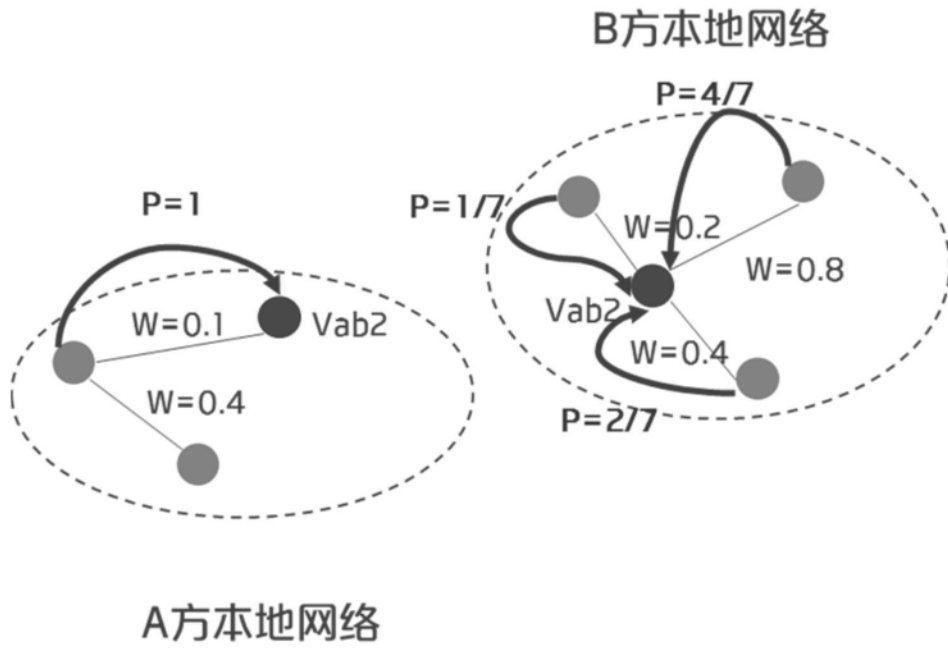


图5

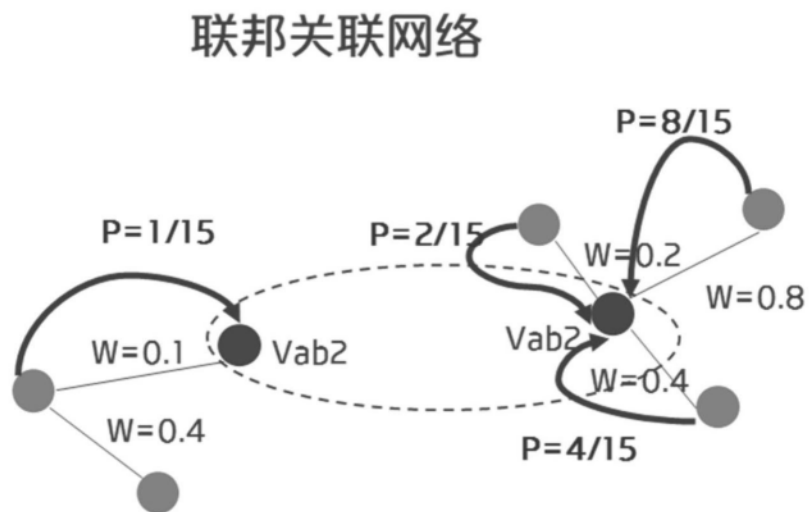


图6

联邦关联网络

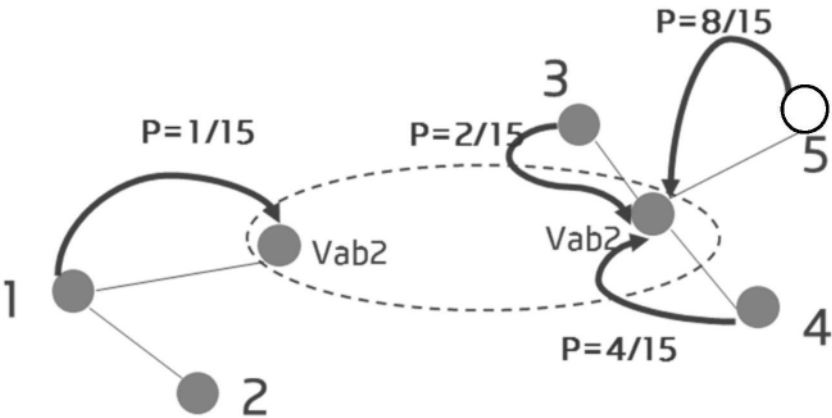


图7

经过一轮传播后的标签分布

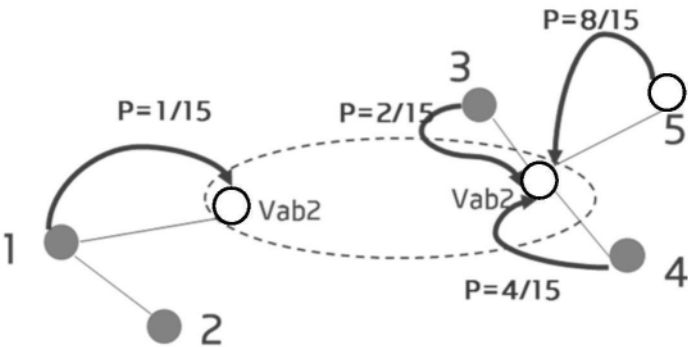


图8

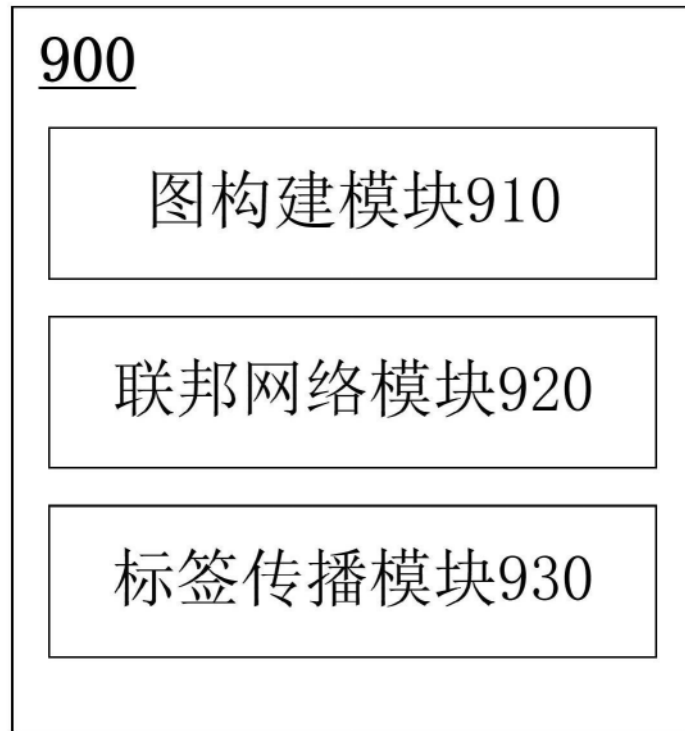


图9