



## (12) 发明专利申请

(10) 申请公布号 CN 117520610 A

(43) 申请公布日 2024. 02. 06

(21) 申请号 202311581597.9

(22) 申请日 2023.11.24

(71) 申请人 北京富算科技有限公司

地址 100070 北京市丰台区南四环西路188  
号十六区18号楼1至15层101内7层  
701-8

(72) 发明人 杜浩 张洪锋 余亚丽 孙小超  
陈立峰 赵华宇 卫骞 尤志强  
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 孙黎生

(51) Int. Cl.

G06F 16/903 (2019.01)

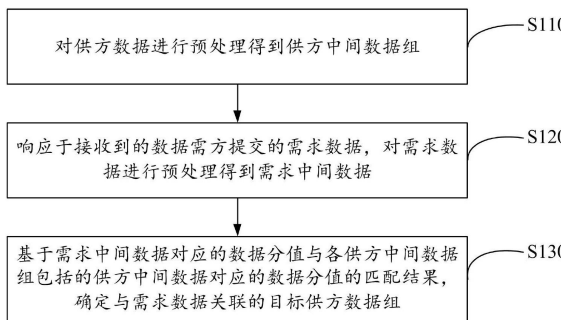
权利要求书2页 说明书12页 附图1页

### (54) 发明名称

基于图计算的数据搜索方法、装置、设备和  
存储介质

### (57) 摘要

本公开的实施例提供的基于图计算的数据搜索方法、装置、设备和存储介质,包括:对供方数据进行预处理得到供方中间数据组;响应于接收到的数据需方提交的需求数据,对需求数据进行预处理得到需求中间数据;基于需求中间数据对应的数据分值与各供方中间数据组包括的供方中间数据对应的数据分值的匹配结果,确定与需求数据关联的目标供方数据组。实现在保证数据不被泄露的基础上,对供方数据和需方数据基于特征信息赋予数据分值,可以基于供方中间数据和需方中间数据对应的数据分值,以图计算的方式,从供方数据中查找到目标供方数据。



1. 一种基于图计算的数据搜索方法,其特征在于,包括:

对供方数据进行预处理得到供方中间数据组,其中,一所述供方中间数据组包括多个供方中间数据,任意一个所述供方中间数据包括特征信息、特征标签标识、特征权重系数和数据分值;

响应于接收到的数据需方提交的需求数据,对所述需求数据进行预处理得到需求中间数据,其中,所述需求中间数据包括特征信息、特征标签标识、特征权重系数和数据分值;

基于所述需求中间数据对应的数据分值与各所述供方中间数据组包括的供方中间数据对应的数据分值的匹配结果,确定与所述需求数据关联的目标供方数据组。

2. 根据权利要求1所述的方法,其特征在于,所述对供方数据进行预处理得到供方中间数据组,包括:

获取所述供方数据的数据标签标识和特征信息;

根据所述供方数据的特征信息,确定所述供方数据的特征标签标识;

根据所述供方数据的特征标签标识,确定所述供方数据的特征权重系数;

根据所述供方数据的特征标签标识和特征权重系数,确定所述供方数据的数据分值;

将所述供方数据对应的特征信息、特征标签标识、特征权重系数和数据分值组合形成所述供方数据对应的供方中间数据;

根据所述供方数据的数据标签标识,对所述供方数据对应的供方中间数据进行分类得到多组供方中间数据组。

3. 根据权利要求2所述的方法,其特征在于,所述根据所述供方数据的特征信息,确定所述供方数据的特征标签标识,包括:

对所述供方数据的特征信息中的关键特征信息添加标注标识;

从所述特征信息中提取出添加标注标识的关键特征信息,并根据所述关键特征信息确定所述供方数据的特征标签标识。

4. 根据权利要求1所述的方法,其特征在于,所述对所述需求数据进行预处理得到需求中间数据,包括:

获取所述需求数据的特征信息;

根据所述需求数据的特征信息,确定所述供方数据的特征标签标识;

根据所述需求数据的特征标签标识,确定所述需求数据的特征权重系数;

根据所述需求数据特征标签标识和特征权重系数,确定所述需求数据的数据分值;

将所述需求数据对应的特征信息、特征标签标识、特征权重系数和数据分值组合形成所述需求数据对应的需求中间数据。

5. 根据权利要求1所述的方法,其特征在于,所述基于所述需求中间数据与各所述供方中间数据组的匹配结果,确定与所述需求数据关联的目标供方数据组,包括:

根据各所述供方中间数据组中各供方中间数据的数据分值,对所述供方中间数据组的各供方中间数据进行降序排序;

依次从各所述供方中间数据组中选取前第一目标数量的供方中间数据所对应的数据分值,并求取平均数据分值;

基于各所述供方中间数据组对应的平均数据分值与所述需求中间数据对应的数据分值的匹配结果,确定与所述需求数据关联的目标供方数据组。

6. 根据权利要求5所述的方法, 其特征在于, 所述基于各所述供方中间数据组对应的平均数据分值与所述需求中间数据对应的数据分值的绝对值差值, 确定与所述需求数据关联的目标供方数据组, 包括:

依次求取各所述供方中间数据组对应的平均数据分值与所述需求中间数据对应的数据分值的绝对值差值;

根据各所述供方中间数据组对应的平均数据分值与所述需求中间数据对应的数据分值的绝对值差值, 确定与所述需求数据关联的目标供方数据组。

7. 根据权利要求6所述的方法, 其特征在于, 所述根据各所述供方中间数据组对应的平均数据分值与所述需求中间数据对应的数据分值的绝对值差值, 确定与所述需求数据关联的目标供方数据组, 包括:

根据各所述供方中间数据组对应的平均数据分值与所述需求中间数据对应的数据分值的绝对值差值, 对各所述供方中间数据组进行降序排序;

基于对各所述供方中间数据组的降序排序结果, 选取前第二目标数量的供方中间数据组;

基于选取的前第二目标数量的供方数据中间组中各供方中间数据对应的供方数据作为与所述需求数据关联的目标供方数据。

8. 一种基于图计算的数据搜索装置, 其特征在于, 包括:

供方中间数据组确定模块, 用于对供方数据进行预处理得到供方中间数据组, 其中, 一所述供方中间数据组包括多个供方中间数据, 任意一个所述供方中间数据包括特征信息、特征标签标识、特征权重系数和数据分值;

需求中间数据确定模块, 用于响应于接收到的数据需求方提交的需求数据, 对所述需求数据进行预处理得到需求中间数据, 其中, 所述需求中间数据包括特征信息、特征标签标识、特征权重系数和数据分值;

目标供方数据组确定模块, 用于基于所述需求中间数据对应的数据分值与各所述供方中间数据组包括的供方中间数据对应的数据分值的匹配结果, 确定与所述需求数据关联的目标供方数据组。

9. 一种计算机设备, 其特征在于, 包括:

一个或多个处理器;

存储装置, 用于存储一个或多个程序,

当所述一个或多个程序被所述一个或多个处理器执行, 使得所述一个或多个处理器实现如权利要求1~7中任一所述的方法。

10. 一种计算机可读存储介质, 其上存储有计算机程序, 其特征在于, 该程序被处理器执行时实现如权利要求1~7中任一所述的方法。

## 基于图计算的数据搜索方法、装置、设备和存储介质

### 技术领域

[0001] 本发明涉及信息处理技术领域以及相关技术领域,具体地,涉及适用于一种基于图计算的数据搜索方法、装置、设备和存储介质。

### 背景技术

[0002] 针对指出的“建立合规高效、场内外结合的数据要素流通和交易制度”等要求,数联网方案逐步完善。

[0003] 在数联网方案中,将数据需求方称为需方,将数据提供方称为供方,在实际数据撮合过程中,由于数据安全考虑,并不会直接提供数据,而是给出数据的元信息和一些描述信息,需方想要找到匹配的数据无疑于大海捞针,因此需要在保证原始数据不暴漏的情况下,找到预期的相关数据,从而利用多组供方数据+需求方数据进行组建虚拟的数据联盟进行联合,训练出更为合理的模型或提供数据查询服务,提高数据要素流通的效率,促成数据的交易。

[0004] 现有技术中,在数联网情况下,如何依据需方的数据“以数找数”,高效的找到与之匹配的数据是亟需解决的问题。

### 发明内容

[0005] 本文中描述的实施例提供了一种基于图计算的数据搜索方法、装置、设备和存储介质,解决现有技术存在的问题。

[0006] 第一方面,根据本公开的内容,提供了一种基于图计算的数据搜索方法,包括:

[0007] 对供方数据进行预处理得到供方中间数据组,其中,一所述供方中间数据组包括多个供方中间数据,任意一个所述供方中间数据包括特征信息、特征标签标识、特征权重系数和数据分值;

[0008] 响应于接收到的数据需方提交的需求数据,对所述需求数据进行预处理得到需求中间数据,其中,所述需求中间数据包括特征信息、特征标签标识、特征权重系数和数据分值;

[0009] 基于所述需求中间数据对应的数据分值与各所述供方中间数据组包括的供方中间数据对应的数据分值的匹配结果,确定与所述需求数据关联的目标供方数据组。

[0010] 在本公开的一些实施例中,所述对供方数据进行预处理得到供方中间数据组,包括:

[0011] 获取所述供方数据的数据标签标识和特征信息;

[0012] 根据所述供方数据的特征信息,确定所述供方数据的特征标签标识;

[0013] 根据所述供方数据的特征标签标识,确定所述供方数据的特征权重系数;

[0014] 根据所述供方数据的特征标签标识和特征权重系数,确定所述供方数据的数据分值;

[0015] 将所述供方数据对应的特征信息、特征标签标识、特征权重系数和数据分值组合

形成所述供方数据对应的供方中间数据；

[0016] 根据所述供方数据的数据标签标识,对所述供方数据对应的供方中间数据进行分类得到多组供方中间数据组。

[0017] 在本公开的一些实施例中,所述根据所述供方数据的特征信息,确定所述供方数据的特征标签标识,包括:

[0018] 对所述供方数据的特征信息中的关键特征信息添加标注标识;

[0019] 从所述特征信息中提取出添加标注标识的关键特征信息,并根据所述关键特征信息确定所述供方数据的特征标签标识。

[0020] 在本公开的一些实施例中,所述对所述需求数据进行预处理得到需求中间数据,包括:

[0021] 获取所述需求数据的特征信息;

[0022] 根据所述需求数据的特征信息,确定所述供方数据的特征标签标识;

[0023] 根据所述需求数据的特征标签标识,确定所述需求数据的特征权重系数;

[0024] 根据所述需求数据特征标签标识和特征权重系数,确定所述需求数据的数据分值;

[0025] 将所述需求数据对应的特征信息、特征标签标识、特征权重系数和数据分值组合形成所述需求数据对应的需求中间数据。

[0026] 在本公开的一些实施例中,所述基于所述需求中间数据与各所述供方中间数据组的匹配结果,确定与所述需求数据关联的目标供方数据组,包括:

[0027] 根据各所述供方中间数据组中各供方中间数据的数据分值,对所述供方中间数据组的各供方中间数据进行降序排序;

[0028] 依次从各所述供方中间数据组中选取前第一目标数量的供方中间数据所对应的数据分值,并求取平均数据分值;

[0029] 基于各所述供方中间数据组对应的平均数据分值与所述需求中间数据对应的数据分值的绝对值差值,确定与所述需求数据关联的目标供方数据组。

[0030] 在本公开的一些实施例中,所述基于各所述供方中间数据组对应的平均数据分值与所述需求中间数据对应的数据分值的绝对值差值,确定与所述需求数据关联的目标供方数据组,包括:

[0031] 依次求取各所述供方中间数据组对应的平均数据分值与所述需求中间数据对应的数据分值的绝对值差值;

[0032] 根据各所述供方中间数据组对应的平均数据分值与所述需求中间数据对应的数据分值的绝对值差值,确定与所述需求数据关联的目标供方数据组。

[0033] 在本公开的一些实施例中,所述根据各所述供方中间数据组对应的平均数据分值与所述需求中间数据对应的数据分值的绝对值差值,确定与所述需求数据关联的目标供方数据组,包括:

[0034] 根据各所述供方中间数据组对应的平均数据分值与所述需求中间数据对应的数据分值的绝对值差值,对各所述供方中间数据组进行降序排序;

[0035] 基于对各所述供方中间数据组的降序排序结果,选取前第二目标数量的供方中间数据组;

[0036] 基于选取的前第二目标数量的供方数据中间组中各供方中间数据对应的供方数据作为与所述需求数据关联的目标供方数据。

[0037] 第二方面,根据本公开的内容,提供了一种基于图计算的数据搜索装置,包括:

[0038] 供方中间数据组确定模块,用于对供方数据进行预处理得到供方中间数据组,其中,一所述供方中间数据组包括多个供方中间数据,任意一个所述供方中间数据包括特征信息、特征标签标识、特征权重系数和数据分值;

[0039] 需求中间数据确定模块,用于响应于接收到的数据需方提交的需求数据,对所述需求数据进行预处理得到需求中间数据,其中,所述需求中间数据包括特征信息、特征标签标识、特征权重系数和数据分值;

[0040] 目标供方数据组确定模块,用于基于所述需求中间数据对应的数据分值与各所述供方中间数据组包括的供方中间数据对应的数据分值的匹配结果,确定与所述需求数据关联的目标供方数据组。

[0041] 第三方面,根据本公开的内容,提供了一种计算机设备,包括存储器和处理器,存储器中存储有计算机程序,处理器执行计算机程序时实现如以上任意一个实施例方法的步骤。

[0042] 第四方面,根据本公开的内容,提供了一种计算机可读存储介质,计算机可读存储介质上存储有计算机程序,计算机程序被处理器执行时实现如以上任意一个实施例中方法的步骤。

[0043] 本公开实施例提供的基于图计算的数据搜索方法、装置、设备和存储介质,首先对供方数据进行预处理得到供方中间数据组;然后响应于接收到的数据需方提交的需求数据,对需求数据进行预处理得到需求中间数据;最后基于需求中间数据对应的数据分值与各供方中间数据组包括的供方中间数据对应的数据分值的匹配结果,确定与需求数据关联的目标供方数据组。本公开实施例中,通过对供方数据和需方数据进行预处理,得到供方中间数据和需方中间数据,由于供方中间数据和需方中间数据中,各中间数据以特征信息、特征标签标识、特征权重系数以及数据分值进行表示,因此,在保证数据不被泄露的基础上,对供方数据和需方数据基于特征信息赋予数据分值,可以基于供方中间数据和需方中间数据对应的数据分值,以图计算的方式,从供方数据中查找到目标供方数据。

[0044] 上述说明仅是本申请实施例技术方案的概述,为了能够更清楚了解本申请实施例的技术手段,而可依照说明书的内容予以实施,并且为了让本申请实施例的上述和其它目的、特征和优点能够更明显易懂,以下特举本申请的具体实施方式。

## 附图说明

[0045] 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:

[0046] 图1是本公开实施例提供的一种基于图计算的数据搜索方法的流程示意图;

[0047] 图2是本公开实施例提供的一种基于图计算的数据搜索装置的结构示意图;

[0048] 图3是本公开实施例提供的一种计算机设备的结构示意图。

[0049] 在附图中,最后两位数字相同的标记对应于相同的元素。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

[0050] 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。

[0051] 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相同含义。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。如在此所使用的,将两个或更多部分“连接”或“耦接”到一起的陈述应指这些部分直接结合到一起或通过一个或多个中间部件结合。

[0052] 在本文中提及“实施例”意味着,结合实施例描述的特定特征、结构或特性可以包含在本申请的至少一个实施例中。在说明书中的各个位置出现该短语“实施例”并不一定均是指相同的实施例,也不是与其它实施例互斥的独立的或备选的实施例。本领域技术人员显式地和隐式地理解的是,本文所描述的实施例可以与其它实施例相结合。

[0053] 本文中术语“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:存在A,同时存在A和B,存在B这三种情况。另外,本文中字符“/”,一般表示前后关联对象是一种“或”的关系。

[0054] 此外,在本公开的所有实施例中,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。

[0055] 在本申请的描述中,除非另有说明,“多个”的含义是指两个以上(包括两个),同理,“多组”指的是两组以上(包括两组)。

[0056] 为了使本技术领域的人员更好地理解本申请方案,下面将结合附图,对本申请实施例中的技术方案进行清楚、完整地描述。

[0057] 基于现有技术存在的问题,图1是本公开实施例提供的一种基于图计算的数据搜索方法的流程示意图,如图1所示,基于图计算的数据搜索方法的具体过程包括:

[0058] S110、对供方数据进行预处理得到供方中间数据组。

[0059] 其中,一供方中间数据组包括多个供方中间数据,任意一个所供方中间数据包括特征信息、特征标签标识、特征权重系数和数据分值。

[0060] 在现有的数联网方案中,将数据需求方称为需方,将数据提供方称为供方,在数据需求方从数据提供方查找需求数据的过程中,出于数据安全考虑,数据提供方并不会直接提供数据,而是给出数据的元信息和一些描述信息,数据需求方想要找到匹配的数据无疑于大海捞针。

[0061] 基于现有技术存在的问题,本公开实施例提供的基于图计算的数据搜索方法中,首先对供方数据进行预处理得到供方中间数据组,其中,对供方数据进行预处理的过程包括获取供方数据的数据标签标识和特征信息,然后基于数据标签标识和特征信息对供方数据进行分类、特征打标、特征权重分配以及数据打分等。

[0062] 在具体的实施方式中,对供方数据进行预处理得到供方中间数据组,包括:获取供方数据的数据标签标识和特征信息;根据供方数据的特征信息,确定供方数据的特征标签标识;根据供方数据的特征标签标识,确定供方数据的特征权重系数;根据供方数据的特征标签标识和特征权重系数,确定供方数据的数据分值;将供方数据对应的特征信息、特征标签标识、特征权重系数和数据分值组合形成供方数据对应的供方中间数据;根据供方数据的数据标签标识,对供方数据对应的供方中间数据进行分类得到多组供方中间数据组。

[0063] 其中,根据供方数据的特征信息,确定供方数据的特征标签标识,包括:对供方数据的特征信息中的关键特征信息添加标注标识;从特征信息中提取出添加标注标识的关键特征信息,并根据关键特征信息确定供方数据的特征标签标识。

[0064] 具体的,首先获取供方数据的数据标签标识和特征信息,其中,数据标签标识示例性包括行业标签标识,例如综合、金融、医疗、交通和教育等,也即通过获取各供方数据对应的数据标签标识,确定各供方数据所属的行业。此外,获取供方数据的特征信息,通过对各供方数据中包括的特征信息的分析,确定各供方数据包括的特征信息所对应的特征标签标识,一种示例性的,供方数据1包括“1岁”特征信息,供方数据2包括“7岁”特征信息,供方数据3包括“13岁”特征信息,则可以确定供方数据1中“1岁”特征信息对应的特征标签标识为儿童,供方数据2中“7岁”特征信息对应的特征标签标识为少年,供方数据3中“13岁”特征信息对应的特征标签标识为青年。由于不同的特征标签标识对应不同的特征权重系数,因此,在确定各供方数据对应的特征标签标识后,可以根据特征标签标识与特征权重系数的对应关系,确定各供方数据对应的特征权重系数。进一步的,基于各供方数据对应的特征标签标识和特征权重系数,确定各供方数据对应的数据分值。

[0065] 需要说明的是,上述实施例,供方数据包括的特征信息可以为一个,也可以为多个,当供方数据包括的特征信息为一个时,根据该特征信息对应的特征标签标识与特征权重系数的对应关系,确定该供方数据的特征权重系数,然后基于该特征信息对应的特征标签标识和特征权重系数确定在条供方数据在该特征信息下的数据分值,当供方数据包括的特征信息为多个时,根据每个特征信息对应的特征标签标识和特征权重系数,确定该供方数据在该特征信息下对应的数据分值,也即形成的供方中间数据包括某一个供方数据在不同特征信息下对应的特征标签标识、特征权重系数和数据分值。

[0066] 在得到每个供方数据在不同特征信息对应的特征标签标识、特征权重系数和数据分值后,将供方数据对应的特征信息、特征标签标识、特征权重系数和数据分值组合形成供方数据对应的供方中间数据,然后基于每个供方数据的数据标签标识,对各供方数据进行分类得到多组供方中间数据组,也即形成的一组供方中间数据组为同一领域的供方中间数据。

[0067] 形成的一组供方中间数据组的形式如下表所示:



$$[0068] \quad \begin{bmatrix} f_0 & p_0 & w_0 & 3 \\ f_0 & p_1 & w_1 & 6 \\ f_1 & p_1 & w_1 & 6 \\ f_2 & p_1 & w_0 & 8 \\ f_2 & p_0 & w_2 & 8 \\ \cdot & \cdot & \cdot & \cdot \\ f_x & p_x & w_x & xx \end{bmatrix}$$

[0069] 其中,  $f_x$  表示一条供方中间数据的特征信息,  $p_x$  表示一条供方中间数据的特征标签标识,  $w_x$  表示一条供方中间数据的特征权重系数,  $x_x$  表示一条供方中间数据的数据分值。

[0070] 由于供方中间数据对应包括数据分值, 供方中间数据的数据分值基于供方数据的特征标签标识和特征权重系数确定, 而供方数据的特征标签标识基于供方数据的特征信息确定, 因此, 基于供方数据的特征标签标识和特征权重系数确定的供方中间数据的数据分值反映供方数据在图计算网络中的顶点和边的关系。

[0071] 本公开实施例中, 通过对供方数据进行预处理, 得到供方中间数据, 并对得到的供方中间数据进行分类得到供方中间数据组, 由于供方中间数据中, 各供方数据以特征信息、特征标签标识、特征权重系数以及数据分值进行表示, 因此, 在保证数据不被泄露的基础上, 对供方数据基于特征信息赋予数据分值, 该数据分值可以作为后续与需方数据进行关联的依据。

[0072] S120、响应于接收到的数据需方提交的需求数据, 对需求数据进行预处理得到需求中间数据。

[0073] 其中, 需求中间数据包括特征信息、特征标签标识、特征权重系数和数据分值。

[0074] 在具体的实施方式中, 对需求数据进行预处理得到需求中间数据的具体过程包括: 获取需求数据的特征信息; 根据需求数据的特征信息, 确定需求数据的特征标签标识; 根据需求数据的特征标签标识, 确定供方数据的特征权重系数; 根据需求数据的特征标签标识和特征权重系数, 确定需求数据的数据分值; 将需求数据对应的特征信息、特征标签标识、特征权重系数和数据分值组合形成需求数据对应的需求中间数据。

[0075] 当执行步骤S110得到供方中间数据组后, 响应于数据需求方提交的需求数据, 对需求数据进行预处理得到需求中间数据, 也即首先获取需求数据的特征信息, 然后确定需求数据的特征信息对应的特征标签标识, 进而根据特征标签标识与特征权重系数的关系, 确定特征权重系数, 再进而根据需求数据的特征标签标识和特征权重系数, 确定需求数据的数据分值, 最后将需求数据对应的特征信息、特征标签标识、特征权重系数和数据分值组合形成需求数据对应的需求中间数据。

[0076] 需要说明的是, 得到需求中间数据的预处理过程与得到供方中间数据的预处理过程相同, 但是数据需求方提交的需求数据一般仅仅是一条数据(需求数据例如包括: 目标数据标签标识、目标特征标签标识对应的特征信息等), 因此, 对需求数据进行预处理得到需求中间数据的过程无对需求数据进行分类的过程。

[0077] 上述实施例中, 需求中间数据的数据分值基于需求数据的特征标签标识和特征权重系数确定, 而需求数据的特征标签标识基于需求数据的特征信息确定, 因此, 基于需求数据的特征标签标识和特征权重系数确定的需求中间数据的数据分值反映需求数据在图计

算网络中的顶点和边的关系。

[0078] 此外,数据需求方提交的需求数据可以是一条数据,也可以是需求数据与描述信息结合的数据,本公开实施例不对此进行具体限定。

[0079] S130、基于需求中间数据对应的数据分值与各供方中间数据组包括的供方中间数据对应的数据分值的匹配结果,确定与需求数据关联的目标供方数据组。

[0080] 在上述步骤S110和步骤S120中,通过对供方数据进行预处理得到供方中间数据,通过对需求数据进行预处理得到需求中间数据,在得到供方中间数据和需求中间数据后,由于需求中间数据和供方中间数据均对应包括数据分值,其中,需求中间数据的数据分值基于需求数据的特征标签标识和特征权重系数确定,而需求数据的特征标签标识基于需求数据的特征信息确定,因此,基于需求数据的特征标签标识和特征权重系数确定的需求中间数据的数据分值反映需求数据在图计算网络中的顶点和边的关系。供方中间数据对应包括数据分值,供方中间数据的数据分值基于供方数据的特征标签标识和特征权重系数确定,而供方数据的特征标签标识基于供方数据的特征信息确定,因此,基于供方数据的特征标签标识和特征权重系数确定的供方中间数据的数据分值反映供方数据在图计算网络中的顶点和边的关系。

[0081] 因此,在得到需求中间数据和供方中间数据后,可以根据需求中间数据对应的数据分值和各供方中间数据组包括的各供方中间数据对应的数据分值的匹配结果,确定与需求数据关联的目标供方数据组。

[0082] 在具体的实施方式中,基于需求中间数据与各供方中间数据组的匹配结果,确定与需求数据关联的目标供方数据组,包括:根据各供方中间数据组中各供方中间数据的数据分值,对供方中间数据组的各供方中间数据进行降序排序;依次从各供方中间数据组中选取前第一目标数量的供方中间数据所对应的数据分值,并求取平均数据分值;基于各供方中间数据组对应的平均数据分值与需求中间数据对应的数据分值的匹配结果,确定与需求数据关联的目标供方数据组。

[0083] 其中,基于各供方中间数据组对应的平均数据分值与需求中间数据对应的数据分值的匹配结果,确定与需求数据关联的目标供方数据组,包括:依次求取各所述供方中间数据组对应的平均数据分值与需求中间数据对应的数据分值的差值;根据各供方中间数据组对应的平均数据分值与需求中间数据对应的数据分值的绝对值差值,确定与需求数据关联的目标供方数据组。

[0084] 而根据各供方中间数据组对应的平均数据分值与需求中间数据对应的数据分值的绝对值差值,确定与需求数据关联的目标供方数据组的一种具体的实现方式为:根据各供方中间数据组对应的平均数据分值与需求中间数据对应的数据分值的绝对值差值,对各供方中间数据组进行降序排序;基于对各供方中间数据组的降序排序结果,选取前第二目标数量的供方中间数据组;基于选取的前第二目标数量的供方数据中间组中各供方中间数据对应的供方数据作为与需求数据关联的目标供方数据。

[0085] 在上述步骤S110和步骤S120中,通过对供方数据进行预处理得到供方中间数据,通过对需求数据进行预处理得到需求中间数据,在得到供方中间数据和需求中间数据后,首先对每一供方中间数据组包括的供方中间数据基于数据分值进行降序排序,然后根据排序结果,依次从各供方中间数据组中选取前第一目标数量的供方中间数据所对应的数据分

值,并求取平均数据分值,然后依次求取各供方中间数据组对应的平均数据分值与需求中间数据对应的数据分值的差值;最后根据各供方中间数据组对应的平均数据分值与需求中间数据对应的数据分值的绝对值差值,确定与需求数据关联的目标供方数据组。

[0086] 由于从每一供方中间数据组包括的供方中间数据中选取的第一目标数量的供方中间数据为该供方中间数据组中各供方中间数据对应的数据分值最大的前第一目标数据的供方中间数据(数据分值反映该供方数据在图计算网络中的顶点和边的关系,数据分值越大,表示该供方数据在图计算网络中的顶点和边的关联关系越大,数据分值越小,表示该供方数据在图计算网络中的顶点和边的关联关系越小),也即选取每一个供方中间数据组中数据分值较大的供方中间数据,可以较好的反应该供方中间数据组中的数据特征。

[0087] 此外,通过比较各供方中间数据组中前第一目标数量的供方中间数据所对应的数据分值的平均数据分值与需求中间数据对应的数据分值的绝对值差值满足预设绝对值差值的供方中间数据组包括的各供方中间数据对应的供方数据作为与需求数据关联的目标供方数据,也即选取数据分值相近的供方数据作为需求数据关联的目标供方数据。

[0088] 由于基于供方数据的特征标签标识和特征权重系数确定的供方中间数据的数据分值反映供方数据在图计算网络中的顶点和边的关系,基于需要数据的特征标签标识和特征权重系数确定的需求中间数据的数据分值反映需求数据在图计算网络中的顶点和边的关系,因此,根据需求数据的数据分值和供方数据的数据分值的绝对值差值的关系,从供方数据中选取目标供方数据的方式也是一种另外形式的基于图计算的方式,从供方数据中查找到目标供方数据,在保证查找效率的基础上避免数据泄露。

[0089] 本公开实施例提供的基于图计算的数据搜索方法,首先对供方数据进行预处理得到供方中间数据组;然后响应于接收到的数据需方提交的需求数据,对需求数据进行预处理得到需求中间数据;最后基于需求中间数据对应的数据分值与各供方中间数据组包括的供方中间数据对应的数据分值的匹配结果,确定与需求数据关联的目标供方数据组。本公开实施例中,通过对供方数据和需方数据进行预处理,得到供方中间数据和需方中间数据,由于供方中间数据和需方中间数据中,各中间数据以特征信息、特征标签标识、特征权重系数以及数据分值进行表示,因此,在保证数据不被泄露的基础上,对供方数据和需方数据基于特征信息赋予数据分值,可以基于供方中间数据和需方中间数据对应的数据分值,以图计算的方式,从供方数据中查找到目标供方数据。

[0090] 在上述实施例的基础上,本公开实施例还提供一种基于图计算的数据搜索装置,如图2所示,基于图计算的数据搜索装置包括:

[0091] 供方中间数据组确定模块210,用于对供方数据进行预处理得到供方中间数据组,其中,一供方中间数据组包括多个供方中间数据,任意一个供方中间数据包括特征信息、特征标签标识、特征权重系数和数据分值;

[0092] 需求中间数据确定模块220,用于响应于接收到的数据需方提交的需求数据,对需求数据进行预处理得到需求中间数据,其中,需求中间数据包括特征信息、特征标签标识、特征权重系数和数据分值;

[0093] 目标供方数据组确定模块230,用于基于需求中间数据对应的数据分值与各供方

中间数据组包括的供方中间数据对应的数据分值的匹配结果,确定与需求数据关联的目标供方数据组。

[0094] 本公开实施例提供的基于图计算的数据搜索装置,首先对供方数据进行预处理得到供方中间数据组;然后响应于接收到的数据需方提交的需求数据,对需求数据进行预处理得到需求中间数据;最后基于需求中间数据对应的数据分值与各供方中间数据组包括的供方中间数据对应的数据分值的匹配结果,确定与需求数据关联的目标供方数据组。本公开实施例中,通过对供方数据和需方数据进行预处理,得到供方中间数据和需方中间数据,由于供方中间数据和需方中间数据中,各中间数据以特征信息、特征标签标识、特征权重系数以及数据分值进行表示,因此,在保证数据不被泄露的基础上,对供方数据和需方数据基于特征信息赋予数据分值,可以基于供方中间数据和需方中间数据对应的数据分值,以图计算的方式,从供方数据中查找到目标供方数据。

[0095] 在具体的实施方式中,所述对供方数据进行预处理得到供方中间数据组,包括:

[0096] 获取所述供方数据的数据标签标识和特征信息;

[0097] 根据所述供方数据的特征信息,确定所述供方数据的特征标签标识;

[0098] 根据所述供方数据的特征标签标识,确定所述供方数据的特征权重系数;

[0099] 根据所述供方数据的特征标签标识和特征权重系数,确定所述供方数据的数据分值;

[0100] 将所述供方数据对应的特征信息、特征标签标识、特征权重系数和数据分值组合形成所述供方数据对应的供方中间数据;

[0101] 根据所述供方数据的数据标签标识,对所述供方数据对应的供方中间数据进行分类得到多组供方中间数据组。

[0102] 在具体的实施方式中,所述根据所述供方数据的特征信息,确定所述供方数据的特征标签标识,包括:

[0103] 对所述供方数据的特征信息中的关键特征信息添加标注标识;

[0104] 从所述特征信息中提取出添加标注标识的关键特征信息,并根据所述关键特征信息确定所述供方数据的特征标签标识。

[0105] 在具体的实施方式中,所述对所述需求数据进行预处理得到需求中间数据,包括:

[0106] 获取所述需求数据的特征信息;

[0107] 根据所述需求数据的特征信息,确定所述供方数据的特征标签标识;

[0108] 根据所述需求数据的特征标签标识,确定所述需求数据的特征权重系数;

[0109] 根据所述需求数据特征标签标识和特征权重系数,确定所述需求数据的数据分值;

[0110] 将所述需求数据对应的特征信息、特征标签标识、特征权重系数和数据分值组合形成所述需求数据对应的需求中间数据。

[0111] 在具体的实施方式中,所述基于所述需求中间数据与各所述供方中间数据组的匹配结果,确定与所述需求数据关联的目标供方数据组,包括:

[0112] 根据各所述供方中间数据组中各供方中间数据的数据分值,对所述供方中间数据组的各供方中间数据进行降序排序;

[0113] 依次从各所述供方中间数据组中选取前第一目标数量的供方中间数据所对应的

数据分值,并求取平均数据分值;

[0114] 基于各所述供方中间数据组对应的平均数据分值与所述需求中间数据对应的数据分值的关系,确定与所述需求数据关联的目标供方数据组。

[0115] 在具体的实施方式中,所述基于各所述供方中间数据组对应的平均数据分值与所述需求中间数据对应的数据分值的绝对值差值,确定与所述需求数据关联的目标供方数据组,包括:

[0116] 依次求取各所述供方中间数据组对应的平均数据分值与所述需求中间数据对应的数据分值的绝对值差值;

[0117] 根据各所述供方中间数据组对应的平均数据分值与所述需求中间数据对应的数据分值的绝对值差值,确定与所述需求数据关联的目标供方数据组。

[0118] 在具体的实施方式中,所述根据各所述供方中间数据组对应的平均数据分值与所述需求中间数据对应的数据分值的绝对值差值,确定与所述需求数据关联的目标供方数据组,包括:

[0119] 根据各所述供方中间数据组对应的平均数据分值与所述需求中间数据对应的数据分值的绝对值差值,对各所述供方中间数据组进行降序排序;

[0120] 基于对各所述供方中间数据组的降序排序结果,选取前第二目标数量的供方中间数据组;

[0121] 基于选取的前第二目标数量的供方数据中间组中各供方中间数据对应的供方数据作为与所述需求数据关联的目标供方数据。

[0122] 本申请实施例还提供了一种计算机设备。具体请参阅图3,图3为本实施例计算机设备基本结构框图。

[0123] 计算机设备包括通过系统总线相互通信连接存储器410和处理器420。需要指出的是,图中仅示出了具有组件410-420的计算机设备,但是应理解的是,并不要求实施所有示出的组件,可以替代的实施更多或者更少的组件。其中,本技术领域技术人员可以理解,这里的计算机设备是一种能够按照事先设定或存储的指令,自动进行数值计算和/或信息处理的设备,其硬件包括但不限于微处理器、专用集成电路(Application Specific Integrated Circuit,ASIC)、可编程门阵列(Field-Programmable Gate Array,FPGA)、数字处理器(Digital Signal Processor,DSP)、嵌入式设备等。

[0124] 计算机设备可以是桌上型计算机、笔记本、掌上电脑及云端服务器等计算设备。计算机设备可以与用户通过键盘、鼠标、遥控器、触摸板或声控设备等方式进行人机交互。

[0125] 存储器410至少包括一种类型的可读存储介质,可读存储介质包括非易失性存储器(non-volatile memory)或易失性存储器,例如,闪存(flashmemory)、硬盘、多媒体卡、卡型存储器(例如,SD或DX存储器等)、随机访问存储器(random access memory,RAM)、只读存储器(read-only memory,ROM)、可擦写可编程只读存储器(erasable programmable read-only memory,EPR0M)、电可擦写可编程只读存储器(electrically erasable programmable read-only memory,EEPROM)、可编程只读存储器(programmable read-only memory,PROM)、磁性存储器、磁盘、光盘等,RAM可以包括静态RAM或动态RAM。在一些实施例中,存储器410可以是计算机设备的内部存储单元,例如,该计算机设备的硬盘或内存。在另一些实施例中,存储器410也可以是计算机设备的外部存储设备,例如该计算机设备上配备的插接式硬盘、

智能存储卡 (Smart Media Card, SMC)、安全数字 (Secure Digital, SD) 卡或闪存卡 (Flash Card) 等。当然, 存储器410还可以既包括计算机设备的内部存储单元也包括其外部存储设备。本实施例中, 存储器410通常用于存储安装于计算机设备的操作系统和各类应用软件, 例如上述方法的程序代码等。此外, 存储器410还可以用于暂时地存储已经输出或者将要输出的各类数据。

[0126] 处理器420通常用于执行计算机设备的总体操作。本实施例中, 存储器410用于存储程序代码或指令, 程序代码包括计算机操作指令, 处理器420用于执行存储器410存储的程序代码或指令或者处理数据, 例如运行上述方法的程序代码。

[0127] 本文中, 总线可以是工业标准体系结构 (Industry Standard Architecture, ISA) 总线、外设部件互连标准 (Peripheral Component Interconnect, PCI) 总线或扩展工业标准结构 (Extended Industry Standard Architecture, EISA) 总线等。该总线系统可以分为地址总线、数据总线、控制总线等。为便于表示, 图中仅用一条粗线表示, 但并不表示仅有一根总线或一种类型的总线。

[0128] 本申请的另一实施例还提供一种计算机可读介质, 计算机可读介质可以是计算机可读信号介质或者计算机可读介质。计算机中的处理器读取存储在计算机可读介质中的计算机可读程序代码, 使得处理器能够执行在上述方法中每个步骤、或各步骤的组合中规定的功能动作; 生成实施在框图的每一块、或各块的组合中规定的功能动作的装置。

[0129] 计算机可读介质包含但不限于电子、磁性、光学、电磁、红外的存储器或半导体系统、设备或者装置, 或者前述的任意适当组合, 存储器用于存储程序代码或指令, 程序代码包括计算机操作指令, 处理器用于执行存储器存储的上述方法的程序代码或指令。

[0130] 存储器和处理器的定义, 可以参考前述计算机设备实施例的描述, 在此不再赘述。

[0131] 在本申请所提供的几个实施例中, 应该理解到, 所揭露的系统、装置和方法, 可以通过其它的方式实现。例如, 以上所描述的装置实施例仅仅是示意性的, 例如, 模块或单元的划分, 仅仅为一种逻辑功能划分, 实际实现时可以有另外的划分方式, 例如多个单元或组件可以结合或者可以集成到另一个系统, 或一些特征可以忽略, 或不执行。另一点, 所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口, 装置或单元的间接耦合或通信连接, 可以是电性, 机械或其它的形式。

[0132] 在本申请各个实施例中的各功能单元或模块可以集成在一个处理单元中, 也可以是各个单元单独物理存在, 也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现, 也可以采用软件功能单元的形式实现。

[0133] 集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用, 可以存储在一个计算机可读存储介质中。基于这样的理解, 本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来, 该计算机软件产品存储在一个存储介质中, 包括若干指令用以使得一台计算机设备 (可以是个人计算机, 服务器, 或者网络设备) 或处理器 (processor) 执行本申请各个实施例方法的全部或部分步骤。而前述的存储介质包括: U盘、移动硬盘、只读存储器 (Read-Only Memory, ROM)、随机存取存储器 (Random Access Memory, RAM)、磁碟或者光盘等各种可以存储程序代码的介质。

[0134] 除非上下文中另外明确地指出, 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。

[0135] 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。

[0136] 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。

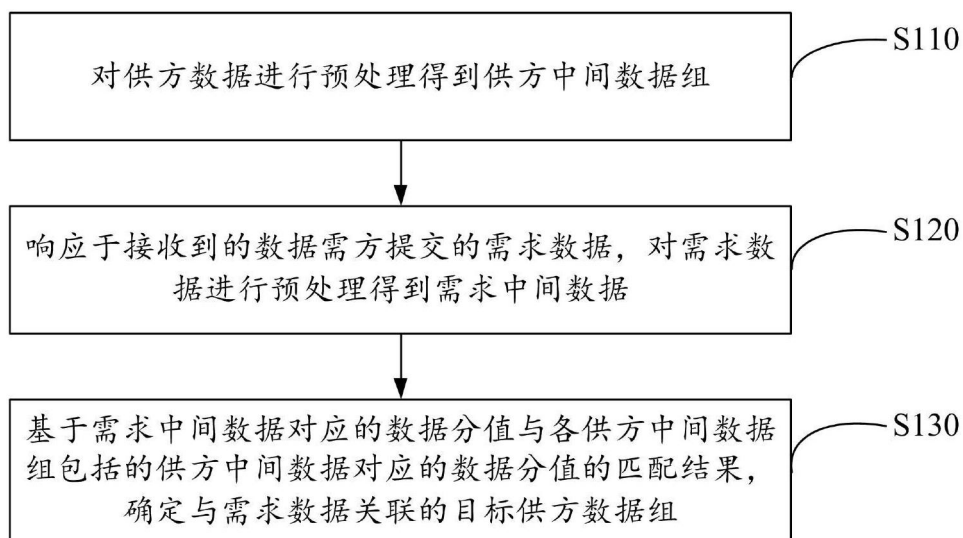


图1

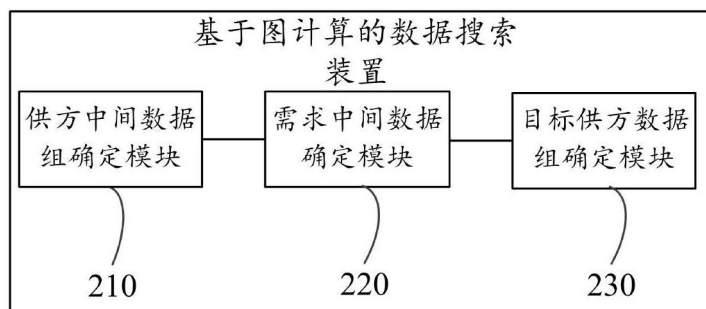


图2

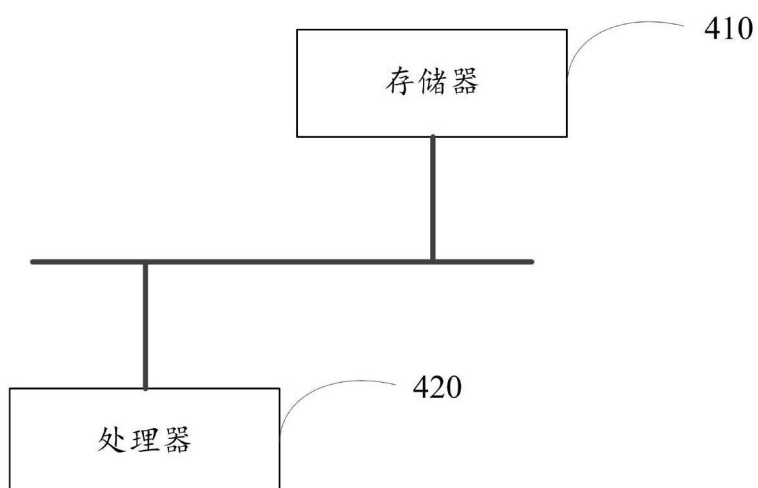


图3