

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2023 年 9 月 7 日 (07.09.2023)



(10) 国际公布号
WO 2023/165271 A1

(51) 国际专利分类号:
G06F 16/36 (2019.01) **G06Q 40/04** (2012.01)

江省杭州市西湖区西溪路 556 号 8 层 B 段
801-11, Zhejiang 310000 (CN)。

(21) 国际申请号: PCT/CN2023/071509

(22) 国际申请日: 2023 年 1 月 10 日 (10.01.2023)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202210191557.2 2022 年 3 月 1 日 (01.03.2022) CN

(71) 申请人: 支付宝 (杭州) 信息技术有限公司 (ALIPAY (HANGZHOU) INFORMATION TECHNOLOGY CO., LTD.) [CN/CN]; 中国浙江省杭州市西湖区西溪路 556 号 8 层 B 段 801-11, Zhejiang 310000 (CN)。

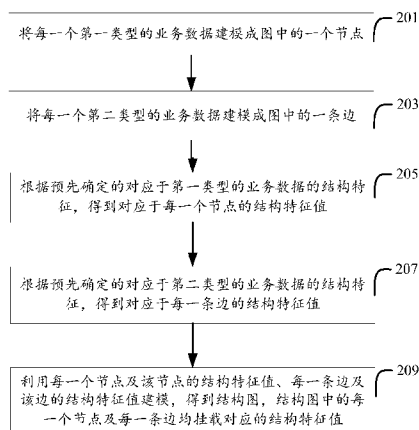
(72) 发明人: 唐坤 (TANG, Kun); 中国浙江省杭州市西湖区西溪路 556 号 8 层 B 段 801-11, Zhejiang 310000 (CN)。 易鹏 (YI, Peng); 中国浙

(74) 代理人: 北京博思佳知识产权代理有限公司 (BEIJING BESTIPR INTELLECTUAL PROPERTY LAW CORPORATION); 中国北京市海淀区上地三街 9 号嘉华大厦 B 座 409, Beijing 100085 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(54) Title: KNOWLEDGE GRAPH CONSTRUCTION AND GRAPH CALCULATION

(54) 发明名称: 知识图谱的构建、和图计算



201 Model each piece of service data of a first type into one vertex in a graph
203 Model each piece of service data of a second type into one edge in the graph
205 According to a predetermined structural feature corresponding to the service data of the first type, obtain a structural feature value corresponding to each vertex
207 According to a predetermined structural feature corresponding to the service data of the second type, obtain a structural feature value corresponding to each edge
209 Perform modeling by using each vertex and the structural feature value of the vertex, and each edge and the structural feature value of the edge, so as to obtain a structure diagram, wherein each vertex and each edge in the structure diagram are mounted with corresponding structural feature values

(57) Abstract: The present description relates to a method for constructing a knowledge graph, and graph calculation. According to an example of the method for constructing a knowledge graph, the method comprises: modelling each piece of service data of a first type into one vertex in a graph; modelling each piece of service data of a second type into one edge in the graph; according to a predetermined structural feature corresponding to the service data of the first type, obtaining a structural feature value corresponding to each vertex; according to a predetermined structural feature corresponding to the service data of the second type, obtaining a structural feature value corresponding to each edge, wherein the structural feature is a universal feature in at least two application scenarios; and performing modeling by using each vertex and the structural feature value of the vertex, and each edge and the structural feature value of the edge, so as to obtain a structure diagram, wherein each vertex and each edge in the structure diagram are mounted with corresponding structural feature values.

(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告 (条约第21条(3))。

(57) 摘要: 本说明书涉及知识图谱的构建方法和图计算。根据一种知识图谱的构建方法的示例, 将每一个第一类型的业务数据建模成图中的一个节点; 将每一个第二类型的业务数据建模成图中的一条边; 根据预先确定的对应于第一类型的业务数据的结构特征, 得到对应于每一个节点的结构特征值; 根据预先确定的对应于第二类型的业务数据的结构特征, 得到对应于每一条边的结构特征值, 所述结构特征为在至少两个应用场景中通用的特征; 利用每一个节点及该节点的结构特征值、每一条边及该边的结构特征值进行建模, 得到结构图, 结构图中的每一个节点及每一条边均挂载有对应的结构特征值。

知识图谱的构建、和图计算

技术领域

[01]本说明书一个或多个实施例涉及计算机技术，尤其涉及用于知识图谱构建、图计算的方法及装置。

背景技术

[02]图 (Graph) 是用于表示对象之间关联关系的一种抽象数据结构，使用节点 (Vertex) 和边 (Edge) 进行描述，其中，节点表示对象，边表示对象之间的关系。随着信息的爆炸式增长，为了体现各种信息之间的语义关系，基于图的思路产生了知识图谱 (Knowledge Graph)。知识图谱本质上是一种揭示实体之间关系的语义网络。在知识图谱中，图中的每一个节点都具有自己的各种特征，每一个边也具有自己的各种特征。

[03]目前构建的知识图谱中，会将一个节点及一条边的所有特征都挂载在知识图谱中，使得构建出的知识图谱异常庞大，缺乏灵活性。在基于此种知识图谱进行图计算的过程中，节点和边的所有特征都会参与计算过程，这样就会导致图计算的效率大大降低。

发明内容

[04]本说明书一个或多个实施例描述了知识图谱的构建方法和装置、图计算方法及装置，能够提高知识图谱构建的灵活性，并提高图计算的效率。

[05]根据第一方面，提供了一种知识图谱的构建方法，其包括：将每一个第一类型的业务数据建模成图中的一个节点；将每一个第二类型的业务数据建模成图中的一条边；根据预先确定的对应于第一类型的业务数据的结构特征，得到对应于每一个节点的结构特征值；根据预先确定的对应于第二类型的业务数据的结构特征，得到对应于每一条边的结构特征值；其中，所述结构特征为在至少两个应用场景中通用的特征；利用每一个节点及该节点的结构特征值、每一条边及该边的结构特征值进行建模，得到结构图。

[06]其中，在所述得到结构图后，进一步包括：针对结构图中的每一个节点，从对应于第一类型的业务数据的各应用特征中得到对应于当前应用场景的当前应用特征；针对结构图中的每一条边，从对应于第二类型的业务数据的各应用特征中得到对应于当前应用场景的当前应用特征；其中，所述应用特征与所述结构特征不同；针对结构图中的每一

个节点，将对应该节点的当前应用特征的特征值挂载到该节点上，针对结构图中的每一条边，将对应该边的当前应用特征的特征值挂载到该边上，以形成对应于当前应用场景的特征图。

[07]其中，该方法进一步包括：对每一个节点及每一条边均设置对应的全局 ID；在图特征库中，保存并动态更新每一个节点的全局 ID 与该节点的各应用特征之间的对应关系，以及保存并动态更新每一条边的全局 ID 与该边的各应用特征之间的对应关系。相应地，所述从对应于该节点各应用特征中得到对应于当前应用场景的当前应用特征，包括：从图特征库中查找到对应于该节点的全局 ID 的各应用特征，从查找到的该各应用特征中筛选出适用于当前应用场景的当前应用特征。所述从对应于该边各应用特征中得到对应于当前应用场景的当前应用特征，包括：从图特征库中查找到对应于该边的全局 ID 的各应用特征，从查找到的该各应用特征中筛选出适用于当前应用场景的当前应用特征。

[08]其中，该方法应用于具有时序性的知识图谱的构建中。

[09]其中，该方法应用于具有时序性的交易类业务的知识图谱的构建中；则所述第一类型的业务数据包括账户信息；所述第二类型的业务数据包括交易行为；所述节点的结构特征包括账户 ID；所述边的结构特征包括如下中的至少一项：时间、交易 ID、金额。

[10]根据第二方面，提供了图计算方法，其中包括：利用上述任一方法得到结构图；加载结构图中的图结构信息；所述图结构信息包括每一个节点、每一条边、每一个节点的结构特征值、每一条边的结构特征值、节点及边的顺序；利用加载的所述图结构信息进行图计算，得到流转路径。

[11]得到结构图之后，该图计算方法进一步包括：利用对应于当前应用场景的特征图及所述流转路径，进行对应于当前应用场景的图计算。

[12]根据第三方面，提供了知识图谱的构建装置，其包括：模型建立模块，配置为将每一个第一类型的业务数据建模成图中的一个节点；将每一个第二类型的业务数据建模成图中的一条边；结构特征筛选模块，配置为根据预先确定的对应于第一类型的业务数据的结构特征，得到对应于每一个节点的结构特征值；根据预先确定的对应于第二类型的业务数据的结构特征，得到对应于每一条边的结构特征值，所述结构特征为在至少两个应用场景中通用的特征；结构图构建模块，配置为利用每一个节点及该节点的结构特征值、每一条边及该边的结构特征值进行建模，得到结构图。

[13]进一步包括：应用特征筛选模块，配置为针对结构图中的每一个节点，从对应于该

节点的各应用特征中得到对应于当前应用场景的当前应用特征；针对结构图中的每一条边，从对应于该边的各应用特征中得到对应于当前应用场景的当前应用特征，所述应用特征与所述结构特征不同；特征图构建模块，配置为针对结构图中的每一个节点，将对应于该节点的当前应用特征的特征值挂载到该节点上，针对结构图中的每一条边，将对应于该边的当前应用特征的特征值挂载到该边上，以形成对应于当前应用场景的特征图。

[14]根据第四方面，提供了图计算装置，其包括：知识图谱的构建装置；以及流转路径计算模块，配置为加载结构图中的图结构信息；所述图结构信息包括每一个节点、每一条边、每一个节点的结构特征值、每一条边的结构特征值、节点及边的顺序；利用加载的所述图结构信息进行图计算，得到流转路径。

[15]所述图计算装置进一步包括：业务分析模块，配置为利用对应于当前应用场景的特征图及所述流转路径，进行对应于当前应用场景的图计算。

[16]根据第五方面，提供了一种计算设备，包括存储器和处理器，所述存储器中存储有可执行代码，所述处理器执行所述可执行代码时，实现本说明书任一实施例所述的方法。

[17]本说明书实施例提供的知识图谱的构建方法及装置、图计算的方法及装置，不是使用一个节点及一条边的所有特征来进行建模及计算，而是仅使用节点及边对应的结构特征来进行建模及计算，因为结构特征是在多个应用场景中通用的特征，因此，结构特征是节点或边的所有特征中的一部分特征，因此得到的结构图是一种能够在各种应用场景中通用的、具有精简结构（或者说具有框架结构）的知识图谱，面对目前信息量的爆炸式增长及诸如百亿级别的图计算，基于本说明书实施例中构建的知识图谱将大大降低图计算过程中所利用的特征的数量，大大提高图计算的效率。

附图说明

[18]为了更清楚地说明本说明书实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图是本说明书的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

[19]图 1 是在现有技术中对于具有时序性的交易类业务的知识图谱的示意图。

[20]图 2 是本说明书一个实施例中知识图谱的构建方法的流程图。

[21]图 3 是在本说明书一个实施例中对于具有时序性的交易类业务的结构图的示意图。

[22]图 4 是本说明书一个实施例在一种应用场景中进行知识图谱的构建方法的流程图。

[23]图 5 是本说明书一个实施例中构建的知识图谱的组成示意图。

[24]图 6 是本说明书一个实施例中基于结构图进行图计算的流程图。

[25]图 7 是本说明书一个实施例在一种应用场景中进行图计算的流程图。

[26]图 8 是本说明书一个实施例中知识图谱的构建装置的结构示意图。

[27]图 9 是本说明书另一个实施例中知识图谱的构建装置的结构示意图。

[28]图 10 是本说明书一个实施例中图计算装置的结构示意图。

[29]图 11 是本说明书另一个实施例中图计算装置的结构示意图。

具体实施方式

[30]如前所述，现有技术中构建知识图谱时，节点和边的所有特征都会参与建模过程，相应地，无论哪一个应用场景，在进行图计算时都会使用节点和边的所有特征，这样就会导致知识图谱过于庞大，图计算的效率大大降低。

[31]比如，以具有时序性的交易类业务的知识图谱为例，参见图 1 所示（可以理解，图 1 中示出的节点的数量仅仅是示意性的，其中 N 为正整数），图中的节点为用户的账户信息，边为用户之间的交易行为，那么，每一个节点包括的特征就涉及到账户的所有特征，比如账户 ID、人群、相关用户的性别、年龄、学历、账户信息、资产信息、历史交易习惯等各种信息，而每一个边包括的特征就涉及到一个交易的所有特征，比如交易 ID、交易发生的时间、交易发生的地点、金额、支付渠道、交易的性质比如是否属于违规交易等。而随着网络信息的爆炸式增长，一个知识图谱中会包括海量的节点及边，因此，导致知识图谱过于庞大，缺乏灵活性，同时，图计算的量级往往是百亿级别以上的量级，如果每一个节点及每一个边的所有特征都参与建模及计算过程，必定会大大降低图计算的效率。比如，在图计算的过程中，计算方需要把节点的所有特征及边的所有特征都存储起来，以便计算时加载使用，这样，就会占用计算方的大量存储资源。再如，每一个节点及每一个边的所有特征都参与图计算过程，会大大占用计算方的计算资源。

[32]下面结合附图，对本说明书提供的方案进行描述。

[33]图 2 是本说明书一个实施例中知识图谱的构建方法的流程图。该方法的执行主体为知识图谱的构建装置。可以理解，该方法也可以通过任何具有计算、处理能力的装置、

设备、平台、设备集群来执行。参见图 2，该方法包括以下步骤。

[34] 步骤 201：将每一个第一类型的业务数据建模成图中的一个节点。

[35] 步骤 203：将每一个第二类型的业务数据建模成图中的一条边。

[36] 步骤 205：根据预先确定的对应于第一类型的业务数据的结构特征，得到对应于每一个节点的结构特征值。

[37] 步骤 207：根据预先确定的对应于第二类型的业务数据的结构特征，得到对应于每一条边的结构特征值。

[38] 其中，结构特征为在至少两个应用场景中通用的特征。

[39] 步骤 209：利用每一个节点及该节点的结构特征值、每一条边及该边的结构特征值进行建模，得到结构图，结构图中的每一个节点及每一条边均挂载有对应的结构特征值。

[40] 可见，在图 2 所示的知识图谱的构建过程中，不是使用一个节点及一条边的所有特征来进行建模，而是仅使用节点及边对应的结构特征来进行建模，因为结构特征是在多个应用场景中通用的特征，因此，结构特征是节点或边的所有特征中的一部分特征，因此得到的结构图是一种能够在各种应用场景中通用的、具有精简结构（或者说具有框架结构）的知识图谱，此种知识图谱更具有灵活性。

[41] 下面结合附图及具体的例子对图 2 中的每一个步骤分别进行说明。

[42] 首先对于步骤 201：将每一个第一类型的业务数据建模成图中的一个节点。

[43] 本步骤中，可以将任意一种能够表征一种对象的业务数据建模成图的节点。比如，对于交易类业务，可以将一个账户信息建模成图中的一个节点。这里，账户可以是以产品/容器为单位划分的，也就是说，同一个用户的不同产品/容器会对应不同的账户信息，也就对应不同的节点。比如，用户 A 的银行账户对应节点 1，用户 A 的微信账户对应节点 2。

[44] 接下来对于步骤 203：将每一个第二类型的业务数据建模成图中的一条边。

[45] 本步骤 203 中，可以将任意一种能够表征两个对象之间的关系的业务数据建模成图的边。比如，对于交易类业务，可以将一笔交易行为建模成图中的一条边。

[46] 本说明书实施例预先定义了结构特征和应用特征。结构特征为在至少两个应用场景中通用的特征。也就是说，结构特征是在多种应用场景中都会被关注，并被用于进行多种应用场景的业务分析计算的特征。应用特征是除了结构特征之外剩余的特征，不同的

应用场景会对应各自的应用特征。

[47]为了提高图计算的效率，本说明书实施例预先从节点及边的各种类型的特征中筛选出结构特征，因为结构特征只是众多类型特征中的一部分，因此能保证图计算过程中使用的特征的数量大大减少从而提高计算效率，同时，因为结构特征是在至少两个应用场景中通用的特征，因此利用图计算过程得到的结构图能够体现适用于各种应用场景的通用的路径及流转情况，可用于后续各种应用场景的分析，即保证能够进行后续的业务分析。

[48]比如，以交易类的业务为例，在建模时，图中的节点为账户信息，边为两个账户之间的交易行为。也就是说，第一类型的业务数据是各种账户信息，第二类型的业务数据是各种交易行为。对应于账户信息这种类型的业务数据，能够在各个应用场景中通用的特征是账户 ID，也就是说，无论后续进行何种应用场景中的业务分析，都会使用该账户 ID。对应于交易行为这种类型的业务数据，能够在各个应用场景中通用的特征是金额、时间、交易 ID 中的至少一种，也就是说，无论后续进行何种应用场景中的业务分析，都会使用金额、时间、交易 ID 中的至少一种。因此，预先定义对应于账户信息（即第一类型的业务数据）的结构特征为：账户 ID。这样，对应于账户信息的应用特征为除了账户 ID 之外的其他特征，比如包括所属人群、该账户对应的用户的姓名、性别、年龄、学历、账户所属银行信息、资产信息、历史交易习惯等各种信息。同时，预先定义对应于交易行为（即第二类型的业务数据）的结构特征包括时间、交易 ID、金额；对应于交易行为的应用特征为除了时间、交易 ID、金额之外的其他特征，比如包括交易发生的地点、支付渠道、交易场景、交易是否成功、交易的性质比如是否被投诉为违规交易等。

[49]接下来对于步骤 205：根据预先确定的对应于第一类型的业务数据的结构特征，得到对应于每一个节点的结构特征值。以及对于步骤 207：根据预先确定的对应于第二类型的业务数据的结构特征，得到对应于每一条边的结构特征值。

[50]比如，仍然以上述具有时序性质的交易类业务为例，参见图 3 所示，在建模时，每一个节点只得到并挂载账户 ID 这种结构特征的特征值，比如对于节点 1，账户 ID 为 2088....0001，对于节点 2，账户 ID 为：5338.....1005；每一条边只得到并挂载金额、时间、交易 ID 这三种结构特征的特征值，比如对于边 1，金额为 200 元，时间为 2021 年 1 月 5 日 10:00，交易 ID 为 10000001，对于边 2，金额为 20 万元，时间为 2021 年 2 月 15 日 21:00，交易 ID 为 16009801。

[51]接下来对于步骤 209: 利用每一个节点及该节点的结构特征值、每一条边及该边的结构特征值进行建模, 得到结构图, 结构图中的每一个节点及每一条边均挂载有对应的结构特征值。

[52]步骤 209 中得到的结构图是一种具有精简结构的、框架形式的知识图谱, 是在多种应用场景中通用的知识图谱。

[53]如前所述, 现有技术中是将节点的所有特征及边的所有特征都构建在知识图谱中, 但是除了结构特征在各应用场景中通用之外, 不同应用场景中用到的应用特征通常是不相同的。因此, 在本说明书实施例中, 可以针对应用场景来构建专用于一个应用场景的特征图, 不同应用场景的特征图通常不同。参见图 4, 在本说明书一个实施例中, 在步骤 209 之后, 构建专用于一个应用场景的特征图的过程包括以下步骤。

[54]步骤 401: 针对结构图中的每一个节点, 从对应于第一类型的业务数据的各应用特征中得到对应于当前应用场景的当前应用特征。

[55]步骤 403: 针对结构图中的每一条边, 从对应于第二类型的业务数据的各应用特征中得到对应于当前应用场景的当前应用特征。其中, 应用特征与所述结构特征不同。

[56]步骤 405: 针对结构图中的每一个节点, 将对应该节点的当前应用特征的特征值挂载到该节点上, 针对结构图中的每一条边, 将对应该边的当前应用特征的特征值挂载到该边上, 以形成对应于当前应用场景的特征图。

[57]下面对图 4 所示的过程进行说明。

[58]如前所述, 预先定义了节点对应的各种应用特征及边对应的应用特征。而不同的应用场景进行分析计算时, 使用的应用特征是不完全相同的。比如, 对于诈骗分析这种应用场景, 在进行图计算时, 一个节点需要用到的应用特征包括账户对应的用户的历史交易习惯, 该节点不需要用到的应用特征包括账户对应的用户的性别, 一条边需要用到的应用特征包括是否被投诉为违规交易, 而该边不需要用到的应用特征包括交易是否成功。然而, 对于洗钱分析这种应用场景, 在进行图计算时, 一个节点需要用到的应用特征包括账户对应的用户的姓名及资产信息, 该节点不需要用到的应用特征包括账户对应的用户的学历, 一条边需要用到的应用特征包括交易发生的地点, 而该边不需要用到的应用特征包括是否被投诉为违规交易。

[59]因此, 当需要针对一个特定的当前应用场景进行分析时, 可以利用上述图 4 所示过程首先得到一个节点对应于当前应用场景的当前应用特征, 而不是节点的全部应用特征,

以及一条边对应于当前应用场景的当前应用特征，而不是边的全部应用特征，在形成了上述的特征图之后，就得到了专门适用于当前应用场景的特征图，可以理解，采用图 4 的方法，对于不同的应用场景，通常会得到不同的特征图，这样，再利用对应于一个应用场景的专用的特征图进行图计算，就可以有针对性地分析，得到对于该应用场景的分析结果，比如是否为赌博，或者是否出现了欺诈。

[60]在本说明书实施例中，可以预先建立图特征库，在建模时在结构图中未使用的所有的应用特征都先保存在该图特征库中，在保存时可以按照 ID 号与应用特征的对应关系的方式进行保存，也就是说，对每一个节点及每一条边均分别设置对应的全局 ID，这可以在全链路中唯一标识一个节点及一条边，在图特征库中，保存并动态更新每一个节点的全局 ID 与该节点的各应用特征之间的对应关系；同时，在图特征库中，保存并动态更新每一条边的全局 ID 与该边的各应用特征之间的对应关系。比如，将上述图 3 中节点 1 的全局 ID 与节点 1 的各应用特征的对应关系保存在图特征库中，将边 1 的全局 ID 与边 1 的各应用特征的对应关系保存在图特征库中。

[61]当一个节点或者边对应的应用特征更新时，本说明书实施例中只需要在图特征库中进行离线方式的动态更新即可，无需更新结构图。而现有技术中，因为构建的是全链路图，一个节点或者边上加载有所有的特征，如果需要增加或者减少一个特征时，需要修改全链路的配置。可见，本说明书实施例动态更新图特征库的做法，大大降低了工作量，提升了图计算业务的灵活性。

[62]这样，上述步骤 401 的一种具体实现过程包括：从图特征库中查找到对应于该节点的全局 ID 的各应用特征，从查找到的该各应用特征中筛选出适用于当前应用场景的当前应用特征。上述步骤 403 的一种具体实现过程包括：从图特征库中查找到对应于该边的全局 ID 的各应用特征，从查找到的该各应用特征中筛选出适用于当前应用场景的当前应用特征。

[63]在本说明书的实施例中，因为将所有的应用特征都先保存在图特征库中，在计算得到结构图的过程中，所有应用特征都无需通过消息传输在各节点之间传递，只需要在针对一种具体的应用场景进行业务分析计算时，再从图特征库中找出对应于此种应用场景的应用特征即可，因此，极大提高了计算效率。

[64]结合上述图 2 及图 4 所示的过程可以看出，在本说明书实施例中，采用了先分离再挂载的方式。即，先对节点及边的所有特征进行分离，即分离出结构特征及应用特征，从而利用精简的特征得到结构图，之后，再分应用场景将分离出的特定应用特征挂载到

结构图上，即进行图结构及特征的结合，从而还原出适用于一个应用场景的完整的特征图，这样就可以进行具体应用场景的图计算。

[65]通过上述图 2 所示过程，得到了结构图，即知识图谱的框架结构，之后通过图 4 所示的过程得到了对应于每一个应用场景的特征图，这样，在本说明书实施例中，构建的知识图谱可以如图 5 所示（可以理解，图 5 中示出的特征图的数量仅仅是示意性的，其中 L 为正整数），包括结构图以及至少一个特征图。

[66]在通过上述图 2 所示过程得到结构图之后，可以基于该结构图进行图计算，得到节点的流转路径，参见图 6，该图计算的过程包括以下步骤。

[67]步骤 601：得到结构图。结构图可以是利用本说明书任一实施例的方法得到的。

[68]步骤 603：加载结构图中的图结构信息；所述图结构信息包括每一个节点、每一条边、每一个节点的结构特征值、每一条边的结构特征值、节点及边的顺序。

[69]步骤 605：利用加载的所述图结构信息进行图计算，得到流转路径。

[70]在本步骤 605 中，可以针对不同的需求，利用图计算的各种方法得到节点之间的流转路径，比如遍历算法以及社区发现（Community Detection）算法等。

[71]在本说明书一个实施例中，步骤 605 的具体实现过程包括以下步骤。

[72]步骤 6051：加载结构图中的图结构信息。所述图结构信息包括每一个节点、每一条边、每一个节点的结构特征值、每一条边的结构特征值、节点及边的顺序。也就是说，不会加载任何节点及边的应用特征。

[73]步骤 6053：只利用加载的图结构信息进行消息传播、存储及计算，而不会利用应用特征进行消息的传播及存储。

[74]面对目前信息量的爆炸式增长及诸如百亿级别的图计算，基于本说明书实施例中构建的知识图谱将大大降低图计算过程中所利用的特征的数量，大大提高图计算的效率。比如，在上述图 6 所示的图计算的过程中，计算方不需要把海量的节点及边的所有特征的值都存储起来，而只需要存储各个节点及边的结构特征的值，因此，大大减少了对存储资源的占用。再如，在上述图 6 所示的图计算过程中，无需将海量的节点及边的所有特征的值在节点之间进行消息传播，而只需要将结构特征的值进行消息传播，因此，大大节约了带宽资源。又如，在上述图 6 所示的图计算过程中，无需将海量的节点及边的所有特征的值都参与计算过程，而只需要将结构特征的值参与计算过程，因此，大大节

约了计算方的计算资源。

[75]在利用图 4 所示的过程得到了对应于一个应用场景的特征图以及利用图 6 所示过程得到了节点之间的流转路径之后，则可以在不同的应用场景中进行不同的业务分析，参见图 7，具体包括以下步骤。

[76]步骤 701：得到对应于当前应用场景的特征图。

[77]步骤 703：得到利用结构图计算出的流转路径。

[78]步骤 705：利用对应于当前应用场景的特征图及流转路径，进行对应于当前应用场景的图计算。

[79]比如，对于具有时序性质的交易类业务的图计算，通过上述步骤 605 的计算过程可以计算出每笔资金的完整时序流转路径，而这种时序流转路径可以在后续的多种不同的应用场景中被使用，比如对于洗钱这种违规业务，基于图 7 所示流程，利用对应于洗钱应用场景的特征图及上述流转路径进行图计算，得到一个用户是否涉及洗钱这种违规业务；再如，对于诈骗这种违规业务，基于图 7 所示流程，利用对应于诈骗应用场景的特征图及上述流转路径进行图计算，得到一个用户是否涉及诈骗这种违规业务等。

[80]本说明书实施例的方法可以应用于各种类型的知识图谱的构建及图计算。

[81]比如，本说明书实施例的方法可以应用于具有时序性的知识图谱的构建及图计算中，如上述的具有时序性的交易类业务的知识图谱的构建及相应的图计算中。

[82]再如，本说明书实施例的方法应用于不具有时序性的知识图谱的构建及图计算中，比如对于事件类的知识图谱的构建及图计算中。在此类知识图谱中，比如，企业可以是节点，一个事件比如发生了某产品的涨价事件可以是边，企业的 ID 可以是节点的结构特征，企业的其他信息比如成立时间、与其他公司是否为子公司的关系、成立地点、法人等可以是节点的应用特征；事件 ID 可以是边的结构特征，事件发生的时间、地点、内容等可以是边的应用特征。基于上述图 2 所示方法可以得到针对事件类业务的知识图谱的框架结构即结构图，然后针对不同的应用场景，比如分析一个企业的股价上涨的原因的应用场景及分析一个企业盈亏情况的应用场景，则可以基于上述图 4 所述的方法得到对应于不同应用场景的特征图。基于图 2 得到的结构图，可以得到企业之间基于事件影响关系的流转路径，基于图 4 得到的特征图及图 6 得到的流转路径，可以针对一个应用场景分析事件影响的根本原因。

[83]在本说明书的一个实施例中，提供了一种知识图谱的构建装置，参见图 8，该装置

包括:模型建立模块 801,配置为将每一个第一类型的业务数据建模成图中的一个节点;将每一个第二类型的业务数据建模成图中的一条边;结构特征筛选模块 802,配置为根据预先确定的对应于第一类型的业务数据的结构特征,得到对应于每一个节点的结构特征值;根据预先确定的对应于第二类型的业务数据的结构特征,得到对应于每一条边的结构特征值;其中,所述结构特征为在至少两个应用场景中通用的特征;结构图构建模块 803,配置为利用每一个节点及该节点的结构特征值、每一条边及该边的结构特征值进行建模,得到结构图,结构图中的每一个节点及每一条边均挂载有对应的结构特征值。

[84]参见图 9,在本说明书装置的一个实施例中,进一步包括:应用特征筛选模块 901,配置为针对结构图中的每一个节点,从对应于该节点的各应用特征中得到对应于当前应用场景的当前应用特征;针对结构图中的每一条边,从对应于该边的各应用特征中得到对应于当前应用场景的当前应用特征;其中,所述应用特征与所述结构特征不同;特征图构建模块 902,配置为针对结构图中的每一个节点,将对应该节点的当前应用特征的特征值挂载到该节点上,针对结构图中的每一条边,将对应该边的当前应用特征的特征值挂载到该边上,以形成对应于当前应用场景的特征图。

[85]在结合图 9 描述的本说明书装置的一个实施例中,可以进一步包括图特征库;其中,图特征库,用于保存并动态更新每一个节点的全局 ID 与该节点各应用特征之间的对应关系,以及保存并动态更新每一条边的全局 ID 与该边各应用特征之间的对应关系;应用特征筛选模块 901,配置为执行:从图特征库中查找到对应于该节点的全局 ID 的各应用特征,从查找到的该各应用特征中筛选出适用于当前应用场景的当前应用特征;从图特征库中查找到对应于该边的全局 ID 的各应用特征,从查找到的该各应用特征中筛选出适用于当前应用场景的当前应用特征。

[86]在本说明书装置的一个实施例中,该装置应用于具有时序性的知识图谱的构建中,具体可以是具有时序性的交易类业务的知识图谱的构建中;所述第一类型的业务数据包括账户信息;所述第二类型的业务数据包括交易行为;节点的结构特征包括账户 ID;边的结构特征包括如下中的至少一项:时间、交易 ID、金额。

[87]在本说明书一个实施例中,还提出了一种图计算装置,参见图 10,该装置包括知识图谱的构建装置 1001 和流转路径计算模块 1002。知识图谱的构建装置 1001 采用本说明书任一实施例提供的如结合图 8 或者图 9 描述的知识图谱的构建装置实现。流转路径计算模块 1002,配置为:加载结构图中的图结构信息,所述图结构信息包括每一个节点、每一条边、每一个节点的结构特征值、每一条边的结构特征值、节点及边的顺序;利用

加载的所述图结构信息进行图计算，得到流转路径。

[88]当图计算装置中采用结合图9描述的知识图谱的构建装置实现时，参见图11，图计算装置可以进一步包括：业务分析模块1101，配置为利用对应于当前应用场景的特征图及所述流转路径，进行对应于当前应用场景的图计算。

[89]本说明书一个实施例提供了一种计算机可读存储介质，其上存储有计算机程序，当所述计算机程序在计算机中执行时，令计算机执行说明书中任一个实施例中的方法。

[90]本说明书一个实施例提供了一种计算设备，包括存储器和处理器，所述存储器中存储有可执行代码，所述处理器执行所述可执行代码时，实现执行说明书中任一个实施例中的方法。

[91]可以理解的是，本说明书实施例示意的结构并不构成对本说明书实施例的装置的具体限定。在说明书的另一些实施例中，上述装置可以包括比图示更多或者更少的部件，或者组合某些部件，或者拆分某些部件，或者不同的部件布置。图示的部件可以以硬件、软件或者软件和硬件的组合来实现。

[92]上述装置、系统内的各模块之间的信息交互、执行过程等内容，由于与本说明书方法实施例基于同一构思，具体内容可参见本说明书方法实施例中的叙述，此处不再赘述。

[93]本说明书中的各个实施例均采用递进的方式描述，各个实施例之间相同相似的部分互相参见即可，每个实施例重点说明的都是与其他实施例的不同之处。尤其，对于装置实施例而言，由于其基本相似于方法实施例，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

[94]本领域技术人员应该可以意识到，在上述一个或多个示例中，本发明所描述的功能可以用硬件、软件、挂件或它们的任意组合来实现。当使用软件实现时，可以将这些功能存储在计算机可读介质中或者作为计算机可读介质上的一个或多个指令或代码进行传输。

[95]以上所述的具体实施方式，对本发明的目的、技术方案和有益效果进行了进一步详细说明，所应理解的是，以上所述仅为本发明的具体实施方式而已，并不用于限定本发明的保护范围，凡在本发明的技术方案的基础之上，所做的任何修改、等同替换、改进等，均应包括在本发明的保护范围之内。

权利要求书

1、知识图谱的构建方法，包括：

将每一个第一类型的业务数据建模成图中的一个节点；

将每一个第二类型的业务数据建模成图中的一条边；

根据预先确定的对应于第一类型的业务数据的结构特征，得到对应于每一个节点的结构特征值；

根据预先确定的对应于第二类型的业务数据的结构特征，得到对应于每一条边的结构特征值；

其中，所述结构特征为在至少两个应用场景中通用的特征；

利用每一个节点及该节点的结构特征值、每一条边及该边的结构特征值进行建模，得到结构图。

2、根据权利要求1所述的方法，其中，在所述得到结构图后，进一步包括：

针对结构图中的每一个节点，从对应于第一类型的业务数据的各应用特征中得到对应于当前应用场景的当前应用特征；

针对结构图中的每一条边，从对应于第二类型的业务数据的各应用特征中得到对应于当前应用场景的当前应用特征；

其中，所述应用特征与所述结构特征不同；

针对结构图中的每一个节点，将对应该节点的当前应用特征的特征值挂载到该节点上，针对结构图中的每一条边，将对应该边的当前应用特征的特征值挂载到该边上，以形成对应于当前应用场景的特征图。

3、根据权利要求2所述的方法，其中，

该方法进一步包括：对每一个节点及每一条边均设置对应的全局ID；在图特征库中，保存并动态更新每一个节点的全局ID与该节点各应用特征之间的对应关系，以及保存并动态更新每一条边的全局ID与该边各应用特征之间的对应关系；

则，所述从对应于该节点的各应用特征中得到对应于当前应用场景的当前应用特征，包括：从图特征库中查找到对应于该节点的全局ID的各应用特征，从查找到的该各应用特征中筛选出适用于当前应用场景的当前应用特征；

则所述从对应于该边的各应用特征中得到对应于当前应用场景的当前应用特征，包括：从图特征库中查找到对应于该边的全局ID的各应用特征，从查找到的该各应用特征中筛选出适用于当前应用场景的当前应用特征。

4、根据权利要求1所述的方法，其中，该方法应用于具有时序性的知识图谱的构

建中。

5、根据权利要求 4 所述的方法，其中，该方法应用于具有时序性的交易类业务的知识图谱的构建中，则

所述第一类型的业务数据包括账户信息；

所述第二类型的业务数据包括交易行为；

所述节点的结构特征包括账户 ID；

所述边的结构特征包括如下中的至少一项：时间、交易 ID、金额。

6、图计算方法，其中包括：

利用权利要求 1 至 5 中任一所述的方法得到结构图；

加载结构图中的图结构信息；所述图结构信息包括每一个节点、每一条边、每一个节点的结构特征值、每一条边的结构特征值、节点及边的顺序；

利用加载的所述图结构信息进行图计算，得到流转路径。

7、根据权利要求 6 所述的方法，当利用权利要求 2 所述的方法得到结构图之后，该图计算方法进一步包括：

利用对应于当前应用场景的特征图及所述流转路径，进行对应于当前应用场景的图计算。

8、知识图谱的构建装置，包括：

模型建立模块，配置为将每一个第一类型的业务数据建模成图中的一个节点；将每一个第二类型的业务数据建模成图中的一条边；

结构特征筛选模块，配置为根据预先确定的对应于第一类型的业务数据的结构特征，得到对应于每一个节点的结构特征值；根据预先确定的对应于第二类型的业务数据的结构特征，得到对应于每一条边的结构特征值；其中，所述结构特征为在至少两个应用场景中通用的特征；

结构图构建模块，配置为利用每一个节点及该节点的结构特征值、每一条边及该边的结构特征值进行建模，得到结构图。

9、根据权利要求 8 所述的装置，进一步包括：

应用特征筛选模块，配置为针对结构图中的每一个节点，从对应于该节点的各应用特征中得到对应于当前应用场景的当前应用特征；针对结构图中的每一条边，从对应于该边的各应用特征中得到对应于当前应用场景的当前应用特征；其中，所述应用特征与所述结构特征不同；

特征图构建模块，配置为针对结构图中的每一个节点，将对应于该节点的当前应用特

征的特征值挂载到该节点上，针对结构图中的每一条边，将对应该边的当前应用特征的特征值挂载到该边上，以形成对应于当前应用场景的特征图。

10、一种图计算装置，包括：

权利要求 8 或 9 所述的知识图谱的构建装置；以及

流转路径计算模块，配置为加载结构图中的图结构信息；所述图结构信息包括每一个节点、每一条边、每一个节点的结构特征值、每一条边的结构特征值、节点及边的顺序；利用加载的所述图结构信息进行图计算，得到流转路径。

11、根据权利要求 10 所述的装置，当包括权利要求 9 所述的知识图谱的构建装置时，所述图计算装置进一步包括：

业务分析模块，配置为利用对应于当前应用场景的特征图及所述流转路径，进行对应于当前应用场景的图计算。

12、一种计算设备，包括存储器和处理器，所述存储器中存储有可执行代码，所述处理器执行所述可执行代码时，实现权利要求 1-7 中任一项所述的方法。

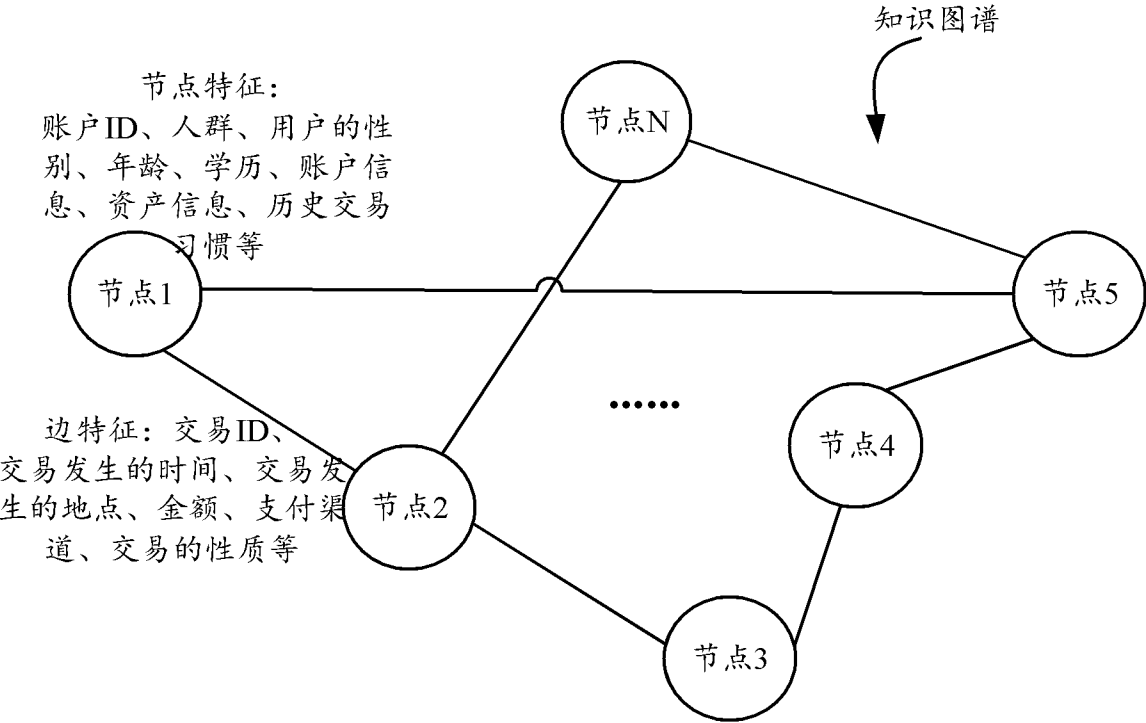


图 1

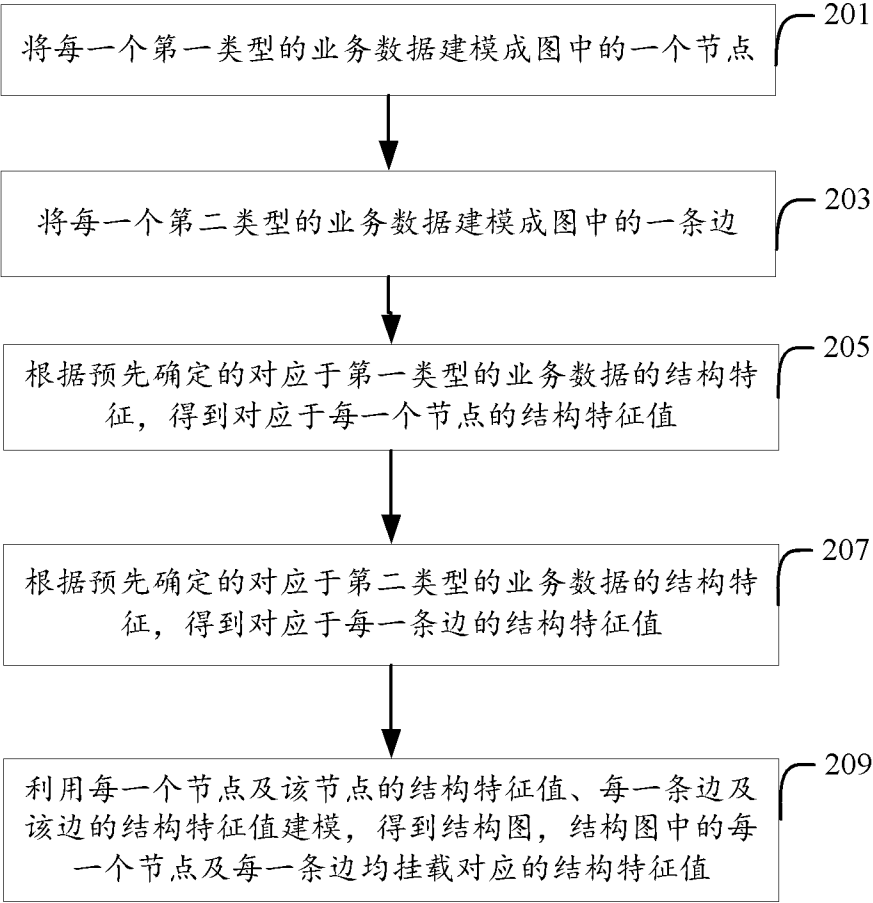


图 2

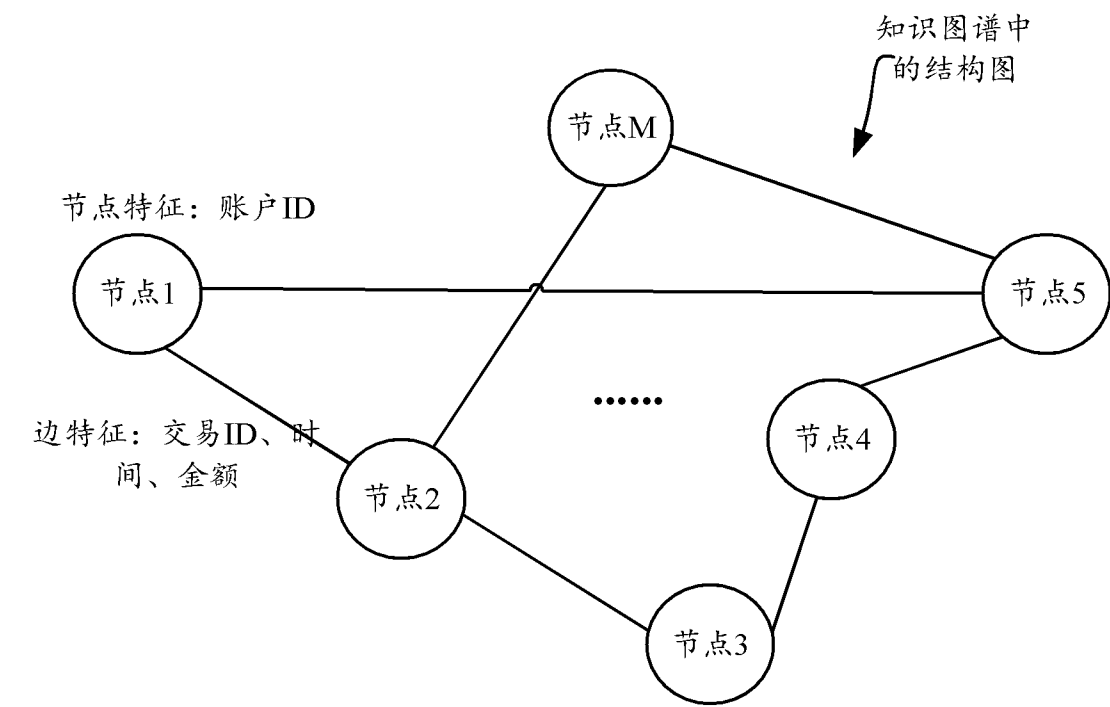


图 3

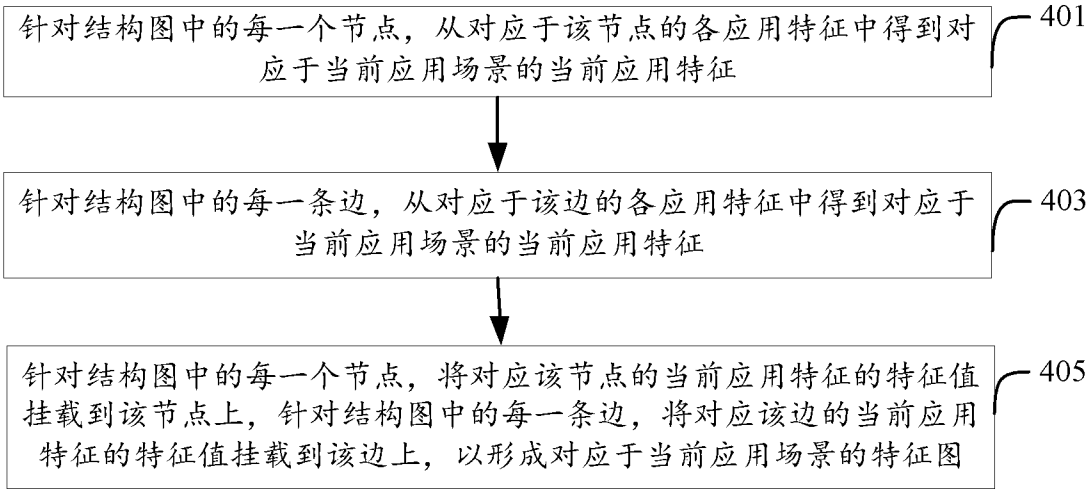


图 4

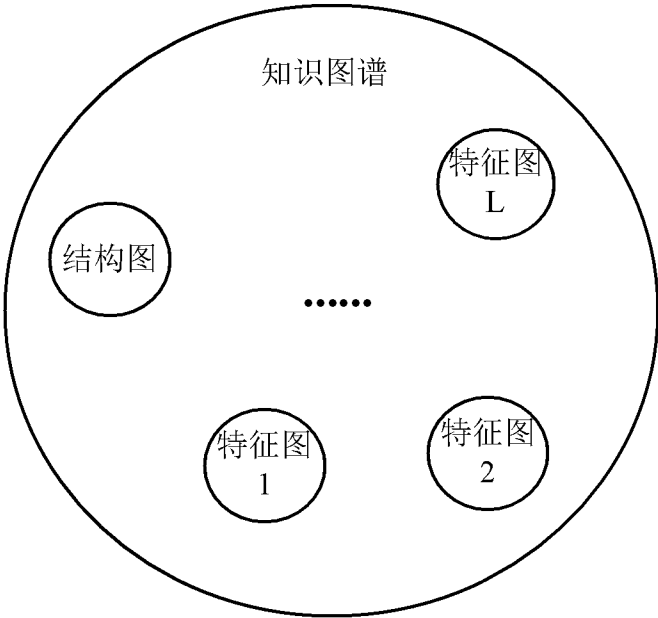


图 5

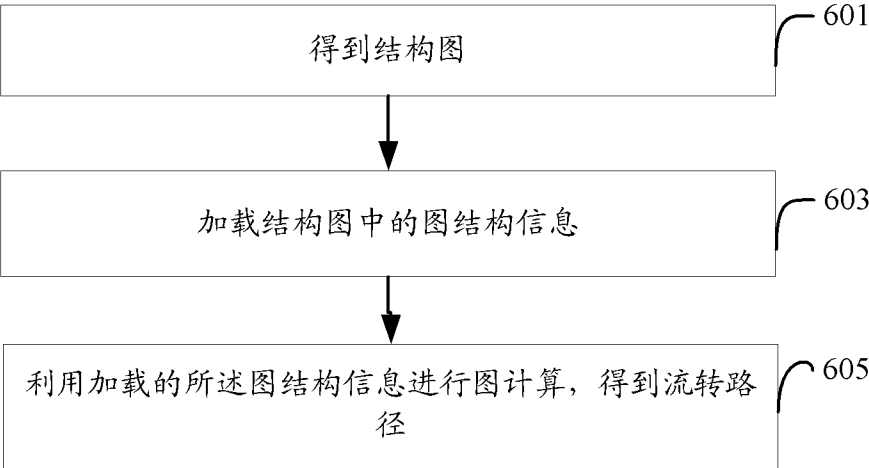


图 6

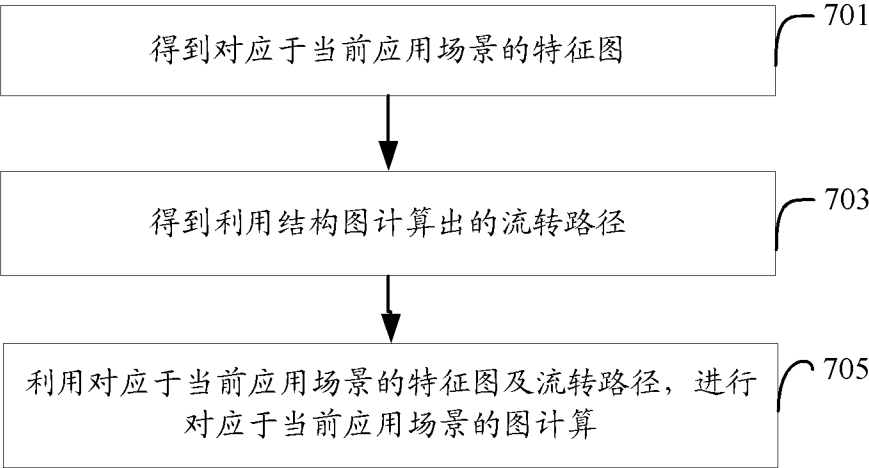


图 7

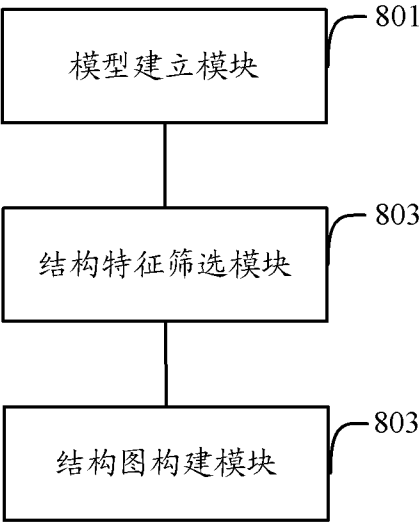


图 8

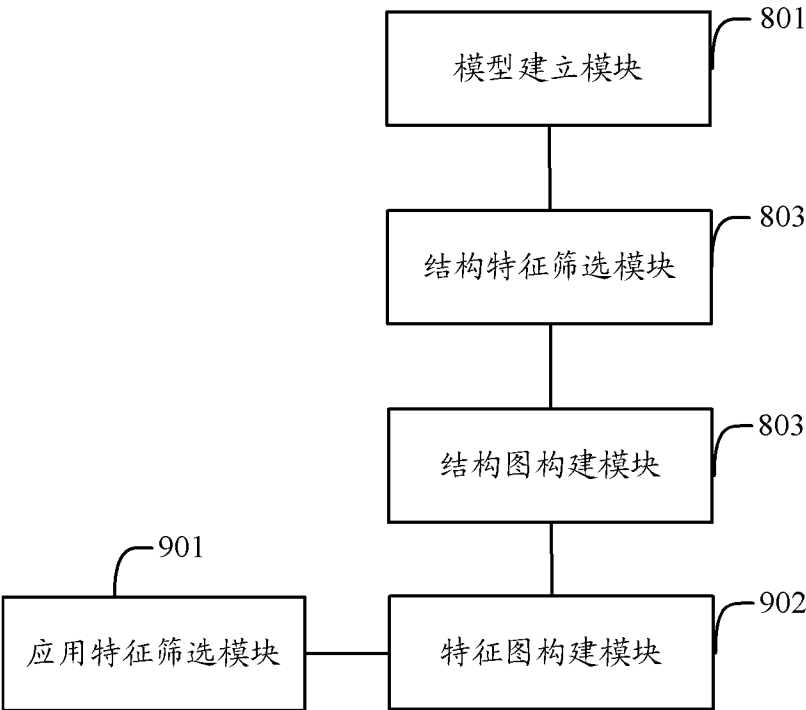


图 9

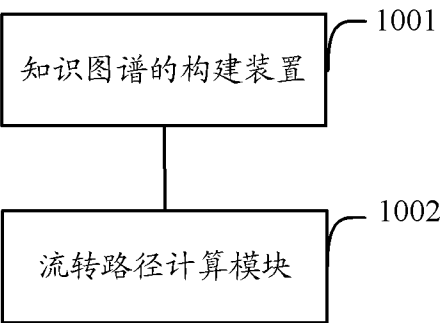


图 10

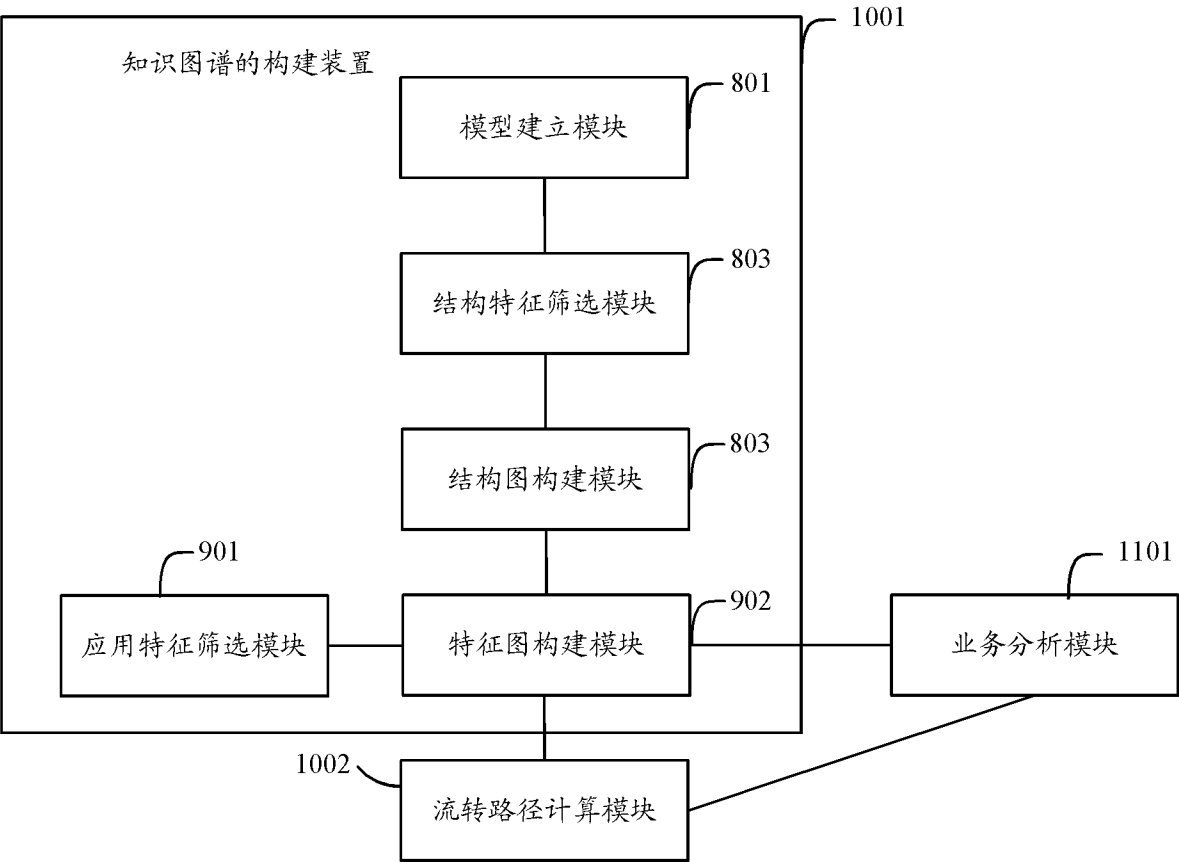


图 11

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/071509

A. CLASSIFICATION OF SUBJECT MATTER

G06F16/36(2019.01)i; G06Q40/04(2012.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: G06

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNTXT, ENTXT, ENTXTC, DWPI, CNKI: 知识图, 知识图谱, 节点, 边, 通用, 公用, 共用, 共同, 共性, 属性, 特征, 参数, 面向对象, 继承, 构建, 生成, 复用, 父, 子, graph, mapping knowledge domain, node, edge, common, characteristic, feature, parameter, object oriented, inherit, generat+, father, son, reuse

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 112966118 A (SINORAIL INFORMATION (BEIJING) NETWORK TECHNOLOGY RESEARCH INSTITUTE CO., LTD. et al.) 15 June 2021 (2021-06-15) claims 1-9, and description, paragraphs 40-106	1-12
PX	CN 114282011 A (ALIPAY (HANGZHOU) INFORMATION TECHNOLOGY CO., LTD.) 05 April 2022 (2022-04-05) claims 1-12	1-12
A	CN 111324643 A (BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO., LTD.) 23 June 2020 (2020-06-23) entire document	1-12
A	CN 111930774 A (GLOBAL ENERGY INTERCONNECTION RESEARCH INSTITUTE CO., LTD. et al.) 13 November 2020 (2020-11-13) entire document	1-12
A	CN 112256927 A (NETEASE (HANGZHOU) NETWORK CO., LTD.) 22 January 2021 (2021-01-22) entire document	1-12



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“D” document cited by the applicant in the international application

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

28 May 2023

Date of mailing of the international search report

29 May 2023

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
China No. 6, Xitucheng Road, Jimenqiao, Haidian District,
Beijing 100088

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/071509

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2021304021 A1 (ACCENTURE GLOBAL SOLUTIONS LTD.) 30 September 2021 (2021-09-30) entire document	1-12

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2023/071509

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	112966118	A	15 June 2021	None			
CN	114282011	A	05 April 2022	CN	114282011	B	23 August 2022
CN	111324643	A	23 June 2020	US	2021319329	A1	14 October 2021
				EP	3859564	A2	04 August 2021
				EP	3859564	A3	15 September 2021
				KR	20210040003	A	12 April 2021
				JP	2021119463	A	12 August 2021
				JP	7098775	B2	11 July 2022
CN	111930774	A	13 November 2020	None			
CN	112256927	A	22 January 2021	None			
US	2021304021	A1	30 September 2021	None			

国际检索报告

国际申请号

PCT/CN2023/071509

A. 主题的分类

G06F16/36 (2019.01) i; G06Q40/04 (2012.01) i

按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献 (标明分类系统和分类号)

IPC: G06

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用))

CNTXT, ENTXT, ENTXTC, DWPI, CNKI: 知识图, 知识图谱, 节点, 边, 通用, 公用, 共用, 共同, 共性, 属性, 特征, 参数, 面向对象, 继承, 构建, 生成, 复用, 父, 子, graph, mapping knowledge domain, node, edge, common, characteristic, feature, parameter, object oriented, inherit, generat+, father, son, reuse

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
X	CN 112966118 A (中铁信 (北京) 网络技术研究院有限公司 等) 2021年6月15日 (2021 - 06 - 15) 权利要求1-9, 说明书第40-106段	1-12
PX	CN 114282011 A (支付宝 (杭州) 信息技术有限公司) 2022年4月5日 (2022 - 04 - 05) 权利要求1-12	1-12
A	CN 111324643 A (北京百度网讯科技有限公司) 2020年6月23日 (2020 - 06 - 23) 全文	1-12
A	CN 111930774 A (全球能源互联网研究院有限公司 等) 2020年11月13日 (2020 - 11 - 13) 全文	1-12
A	CN 112256927 A (网易 (杭州) 网络有限公司) 2021年1月22日 (2021 - 01 - 22) 全文	1-12
A	US 2021304021 A1 (ACCENTURE GLOBAL SOLUTIONS LTD) 2021年9月30日 (2021 - 09 - 30) 全文	1-12

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“D” 申请人在国际申请中引证的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2023年5月28日

国际检索报告邮寄日期

2023年5月29日

ISA/CN的名称和邮寄地址

中国国家知识产权局

中国北京市海淀区蓟门桥西土城路6号 100088

授权官员

郑嘉青

电话号码 (+86) 62411684

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2023/071509

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	112966118	A	2021年6月15日	无	
CN	114282011	A	2022年4月5日	CN	114282011 B 2022年8月23日
CN	111324643	A	2020年6月23日	US	2021319329 A1 2021年10月14日
				EP	3859564 A2 2021年8月4日
				EP	3859564 A3 2021年9月15日
				KR	20210040003 A 2021年4月12日
				JP	2021119463 A 2021年8月12日
				JP	7098775 B2 2022年7月11日
CN	111930774	A	2020年11月13日	无	
CN	112256927	A	2021年1月22日	无	
US	2021304021	A1	2021年9月30日	无	