



(12) 发明专利申请

(10) 申请公布号 CN 117829993 A

(43) 申请公布日 2024. 04. 05

(21) 申请号 202311727253.4

(22) 申请日 2023.12.15

(71) 申请人 杭州富算科技有限公司

地址 310051 浙江省杭州市滨江区西兴街
道缤纷街615号4楼401室

(72) 发明人 尤志强 赵东 杜吉锋 杨云波
蔡晓娟 王兆凯 卞阳 张伟奇

(74) 专利代理机构 北京慧加伦知识产权代理有
限公司 16035

专利代理师 李永敏

(51) Int. Cl.

G06Q 40/04 (2012.01)

G06Q 10/063 (2023.01)

G06F 16/9535 (2019.01)

权利要求书3页 说明书12页 附图5页

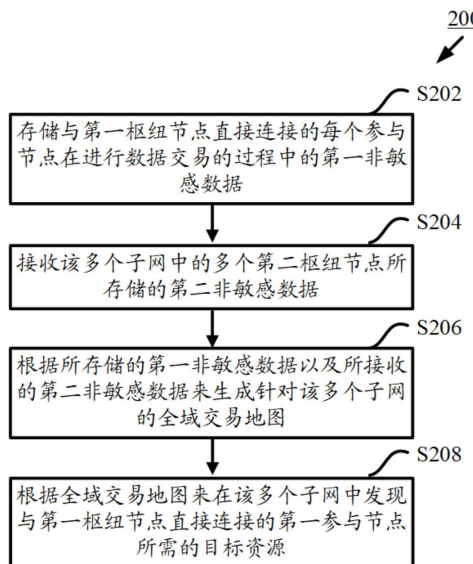
(54) 发明名称

在相互连接的多个子网中发现目标资源的方法及装置

(57) 摘要

本公开的实施例提供一种在相互连接的多个子网中发现目标资源的方法和装置。每个子网包括枢纽节点和与枢纽节点直接连接的多个参与节点。该多个子网中的枢纽节点相互直接连接。每个参与节点经由与其直接连接的枢纽节点与其它参与节点进行数据交易。该方法由第一枢纽节点执行。该方法包括：存储与第一枢纽节点直接连接的每个参与节点在进行数据交易的过程中的第一非敏感数据；接收该多个子网中的多个第二枢纽节点所存储的第二非敏感数据；根据所存储的第一非敏感数据以及所接收的第二非敏感数据来生成针对该多个子网的全域交易地图；根据第一和第二非敏感数据生成全域交易地图；根据全域交易地图来在该多个子网中发现与第一枢纽节点直接连接的第一参与节点所需的目标资源。第一和第二非敏感数据二者都不泄露任一参与节点的原始数据本身。

CN 117829993 A



1. 一种在相互连接的多个子网中发现目标资源的方法,其特征在于,每个子网包括枢纽节点和与所述枢纽节点直接连接的多个参与节点,所述多个子网中的枢纽节点相互直接连接,每个参与节点经由与其直接连接的枢纽节点与所述多个子网中的其它参与节点进行数据交易,所述方法由所述多个子网中的第一枢纽节点执行,所述方法包括:

存储与所述第一枢纽节点直接连接的每个参与节点在进行数据交易的过程中的第一非敏感数据;

接收所述多个子网中的多个第二枢纽节点所存储的第二非敏感数据,其中,每个第二枢纽节点存储与其直接连接的每个参与节点在进行数据交易的过程中的第二非敏感数据;

根据所存储的第一非敏感数据以及所接收的第二非敏感数据来生成针对所述多个子网的全域交易地图;以及

根据所述全域交易地图来在所述多个子网中发现与所述第一枢纽节点直接连接的第一参与节点所需的目标资源;

其中,所述第一非敏感数据和所述第二非敏感数据二者都不泄露任一参与节点的原始数据本身。

2. 根据权利要求1所述的方法,其特征在于,所述第一非敏感数据和所述第二非敏感数据中的每一个包括相关数据交易过程中涉及的数据源的类型,根据所存储的第一非敏感数据以及所接收的第二非敏感数据来生成针对所述多个子网的全域交易地图包括:

从所述第一非敏感数据和所述第二非敏感数据中获取所述多个子网中的所有数据交易过程中涉及的数据源的类型;

按照所获取的数据源的类型初始化全域交易地图,其中,同一类型的数据源对应所述全域交易地图中的一个顶点;

在所述全域交易地图中将存在数据交易的顶点两两之间通过边来连接;

计算所述全域交易地图中每条边所连接的两个顶点之间进行数据交易的次数;以及

根据每条边对应的数据交易的次数来计算该条边的权重,其中,每条边的权重等于该条边对应的数据交易的次数除以所述全域交易地图中所有边对应的数据交易的次数之和。

3. 根据权利要求2所述的方法,其特征在于,根据所述全域交易地图来在所述多个子网中发现与所述第一枢纽节点直接连接的第一参与节点所需的目标资源包括:

根据所述第一参与节点拥有的数据源的类型来在所述全域交易地图中查找与所述第一参与节点相对应的第一顶点;

计算所述第一顶点与所述全域交易地图中的其它顶点之间的关系强度;

按照所计算的关系强度的降序来向所述第一参与节点推荐所述其它顶点上的资源作为所述目标资源;

其中,所述第一顶点与所述其它顶点中的第二顶点之间的关系强度被计算为:

$$c = \sum_{k=1}^m wk / m,$$

其中,c表示所述第一顶点与所述第二顶点之间的关系强度,m表示所述第一顶点与所述第二顶点在所述全域交易地图中的最短路径的跳数,wk表示所述最短路径中的第k边的权重。

4. 根据权利要求1所述的方法,其特征在于,所述第一非敏感数据和所述第二非敏感数据中的一个或多个包括相关数据交易过程中的数据交易双方之间的数据交集大小以及数据交易双方自身的数据集大小,根据所存储的第一非敏感数据以及所接收的第二非敏感数据来生成针对所述多个子网的全域交易地图包括:

将每个参与节点作为所述全域交易地图中的一个顶点;以及

在所述全域交易地图中将数据交集大小已知的顶点两两之间通过边来连接,其中,所述边的权重等于所述数据交集大小。

5. 根据权利要求4所述的方法,其特征在于,根据所述全域交易地图来在所述多个子网中发现与所述第一枢纽节点直接连接的第一参与节点所需的目标资源包括:

确定所述全域交易地图中与所述第一参与节点所对应的第一顶点不相连的所有第三顶点;

针对每个第三顶点,在所述全域交易地图中寻找与所述第一顶点和所述第三顶点二者都相连的第四顶点,其中,所述第四顶点的数据集大小大于所述第一顶点的数据集大小和所述第三顶点的数据集大小中的每一个;

根据所述第一顶点与所述第四顶点之间的数据交集大小以及所述第三顶点与所述第四顶点之间的数据交集大小来估计所述第一顶点与所述第三顶点之间的数据交集大小;以及

按照每个第三顶点与所述第一顶点之间的估计的数据交集大小以及每个第四顶点与所述第一顶点之间的数据交集大小的降序来向所述第一参与节点推荐所述第三顶点和所述第四顶点上的资源作为所述目标资源。

6. 根据权利要求1所述的方法,其特征在于,所述第一非敏感数据和所述第二非敏感数据中的每一个包括以下中的至少一个:相关数据源的交易成功转化率、所述相关数据源的第一复购率、所述相关数据源的第二复购率,

其中,所述相关数据源的交易成功转化率被计算为所述相关数据源的试使用次数与所述相关数据源的成功交易次数之商;

所述相关数据源的第一复购率被计算为复购相关数据源的用户数量与购买过相关数据源的用户数量之商;

所述相关数据源的第二复购率被计算为复购相关数据源的交易次数与购买过相关数据源的用户数量之商。

7. 根据权利要求6所述的方法,其特征在于,根据所述全域交易地图来在所述多个子网中发现与所述第一枢纽节点直接连接的第一参与节点所需的目标资源包括:

针对每个参与节点,根据以下中的至少一个来生成数据价值评估值:所述相关数据源的交易成功转化率、所述相关数据源的第一复购率、所述相关数据源的第二复购率;以及

按照所述数据价值评估值的降序来向所述第一参与节点推荐所述多个子网中其它参与节点上的资源作为所述目标资源。

8. 根据权利要求1至7中任一项所述的方法,其特征在于,所述第一非敏感数据和所述第二非敏感数据中的一个或多个还包括以下中的至少一个:数据交易双方的节点标识符、数据交易的任务标识符、数据交易的开始时间、数据交易的结束时间、数据交易的中间结果、数据交易结果的转售信息、关于数据交易是否达成的信息、数据交易中涉及的数据类别

及简要描述信息、参与节点的计算任务稳定性信息、参与节点的履约效率。

9. 一种在相互连接的多个子网中发现目标资源的装置,其特征在于,每个子网包括枢纽节点和与所述枢纽节点直接连接的多个参与节点,所述多个子网中的枢纽节点相互直接连接,每个参与节点经由与其直接连接的枢纽节点与所述多个子网中的其它参与节点进行数据交易,所述装置被布置在所述多个子网中的第一枢纽节点处,所述装置包括:

至少一个处理器;以及

存储有计算机程序的至少一个存储器;

其中,当所述计算机程序由所述至少一个处理器执行时,使得所述装置执行根据权利要求1至8中任一项所述的方法的步骤。

10. 一种存储有计算机程序的计算机可读存储介质,其特征不在于,所述计算机程序在由处理器执行时实现根据权利要求1至8中任一项所述的方法的步骤。

在相互连接的多个子网中发现目标资源的方法及装置

技术领域

[0001] 本公开的实施例涉及电数字数据处理技术领域,具体地,涉及在相互连接的多个子网中发现目标资源的方法及装置。

背景技术

[0002] 随着互联网的发展,各类政务主体、行业主体、公司主体、机构主体可经由互联网被关联起来。每个主体可被看作一个节点。在互联的主体数量较大的场景下要执行安全的批量式的高质量数据探查,存在很大的困难。现有的探查方式,往往只能通过点对点的质量评估技术来进行单点识别,或者利用联邦模型来进行数据试使用,然后从该联邦模型的指标效果以及对端节点的特征贡献度来判断数据的质量和和价值。这类方式效率低、单点化,不能捕捉更大范围甚至全局相关节点的数据价值影响力信息和质量评估信息,且无法有效发挥网络拓扑能力,难以提供高效的批量化的高质量资源发现方法。

发明内容

[0003] 本文中描述的实施例提供了一种在相互连接的多个子网中发现目标资源的方法、装置以及存储有计算机程序的计算机可读存储介质。

[0004] 根据本公开的第一方面,提供了一种在相互连接的多个子网中发现目标资源的方法。每个子网包括枢纽节点和与枢纽节点直接连接的多个参与节点。该多个子网中的枢纽节点相互直接连接。每个参与节点经由与其直接连接的枢纽节点与该多个子网中的其它参与节点进行数据交易。该方法由该多个子网中的第一枢纽节点执行。该方法包括:存储与第一枢纽节点直接连接的每个参与节点在进行数据交易的过程中的第一非敏感数据;接收该多个子网中的多个第二枢纽节点所存储的第二非敏感数据,其中,每个第二枢纽节点存储与其直接连接的每个参与节点在进行数据交易的过程中的第二非敏感数据;根据所存储的第一非敏感数据以及所接收的第二非敏感数据来生成针对该多个子网的全域交易地图;以及根据全域交易地图来在该多个子网中发现与第一枢纽节点直接连接的第一参与节点所需的目标资源。其中,第一非敏感数据和第二非敏感数据二者都不泄露任一参与节点的原始数据本身。

[0005] 在本公开的一些实施例中,第一非敏感数据和第二非敏感数据中的每一个包括相关数据交易过程中涉及的数据源的类型。根据所存储的第一非敏感数据以及所接收的第二非敏感数据来生成针对多个子网的全域交易地图包括:从第一非敏感数据和第二非敏感数据中获取多个子网中的所有数据交易过程中涉及的数据源的类型;按照所获取的数据源的类型初始化全域交易地图,其中,同一类型的数据源对应全域交易地图中的一个顶点;在全域交易地图中将存在数据交易的顶点两两之间通过边来连接;计算全域交易地图中每条边所连接的两个顶点之间进行数据交易的次数;以及根据每条边对应的数据交易的次数来计算该条边的权重,其中,每条边的权重等于该条边对应的数据交易的次数除以全域交易地图中所有边对应的数据交易的次数之和。

[0006] 在本公开的一些实施例中,根据全域交易地图来在多个子网中发现与第一枢纽节点直接连接的第一参与节点所需的目标资源包括:根据第一参与节点拥有的数据源的类型来在全域交易地图中查找与第一参与节点相对应的第一顶点;计算第一顶点与全域交易地图中的其它顶点之间的关系强度;按照所计算的关系强度的降序来向第一参与节点推荐其它顶点上的资源作为目标资源;其中,第一顶点与其它顶点中的第二顶点之间的关系强度被计算为:

$$[0007] \quad c = \sum_{k=1}^m wk / m,$$

[0008] 其中,c表示第一顶点与第二顶点之间的关系强度,m表示第一顶点与第二顶点在全域交易地图中的最短路径的跳数,wk表示最短路径中的第k边的权重。

[0009] 在本公开的一些实施例中,第一非敏感数据和第二非敏感数据中的一个或多个包括相关数据交易过程中的数据交易双方之间的数据交集大小以及数据交易双方自身的数据集大小。根据所存储的第一非敏感数据以及所接收的第二非敏感数据来生成针对多个子网的全域交易地图包括:将每个参与节点作为全域交易地图中的一个顶点;以及在全域交易地图中将数据交集大小已知的顶点两两之间通过边来连接。其中,边的权重等于数据交集大小。

[0010] 在本公开的一些实施例中,根据全域交易地图来在多个子网中发现与第一枢纽节点直接连接的第一参与节点所需的目标资源包括:确定全域交易地图中与第一参与节点所对应的第一顶点不相连的所有第三顶点;针对每个第三顶点,在全域交易地图中寻找与第一顶点和第三顶点二者都相连的第四顶点,其中,第四顶点的数据集大小大于第一顶点的数据集大小和第三顶点的交集大小中的每一个;根据第一顶点与第四顶点之间的数据交集大小以及第三顶点与第四顶点之间的数据交集大小来估计第一顶点与第三顶点之间的数据交集大小;以及按照每个第三顶点与第一顶点之间的估计的数据交集大小以及每个第四顶点与第一顶点之间的数据交集大小的降序来向第一参与节点推荐第三顶点和第四顶点上的资源作为目标资源。

[0011] 在本公开的一些实施例中,第一非敏感数据和第二非敏感数据中的每一个包括以下中的至少一个:相关数据源的交易成功转化率、相关数据源的第一复购率、相关数据源的第二复购率。其中,相关数据源的交易成功转化率被计算为相关数据源的试使用次数与相关数据源的成功交易次数之商。相关数据源的第一复购率被计算为复购相关数据源的用户数量与购买过相关数据源的用户数量之商。相关数据源的第二复购率被计算为复购相关数据源的交易次数与购买过相关数据源的用户数量之商。

[0012] 在本公开的一些实施例中,根据全域交易地图来在多个子网中发现与第一枢纽节点直接连接的第一参与节点所需的目标资源包括:针对每个参与节点,根据以下中的至少一个来生成数据价值评估值:相关数据源的交易成功转化率、相关数据源的第一复购率、相关数据源的第二复购率;以及按照数据价值评估值的降序来向第一参与节点推荐多个子网中其它参与节点上的资源作为目标资源。

[0013] 在本公开的一些实施例中,第一非敏感数据和第二非敏感数据中的一个或多个还包括以下中的至少一个:数据交易双方的节点标识符、数据交易的任务标识符、数据交易的

开始时间、数据交易的结束时间、数据交易的中间结果、数据交易结果的转售信息、关于数据交易是否达成的信息、数据交易中涉及的数据类别及简要描述信息、参与节点的计算任务稳定性信息、参与节点的履约效率。

[0014] 根据本公开的第二方面,提供了一种在相互连接的多个子网中发现目标资源的装置。每个子网包括枢纽节点和与枢纽节点直接连接的多个参与节点。该多个子网中的枢纽节点相互直接连接。每个参与节点经由与其直接连接的枢纽节点与该多个子网中的其它参与节点进行数据交易。该装置被布置在该多个子网中的第一枢纽节点处。该装置包括至少一个处理器;以及存储有计算机程序的至少一个存储器。当计算机程序由至少一个处理器执行时,使得装置:存储与第一枢纽节点直接连接的每个参与节点在进行数据交易的过程中的第一非敏感数据;接收该多个子网中的多个第二枢纽节点所存储的第二非敏感数据,其中,每个第二枢纽节点存储与其直接连接的每个参与节点在进行数据交易的过程中的第二非敏感数据;根据所存储的第一非敏感数据以及所接收的第二非敏感数据来生成针对该多个子网的全域交易地图;以及根据全域交易地图来在该多个子网中发现与第一枢纽节点直接连接的第一参与节点所需的目标资源。其中,第一非敏感数据和第二非敏感数据二者都不泄露任一参与节点的原始数据本身。

[0015] 在本公开的一些实施例中,第一非敏感数据和第二非敏感数据中的每一个包括相关数据交易过程中涉及的数据源的类型。计算机程序在由至少一个处理器执行时使得装置通过以下操作来根据所存储的第一非敏感数据以及所接收的第二非敏感数据来生成针对多个子网的全域交易地图:从第一非敏感数据和第二非敏感数据中获取多个子网中的所有数据交易过程中涉及的数据源的类型;按照所获取的数据源的类型初始化全域交易地图,其中,同一类型的数据源对应全域交易地图中的一个顶点;在全域交易地图中将存在数据交易的顶点两两之间通过边来连接;计算全域交易地图中每条边所连接的两个顶点之间进行数据交易的次数;以及根据每条边对应的数据交易的次数来计算该条边的权重,其中,每条边的权重等于该条边对应的数据交易的次数除以全域交易地图中所有边对应的数据交易的次数之和。

[0016] 在本公开的一些实施例中,计算机程序在由至少一个处理器执行时使得装置通过以下操作来根据全域交易地图来在多个子网中发现与第一枢纽节点直接连接的第一参与节点所需的目标资源:根据第一参与节点拥有的数据源的类型来在全域交易地图中查找与第一参与节点相对应的第一顶点;计算第一顶点与全域交易地图中的其它顶点之间的关系强度;按照所计算的关系强度的降序来向第一参与节点推荐其它顶点上的资源作为目标资源;其中,第一顶点与其它顶点中的第二顶点之间的关系强度被计算为:

$$[0017] \quad c = \sum_{k=1}^m wk / m,$$

[0018] 其中,c表示第一顶点与第二顶点之间的关系强度,m表示第一顶点与第二顶点在全域交易地图中的最短路径的跳数,wk表示最短路径中的第k边的权重。

[0019] 在本公开的一些实施例中,第一非敏感数据和第二非敏感数据中的一个或多个包括相关数据交易过程中的数据交易双方之间的数据交集大小以及数据交易双方自身的数据集大小。计算机程序在由至少一个处理器执行时使得装置通过以下操作来根据所存储的

第一非敏感数据以及所接收的第二非敏感数据来生成针对多个子网的全域交易地图:将每个参与节点作为全域交易地图中的一个顶点;以及在全域交易地图中将数据交集大小已知的顶点两两之间通过边来连接。其中,边的权重等于数据交集大小。

[0020] 在本公开的一些实施例中,计算机程序在由至少一个处理器执行时使得装置通过以下操作来根据全域交易地图来在多个子网中发现与第一枢纽节点直接连接的第一参与节点所需的目标资源:确定全域交易地图中与第一参与节点所对应的第一顶点不相连的所有第三顶点;针对每个第三顶点,在全域交易地图中寻找与第一顶点和第三顶点二者都相连的第四顶点,其中,第四顶点的数据集大小大于第一顶点的数据集大小和第三顶点的数据集大小中的每一个;根据第一顶点与第四顶点之间的数据交集大小以及第三顶点与第四顶点之间的数据交集大小来估计第一顶点与第三顶点之间的数据交集大小;以及按照每个第三顶点与第一顶点之间的估计的数据交集大小以及每个第四顶点与第一顶点之间的数据交集大小的降序来向第一参与节点推荐第三顶点和第四顶点上的资源作为目标资源。

[0021] 在本公开的一些实施例中,第一非敏感数据和第二非敏感数据中的每一个包括以下中的至少一个:相关数据源的交易成功转化率、相关数据源的第一复购率、相关数据源的第二复购率。其中,相关数据源的交易成功转化率被计算为相关数据源的试使用次数与相关数据源的成功交易次数之商。相关数据源的第一复购率被计算为复购相关数据源的用户数量与购买过相关数据源的用户数量之商。相关数据源的第二复购率被计算为复购相关数据源的交易次数与购买过相关数据源的用户数量之商。

[0022] 在本公开的一些实施例中,计算机程序在由至少一个处理器执行时使得装置通过以下操作来根据全域交易地图来在多个子网中发现与第一枢纽节点直接连接的第一参与节点所需的目标资源:针对每个参与节点,根据以下中的至少一个来生成数据价值评估值:相关数据源的交易成功转化率、相关数据源的第一复购率、相关数据源的第二复购率;以及按照数据价值评估值的降序来向第一参与节点推荐多个子网中其它参与节点上的资源作为目标资源。

[0023] 根据本公开的第三方面,提供了一种存储有计算机程序的计算机可读存储介质,其中,计算机程序在由处理器执行时实现根据本公开的第一方面所述的方法的步骤。

附图说明

[0024] 为了更清楚地说明本公开的实施例的技术方案,下面将对实施例的附图进行简要说明,应当知道,以下描述的附图仅仅涉及本公开的一些实施例,而非对本公开的限制,其中:

[0025] 图1是数联网的示意性拓扑图;

[0026] 图2是根据本公开的实施例的在相互连接的多个子网中发现目标资源的方法的示意性流程图;

[0027] 图3是根据本公开的实施例的全域交易地图的示例性拓扑图;

[0028] 图4是根据本公开的实施例的全域交易地图的另一示例性拓扑图;

[0029] 图5是根据本公开的实施例的在相互连接的多个子网中发现目标资源的装置的示意性框图。

[0030] 需要注意的是,附图中的元素是示意性的,没有按比例绘制。

具体实施方式

[0031] 为了使本公开的实施例的目的、技术方案和优点更加清楚,下面将结合附图,对本公开的实施例的技术方案进行清楚、完整的描述。显然,所描述的实施例是本公开的一部分实施例,而不是全部的实施例。基于所描述的本公开的实施例,本领域技术人员在无需创造性劳动的前提下所获得的所有其它实施例,也都属于本公开保护的范围。

[0032] 除非另外定义,否则在此使用的所有术语(包括技术和科学术语)具有与本公开主题所属领域的技术人员所通常理解的含义。进一步将理解的是,诸如在通常使用的词典中定义的那些的术语应解释为具有与说明书上下文和相关技术中它们的含义一致的含义,并且将不以理想化或过于正式的形式来解释,除非在此另外明确定义。另外,诸如“第一”和“第二”的术语仅用于将一个部件(或部件的一部分)与另一个部件(或部件的另一部分)区分开。

[0033] 如上所述,各行各业中的各类政务主体、行业主体、公司主体、机构主体可组合成网络,而此类网络中的主体数量可能非常大。数联网就是此类网络的一个示例。数联网作为开放性的网络体系,可以支持各类主体通过被允许的方式接入数联网,接入数联网后就形成了相应数据节点。成熟的数联网中数据节点数量级可以达到万、十万甚至是百万量级,规模跨越多个数量级。

[0034] 图1示出数联网的示意性拓扑图。数联网可包括多个子网10。每个子网10包括枢纽节点11和与枢纽节点直接连接的多个参与节点12。该多个子网10中的枢纽节点11相互直接连接。枢纽节点11与枢纽节点11之间可以通过专网进行互联。枢纽节点11承担对参与节点12进行信息聚合、寻址导航等功能。参与节点12可以是各类政务主体、行业主体、公司主体、机构主体等。直接连接到同一个枢纽节点11的参与节点12通过该枢纽节点11进行通信。直接连接到不同枢纽节点11的参与节点12通过它们各自直接连接的枢纽节点11进行通信。也就是说,参与节点12只与其直接连接的枢纽节点11直接通信,枢纽节点11之间可直接通信,而参与节点12之间需经由相应的枢纽节点11进行通信。

[0035] 在实践中,数联网中可能存在海量的子网10。单个子网10中可能存在海量的参与节点12。如果将每个参与节点12看作一个节点,那么数联网中的节点的数量可能是非常庞大的。而要在海量的节点中批量发现目标资源是存在困难的。

[0036] 本公开的发明人留意到当前在类似数联网的网络中,枢纽节点仅仅具备透传信息、路由信息、链接其它枢纽节点等功能,没有针对性记录其透传或路由的相关交易的行为信息,也就是说,枢纽节点是无记忆的。因此在每个交易过程中,枢纽节点都需要重新链接任务节点,执行数据的联合探查等操作,浪费了历史交易中很多有价值的中间过程信息。由于枢纽节点没有知识的累积,因此无法支撑更有价值和更大范围的针对业务、数据生态等变化的分析。

[0037] 因此本公开提出一种在相互连接的多个子网中发现目标资源的方法。该多个子网可以是数联网的一部分或者全部。每个子网包括枢纽节点和与枢纽节点直接连接的多个参与节点。该多个子网中的枢纽节点相互直接连接。每个参与节点经由与其直接连接的枢纽节点与该多个子网中的其它参与节点进行数据交易。图2示出根据本公开的实施例的在相互连接的多个子网中发现目标资源的方法200的示意性流程图。该方法200由该多个子网中的第一枢纽节点执行。第一枢纽节点在上下文中指的是该多个子网中的任意一个枢纽节

点,而非某个特定的枢纽节点。

[0038] 在图2的框S202处,第一枢纽节点存储与第一枢纽节点直接连接的每个参与节点在进行数据交易的过程中的第一非敏感数据。其中,第一非敏感数据不泄露任一参与节点的原始数据本身。在本公开的一些实施例中,第一非敏感数据可包括以下中的一个或多个:数据交易双方的节点标识符(节点ID)、数据交易的任务标识符(任务ID)、数据交易的开始时间、数据交易的结束时间、数据交易的中间结果、数据交易结果的转售信息、关于数据交易是否达成的信息、数据交易中涉及的数据类别及简要描述信息、参与节点的计算任务稳定性信息、参与节点的履约效率。

[0039] 在框S204处,第一枢纽节点接收该多个子网中的多个第二枢纽节点所存储的第二非敏感数据。第二枢纽节点在上下文中指的是该多个子网中除了第一枢纽节点之外的任意一个枢纽节点。每个第二枢纽节点存储与其直接连接的每个参与节点在进行数据交易的过程中的第二非敏感数据。第二非敏感数据不泄露任一参与节点的原始数据本身。在本公开的一些实施例中,第二非敏感数据可包括以下中的一个或多个:数据交易双方的节点标识符(节点ID)、数据交易的任务标识符(任务ID)、数据交易的开始时间、数据交易的结束时间、数据交易的中间结果、数据交易结果的转售信息、关于数据交易是否达成的信息、数据交易中涉及的数据类别及简要描述信息、参与节点的计算任务稳定性信息、参与节点的履约效率。

[0040] 应注意,在框S204处执行的操作可以与在框S202处执行的操作并行地执行,也可以先后执行,本公开不限制二者执行的先后顺序。经过在框S204处执行的操作和在框S202处执行的操作之后,第一枢纽节点汇总了该多个子网中的所有非敏感数据(即,第一非敏感数据和第二非敏感数据的集合)。

[0041] 在框S206处,第一枢纽节点根据所存储的第一非敏感数据以及所接收的第二非敏感数据来生成针对该多个子网的全域交易地图。全域交易地图可以体现在该多个子网中发生的数据交易所涉及的非敏感数据的统计信息。

[0042] 在框S208处,第一枢纽节点根据全域交易地图来在该多个子网中发现与第一枢纽节点直接连接的第一参与节点所需的目标资源。

[0043] 在本公开的一些实施例中,目标资源是高质量的数据资源。可以从以下五个方面来判断数据是否是高质量数据。

[0044] (1) 供需关系:数联网中的高质量数据必须是能够流通的,是符合数联网中客户生态需求的,否则即使数据再丰富、再完整,因为不受客户所需,导致不能有效流通,造成数联网资源浪费。因此高质量的数据,首先是能够满足合理的供需关系。需求和供应都处于不断变化之中,因此也是一个无休止动态迭代的过程。确定高质量数据的一个关键是可以及时捕捉供需关系的变化趋势。

[0045] (2) 数据样本丰富度:这里针对的是纵向联邦场景,需要样本特征互补,因此对于不同节点之间的数据样本重叠度有一定的要求。数据样本丰富度是指在数据联合建模或者联合统计的时候,样本的重合度高,可以向需求方提供大量的互补特征数据,以有效提升需求方的模型指标或者使得统计任务更具有统计意义且结果更具代表性。

[0046] (3) 数据源节点的可靠性:对于历史履约能力强的节点,其提供的的数据质量、计算质量都能得到较高的保证。通过历史的交易或者任务行为,可以筛选出信誉度高的数据源

节点,增强对高质量数据的背书。

[0047] (4) 价格因素:在一些应用场景下,高质量数据必须在一个合理的且可被接受的价格区间内,否则价格畸高的数据,即使需求度和质量都很高,也难以被成功交易。从健康生态的维度,希望筛选出的数据源是具备合理定价且能够完成交易的。

[0048] (5) 数据自身的质量:本公开可以不直接衡量数据自身的质量(异常、重复性、完整性、贡献度等),而是采用大量数联网用户自身的行为数据,利用群体智慧,间接实现对于高质量数据自身的筛选。比如用户之间进行试运行(例如,利用“隔空测一测”软件进行试运行)联邦计算任务的次数很多,但往往不能达成交易,大概率是因为数据的互补价值、数据质量等得不到保证。通过枢纽节点的任务行为分析,可以得到一系列的评价指标。反之,如果某类数据在进行少量的试运行后就迅速达成后续交易,说明这类数据更容易满足客户的需求和预期。

[0049] 针对上述判断条件,本公开提出多个利用第一非敏感数据和第二非敏感数据来发现目标资源的实施方式。

[0050] 本公开的一些实施例提出利用相关数据交易过程中涉及的数据源的类型来帮助发现目标资源。第一非敏感数据和第二非敏感数据中的每一个包括相关数据交易过程中涉及的数据源的类型。

[0051] 在根据所存储的第一非敏感数据以及所接收的第二非敏感数据来生成针对多个子网的全域交易地图的过程中,从第一非敏感数据和第二非敏感数据中获取多个子网中的所有数据交易过程中涉及的数据源的类型。然后,按照所获取的数据源的类型初始化全域交易地图。其中,同一类型的数据源对应全域交易地图中的一个顶点。图3示出此类全域交易地图的示例性拓扑图。图1中所示的各个参考节点可按照图3的示例分成能源类、证券类、运营商类等。在全域交易地图中将存在数据交易(即,进行过数据交易)的顶点两两之间通过边来连接。不存在数据交易的顶点则不连接。在连接顶点的过程中可并行地计算全域交易地图中每条边所连接的两个顶点之间进行数据交易的次数。之后,可根据每条边对应的数据交易的次数来计算该条边的权重。在一个示例中,每条边的权重等于该条边对应的数据交易的次数除以全域交易地图中所有边对应的数据交易的次数之和。

[0052] 在图3的示例中,可先获得20条边中每条边对应的交易次数。例如, w_1 表示证券类参考节点与能源类参考节点之间的交易次数, w_2 表示能源类参考节点与泛金融类参考节点之间的交易次数,以此类推。 w_1 至 w_{20} 之和等于全域交易地图中所有边对应的数据交易的次数之和。那么, $w_1 / \sum_{k=1}^{20} w_k$ 等于证券类参考节点与能源类参考节点之间的边的权重, $w_2 / \sum_{k=1}^{20} w_k$ 等于能源类参考节点与泛金融类参考节点之间的边的权重,以此类推。在计算出各条边的权重之后,可将 w_1 的值更新为 $w_1 / \sum_{k=1}^{20} w_k$ (即, w_1 被更新为相应边的权重),将 w_2 的值更新为 $w_2 / \sum_{k=1}^{20} w_k$ (即, w_2 被更新为相应边的权重),以此类推。也就是说,在全域交易地图中可只存储边的权重 w_k 。

[0053] 这样,在根据全域交易地图来在多个子网中发现与第一枢纽节点直接连接的第一参与节点所需的目标资源的过程中,可根据第一参与节点拥有的数据源的类型来在全域交易地图中查找与第一参与节点相对应的第一顶点。假设第一参考节点拥有能源类的数据,

则第一参考节点对应的第一顶点是图3中的能源类顶点。然后,可计算第一顶点与全域交易地图中的其它顶点之间的关系强度。例如,图3中的能源类顶点与其它19个顶点之间的关系强度。之后,可按照所计算的关系强度的降序来向第一参与节点推荐其它顶点上的资源作为目标资源。

[0054] 在本公开的一些实施例中,第一顶点与其它顶点中的第二顶点之间的关系强度被计算为:

$$[0055] \quad c = \sum_{k=1}^m w_k / m,$$

[0056] 其中,第二顶点表示其它顶点中的任一顶点,c表示第一顶点与第二顶点之间的关系强度,m表示第一顶点与第二顶点在全域交易地图中的最短路径的跳数, w_k 表示最短路径中的第k边的权重。

[0057] 参考图3的示例,在第一顶点是能源类顶点,而第二顶点是泛金融类顶点的情况下,第一顶点与第二顶点的最短路径的跳数为1。能源类顶点与泛金融类顶点之间的边权重为 w_2 。因此,能源类顶点与泛金融类顶点之间的关系强度为 $w_2/1=w_2$ 。

[0058] 在第一顶点是能源类顶点,而第二顶点是医疗类顶点的情况下,第一顶点到第二顶点的最短路径是能源类顶点->泛金融类顶点->车企类顶点->政务类顶点->医疗类顶点。第一顶点与第二顶点的最短路径的跳数为4。能源类顶点与泛金融类顶点之间的边权重为 w_2 。泛金融类顶点与车企类顶点之间的边权重为 w_5 。车企类顶点与政务类顶点之间的边权重为 w_7 。政务类顶点与医疗类顶点之间的边权重为 w_8 。因此,能源类顶点与医疗类顶点之间的关系强度为 $(w_2+w_5+w_7+w_8)/4$ 。

[0059] 针对前述判断高质量数据的条件,本公开的另一些实施例提出利用相关数据交易过程中涉及的数据交集大小以及数据交易双方自身的数据集大小来帮助发现目标资源。第一非敏感数据可包括相关数据交易过程中的数据交易双方之间的数据交集大小以及数据交易双方自身的数据集大小。第二非敏感数据也可包括相关数据交易过程中的数据交易双方之间的数据交集大小以及数据交易双方自身的数据集大小。

[0060] 在根据所存储的第一非敏感数据以及所接收的第二非敏感数据来生成针对多个子网的全域交易地图的过程中,将每个参与节点作为全域交易地图中的一个顶点,在全域交易地图中将数据交集大小已知的顶点两两之间通过边来连接。其中,边的权重等于数据交集大小。图4示出此类全域交易地图的示例性拓扑图。为便于描述,仅在图4中示出4个顶点。在图4中,实线绘制的边表示边所对应的两个顶点之间的数据交集大小已知,虚线绘制的边表示边所对应的两个顶点之间的数据交集大小未知且需要估计数据交集大小。

[0061] 在本公开的一些实施例中,在根据全域交易地图来在多个子网中发现与第一枢纽节点直接连接的第一参与节点所需的目标资源的过程中,确定全域交易地图中与第一参与节点所对应的第一顶点不相连的所有第三顶点。第三顶点在上下文中指的是与第一顶点不相连的任意一个顶点且不包括第一顶点本身。针对每个第三顶点,在全域交易地图中寻找与第一顶点和第三顶点二者都相连的第四顶点。其中,第四顶点的数据集大小大于第一顶点的数据集大小和第三顶点的数据集大小中的每一个。然后,可根据第一顶点与第四顶点之间的数据交集大小以及第三顶点与第四顶点之间的数据交集大小来估计第一顶点与第

三顶点之间的数据交集大小。之后,可按照每个第三顶点与第一顶点之间的估计的数据交集大小以及每个第四顶点与第一顶点之间的数据交集大小的降序来向第一参与节点推荐第三顶点和第四顶点上的资源作为目标资源。

[0062] 在图4的示例中,假设第一参与节点对应第一顶点C1,第三顶点C2与第一顶点C1不相连,因此需要估计第三顶点C2与第一顶点C1之间的数据交集大小。已知第一顶点C1的数据集大小为N1,第三顶点C2的数据集大小为N2,顶点C3的数据集大小为N3,顶点C4的数据集大小为N4,第一顶点C1与顶点C3的数据交集大小为PSI1,第一顶点C1与顶点C4的数据交集大小为PSI4,第三顶点C2与顶点C3的数据交集大小为PSI2,第三顶点C2与顶点C4的数据交集大小为PSI5。针对第三顶点C2,可在全域交易地图中找到与第一顶点C1和第三顶点C2二者都相连的顶点C3和C4。假设顶点C3的数据集大小N3大于N1和N2,而顶点C4的数据集大小N4小于N1和N2,则将顶点C3确定为第四顶点。假如PSI1与PSI2较大,由于顶点C3的数据集为第一顶点C1和第三顶点C2的共同求交对象,那么大概率第一顶点C1的数据集与第三顶点C2的数据集有更大的交集集合。反之,如果PSI1与PSI2较小,那么大概率第一顶点C1的数据集与第三顶点C2的数据集是低交集的结果。这样,借助于已知的数据交集大小可推测未知的数据交集大小,无需进行再次求交,从而能够降低实际求交的运算量。

[0063] 针对前述判断高质量数据的条件,本公开的又一些实施例提出利用相关数据源的交易成功转化率和/或相关数据源的复购率来帮助发现目标资源。交易成功转化率和相关数据源的复购率越高,则说明相应的数据质量越高。第一非敏感数据和第二非敏感数据中的每一个可包括以下中的至少一个:相关数据源的交易成功转化率、相关数据源的第一复购率、相关数据源的第二复购率。

[0064] 相关数据源的交易成功转化率被计算为相关数据源的试使用次数与相关数据源的成功交易次数之商。数据在正式交易之前,会首先试运行(例如,通过“隔空测一测”软件来试运行)小量的数据,给予用户一定的评估支撑,判断所选数据是否满足其需求。每一个数据源都可能获得试运行的机会,因此试运行次数可以作为分母。如果试运行之后,用户购买该数据源,则表示该用户被成功转化,也就是该数据源得到一次成功的交易,那么将交易成功的次数作为分子,就可以得到交易行为的成功率。

[0065] 以 t_i 表示数据源i的试跑次数,以 b_i 表示数据源i被交易成功的次数,那么交易成功转化率 s_i 的计算公式如下:

$$[0066] \quad s_i = \frac{b_i}{t_i}。$$

[0067] 相关数据源的第一复购率被计算为复购相关数据源的用户数量与购买过相关数据源的用户数量之商。通过枢纽节点所记录的交易历史行为数据可以得到重复购买数据源i的用户数量。在这里,只要购买超过2次的用户,都归类为重复购买客户(即,复购用户),以 mu_i 表示多次购买数据源i的复购用户数量。只要发生过至少1次购买的用户,归类为购买过相关数据源i的用户(交易客户),以 u_i 表示至少购买过1次数据源i的用户数量。那么第一复购率 $rpr1_i$ 公式如下:

$$[0068] \quad rpr1_i = \frac{mu_i}{u_i}。$$

[0069] 相关数据源的第二复购率被计算为复购相关数据源的交易次数与购买过相关数据源的用户数量之商。通过枢纽节点所记录的交易历史行为数据可以得到重复购买数据源*i*的用户数量。在这里,只要购买超过2次的用户,都归类为重复购买客户(即,复购用户),以 mp_i 表示复购用户购买数据源*i*的复购交易次数。只要发生过至少1次购买的用户,归类为购买过相关数据源*i*的用户(交易客户),以 u_i 表示至少购买过1次数据源*i*的用户数量。那么第二复购率 $rpr2_i$ 公式如下:

$$[0070] \quad rpr2_i = \frac{mp_i}{u_i}。$$

[0071] 在根据全域交易地图来在多个子网中发现与第一枢纽节点直接连接的第一参与节点所需的目标资源的过程中,针对每个参与节点,根据以下中的至少一个来生成数据价值评估值:相关数据源的交易成功转化率、相关数据源的第一复购率、相关数据源的第二复购率。数据价值评估值例如可以等于相关数据源的交易成功转化率、相关数据源的第一复购率、相关数据源的第二复购率的加权和。然后,可按照数据价值评估值的降序来向第一参与节点推荐多个子网中其它参与节点上的资源作为目标资源。

[0072] 在本公开的一些实施例中,可根据参考节点之间的关系强度、参考节点之间的数据交集大小(包括已知数据交集大小和估计的数据交集大小)、数据价值评估值中的至少两个的加权和的降序来向第一参与节点推荐多个子网中其它参与节点上的资源作为目标资源。

[0073] 此外,除了上文提到的第一非敏感数据和第二非敏感数据,本公开的一些实施例还可以利用各个枢纽节点能够获得的其它非敏感数据来发现目标资源。

[0074] 图5示出根据本公开的实施例的在相互连接的多个子网中发现目标资源的装置500的示意性框图。如图5所示,该装置500可包括处理器510和存储有计算机程序的存储器520。当计算机程序由处理器510执行时,使得装置500可执行如图2所示的方法200的步骤。在一个示例中,装置500可以是计算机设备或云计算节点。装置500可存储与第一枢纽节点直接连接的每个参与节点在进行数据交易的过程中的第一非敏感数据。装置500可接收该多个子网中的多个第二枢纽节点所存储的第二非敏感数据。其中,每个第二枢纽节点存储与其直接连接的每个参与节点在进行数据交易的过程中的第二非敏感数据。装置500可根据所存储的第一非敏感数据以及所接收的第二非敏感数据来生成针对该多个子网的全域交易地图。装置500可根据全域交易地图来在该多个子网中发现与第一枢纽节点直接连接的第一参与节点所需的目标资源。其中,第一非敏感数据和第二非敏感数据二者都不泄露任一参与节点的原始数据本身。

[0075] 在本公开的一些实施例中,第一非敏感数据和第二非敏感数据中的每一个包括相关数据交易过程中涉及的数据源的类型。装置500可从第一非敏感数据和第二非敏感数据中获取多个子网中的所有数据交易过程中涉及的数据源的类型。装置500可按照所获取的数据源的类型初始化全域交易地图。其中,同一类型的数据源对应全域交易地图中的一个顶点。装置500可在全域交易地图中将存在数据交易的顶点两两之间通过边来连接。装置500可计算全域交易地图中每条边所连接的两个顶点之间进行数据交易的次数。装置500可根据每条边对应的数据交易的次数来计算该条边的权重。其中,每条边的权重等于该条边对应的数据交易的次数除以全域交易地图中所有边对应的数据交易的次数之和。

[0076] 在本公开的一些实施例中,装置500可根据第一参与节点拥有的数据源的类型来在全域交易地图中查找与第一参与节点相对应的第一顶点。装置500可计算第一顶点与全域交易地图中的其它顶点之间的关系强度。装置500可按照所计算的关系强度的降序来向第一参与节点推荐其它顶点上的资源作为目标资源。其中,第一顶点与其它顶点中的第二顶点之间的关系强度被计算为:

$$[0077] \quad c = \sum_{k=1}^m wk / m,$$

[0078] 其中,c表示第一顶点与第二顶点之间的关系强度,m表示第一顶点与第二顶点在全域交易地图中的最短路径的跳数,wk表示最短路径中的第k边的权重。

[0079] 在本公开的一些实施例中,第一非敏感数据和第二非敏感数据中的一个或多个包括相关数据交易过程中的数据交易双方之间的数据交集大小以及数据交易双方自身的数据集大小。装置500可将每个参与节点作为全域交易地图中的一个顶点。装置500可在全域交易地图中将数据交集大小已知的顶点两两之间通过边来连接。其中,边的权重等于数据交集大小。

[0080] 在本公开的一些实施例中,装置500可确定全域交易地图中与第一参与节点所对应的第一顶点不相连的所有第三顶点。装置500可针对每个第三顶点,在全域交易地图中寻找与第一顶点和第三顶点二者都相连的第四顶点。其中,第四顶点的数据集大小大于第一顶点的数据集大小和第三顶点的数据集大小中的每一个。装置500可根据第一顶点与第四顶点之间的数据交集大小以及第三顶点与第四顶点之间的数据交集大小来估计第一顶点与第三顶点之间的数据交集大小。装置500可按照每个第三顶点与第一顶点之间的估计的数据交集大小以及每个第四顶点与第一顶点之间的数据交集大小的降序来向第一参与节点推荐第三顶点和第四顶点上的资源作为目标资源。

[0081] 在本公开的一些实施例中,根据全域交易地图来在多个子网中发现与第一枢纽节点直接连接的第一参与节点所需的目标资源包括:针对每个参与节点,根据以下中的至少一个来生成数据价值评估值:相关数据源的交易成功转化率、相关数据源的第一复购率、相关数据源的第二复购率;以及按照数据价值评估值的降序来向第一参与节点推荐多个子网中其它参与节点上的资源作为目标资源。

[0082] 在本公开的实施例中,处理器510可以是例如中央处理单元(CPU)、微处理器、数字信号处理器(DSP)、基于多核的处理器架构的处理器等。存储器520可以是使用数据存储技术实现的任何类型的存储器,包括但不限于随机存取存储器、只读存储器、基于半导体的存储器、闪存、磁盘存储器等。

[0083] 此外,在本公开的实施例中,装置500也可包括输入设备530,例如键盘、鼠标等,用于输入第一非敏感数据和第二非敏感数据。另外,装置500还可包括输出设备540,例如显示器等,用于发现的目标资源。

[0084] 在本公开的其它实施例中,还提供了一种存储有计算机程序的计算机可读存储介质,其中,计算机程序在由处理器执行时能够实现如图2所示的方法的步骤。

[0085] 综上所述,根据本公开的实施例的在相互连接的多个子网中发现目标资源的方法利用子网中各个枢纽节点的记忆来批量发现目标资源(高质量数据),无需持续地付出昂贵

的计算代价就可以完成大批量数据分析,可以利用多个子网的群体智慧来筛选数据资源,还可以实现数据资源的动态供需趋势实时分析。

[0086] 附图中的流程图和框图显示了根据本公开的多个实施例的装置和方法的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段或指令的一部分,所述模块、程序段或指令的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。在有些作为替换的实现中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个连续的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意的,框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合,可以用执行规定的功能或动作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0087] 除非上下文另外明确地指出,否则在本文和所附权利要求中所使用的词语的单数形式包括复数,反之亦然。因而,当提及单数时,通常包括相应术语的复数。相似地,措辞“包含”和“包括”将解释为包含在内而不是独占性地。同样地,术语“包括”和“或”应当解释为包括在内的,除非本文中明确禁止这样的解释。在本文中使用术语“示例”之处,特别是当其位于一组术语之后时,所述“示例”仅仅是示例性的和阐述性的,且不应当被认为是独占性的或广泛性的。

[0088] 适应性的进一步的方面和范围从本文中提供的描述变得明显。应当理解,本申请的各个方面可以单独或者与一个或多个其它方面组合实施。还应当理解,本文中的描述和特定实施例旨在仅说明的目的并不旨在限制本申请的范围。

[0089] 以上对本公开的若干实施例进行了详细描述,但显然,本领域技术人员可以在不脱离本公开的精神和范围的情况下对本公开的实施例进行各种修改和变型。本公开的保护范围由所附的权利要求限定。

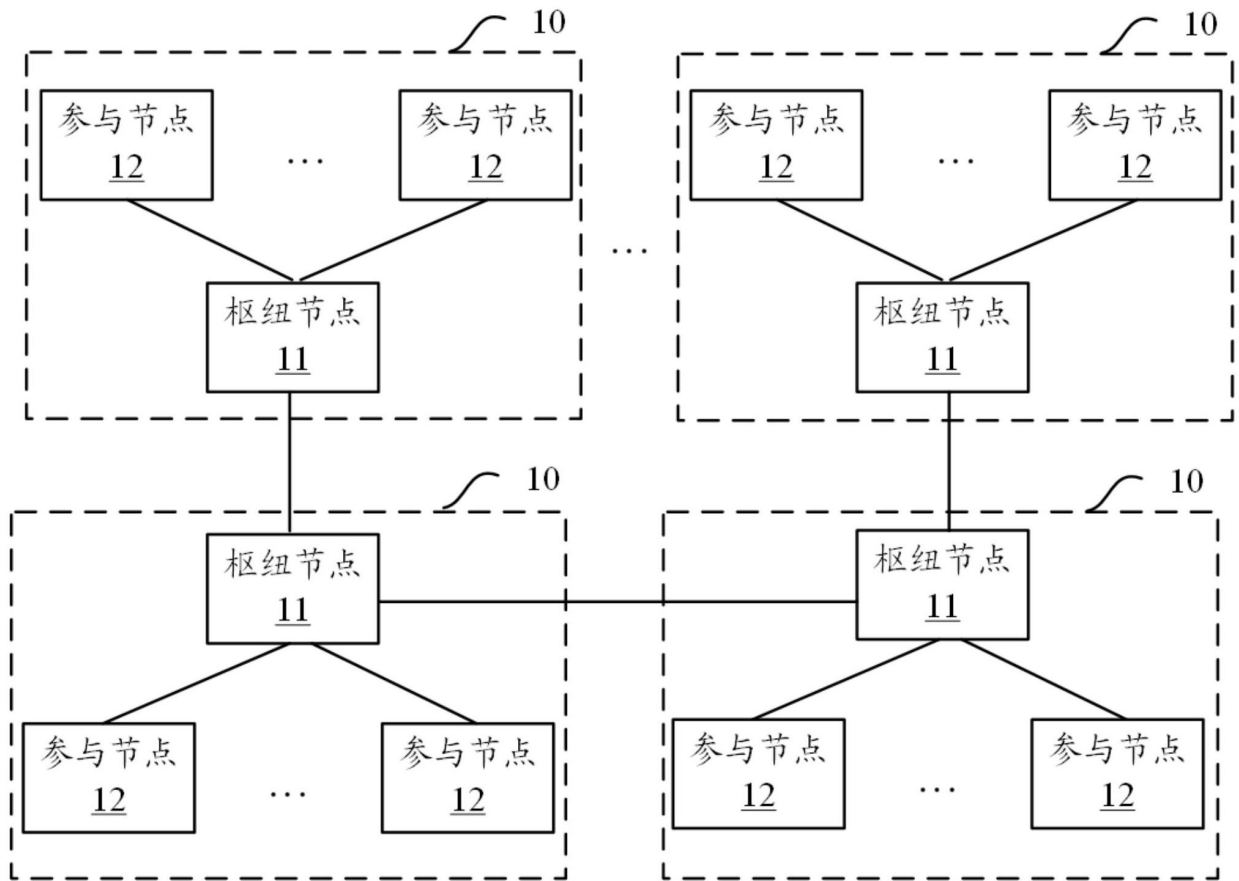


图1

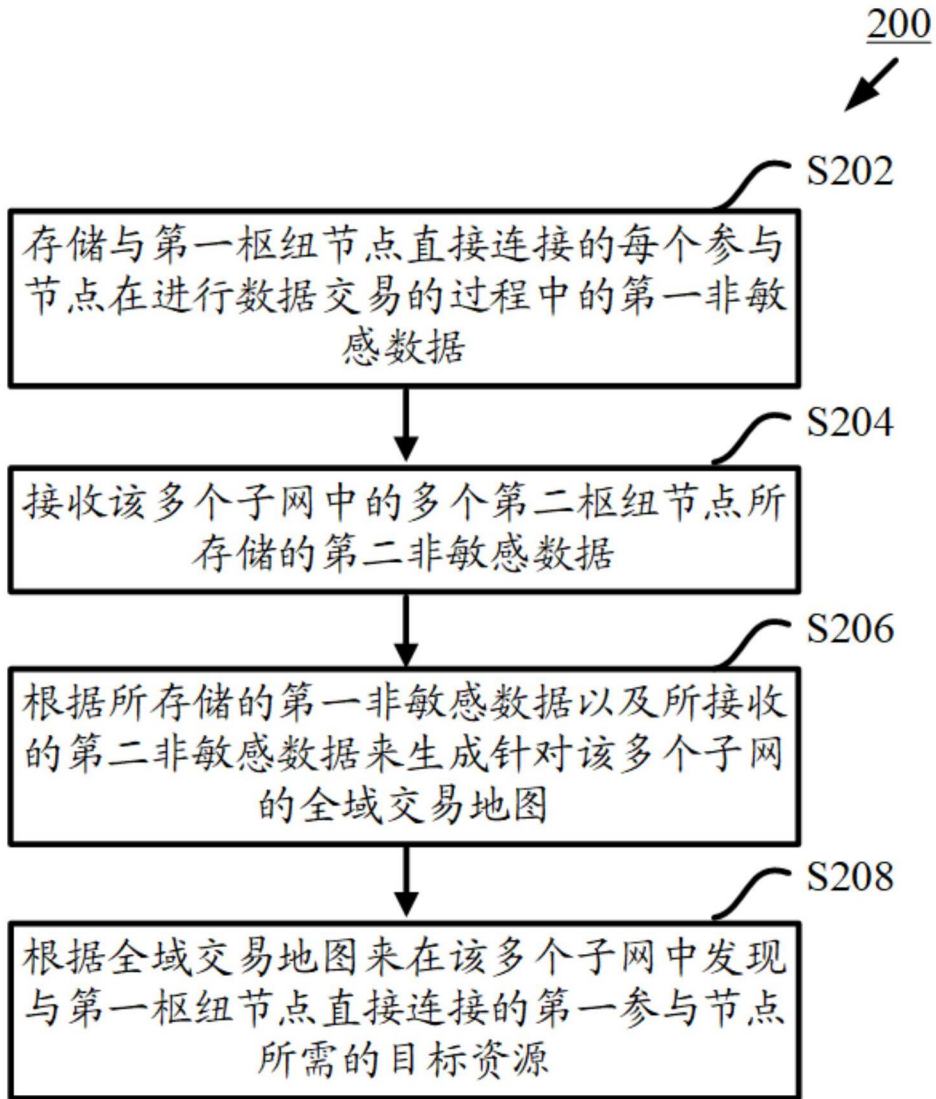


图2

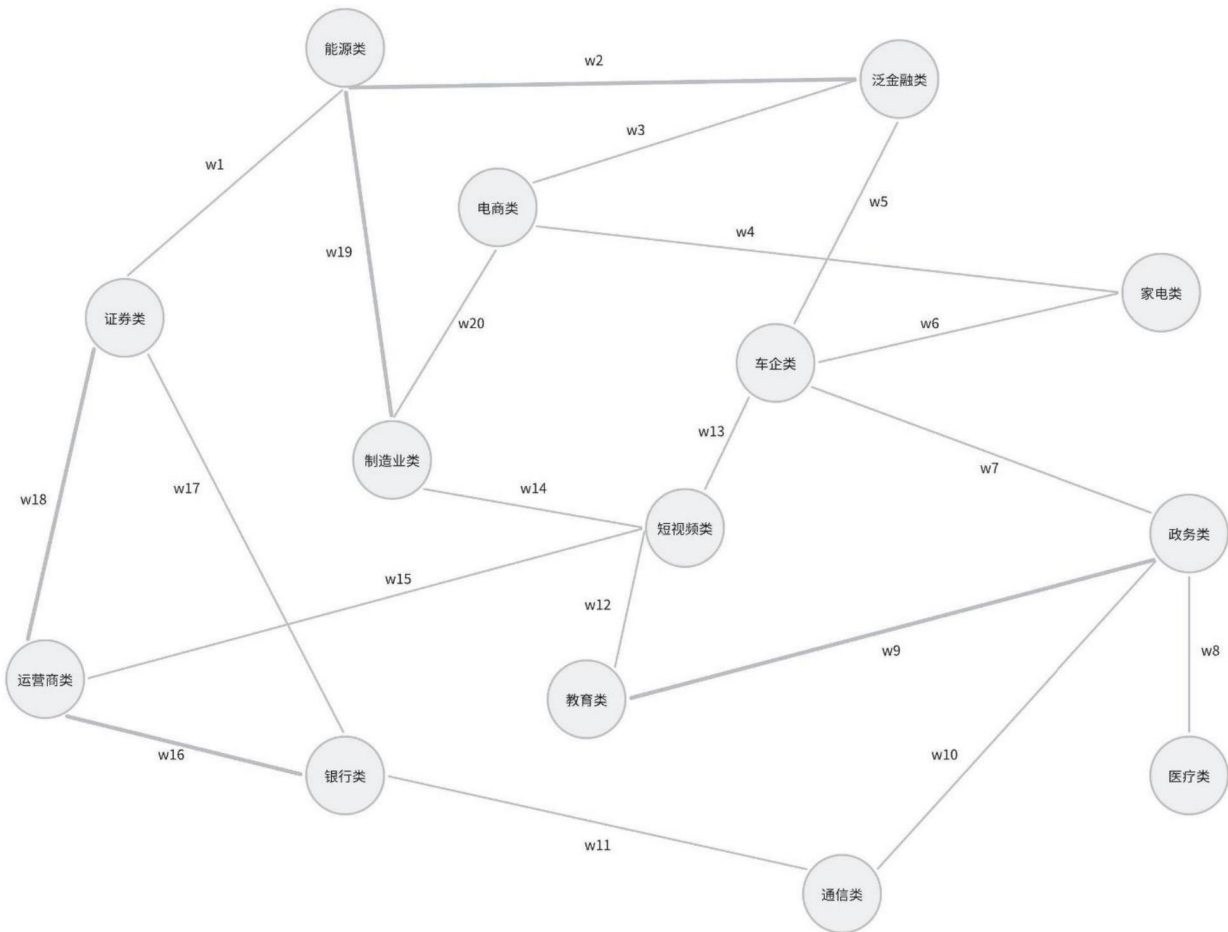


图3

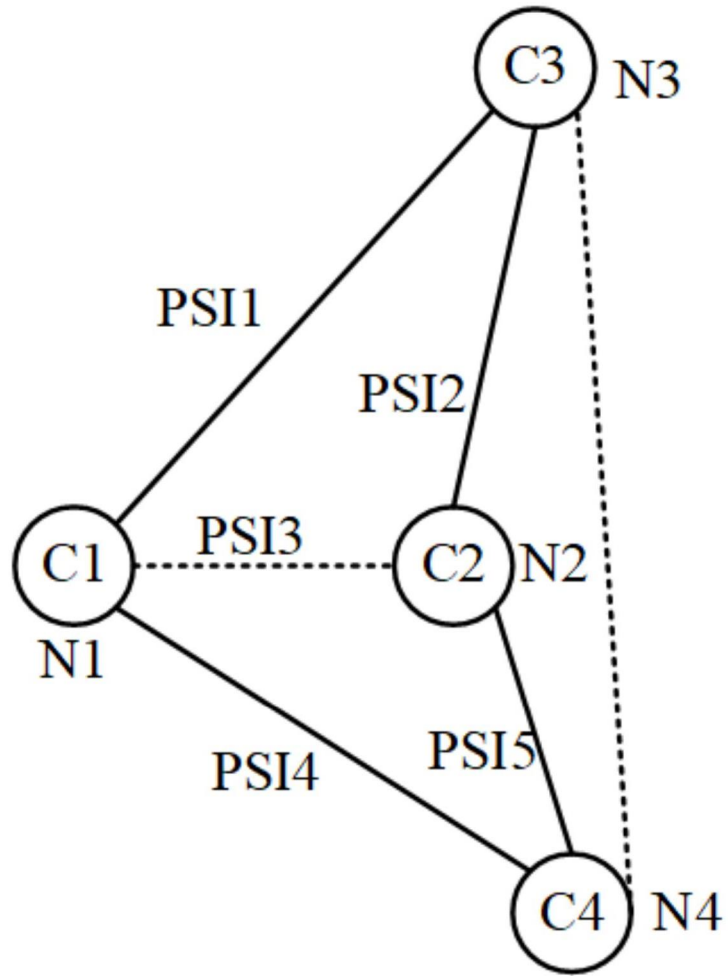


图4

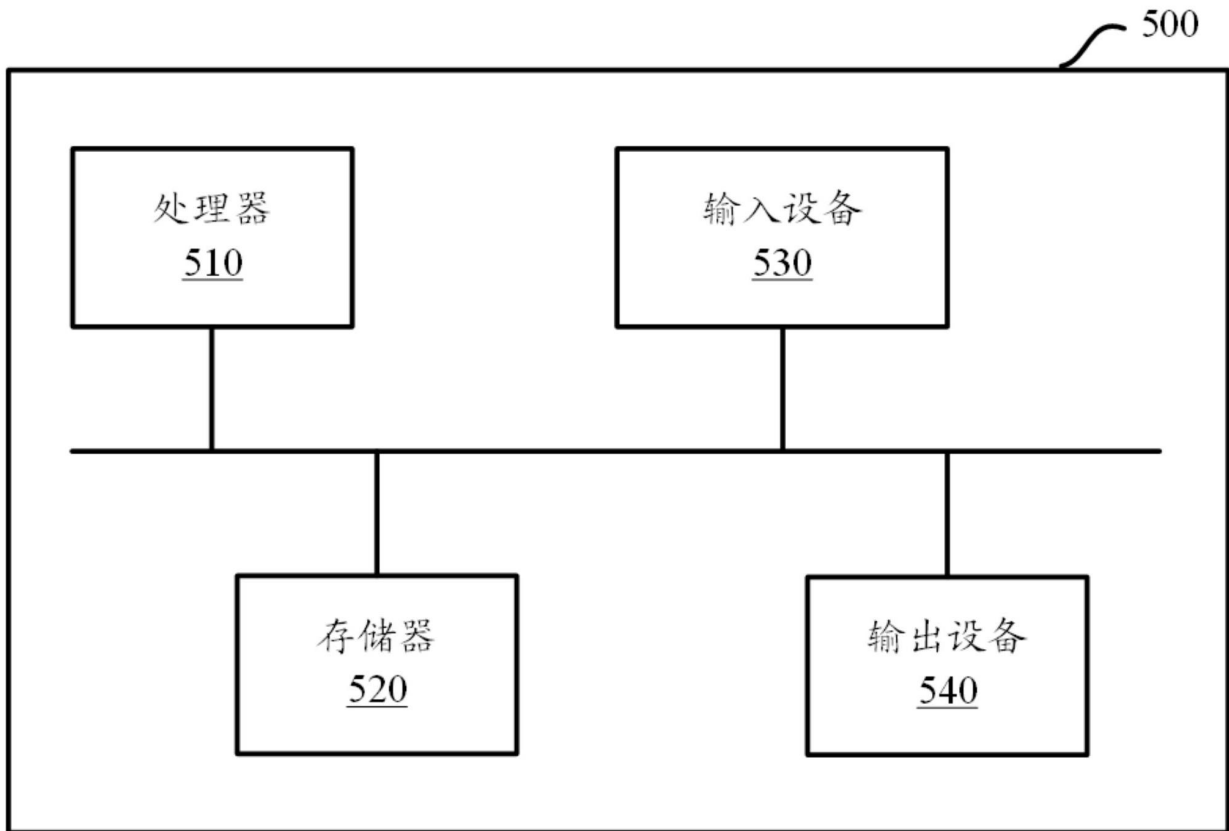


图5