

(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
2024年2月1日 (01.02.2024)



(10) 国际公布号
WO 2024/021738 A1

- (51) 国际专利分类号:
G06V 10/86 (2022.01)
- (21) 国际申请号: PCT/CN2023/092130
- (22) 国际申请日: 2023年5月5日 (05.05.2023)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202210909021.X 2022年7月29日 (29.07.2022) CN
- (71) 申请人: 腾讯科技(深圳)有限公司 (TENCENT TECHNOLOGY (SHENZHEN) COMPANY LIMITED) [CN/CN]; 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层, Guangdong 518057 (CN)。

- (72) 发明人: 张杰(ZHANG, Jie); 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层, Guangdong 518057 (CN)。 黄文(HUANG, Wen); 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层, Guangdong 518057 (CN)。 董井然(DONG, Jingran); 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层, Guangdong 518057 (CN)。 陈守志(CHEN, Shouzhi); 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层, Guangdong 518057 (CN)。 陈川(CHEN, Chuan); 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层, Guangdong 518057 (CN)。 张梓暘(ZHANG, Ziyang); 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层, Guangdong 518057 (CN)。
- (74) 代理人: 华进联合专利商标代理有限公司 (ADVANCE CHINA IP LAW OFFICE); 中国广东省广州市天河区珠江东路6号4501房(部)

(54) Title: DATA NETWORK GRAPH EMBEDDING METHOD AND APPARATUS, COMPUTER DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 数据网络图的嵌入方法、装置、计算机设备和存储介质

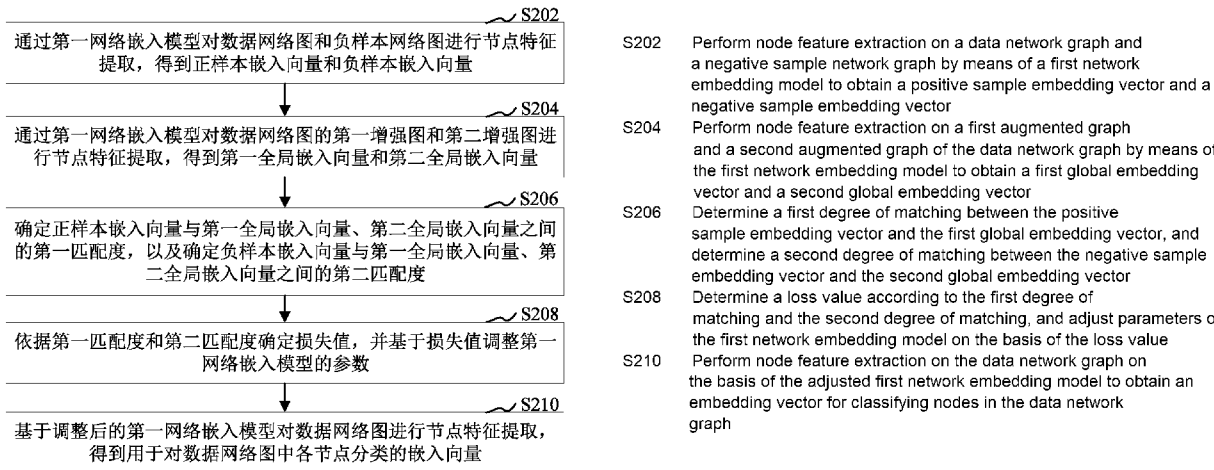


图2

(57) Abstract: The present application relates to a data network graph embedding method and apparatus, a computer device, a storage medium, and a computer program product. The method can be applied to the field of artificial intelligence and intelligent traffic networks, and comprises: performing node feature extraction on a data network graph and a negative sample network graph by means of a first network embedding model to obtain a positive sample embedding vector and a negative sample embedding vector (S202); performing node feature extraction on a first augmented graph and a second augmented graph of the data network graph by means of the first network embedding model to obtain a first global embedding vector and a second global embedding vector (S204); determining a first degree of matching and a second degree of matching (S206); adjusting parameters of the first network embedding model on the basis of a loss value determined according to the first degree of matching and the second degree of matching (S208); and performing node feature extraction on the data network graph on the basis of the adjusted first network embedding model to obtain an embedding vector for classifying nodes in the data network graph (S210).

位：自编01-03和08-12单元）（仅限办公用途），Guangdong 510623（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, MU, 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。
- (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: 本申请涉及一种数据网络图的嵌入方法、装置、计算机设备、存储介质和计算机程序产品。所述方法可应用于人工智能和智能交通网领域，包括：通过第一网络嵌入模型对数据网络图和负样本网络图进行节点特征提取，得到正样本嵌入向量和负样本嵌入向量(S202)；通过第一网络嵌入模型对数据网络图的第一增强图和第二增强图进行节点特征提取，得到第一全局嵌入向量和第二全局嵌入向量(S204)；确定第一匹配度和第二匹配度(S206)；依据第一匹配度和第二匹配度确定的损失值调整第一网络嵌入模型的参数(S208)；基于调整后的第一网络嵌入模型对数据网络图进行节点特征提取，得到用于对数据网络图中各节点分类的嵌入向量(S210)。

发明名称：数据网络图的嵌入方法、装置、计算机设备和存储介质

本申请要求于 2022 年 07 月 29 日提交中国专利局，申请号为 202210909021X，发明名称为“数据网络图的嵌入方法、装置、计算机设备和存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及人工智能技术领域，特别是涉及一种数据网络图的嵌入方法、装置、计算机设备、存储介质和计算机程序产品。

背景技术

在一些应用场景中，在获得数据集之后，需要对数据集内的数据进行分类。传统的数据分类方案中，通常是获得的数据集转换成数据网络图，然后采用网络嵌入模型对数据网络图进行节点嵌入，得到该数据网络图的嵌入向量，然后利用该嵌入向量进行分类。然而，获得的数据集通常可能是不平衡数据集，因此对应的数据网络图中不同类别节点间的特征存在差异，采用传统方案中所得的数据网络图的嵌入向量在进行节点分类时，导致分类效果较差。

发明内容

根据本申请的各种实施例，提供了一种数据网络图的嵌入方法、装置、计算机设备、计算机可读存储介质和计算机程序产品。

第一方面，本申请提供了一种数据网络图的嵌入方法，由计算机设备执行，所述方法包括：

通过第一网络嵌入模型对数据网络图和负样本网络图进行节点特征提取，得到正样本嵌入向量和负样本嵌入向量；所述数据网络图为正样本网络图，是基于不平衡的对象数据集构建所得的不平衡网络图；

通过所述第一网络嵌入模型对所述数据网络图的第一增强图和第二增强图进行节点特征提取，得到第一全局嵌入向量和第二全局嵌入向量；

确定所述正样本嵌入向量与所述第一全局嵌入向量、所述第二全局嵌入向量之间的第一匹配度，以及确定所述负样本嵌入向量与所述第一全局嵌入向量、所述第二全局嵌入向量之间的第二匹配度；

依据所述第一匹配度和所述第二匹配度确定损失值，并基于所述损失值调整所述第一网络嵌入模型的参数；

基于调整后的所述第一网络嵌入模型对所述数据网络图进行节点特征提取，得到用于对所述数据网络图中各节点分类的嵌入向量。

第二方面，本申请还提供了一种数据网络图的嵌入装置。所述装置包括：

第一提取模块，用于通过第一网络嵌入模型对数据网络图和负样本网络图进行节点特征提取，得到正样本嵌入向量和负样本嵌入向量；所述数据网络图为正样本网络图，是基于不平衡的对象数据集构建所得的不平衡网络图；

第二提取模块，用于通过所述第一网络嵌入模型对所述数据网络图的第一增强图和第二增强图进行节点特征提取，得到第一全局嵌入向量和第二全局嵌入向量；

确定模块，用于确定所述正样本嵌入向量与所述第一全局嵌入向量、所述第二全局嵌入向量之间的第一匹配度，以及确定所述负样本嵌入向量与所述第一全局嵌入向量、所述第二全局嵌入向量之间的第二匹配度；

调整模块，用于依据所述第一匹配度和所述第二匹配度确定损失值，并基于所述损失值调整所述第一网络嵌入模型的参数；

第三提取模块，用于基于调整后的所述第一网络嵌入模型对所述数据网络图进行节点特征提取，得到用于对所述数据网络图中各节点分类的嵌入向量。

第三方面，本申请还提供了一种计算机设备。所述计算机设备包括存储器和处理器，所述存储器存储有计算机程序，所述处理器执行所述计算机程序时实现所述数据网络图的嵌入方法的步骤。

第四方面，本申请还提供了一种计算机可读存储介质。所述计算机可读存储介质，其上存储有计算机程序，所述计算机程序被处理器执行时实现所述数据网络图的嵌入方法的步骤。

第五方面，本申请还提供了一种计算机程序产品。所述计算机程序产品，包括计算机程序，该计算机程序被处理器执行时实现所述数据网络图的嵌入方法的步骤。

本申请的一个或多个实施例的细节在下面的附图和描述中提出。本申请的其它特征和优点将从说明书、附图以及权利要求书变得明显。

附图说明

图 1 为一个实施例中数据网络图的嵌入方法的应用环境图；

图 2 为一个实施例中数据网络图的嵌入方法的流程示意图；

图 3 为一个实施例中将网络数据图转换负样本网络图的示意图；

图 4 为一个实施例中对网络数据图进行数据增强处理，并对所得的增强图进行低维映射的示意图；

图 5 为一个实施例中对第二网络嵌入模型进行训练和提取结构信息，以及基于结构信息和嵌入向量获得目标嵌入向量的流程示意图；

图 6 为一个实施例中对图卷积网络模型 1、图卷积网络模型 2 和分类器进行训练的示意图；

图 7 为一个实施例中数据网络图的嵌入装置的结构框图；

图 8 为另一个实施例中数据网络图的嵌入装置的结构框图；

图 9 为一个实施例中计算机设备的内部结构图。

具体实施方式

为了使本申请的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本申请进行进一步详细说明。应当理解，此处描述的具体实施例仅仅用以解释本申请，并不用于限定本申请。

本申请实施例提供的数据网络图的嵌入方法，可以应用于如图 1 所示的应用环境中。其中，终端 102 通过网络与服务器 104 进行通信。数据存储系统可以存储服务器 104 需要处理的数据。数据存储系统可以集成在服务器 104 上，也可以放在云上或其他网络服务器上。

服务器 104 通过第一网络嵌入模型对数据网络图和负样本网络图进行节点特征提取，得到正样本嵌入向量和负样本嵌入向量；数据网络图为正样本网络图；通过第一网络嵌入模型

对数据网络图的第一增强图和第二增强图进行节点特征提取，得到第一全局嵌入向量和第二全局嵌入向量；确定正样本嵌入向量与第一全局嵌入向量、第二全局嵌入向量之间的第一匹配度，以及确定负样本嵌入向量与第一全局嵌入向量、第二全局嵌入向量之间的第二匹配度；依据第一匹配度和第二匹配度确定损失值，并基于损失值调整第一网络嵌入模型的参数；基于调整后的第一网络嵌入模型对数据网络图进行节点特征提取，得到用于对数据网络图中各节点分类的嵌入向量。此外，还可以通过第二网络嵌入模型构建邻接矩阵，依据邻接矩阵与真实邻接矩阵之间的损失值调整第二网络嵌入模型的参数，从而最小化邻接矩阵与真实邻接矩阵之间的损失值，使模型能学习到与真实邻接矩阵一致或接近的结构信息，将该结构信息与嵌入向量进行拼接，得到新的用于对数据网络图中各节点分类的目标嵌入向量，利用该目标嵌入向量训练分类器，并将训练后的第一网络嵌入模型、第二网络嵌入模型和分类器进行部署。在需要执行分类任务时，终端 102 可以发起分类请求，服务器 104 响应该分类请求，调用第一网络嵌入模型和第二网络嵌入模型进行特征提取和拼接，并通过分类器对拼接所得的目标嵌入向量进行分类处理，得到分类结果，如图 1 所示。

或者，在得到用于对数据网络图中各节点分类的嵌入向量之后，服务器 104 也可以直接利用该嵌入向量训练分类器，并将训练后的第一网络嵌入模型和分类器进行部署。在需要执行分类任务时，终端 102 可以发起分类请求，服务器 104 响应该分类请求，调用第一网络嵌入模型进行特征提取，并通过分类器对提取的目标嵌入向量进行分类处理，得到分类结果。

其中，终端 102 可以是智能手机、平板电脑、笔记本电脑、台式计算机、智能音箱、智能手表、物联网设备和便携式可穿戴设备，物联网设备可为智能音箱、智能电视、智能空调和智能车载设备等。便携式可穿戴设备可为智能手表、智能手环、头戴设备等。

服务器 104 可以是独立的物理服务器，也可以是区块链系统中的服务节点，该区块链系统中的各服务节点之间形成点对点（P2P, Peer To Peer）网络，P2P 协议是一个运行在传输控制协议（TCP, Transmission Control Protocol）协议之上的应用层协议。

此外，服务器 104 还可以是多个物理服务器构成的服务器集群，可以是提供云服务、云数据库、云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全服务、内容分发网络（Content Delivery Network, CDN）、以及大数据和人工智能平台等基础云计算服务的云服务器。

终端 102 与服务器 104 之间可以通过蓝牙、USB(Universal Serial Bus, 通用串行总线) 或者网络等通讯连接方式进行连接，本申请在此不做限制。

在一个实施例中，如图 2 所示，提供了一种数据网络图的嵌入方法，以该方法应用于图 1 中的服务器 104 为例进行说明，包括以下步骤：

S202，通过第一网络嵌入模型对数据网络图和负样本网络图进行节点特征提取，得到正样本嵌入向量和负样本嵌入向量。

其中，该数据网络图为正样本网络图，是基于不平衡的对象数据集构建所得的不平衡网络图。具体地，该数据网络图是以不平衡的对象数据集内的各对象数据为节点、以关联关系为各节点的边构建的不平衡网络图。在文献引用的场景中，该对象数据可以是文献数据和文献交互对象对应的引用数据，因此该数据网络图可以是正样本文献引用关系图；在媒体交互的场景中，该对象数据可以是媒体数据和媒体交互对象对应的交互数据，因此该数据网络图可以是正样本媒体交互图；在社交场景，该对象数据可以是社交对象数据以及社交关系数据，因此该数据网络图可以是正样本社交关系图。该数据网络图是图形化的数据集，因此也可称

为图数据集。该对象数据集属于不平衡的数据集（简称不平衡数据集），表示对象数据集内各不同类型的对象数据的数量差异较大。数据网络图的数量可以是多个。

负样本网络图可以是与数据网络图之间存在特征差异的网络图，该负样本网络图的节点结构可以与数据网络图的节点结构一致，如图 3 所示。

第一网络嵌入模型属于自监督学习模块，用于将数据网络图和负样本网络图中的各节点映射到低维空间，具体可以是图卷积网络（Graph Convolutional Networks, GCN）模型、图注意力网络（Graph Attention Networks, GAN）模型或图同构网络（Graph Isomorphism Networks, GIN）模型。该图卷积网络模型可以是包括至少一层图卷积网络的神经网络模型。其中，该第一网络嵌入模型提取的正样本嵌入向量和负样本嵌入向量分别是数据网络图和负样本网络图中各节点的局部嵌入向量，属于低维空间的特征向量，对应的数据网络图和负样本网络图中各节点的特征矩阵属于高维空间的特征向量。

在一个实施例中，在 S202 之前，服务器获取对象数据集和对象数据集内各对象数据之间的关联关系；该对象数据集属于不平衡数据集；以对象数据集内的各对象数据为节点、以关联关系为各节点的边构建数据网络图。

其中，该对象数据集内的对象数据可以是文献数据，对应的关联关系可以是引用关系；此外，该对象数据集内的对象数据也可以是媒体数据和对象信息，对应的关联关系可以是交互关系，如对象点击了媒体数据，从而媒体数据与对象之间存在交互关系；此外，该对象数据集内的对象数据还可以是社交对象数据，对应的关联关系可以是社交对象之间存在好友关系。

在一个实施例中，在构建完数据网络图之后，服务器还可以对数据网络图中的节点对应的特征进行乱序处理，得到负样本网络图。例如，服务器可以将数据网络图中各节点初始的特征矩阵和邻接矩阵（即节点的结构信息）输入至腐蚀函数中，从而可以生成负样本网络图，该腐蚀函数的表达式如下：

$$(X', A') = C(X, A)$$

其中， $A' = A$ ， A 表示数据网络图中各节点的邻接矩阵， A' 表示负样本网络图中各节点的邻接矩阵， $X' = \text{Shuffle}(X)$ ， X 表示数据网络图中各节点的特征矩阵， X' 表示负样本网络图中各节点的特征矩阵， $\text{Shuffle}()$ 表示对 X 进行乱序处理。

因此，利用腐蚀函数对数据结构图进行处理的示意图可参考图 3，该腐蚀函数保留数据网络图中的节点结构不变，而对数据网络图中各节点的特征进行随机乱序处理。

在一个实施例中，服务器通过第一网络嵌入模型在数据网络图中提取各节点的嵌入向量，得到数据网络图中各节点的正样本嵌入向量；以及，服务器通过第一网络嵌入模型在负样本网络图中提取各节点的嵌入向量，得到负样本网络图中各节点的负样本嵌入向量。

具体地，服务器获取数据网络图中各节点的邻接矩阵和特征矩阵；将数据网络图中各节点的邻接矩阵和特征矩阵输入至第一网络嵌入模型，以使第一网络嵌入模型基于输入的邻接矩阵、邻接矩阵的度矩阵、特征矩阵和第一网络嵌入模型的权重矩阵，生成数据网络图中各节点的正样本嵌入向量；获取负样本网络图中各节点的邻接矩阵和特征矩阵；将负样本网络图中各节点的邻接矩阵和特征矩阵输入至第一网络嵌入模型，以使第一网络嵌入模型基于输入的邻接矩阵、邻接矩阵的度矩阵、特征矩阵和第一网络嵌入模型的权重矩阵，生成负样本网络图中各节点的负样本嵌入向量。其中，该第一网络嵌入模型可以包括两个网络嵌入分支，这两个网络嵌入分支分别对不同网络图进行节点特征提取。

例如，针对数据网络图的节点特征提取，当第一网络嵌入模型包括一层图卷积网络的网络模型时，该第一网络嵌入模型在数据网络图中各节点的邻接矩阵中添加自环，得到具有自环的邻接矩阵，然后基于添加自环的邻接矩阵、邻接矩阵的度矩阵、特征矩阵和图卷积网络的权重矩阵确定正样本嵌入向量。

当第一网络嵌入模型包括多层图卷积网络的网络模型时，该第一网络嵌入模型的第一层图卷积网络在数据网络图中各节点的邻接矩阵中添加自环，得到具有自环的邻接矩阵，然后基于添加自环的邻接矩阵、度矩阵、特征矩阵和第一层图卷积网络的权重矩阵确定第一层图卷积网络输出的嵌入向量；接着将第一层图卷积网络输出的嵌入向量作为第二层图卷积网络的输入数据，并基于添加自环的邻接矩阵、度矩阵、第二层图卷积网络的输入数据和第二层图卷积网络的权重矩阵确定第二层图卷积网络输出的嵌入向量，依此类推，得到最后一层图卷积网络输出的嵌入向量，并将最后一层图卷积网络输出的嵌入向量作为正样本嵌入向量。为了清楚说明上述计算过程，这里给出各层图卷积网络的计算公式，具体如下：

$$H^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)})$$

其中， $H^{(l)}$ 表示在处理数据网络图的过程中，第 l 层图卷积网络输出的嵌入向量； A 为数据网络图中各节点的邻接矩阵， $\tilde{A}=A+I$ 为添加自环 I 的邻接矩阵； \tilde{D} 为 \tilde{A} 的度矩阵； $W^{(l)}$ 为第 l 层图卷积网络的权重矩阵； $\sigma(\cdot)$ 为激活函数。特别地，当 $l=0$ 时， $H^{(0)}=X$ ， X 表示数据网络图中各节点的特征矩阵。若第一网络嵌入模型总共有 N 层图卷积网络，当 $l=N-1$ 时， $H^{(N)} = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(N-1)} W^{(N-1)})$ 即为数据网络图中各节点的正样本嵌入向量。

需要指出的是，对于数据网络图中第 i 个节点的正样本嵌入向量 $h_i = H_i^N = \sigma(\tilde{D}_i^{-\frac{1}{2}} \tilde{A}_i \tilde{D}_i^{-\frac{1}{2}} H_i^{(N-1)} W^{(N-1)})$ ，其中， \tilde{A}_i 为第 i 节点的添加自环的邻接矩阵， \tilde{D}_i 为 \tilde{A}_i 的度矩阵； $H_i^{(N-1)}$ 为第 $N-1$ 层图卷积网络输出关于数据网络图中第 i 个节点的嵌入向量。

同理，可参考如下计算公式计算出负样本嵌入向量：

$$H'^{(l+1)} = \sigma(\tilde{D}'^{-\frac{1}{2}} \tilde{A}' \tilde{D}'^{-\frac{1}{2}} H'^{(l)} W^{(l)})$$

其中， $H'^{(l)}$ 表示在处理负样本网络图的过程中，第 l 层图卷积网络输出的嵌入向量； A' 为负样本网络图中各节点的邻接矩阵， $\tilde{A}'=A'+I$ 为添加自环 I 的邻接矩阵； \tilde{D}' 为 \tilde{A}' 的度矩阵。特别地，当 $l=0$ 时， $H'^{(0)}=X'$ ， X' 表示负样本网络图中各节点的特征矩阵。若第一网络嵌入模型总共有 N 层图卷积网络，当 $l=N-1$ 时， $H'^{(N)} = \sigma(\tilde{D}'^{-\frac{1}{2}} \tilde{A}' \tilde{D}'^{-\frac{1}{2}} H'^{(N-1)} W^{(N-1)})$ 即为负样本网络图中各节点的正样本嵌入向量。

需要指出的是，对于负样本网络图中第 i 个节点的负样本嵌入向量 $h'_i = H_i'^N = \sigma(\tilde{D}'_i^{-\frac{1}{2}} \tilde{A}'_i \tilde{D}'_i^{-\frac{1}{2}} H_i'^{(N-1)} W^{(N-1)})$ ，其中， \tilde{A}'_i 为负样本网络图中第 i 节点的邻接矩阵， \tilde{D}'_i 为 \tilde{A}'_i 的度矩阵； $H_i'^{(N-1)}$ 为第 $N-1$ 层图卷积网络输出关于负样本网络图中第 i 个节点的正样本嵌入向量。

S204，通过第一网络嵌入模型对数据网络图的第一增强图和第二增强图进行节点特征提取，得到第一全局嵌入向量和第二全局嵌入向量。

其中，第一增强图和第二增强图分别是对数据网络图进行数据增强处理所得的增强图。第一全局嵌入向量和第二全局嵌入向量分别是第一增强图和第二增强图中各节点的全局嵌入

向量，属于低维空间的特征向量。

在一个实施例中，S204 具体可以包括：服务器通过第一网络嵌入模型，分别从第一增强图和第二增强图中提取各节点的第一局部嵌入向量和第二局部嵌入向量；分别对第一局部嵌入向量和第二局部嵌入向量进行池化处理，得到第一全局嵌入向量和第二全局嵌入向量。

其中，第一局部嵌入向量和第二局部嵌入向量分别是第一增强图和第二增强图中各节点的局部嵌入向量，也属于低维空间的特征向量。上述的池化处理可以是平均池化处理或最大池化处理等。

对于第一局部嵌入向量和第二局部嵌入向量的提取，其步骤包括：服务器获取第一增强图中各节点的第一邻接矩阵和第一特征矩阵；将第一邻接矩阵和第一特征矩阵输入至第一网络嵌入模型，以使第一网络嵌入模型在第一邻接矩阵添加自环，并基于具有自环的第一邻接矩阵、第一度矩阵、第一特征矩阵和第一网络嵌入模型的权重矩阵生成第一增强图中各节点的第一局部嵌入向量；获取第二增强图中各节点的第二邻接矩阵和第二特征矩阵；此外，服务器还将第二邻接矩阵和第二特征矩阵输入至第一网络嵌入模型，以使第一网络嵌入模型在第二邻接矩阵添加自环，并基于具有自环的第二邻接矩阵、第二度矩阵、第二特征矩阵和权重矩阵生成第二增强图中各节点的第二局部嵌入向量。

例如，当第一网络嵌入模型是包括一层图卷积网络的网络模型时，该第一网络嵌入模型在第一邻接矩阵添加自环，并基于具有自环的第一邻接矩阵、第一度矩阵、第一特征矩阵和图卷积网络的权重矩阵确定第一增强图中各节点的第一局部嵌入向量。

当第一网络嵌入模型是包括多层图卷积网络的网络模型时，该第一网络嵌入模型的第一层图卷积网络在第一邻接矩阵添加自环，并基于添加自环的第一邻接矩阵、第一度矩阵、第一特征矩阵和第一层图卷积网络的权重矩阵确定第一层图卷积网络输出的嵌入向量；然后将第一层图卷积网络输出的嵌入向量作为第二层图卷积网络的输入数据，并基于添加自环的第一邻接矩阵、第一度矩阵、第二层图卷积网络的输入数据和第二层图卷积网络的权重矩阵确定第二层图卷积网络输出的嵌入向量，依此类推，得到最后一层图卷积网络输出的嵌入向量，并将最后一层图卷积网络输出的嵌入向量作为第一增强图中各节点的第一局部嵌入向量。为了清楚说明上述计算过程，这里给出各层图卷积网络的计算公式，具体如下：

$$H_a^{(l+1)} = \sigma(\tilde{D}_a^{-\frac{1}{2}} \tilde{A}_a \tilde{D}_a^{-\frac{1}{2}} \tilde{A}_a H_a^{(l)} W^{(l)})$$

其中， $H_a^{(l)}$ 表示在处理第一增强图的过程中，第 l 层图卷积网络输出的嵌入向量； A_a 为第一增强图中各节点的第一邻接矩阵， \tilde{A}_a 为加了自环的第一邻接矩阵； \tilde{D} 为 \tilde{A}_a 的第一度矩阵； $W^{(l)}$ 为第 l 层图卷积网络的权重矩阵； $\sigma(\cdot)$ 为激活函数。特别地，当 $l=0$ 时， $H_a^{(0)}=X_a$ ， X_a 表示第一增强图中各节点的第一特征矩阵。若第一网络嵌入模型总共有 N 层图卷积网络，当 $l=N-1$ 时， $H_a^{(N)} = H_a = \sigma(\tilde{D}_a^{-\frac{1}{2}} \tilde{A}_a \tilde{D}_a^{-\frac{1}{2}} \tilde{A}_a H_a^{(N-1)} W^{(N-1)})$ 即为第一增强图中各节点的第一局部嵌入向量。

同理，可参考如下计算公式计算出第二局部嵌入向量：

$$H_b^{(l+1)} = \sigma(\tilde{D}_b^{-\frac{1}{2}} \tilde{A}_b \tilde{D}_b^{-\frac{1}{2}} \tilde{A}_b H_b^{(l)} W^{(l)})$$

其中， $H_b^{(l)}$ 表示在处理第二增强图的过程中，第 l 层图卷积网络输出的嵌入向量； A_b 为

第二增强图中各节点的第二邻接矩阵， \tilde{A}_b 为添加自环的第二邻接矩阵； \tilde{D} 为 \tilde{A}_b 的第二度矩阵。特别地，当 $l=0$ 时， $H_b^{(0)}=X_b$ ， X_b 表示第二增强图中各节点的第二特征矩阵。若第二网络嵌入模型总共有 N 层图卷积网络，当 $l=N-1$ 时， $H_b^{(N)}=H_b=\sigma(\tilde{D}_b^{-\frac{1}{2}}\tilde{A}_b\tilde{D}_b^{-\frac{1}{2}}\tilde{A}_bH_b^{(N-1)}W^{(N-1)})$ 即为第二增强图中各节点的第二局部嵌入向量。

在计算出第一局部嵌入向量和第二局部嵌入向量之后，服务器可以通过转换函数将第一局部嵌入向量和第二局部嵌入向量分别转换为第一全局嵌入向量和第二全局嵌入向量。不妨令转换函数为Readout()函数，则：

第一全局嵌入向量 $s_a = \text{Readout}(H_a)$ ；

第二全局嵌入向量 $s_b = \text{Readout}(H_b)$ 。

其中， $\text{Readout}(H_a)$ 和 $\text{Readout}(H_b)$ 可以是对 H_a 和 H_b 进行平均池化或最大池化处理，从而分别得到第一全局嵌入向量和第二全局嵌入向量。由于全局嵌入向量是为图中的所有节点所共有，因此第一增强图中各节点的第一全局嵌入向量均是相同的，第二增强图中各节点的第二全局嵌入向量也均是相同的。

S206，确定正样本嵌入向量与第一全局嵌入向量、第二全局嵌入向量之间的第一匹配度，以及确定负样本嵌入向量与第一全局嵌入向量、第二全局嵌入向量之间的第二匹配度。

其中，由于第一增强图和第二增强图均是对数据网络图进行数据增强所得的，因此正样本嵌入向量与第一全局嵌入向量、第二全局嵌入向量之间具有较高的匹配度，而负样本嵌入向量与第一全局嵌入向量、第二全局嵌入向量之间匹配度较低，因此第一匹配度大于第二匹配度。

第一匹配度可以指正样本嵌入向量与第一全局嵌入向量、第二全局嵌入向量之间的匹配程度。第二匹配度可以指负样本嵌入向量与第一全局嵌入向量、第二全局嵌入向量之间的匹配程度。

在一个实施例中，服务器可以利用鉴别器计算正样本嵌入向量与第一全局嵌入向量之间的相似分值，以及计算正样本嵌入向量与第二全局嵌入向量之间的相似分值，并将计算出来的相似分值分别作为正样本嵌入向量与第一全局嵌入向量、第二全局嵌入向量之间的第一匹配度。此外，服务器还可以利用鉴别器计算负样本嵌入向量与第一全局嵌入向量之间的相似分值，以及计算负样本嵌入向量与第二全局嵌入向量之间的相似分值，并将计算出来的相似分值分别作为负样本嵌入向量与第一全局嵌入向量、第二全局嵌入向量间的第二匹配度。

其中，鉴别器可以视为评分函数，通过该鉴别器可以计算出相似分值，从而可以反映出数据网络图的局部嵌入向量与增强图的全局嵌入向量之间的匹配度，以及反映出负样本网络图的局部嵌入向量与增强图的全局嵌入向量之间的匹配度。该鉴别器的函数表达式如下：

$$D(h_i, s) = \sigma(h_i^T W_b s)$$

其中， h_i 可以表示数据网络图中第 i 个节点的正样本嵌入向量，或负样本网络图中第 i 个节点的负样本嵌入向量； s 可以表示第一增强图的第一全局嵌入向量，或第二增强图的第二全局嵌入向量； W_b 为可学习的映射矩阵。

S208，依据第一匹配度和第二匹配度确定损失值，并基于损失值调整第一网络嵌入模型的参数。

其中，第一网络嵌入模型的参数可以是第一网络嵌入模型的权重参数，该第一网络嵌入

模型中每一层网络均有对应的权重参数，将每一层网络的权重参数组合起来可以得到该层网络的权重矩阵。

具体地，服务器将该损失值在第一网络嵌入模型中进行反向传播，得到第一网络嵌入模型中各参数的梯度，根据该梯度调整第一网络嵌入模型的参数。

对于损失值的计算，其计算步骤具体可以包括：服务器确定数据网络图中节点的数量，以及负样本网络图中节点的数量，然后将数据网络图中节点的数量、负样本网络图中节点的数量、第一匹配度和第二匹配度输入至目标函数中，得到损失值。在获得损失值之后，服务器可以根据该损失值调整第一网络嵌入模型的参数，从而优化第一网络嵌入模型的参数，使目标函数的取值最小化。

需要强调的是，在无监督训练形式下，若要学习到高质量的嵌入向量，不是将初始的特征矩阵与重构的特征矩阵之间的误差值最小化，而是使上述两个变量之间的互信息最大化，例如不是将数据网络图中各节点初始的特征矩阵与数据网络图中各节点的正样本嵌入向量之间的损失值最小化，而是使上述两个变量之间的互信息最大化，从而使第一网络嵌入模型学习到的嵌入向量尽可能多地包含数据网络图中的关键信息（如最独特且重要的信息）。此外，由于互信息是指两个变量的联合分布和它们边缘分布乘积的 KL（Kullback Leibler）散度，要使互信息最大化，就要拉大联合分布与边缘分布乘积的距离。为了简化求解难度，可以将 KL 散度转换为 JS（Jensen Shannon）散度。其中，KL 散度与 JS 散度的转换式如下：

$$JS(X, Y) = \frac{1}{2}KL\left(X \parallel \frac{X+Y}{2}\right) + \frac{1}{2}KL\left(Y \parallel \frac{X+Y}{2}\right)$$

上述转换式又可以通过负采样以及网络模型简化和近似，得到一个类似于损失函数的函数 L' ，该函数 L 如下所示：

$$L' = \frac{1}{N+M} \left(\sum_{i=1}^N E_{(X,A)} [\log D(h_i, s)] + \sum_{i=1}^M E_{(X',A')} [\log(1 - D(h'_i, s))] \right)$$

其中， $E_{(X,A)}$ 和 $E_{(X',A')}$ 分别为期望函数， $E_{(X,A)}$ 表示计算 $\log D(h_i, s)$ 的期望值， $E_{(X',A')}$ 表示计算 $1 - D(h'_i, s)$ 的期望值。在实际应用中， $E_{(X,A)}[\log D(h_i, s)] = \log D(h_i, s)$ ，以及 $E_{(X',A')}[\log(1 - D(h'_i, s))] = \log(1 - D(h'_i, s))$ ，即 $L' = \frac{1}{N+M} (\sum_{i=1}^N \log D(h_i, s) + \sum_{i=1}^M \log(1 - D(h'_i, s)))$ 。

由于 s 可以表示第一增强图的第一全局嵌入向量，或第二增强图的第二全局嵌入向量，根据上述函数 L' 可以得到目标函数，该目标函数如下所示：

$$L = \frac{1}{N+M} \left(\sum_{i=1}^N E_{(X,A)} [\log D(h_i, s_a)] + \sum_{i=1}^M E_{(X',A')} [\log(1 - D(h'_i, s_a))] \right) + \frac{1}{N+M} \left(\sum_{i=1}^N E_{(X,A)} [\log D(h_i, s_b)] + \sum_{i=1}^M E_{(X',A')} [\log(1 - D(h'_i, s_b))] \right)$$

根据 $E_{(X,A)}[\log D(h_i, s)] = \log D(h_i, s)$ ，以及 $E_{(X',A')}[\log(1 - D(h'_i, s))] = \log(1 - D(h'_i, s))$ 可化简上述表达式，得到：

$$L = \frac{1}{N+M} \left(\sum_{i=1}^N \log D(h_i, s_a) + \sum_{i=1}^M \log(1 - D(h'_i, s_a)) \right) + \frac{1}{N+M} \left(\sum_{i=1}^N \log D(h_i, s_b) + \sum_{i=1}^M \log(1 - D(h'_i, s_b)) \right)$$

因此，基于数据网络图中节点的数量、负样本网络图中节点的数量、第一匹配度和第二匹配度即可确定损失值，通过不断调整第一网络嵌入模型的参数，可以使损失函数的取值最小化，通过最小化目标函数的取值，可以使原始的特征矩阵与重构的特征矩阵之间的互信息最大化，还可以使数据网络图中各节点在两个不同视角下的增强图中嵌入的一致性最大化。例如，最小化目标函数的取值，可以使数据网络图中各节点初始的特征矩阵与数据网络图中各节点的正样本嵌入向量之间的互信息最大化，还可以使第一增强图中各节点初始的特征矩阵与第一增强图中各节点的第一局部嵌入向量之间的互信息最大化。

S210，基于调整后的第一网络嵌入模型对数据网络图进行节点特征提取，得到用于对数据网络图中各节点分类的嵌入向量。

其中，利用上述方式训练的第一网络嵌入模型，可以提取出平衡特征空间的包含重要特征且更加鲁棒的嵌入向量。

在一个实施例中，服务器可以利用该嵌入向量和分类标签对分类器进行训练，直至预测结果与分类标签一致或相近，停止分类器的训练。在完成训练后，服务器还可以将训练后的第一网络嵌入模型和分类器进行部署。在需要执行分类任务时，服务器响应于终端可以发起分类请求，调用第一网络嵌入模型对分类请求对应的文献引用关系图、媒体交互图或社交关系图进行特征提取，并通过分类器对提取的目标嵌入向量进行分类处理，得到最终的分类结果。

在一个实施例中，上述利用该嵌入向量和分类标签对分类器进行训练的步骤，具体可以包括：服务器通过分类器对嵌入向量进行分类处理，得到预测结果；基于预测结果与分类标签之间的损失值，对分类器进行参数调整；当调整后的分类器达到收敛条件时，停止训练过程。在完成训练之后，服务器可以将训练后的第一网络嵌入模型和分类器进行部署。

在需要执行分类任务时，服务器响应于终端可以发起分类请求，执行分类处理过程。其中，结合几个具体的应用场景对分类模型的处理过程进行进一步描述，具体如下所述：

应用场景 1，文献分类的场景。

在一个实施例中，服务器接收终端发起的文献分类请求，获取文献引用关系图；通过第一网络嵌入模型提取文献引用关系图的第一嵌入向量；通过分类器对第一嵌入向量进行分类处理，得到各文献的主题或所属领域。

应用场景 2，对媒体的兴趣爱好分类及推送的场景。

在一个实施例中，服务器接收终端发起的媒体推荐请求，获取媒体交互图；通过第一网络嵌入模型提取媒体交互图的第二嵌入特征；通过分类器对第二嵌入特征进行分类处理，得到对象节点对应的兴趣类型；按照兴趣类型向对象节点对应的媒体账号推荐目标媒体。

应用场景 3，对感兴趣的通信群组分类及推送的场景。

在一个实施例中，服务器接收终端发起的群组推荐请求，获取社交关系图；通过第一网络嵌入模型提取社交关系图的第三嵌入特征；通过分类器对第三嵌入特征进行分类处理，得到社交对象感兴趣的通信群组；向社交对象推送感兴趣的通信群组。

上述实施例中，通过第一网络嵌入模型对数据网络图和负样本网络图进行节点特征提取，得到正样本嵌入向量和负样本嵌入向量；此外还通过第一网络嵌入模型对数据网络图的两个不同增强图进行节点特征提取，得到第一全局嵌入向量和第二全局嵌入向量；确定正样本嵌入向量与第一全局嵌入向量、第二全局嵌入向量之间的第一匹配度，以及确定负样本嵌入向量与第一全局嵌入向量、第二全局嵌入向量之间的第二匹配度，由于上述的增强图是由数据网络图增强所得的，因此正样本嵌入向量与第一全局嵌入向量、第二全局嵌入向量之间具有较高的匹配度，而负样本嵌入向量与第一全局嵌入向量、第二全局嵌入向量之间匹配度较低，因此依据第一匹配度和第二匹配度调整第一网络嵌入模型的参数，可以使调整后的第一网络嵌入模型学习到鲁棒的且能准确分类出数据网络图中各节点的嵌入向量。此外，在训练过程中，并没有使用节点的标签，因此模型学习过程中并不会受数据网络图中多数类的影响，从而即便数据网络图为不平衡网络图，模型也可以学习到平衡的特征空间，使嵌入向量包含重要的特征且更加鲁棒，进而能够在分类过程中有效提高分类效果。此外，将训练后的第一网络嵌入模型和分类器应用在不同的应用场景中，可以实现相应的分类过程，如通过第一网络嵌入模型可以获得包含节点特征的嵌入向量，利用该嵌入向量对文献引用关系图、媒体交互图或社交关系图中的节点进行精确地分类，分别得到各文献的主题或所属领域、对象的兴趣类型和感兴趣的通信群组，有效地提高了分类效果，而且还可以准确地推送目标媒体或感兴趣的通信群组。

在一个实施例中，服务器对数据网络图进行第一数据增强处理，得到第一增强图；对数据网络图进行第二数据增强处理，得到第二增强图，如图 4 所示；其中，第一数据增强处理与第二数据增强处理分别为特征掩盖、边扰动或子图提取。需要指出的是，第一数据增强处理与第二数据增强处理可以是相同方式的数据增强处理，也可以是不同方式的数据增强处理。第一增强图和第二增强图是数据网络图的增强图，也可称为子图或增强子图。

由于第一数据增强处理和第二数据增强处理均可以是特征掩盖、边扰动或子图提取，因此可以对上述数据增强处理的方案分为以下四种场景进行描述：

场景 1，通过特征掩盖的方式获得第一增强图和第二增强图。

在一个实施例中，服务器对数据网络图中的图像区块进行特征掩盖，得到第一增强图和第二增强图。其中，特征掩盖的图像区块内的特征数值置为 0。在对第一网络嵌入模型进行训练时，可以利用数据网络图中未被掩盖的特征推断出被掩盖的特征。

场景 2，通过边扰动的方式获得第一增强图和第二增强图。

在一个实施例中，服务器对数据网络图中的边进行随机增加或删除，得到第一增强图和第二增强图。其中，对于增加数据网络图中的边或删除数据网络图中的边，可以遵循独立同分布的原则进行均匀的采样。例如，按照一定的比例对数据网络图中的边进行随机增加或删除，如随机删除 5%或 10%的边，又如随机增加 5%或 10%的边。

场景 3，通过子图提取的方式获得第一增强图和第二增强图。

在一个实施例中，服务器可以在数据网络图中进行节点采样，得到第一采样节点和第二采样节点；在数据网络图中，以第一采样节点为中心点逐级扩散采样，并在逐级扩散采样过程中，将每次采样的邻居节点置于第一采样集内；当第一采样集内的节点数量达到目标值时，停止采样，得到第一增强图；在数据网络图中，以第二采样节点为中心点逐级扩散采样，并在逐级扩散采样过程中，将每次采样的邻居节点置于第二采样集内；当第二采样集内的节点数量达到目标值时，停止采样，得到第二增强图。

其中，第一采样节点和第二采样节点可以是随机采样的节点，也可以定点采样的节点。对于第一增强图和第二增强图的采集流程，具体可以参考表 1 的算法流程：

表 1

Input: 原图 $g=(V,E)$, 图增强率 k , 采样子图 $g_s=(V_s,E_s)$, 其中, $V_s=E_s=\{\}$, 邻域节点集 $V_{neigh}=\{\}$
Output: 采样后的子图 g_s
1: 在原图中采样某一个节点 $v \in V$, 则 $V_s=\{v\}$, $V_{neigh}=\{v\}$;
2: while $ V_s \leq (1-k) V $ do
3: 从其邻居节点中采样出一个节点 $v \in V_{neigh}$;
4: if $v \in V_s$ then
5: 重新开始循环;
6: 更新采样集和邻域节点集: $V_s = V_s \cup \{v\}$, $V_{neigh} = N(v)$;
7: 更新编辑: $E_s = \{e e \in E \text{ and } (e[0] \in V_s \text{ or } e[1] \in V_s)\}$
8: 返回 g_s

场景 4, 通过混合方式获得第一增强图和第二增强图。

在一个实施例中, 服务器在数据网络图中选取采样节点, 以第一采样节点为中心点逐级扩散采样, 并在逐级扩散采样过程中, 将每次采样的邻居节点置于第一采样集内; 当第一采样集内的节点数量达到目标值时, 停止采样, 得到第一增强图; 以及, 对数据网络图进行特征掩盖, 得到第二增强图。

在另一个实施例中, 服务器在数据网络图中选取采样节点, 以第一采样节点为中心点逐级扩散采样, 并在逐级扩散采样过程中, 将每次采样的邻居节点置于第一采样集内; 当第一采样集内的节点数量达到目标值时, 停止采样, 得到第一增强图; 以及, 对数据网络图进行边扰动, 得到第二增强图。

在另一个实施例中, 服务器对数据网络图进行特征掩盖, 得到第一增强图; 以及, 对数据网络图进行边扰动, 得到第二增强图。

上述实施例中, 通过对数据网络图进行数据增强处理, 可以得到不同角度的增强图, 从而利用增强图进行模型训练时, 可以使模型具有普适性, 能够适应各种场景。

在一个实施例中, 为了更进一步提高分类效果, 还可以将第一网络嵌入模型提取的嵌入向量与数据网络图的结构信息进行拼接, 并将拼接所得的拼接向量作为用于对数据网络图中各节点分类的目标嵌入向量。具体地, 如图 5 所示, 该方法还包括:

S502, 通过第二网络嵌入模型对数据网络图进行节点特征提取, 并基于提取的节点特征重构出目标邻接矩阵。

其中, 第二网络嵌入模型属于结构保留模块, 用于对数据网络图进行结构重构。该第二网络嵌入模型可以是图卷积网络模型、图注意力网络模型或图同构网络模型。如图卷积网络模型可以是包括至少一层图卷积网络的神经网络模型。

在一个实施例中, S502 具体可以包括: 服务器获取数据网络图中各节点的特征矩阵和邻接矩阵, 将数据网络图中各节点的特征矩阵和邻接矩阵输入至第二网络嵌入模型中, 通过第二网络嵌入模型提取数据网络图中各节点的邻接矩阵对应的度矩阵, 并基于数据网络图中各节点的邻接矩阵、度矩阵、特征矩阵和第二网络嵌入模型的权重矩阵确定节点特征; 然后, 基于该节点特征和该节点特征的转置矩阵重构出目标邻接矩阵。

例如，当第二网络嵌入模型是包括一层图卷积网络的网络模型时，该第二网络嵌入模型提取数据网络图中各节点的邻接矩阵对应的度矩阵，并基于数据网络图中各节点的邻接矩阵、度矩阵、特征矩阵和图卷积网络的权重矩阵确定节点特征。

为了清楚说明上述计算过程，这里给出图卷积网络的计算公式，具体如下：

$$H_s = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} X U)$$

其中， H_s 表示图卷积网络输出的节点特征； \tilde{A} 为数据网络图中各节点的邻接矩阵，该邻接矩阵是加了自环的邻接矩阵； \tilde{D} 为 \tilde{A} 的度矩阵； U 为图卷积网络的可学习的权重矩阵； $\sigma()$ 为激活函数。

在提取出节点特征之后，服务器通过重构目标邻接矩阵的形式让模型的嵌入能够保留数据网络图中原有的结构信息，重构的表达式如下：

$$\hat{A} = H_s^T H_s$$

其中， \hat{A} 为重构的目标邻接矩阵， H_s^T 为节点特征的转置矩阵。

S504，依据目标邻接矩阵和矩阵标签之间的损失值，对第二网络嵌入模型的参数进行调整。

其中，矩阵标签指的是数据网络图的真实邻接矩阵，例如可以是数据网络图中各节点的添加自环的邻接矩阵，或者是未添加自环的邻接矩阵。

在一个实施例中，服务器基于目标损失函数计算目标邻接矩阵和矩阵标签之间的损失值，然后利用该损失值对第二网络嵌入模型的参数进行调整。该目标损失函数的表达式如下所示：

$$L = -\frac{1}{N} \sum_i \sum_j (\hat{a}_{ij} - \tilde{a}_{ij})^2$$

其中， L 表示损失值， N 为数据网络图中节点的数量， i 和 j 分别表示数据网络图中第 i 行第 j 列， \hat{a}_{ij} 为重构的第 i 行第 j 列的节点的目标邻接矩阵， \tilde{a}_{ij} 为数据网络图中第 i 行第 j 列的节点的真实邻接矩阵。

S506，当调整后的第二网络嵌入模型达到收敛条件时，通过调整后的第二网络嵌入模型获得数据网络图中各节点的结构信息。

其中，通过最小化目标损失函数，使第二网络嵌入模型达到收敛条件，从而可以使第二网络嵌入模型能学习到与真实邻接矩阵最接近的邻接矩阵。因此，在完成第二网络嵌入模型的训练之后，利用该第二网络嵌入模型获得保留数据网络图原有结构的结构信息。

S508，将嵌入向量和结构信息之间的拼接向量，作为用于对数据网络图中各节点分类的目标嵌入向量。

在一个实施例中，服务器通过第一网络嵌入模型和第二网络嵌入模型，可以分别得到包含节点特征的嵌入向量和各节点的结构信息，为了使节点获得更全面的表达能力，对上述嵌入向量和结构信息进行拼接，得到用于对数据网络图中各节点分类的目标嵌入向量。其中，目标嵌入向量的表达式如下：

$$H_f = (H_{tf} \parallel H_{sf})$$

其中， H_f 表示目标嵌入向量， H_{tf} 表示第一网络嵌入模型提取的数据网络图中各节点的嵌入向量， H_{sf} 表示第二网络嵌入模型提取的结构信息。

在一个实施例中，S408之后，该方法还包括：服务器通过分类器对目标嵌入向量进行

分类处理，得到预测结果；基于预测结果与分类标签之间的损失值，对分类器进行参数调整；当调整后的分类器达到收敛条件时，停止训练过程。

其中，对于分类器，可以选择线性模型作为分类器，如单层神经网络或支持向量机。需要指出的是，选择线性模型作为分类器，可以有效减少由分类器本身带来影响，使分类效果主要取决于模型学习的目标嵌入向量的质量。该分类器的线性映射式如下：

$$\hat{Y} = g(WH_f + b)$$

其中， $\hat{Y} \in R^{N \times C}$ 表示分类器输出的预测结果，该预测结果可以是矩阵形式的预测结果； $g()$ 为可选的放缩函数，如 $\text{softmax}()$ 等， W 和 b 是可学习的映射矩阵和偏差。接下来，通过最小化损失函数进行分类器的训练：

$$L = l(Y, \hat{Y})$$

其中， Y 是数据网络图中节点真实的分类标签。这里，对于不同分类器，可以使用不同的损失函数，如交叉熵损失（Cross Entropy Loss）函数或合页损失（Hinge Loss）函数等。

上述实施例中，通过对第二网络嵌入模型进行训练，使第二网络嵌入模型能学习到结构信息的提取，从而提取出与数据网络图原始的结构一致或接近的结构信息。将该结构信息与第一网络嵌入模型提取的包含节点关键特征的嵌入向量进行拼接，从而可以得到包含节点关键特征和结构信息的目标嵌入向量，使该目标嵌入向量具有更全面的表达能力，具有鲁棒性，能够有效地提高分类效果。

为了更加清楚本申请的方案，这里结合图6进行进一步说明，具体如下：

本申请的训练流程为分别对分类模型中的三个模块进行训练，即对自监督学习模块、网络保留模块和分类器进行训练。假设自监督学习模块和网络保留模块均采用的是图卷积网络模型（即为图卷积网络模型1和图卷积网络模型2），那么在训练时，可以同时对比图卷积网络模型1和图卷积网络模型2进行训练，然后对分类器进行训练，具体训练过程如下：

首先使用预定义的图增强算法对原图（如文献引用关系图）进行数据增强，得到不同视角下的两张增强子图，再使用图卷积网络模型1分别在增强子图、原图和负样本图中进行特征的提取，得到对应图下的嵌入向量；之后利用对比学习结合互信息最大化的方式来对图卷积网络模型1进行优化，使学习的嵌入向量中包含鲁棒且关键的特征信息。

然后利用图卷积网络模型2对原图中的节点进行卷积和变换操作，得到对应的节点特征；再根据节点特征重构出邻接矩阵，使重构出的邻接矩阵与真实图邻接矩阵之间的损失值最小化，从而使训练好的图卷积网络模型2能够提取到丰富的结构信息。

最后将图卷积网络模型1获得的包含节点特征的嵌入向量，与图卷积网络模型2获得的结构信息进行拼接，得到最终的目标嵌入向量，该目标嵌入向量包含重要的节点特征和丰富的结构信息。利用该目标嵌入向量和节点的标签信息训练分类器。

特别地，由于自监督模块和结构保留模块并无固定的执行顺序，因此二者的运算可以并行进行，提高了模型的时效性。

为了验证本申请实施例的技术效果，采用了以下数据和对比方式，可参考表2至表5：

Cora图数据集：是从学术引用网络中抽象出的图数据集，是由机器学习的论文为节点所构成的图数据集，包含2708个节点，5429条边和7个标签。其中，Cora图数据集内的每一个节点表示一篇论文，节点之间的边表示论文之间的引用关系，每篇论文的初始特征由词袋模型（bags-of-words）所生成，每个节点的标签指代了这篇论文的研究主题。

Citeseer 图数据集：是关于学术引文网络的图数据集，包含 3327 个节点，5429 条边和 6 个标签。其中节点和边分别表示文献和文献之间的引用关系，其节点特征由词袋模型生成，每个节点的标签表示这篇文献所属的研究领域。

Pubmed 图数据集：是根据生物学论文所构成的图数据集，包含 19717 个节点，44338 条边和 3 个标签。该图数据集内节点的标签表示对应生物学论文所讨论的疾病类型（如糖尿病类型），其节点特征由词袋模型所生成。

Flickr 图数据集：是提取自图片和视频分享网站的图数据集，在该分享网站中用户通过图片和视频共享的方式进行互动交流，该图数据集内包含了 7575 个节点，239738 条边和 9 种标签。其中的节点表示用户，节点间的边表示用户之间的关系，节点标签表示用户对应的兴趣组。

BlogCatalog 图数据集：源于社交媒体网站的图数据集，其中的节点表示用户，节点之间的边表示用户之间的关注关系，其节点特征由 word2vec 模型所生成，节点的标签表示用户加入的兴趣小组。数据集中包含了 5196 个节点，171743 条边和 6 个标签。

表 2

数据集	节点数	边数	类别数	特征维度
Cora图数据集	2708	5429	7	1433
Citeseer图数据集	3327	4732	6	3703
Pubmed图数据集	19717	44338	3	500
Flickr图数据集	7575	239738	9	12047
BlogCatalog图数据集	5196	171743	6	8189

为了证明本申请的模型的有效性，一方面会与常用的网络嵌入模型和处理不平衡问题的常用方法进行比较，另一方面也会与一些最近发表的针对网络数据上的不平衡问题所设计的模型进行比较，本申请中所用的对比方法具体介绍如下：

(1) 传统的网络嵌入模型：

GCN：是网络嵌入中使用最广泛的基准模型，目前大多数的网络模型均是根据它改进而来。其通过邻接矩阵表示的拓扑关系聚合邻域的嵌入，为每个节点学习到对应的嵌入向量。

APPNP：是网络解耦模型中的代表，一方面其通过解构特征传播和特征变换来减少参数数量，另一方面其基于个性化 PageRank 来改进特征传递方式，扩大了模型的感知域。

SGC：将非线性的 GCN 模型转化为了一个简单的线性模型，其通过去除 GCN 层之间的非线性计算来将函数折叠成一个线性变换从而减少 GCNs 的额外复杂度，且效果在某些实验中优于 GCN。

(2) 针对不平衡问题的通用方法：

重加权方法 (re-weight)：属于代价敏感 (cost-sensitive) 类的算法。其为少数类分配更高的损失权重同时为多数类分配较低的权重，来缓解由多数类主导函数损失下降方向的问题。

过采样 (over-sampling)：过采样的具体方法是从少数类样本中进行重复抽样，然后将抽取的样本重新加入少数类样本集从而使数据集变得相对平衡。在实验中，被抽取的节点依然会保留其原先的邻接关系。

(3) 最近的不平衡网络嵌入方法：

RECT：是基于图卷积网络的嵌入模型，针对完全不平衡问题所设计。其通过特征分解、

建模类间关系和网络结构来使模型学习到每类样本对应的语义信息，辅助不平衡模型的学习。

GraphSMOTE: 先通过插值的方法生成少数类的新节点，再训练一个边分类器来为这些节点增加连边来将网络变得平衡，最后再进行节点嵌入的生成。

上述模型对不同的不平衡率的图数据集进行节点分类，得到如下结果：

表 3 关于不平衡率为 0.1 的图数据集

模型	指标	Cora图数据集	Citeseer图数据集	Pubmed图数据集	Flickr图数据集	BlogCatalog图数据集
GCN	Micro-F1	0.5784	0.4542	0.6317	0.4622	0.6527
	Micro-F1	0.5382	0.4136	0.5397	0.3813	0.6363
Re-weight	Micro-F1	0.6345	0.4765	0.6425	0.4757	0.6657
	Micro-F1	0.6132	0.4443	0.5644	0.4103	0.6512
Over-sampling	Micro-F1	0.5896	0.4602	0.6422	0.4572	0.6651
	Micro-F1	0.5486	0.4187	0.5877	0.3895	0.6528
APPNP	Micro-F1	0.6086	0.4635	0.6533	0.5062	0.6852
	Micro-F1	0.5621	0.3876	0.5761	0.4378	0.6611
SGC	Micro-F1	0.5645	0.4768	0.6538	0.4279	0.6431
	Micro-F1	0.5026	0.4119	0.5702	0.3609	0.6271
RECT	Micro-F1	0.6551	0.5442	0.6298	0.5275	0.6713
	Micro-F1	0.6248	0.5209	0.5325	0.4475	0.6602
grapSMOTE	Micro-F1	0.6357	0.4783	0.6883	0.3472	0.6831
	Micro-F1	0.6273	0.4444	0.6496	0.2837	0.6321
本申请的方案	Micro-F1	0.6807	0.5611	0.6831	0.5261	0.8146
	Micro-F1	0.6547	0.5407	0.6321	0.4749	0.8069

表 4 关于不平衡率为 0.3 的图数据集

模型	指标	Cora图数据集	Citeseer图数据集	Pubmed图数据集	Flickr图数据集	BlogCatalog图数据集
GCN	Micro-F1	0.7385	0.5417	0.7397	0.5181	0.7197
	Micro-F1	0.7372	0.5438	0.7229	0.4875	0.7157
Re-weight	Micro-F1	0.7402	0.5512	0.7519	0.5359	0.7433
	Micro-F1	0.7402	0.5511	0.7419	0.5173	0.7433
Over-sampling	Micro-F1	0.7387	0.5338	0.7283	0.5196	0.7397
	Micro-F1	0.7394	0.5327	0.7196	0.5026	0.7297
APPNP	Micro-F1	0.7625	0.5785	0.7616	0.5391	0.7259
	Micro-F1	0.7619	0.5751	0.7434	0.5131	0.8367
SGC	Micro-F1	0.7472	0.5913	0.7294	0.4951	0.8376
	Micro-F1	0.7448	0.5905	0.7039	0.4788	0.7187
RECT	Micro-F1	0.7863	0.6338	0.7136	0.5374	0.7155
	Micro-F1	0.7875	0.6371	0.6861	0.5281	0.7052
grapSMOTE	Micro-F1	0.7488	0.5517	0.7561	0.3651	0.7218
	Micro-F1	0.7515	0.5556	0.7502	0.2929	0.7202
本申请的方案	Micro-F1	0.8086	0.6679	0.7739	0.5544	0.8767
	Micro-F1	0.8071	0.6712	0.7522	0.5369	0.8765

表 5 关于不平衡率为 0.5 的图数据集

模型	指标	Cora图数据集	Citeseer图数据集	Pubmed图数据集	Flickr图数据集	BlogCatalog图数据集
GCN	Micro-F1	0.7482	0.6008	0.7828	0.5784	0.7342
	Micro-F1	0.7512	0.6015	0.7765	0.5667	0.7292
Re-weight	Micro-F1	0.7748	0.6075	0.7965	0.5895	0.7488
	Micro-F1	0.7766	0.6074	0.7918	0.5845	0.7449
Over-sampling	Micro-F1	0.7739	0.5987	0.7749	0.5788	0.7471
	Micro-F1	0.7752	0.5998	0.7727	0.5747	0.7432
APPNP	Micro-F1	0.8114	0.6438	0.7961	0.5408	0.8435
	Micro-F1	0.8113	0.6472	0.7896	0.5298	0.8435
SGC	Micro-F1	0.7957	0.6521	0.7845	0.5408	0.7405
	Micro-F1	0.7971	0.6544	0.7769	0.5267	0.7385
RECT	Micro-F1	0.8116	0.6647	0.7498	0.5855	0.7379
	Micro-F1	0.8124	0.6679	0.7399	0.5835	0.7366
grapSMOTE	Micro-F1	0.7835	0.6076	0.7963	0.4251	0.7049
	Micro-F1	0.7855	0.6118	0.7949	0.3989	0.7005
本申请的方案	Micro-F1	0.8289	0.6839	0.8004	0.6109	0.8802
	Micro-F1	0.8298	0.6837	0.7954	0.6032	0.8796

需要指出的是，在 Micro-F 和 Macro-F 这两个指标下，表 3~5 中的数据越大，对应的效果最优。因此，通过表 3~5 中的数据可知，在 Micro-F 和 Macro-F 这两个指标下，本申请的

方案取得了最优的实验效果。

在获得训练后的第一网络嵌入模型、第二网络嵌入模型和分类器之后，可以将第一网络嵌入模型、第二网络嵌入模型和分类器组合成一个分类模型，部署在相应的业务服务平台，以便在接收到分类请求时，执行分类处理过程。其中，结合几个具体的应用场景对分类模型的处理过程进行进一步描述，具体如下所述：

应用场景 1，文献分类的场景。

在一个实施例中，服务器接收终端发起的文献分类请求，获取与文献分类请求对应的文献引用关系图；通过第一网络嵌入模型提取文献引用关系图的第一嵌入向量；通过第二网络嵌入模型提取文献引用关系图的第一结构数据；通过分类器，对拼接第一嵌入向量和第一结构数据所得的目标嵌入向量进行分类处理，得到各文献的主题或所属领域。

其中，该文献引用关系图可以是根据从学术引用网络获得的数据集构建的网络图。该文献引用关系图中的节点对应一篇文献，如论文；该文献引用关系图中各节点之间的边对应的是引用关系，当文献 1 引用了文献 2，则文献 1 的节点与文献 2 的节点之间相连。

应用场景 2，对媒体的兴趣爱好分类及推送的场景。

在一个实施例中，服务器接收终端发起的媒体推荐请求，获取媒体推荐请求对应的媒体交互图；通过第一网络嵌入模型提取媒体交互图的第二嵌入特征；通过第二网络嵌入模型提取媒体交互图的第二结构数据；通过分类器，对拼接第二嵌入特征和第二结构数据所得的目标嵌入向量进行分类处理，得到对象节点对应的兴趣类型；按照兴趣类型向对象节点对应的媒体账号推荐目标媒体。

其中，该媒体交互图可以是媒体分享平台获得的用于反映对象与媒体之间交互的网络图，该媒体可以是图片、音乐、视频和直播间中的任一种；对象与媒体之间存在交互，可以指对象点击浏览了某个图片，或播放了某个音乐或视频，或观看了某个直播间等。该媒体交互图中包括对象节点和媒体节点。

通过上述方式可以准确推断出对象的兴趣类型，如对什么类型的媒体感兴趣，如对科幻类型的电影感兴趣，又如对摇滚音乐感兴趣等，然后向该对象推荐其感兴趣的目标媒体，从而可以提高媒体的点播率。

应用场景 3，对感兴趣的通信群组分类及推送的场景。

在一个实施例中，服务器接收终端发起的群组推荐请求，获取与群组推荐请求对应的社交关系图；通过第一网络嵌入模型提取社交关系图的第三嵌入特征；通过第二网络嵌入模型提取社交关系图的第三结构数据；通过分类器，对拼接第三嵌入特征和第三结构数据所得的目标嵌入向量进行分类处理，得到社交对象感兴趣的通信群组；向社交对象推送感兴趣的通信群组。

其中，该社交关系图中包括社交对象的对象节点，若社交对象之间存在关注关系，则该社交对象对应的对象节点之间相连。通过对该社交关系图进行分类，可以得到各社交对象感兴趣的通信群组（如群聊的兴趣小组）。

上述实施例中，将训练后的第一网络嵌入模型、第二网络嵌入模型和分类器应用在不同的应用场景中，可以实现相应的分类过程，如通过第一网络嵌入模型和第二网络嵌入模型可以获得包含节点特征和结构数据的目标嵌入向量，利用该目标嵌入向量对文献引用关系图、媒体交互图或社交关系图中的节点进行精确地分类，分别得到各文献的主题或所属领域、对象的兴趣类型和感兴趣的通信群组，有效地提高了分类效果，而且还可以准确地推送目标媒

体或感兴趣的通信群组。

应该理解的是，虽然如上所述的各实施例所涉及的流程图中的各个步骤按照箭头的指示依次显示，但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明，这些步骤的执行并没有严格的顺序限制，这些步骤可以以其它的顺序执行。而且，如上所述的各实施例所涉及的流程图中的至少一部分步骤可以包括多个步骤或者多个阶段，这些步骤或者阶段并不必然是在同一时刻执行完成，而是可以在不同的时刻执行，这些步骤或者阶段的执行顺序也不必然是依次进行，而是可以与其它步骤或者其它步骤中的步骤或者阶段的至少一部分轮流或者交替地执行。

基于同样的发明构思，本申请实施例还提供了一种用于实现上述所涉及的数据网络图的嵌入方法的数据网络图的嵌入装置。该装置所提供的解决问题的实现方案与上述方法中所记载的实现方案相似，故下面所提供的一个或多个数据网络图的嵌入装置实施例中的具体限定可以参见上文中对于数据网络图的嵌入方法的限定，在此不再赘述。

在一个实施例中，如图 7 所示，提供了一种数据网络图的嵌入装置，包括：第一提取模块 702、第二提取模块 704、确定模块 706、调整模块 708 和第三提取模块 710，其中：

第一提取模块 702，用于通过第一网络嵌入模型对数据网络图和负样本网络图进行节点特征提取，得到正样本嵌入向量和负样本嵌入向量；数据网络图为正样本网络图，是基于不平衡的对象数据集构建所得的不平衡网络图；

第二提取模块 704，用于通过第一网络嵌入模型对数据网络图的第一增强图和第二增强图进行节点特征提取，得到第一全局嵌入向量和第二全局嵌入向量；

确定模块 706，用于确定正样本嵌入向量与第一全局嵌入向量、第二全局嵌入向量之间的第一匹配度，以及确定负样本嵌入向量与第一全局嵌入向量、第二全局嵌入向量之间的第二匹配度；

调整模块 708，用于依据第一匹配度和第二匹配度确定损失值，并基于损失值调整第一网络嵌入模型的参数；

第三提取模块 710，用于基于调整后的第一网络嵌入模型对数据网络图进行节点特征提取，得到用于对数据网络图中各节点分类的嵌入向量。

上述实施例中，通过第一网络嵌入模型对数据网络图和负样本网络图进行节点特征提取，得到正样本嵌入向量和负样本嵌入向量；此外还通过第一网络嵌入模型对数据网络图的两个不同增强图进行节点特征提取，得到第一全局嵌入向量和第二全局嵌入向量；确定正样本嵌入向量与第一全局嵌入向量、第二全局嵌入向量之间的第一匹配度，以及确定负样本嵌入向量与第一全局嵌入向量、第二全局嵌入向量之间的第二匹配度，由于上述的增强图是由数据网络图增强所得的，因此正样本嵌入向量与第一全局嵌入向量、第二全局嵌入向量之间具有较高的匹配度，而负样本嵌入向量与第一全局嵌入向量、第二全局嵌入向量之间匹配度较低，因此依据第一匹配度和第二匹配度调整第一网络嵌入模型的参数，可以使调整后的第一网络嵌入模型学习到鲁棒的且能准确分类出数据网络图中各节点的嵌入向量。此外，在训练过程中，并没有使用节点的标签，因此模型学习过程中并不会受数据网络图中多数类的影响，从而即便数据网络图为不平衡网络图，模型也可以学习到平衡的特征空间，使嵌入向量包含重要的特征且更加鲁棒，进而能够在分类过程中有效提高分类效果。

在其中的一个实施例中，如图 8 所示，该装置还包括：

增强模块 712，用于对数据网络图进行第一数据增强处理，得到第一增强图；对数据网

络图进行第二数据增强处理，得到第二增强图；其中，第一数据增强处理与第二数据增强处理分别为特征掩盖、边扰动或子图提取。

在其中的一个实施例中，增强模块 712，还用于在数据网络图中选取采样节点，以第一采样节点为中心点逐级扩散采样，并在逐级扩散采样过程中，将每次采样的邻居节点置于第一采样集内；当第一采样集内的节点数量达到目标值时，停止采样，得到第一增强图；对数据网络图进行特征掩盖，得到第二增强图。

上述实施例中，通过对数据网络图进行数据增强处理，可以得到不同角度的增强图，从而利用增强图进行模型训练时，可以使模型具有普适性，能够适应各种场景。

在其中的一个实施例中，如图 8 所示，该装置还包括：

打乱模块 714，用于对数据网络图中的节点对应的特征进行乱序处理，得到负样本网络图；其中，负样本网络图的节点结构与数据网络图的节点结构一致。

在其中的一个实施例中，如图 8 所示，该装置还包括：

构建模块 716，用于获取对象数据集和对象数据集内各对象数据之间的关联关系；以对象数据集内的各对象数据为节点、以关联关系为各节点的边构建数据网络图。

在其中的一个实施例中，第一增强图和第二增强图分别是对数据网络图进行数据增强所得的增强图；

第二提取模块 704，还用于通过第一网络嵌入模型，分别从第一增强图和第二增强图中提取各节点的第一局部嵌入向量和第二局部嵌入向量；分别对第一局部嵌入向量和第二局部嵌入向量进行池化处理，得到第一全局嵌入向量和第二全局嵌入向量。

在其中的一个实施例中，第二提取模块 704，还用于获取第一增强图中各节点的第一邻接矩阵和第一特征矩阵；将第一邻接矩阵和第一特征矩阵输入至第一网络嵌入模型，以使第一网络嵌入模型基于第一邻接矩阵、第一邻接矩阵的度矩阵、第一特征矩阵和第一网络嵌入模型的权重矩阵，生成第一增强图中各节点的第一局部嵌入向量；获取第二增强图中各节点的第二邻接矩阵和第二特征矩阵；将第二邻接矩阵和第二特征矩阵输入至第一网络嵌入模型，以使第一网络嵌入模型基于第二邻接矩阵、第二邻接矩阵的度矩阵、第一特征矩阵和第一网络嵌入模型的权重矩阵，生成第二增强图中各节点的第二局部嵌入向量。

在其中的一个实施例中，如图 8 所示，该装置还包括：

第四提取模块 718，用于通过第二网络嵌入模型对数据网络图进行节点特征提取，并基于提取的节点特征重构出目标邻接矩阵；

调整模块 708，还用于依据目标邻接矩阵和矩阵标签之间的损失值，对第二网络嵌入模型的参数进行调整；

第四提取模块 718，还用于当调整后的第二网络嵌入模型达到收敛条件时，通过调整后的第二网络嵌入模型获得数据网络图中各节点的结构信息；将嵌入向量和结构信息之间的拼接向量，作为用于对数据网络图中各节点分类的目标嵌入向量。

在其中的一个实施例中，如图 8 所示，该装置还包括：

分类模块 720，用于通过分类器对目标嵌入向量进行分类处理，得到预测结果；

调整模块 708，还用于基于预测结果与分类标签之间的损失值，对分类器进行参数调整；当调整后的分类器达到收敛条件时，停止训练过程。

上述实施例中，通过对第二网络嵌入模型进行训练，使第二网络嵌入模型能学习到结构信息的提取，从而提取出与数据网络图原始的结构一致或接近的结构信息。将该结构信息与

第一网络嵌入模型提取的包含节点关键特征的嵌入向量进行拼接，从而可以得到包含节点关键特征和结构信息的目标嵌入向量，使该目标嵌入向量具有更全面的表达能力，具有鲁棒性，能够有效地提高分类效果。

在其中的一个实施例中，如图 8 所示，该装置还包括：

第一应用模块 722，用于获取文献引用关系图；通过第一网络嵌入模型提取文献引用关系图的第一嵌入向量；通过第二网络嵌入模型提取文献引用关系图的第一结构数据；通过分类器，对拼接第一嵌入向量和第一结构数据所得的目标嵌入向量进行分类处理，得到各文献的主题或所属领域。

在其中的一个实施例中，如图 8 所示，该装置还包括：

第二应用模块 724，用于获取媒体交互图；通过第一网络嵌入模型提取媒体交互图的第二嵌入特征；通过第二网络嵌入模型提取媒体交互图的第二结构数据；通过分类器，对拼接第二嵌入特征和第二结构数据所得的目标嵌入向量进行分类处理，得到对象节点对应的兴趣类型；按照兴趣类型向对象节点对应的媒体账号推荐目标媒体。

在其中的一个实施例中，如图 8 所示，该装置还包括：

第三应用模块 726，用于获取社交关系图；通过第一网络嵌入模型提取社交关系图的第三嵌入特征；通过第二网络嵌入模型提取社交关系图的第三结构数据；通过分类器，对拼接第三嵌入特征和第三结构数据所得的目标嵌入向量进行分类处理，得到社交对象感兴趣的通信群组；向社交对象推送感兴趣的通信群组。

上述实施例中，将训练后的第一网络嵌入模型、第二网络嵌入模型和分类器应用在不同的应用场景中，可以实现相应的分类过程，如通过第一网络嵌入模型和第二网络嵌入模型可以获得包含节点特征和结构数据的目标嵌入向量，利用该目标嵌入向量对文献引用关系图、媒体交互图或社交关系图中的节点进行精确地分类，分别得到各文献的主题或所属领域、对象的兴趣类型和感兴趣的通信群组，有效地提高了分类效果，而且还可以准确地推送目标媒体或感兴趣的通信群组。

上述数据网络图的嵌入装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中，也可以以软件形式存储于计算机设备中的存储器中，以便于处理器调用执行以上各个模块对应的操作。

在一个实施例中，提供了一种计算机设备，该计算机设备可以是服务器，其内部结构图可以如图 9 所示。该计算机设备包括处理器、存储器、输入/输出接口(Input/Output，简称 I/O)和通信接口。其中，处理器、存储器和输入/输出接口通过系统总线连接，通信接口通过输入/输出接口连接到系统总线。其中，该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质和内存储器。该非易失性存储介质存储有操作系统、计算机程序和数据库。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该计算机设备的数据库用于存储对数据网络图、负样本网络图和增强图。该计算机设备的输入/输出接口用于处理器与外部设备之间交换信息。该计算机设备的通信接口用于与外部的终端通过网络连接通信。该计算机程序被处理器执行时以实现一种数据网络图的嵌入方法。

本领域技术人员可以理解，图 9 中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的计算机设备的限定，具体的计算机设备可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

在一个实施例中，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，该处理器执行计算机程序时实现上述数据网络图的嵌入方法的步骤。

在一个实施例中，提供了一种计算机可读存储介质，其上存储有计算机程序，计算机程序被处理器执行时实现上述数据网络图的嵌入方法的步骤。

在一个实施例中，提供了一种计算机程序产品，包括计算机程序，该计算机程序被处理器执行时实现上述数据网络图的嵌入方法的步骤。

需要说明的是，本申请所涉及的用户信息（包括但不限于用户设备信息、用户个人信息等）和数据（包括但不限于用于分析的数据、存储的数据、展示的数据等），均为经用户授权或者经过各方充分授权的信息和数据，且相关数据的收集、使用和处理需要遵守相关国家和地区的相关法律法规和标准。

本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机程序来指令相关的硬件来完成，所述的计算机程序可存储于一非易失性计算机可读存储介质中，该计算机程序在执行时，可包括如上述各方法的实施例的流程。其中，本申请所提供的各实施例中所使用的对存储器、数据库或其它介质的任何引用，均可包括非易失性和易失性存储器中的至少一种。非易失性存储器可包括只读存储器（Read-Only Memory, ROM）、磁带、软盘、闪存、光存储器、高密度嵌入式非易失性存储器、阻变存储器（ReRAM）、磁变存储器（Magnetoresistive Random Access Memory, MRAM）、铁电存储器（Ferroelectric Random Access Memory, FRAM）、相变存储器（Phase Change Memory, PCM）、石墨烯存储器等。易失性存储器可包括随机存取存储器（Random Access Memory, RAM）或外部高速缓冲存储器等。作为说明而非局限，RAM可以是多种形式，比如静态随机存取存储器（Static Random Access Memory, SRAM）或动态随机存取存储器（Dynamic Random Access Memory, DRAM）等。本申请所提供的各实施例中所涉及的数据库可包括关系型数据库和非关系型数据库中至少一种。非关系型数据库可包括基于区块链的分布式数据库等，不限于此。本申请所提供的各实施例中所涉及的处理器可为通用处理器、中央处理器、图形处理器、数字信号处理器、可编程逻辑器、基于量子计算的数据处理逻辑器等，不限于此。

以上实施例的各技术特征可以进行任意的组合，为使描述简洁，未对上述实施例中的各个技术特征所有可能的组合都进行描述，然而，只要这些技术特征的组合不存在矛盾，都应当认为是本说明书记载的范围。

以上所述实施例仅表达了本申请的几种实施方式，其描述较为具体和详细，但并不能因此而理解为对本申请专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本申请构思的前提下，还可以做出若干变形和改进，这些都属于本申请的保护范围。因此，本申请的保护范围应以所附权利要求为准。

权利要求书

1、一种数据网络图的嵌入方法，由计算机设备执行，其特征在于，所述方法包括：

通过第一网络嵌入模型对数据网络图和负样本网络图进行节点特征提取，得到正样本嵌入向量和负样本嵌入向量；所述数据网络图为正样本网络图，是基于不平衡的对象数据集构建所得的不平衡网络图；

通过所述第一网络嵌入模型对所述数据网络图的第一增强图和第二增强图进行节点特征提取，得到第一全局嵌入向量和第二全局嵌入向量；

确定所述正样本嵌入向量与所述第一全局嵌入向量、所述第二全局嵌入向量之间的第一匹配度，以及确定所述负样本嵌入向量与所述第一全局嵌入向量、所述第二全局嵌入向量之间的第二匹配度；

依据所述第一匹配度和所述第二匹配度确定损失值，并基于所述损失值调整所述第一网络嵌入模型的参数；

基于调整后的所述第一网络嵌入模型对所述数据网络图进行节点特征提取，得到用于对所述数据网络图中各节点分类的嵌入向量。

2、根据权利要求1所述的方法，其特征在于，所述方法还包括：

对所述数据网络图进行第一数据增强处理，得到所述第一增强图；

对所述数据网络图进行第二数据增强处理，得到所述第二增强图；

其中，所述第一数据增强处理与所述第二数据增强处理分别为特征掩盖、边扰动或子图提取中的至少一种处理。

3、根据权利要求2所述的方法，其特征在于，所述对所述数据网络图进行第一数据增强处理，得到所述第一增强图包括：

在所述数据网络图中选取采样节点，以所述第一采样节点为中心点逐级扩散采样，并在逐级扩散采样过程中，将每次采样的邻居节点置于第一采样集内；

当所述第一采样集内的节点数量达到目标值时，停止采样，得到所述第一增强图；

所述对所述数据网络图进行第二数据增强处理，得到所述第二增强图包括：

对所述数据网络图进行特征掩盖，得到所述第二增强图。

4、根据权利要求1所述的方法，其特征在于，所述方法还包括：

对所述数据网络图中的节点对应的特征进行乱序处理，得到负样本网络图；

其中，所述负样本网络图的节点结构与所述数据网络图的节点结构一致。

5、根据权利要求1所述的方法，其特征在于，所述方法还包括：

获取所述对象数据集和所述对象数据集内各对象数据之间的关联关系；

以所述对象数据集内的各对象数据为节点、以所述关联关系为各所述节点的边构建数据网络图。

6、根据权利要求1所述的方法，其特征在于，所述通过所述第一网络嵌入模型对所述数据网络图的第一增强图和第二增强图进行节点特征提取，得到第一全局嵌入向量和第二全局嵌入向量包括：

通过所述第一网络嵌入模型，分别从所述第一增强图和所述第二增强图中提取各节点的第一局部嵌入向量和第二局部嵌入向量；

分别对所述第一局部嵌入向量和所述第二局部嵌入向量进行池化处理，得到所述第一全局嵌入向量和所述第二全局嵌入向量。

7、根据权利要求 6 所述的方法，其特征在于，所述通过所述第一网络嵌入模型，分别从所述第一增强图和所述第二增强图中提取各节点的第一局部嵌入向量和第二局部嵌入向量包括：

获取所述第一增强图中各节点的第一邻接矩阵和第一特征矩阵；将所述第一邻接矩阵和所述第一特征矩阵输入至所述第一网络嵌入模型，以使所述第一网络嵌入模型基于所述第一邻接矩阵、所述第一邻接矩阵的度矩阵、所述第一特征矩阵和所述第一网络嵌入模型的权重矩阵，生成所述第一增强图中各节点的第一局部嵌入向量；

获取所述第二增强图中各节点的第二邻接矩阵和第二特征矩阵；将所述第二邻接矩阵和所述第二特征矩阵输入至所述第一网络嵌入模型，以使所述第一网络嵌入模型基于所述第二邻接矩阵、所述第二邻接矩阵的度矩阵、所述第一特征矩阵和所述第一网络嵌入模型的权重矩阵，生成所述第二增强图中各节点的第二局部嵌入向量。

8、根据权利要求 1 至 7 任一项所述的方法，其特征在于，所述方法还包括：

通过分类器对所述嵌入向量进行分类处理，得到预测结果；

基于所述预测结果与分类标签之间的损失值，对所述分类器进行参数调整；

当调整后的所述分类器达到所述收敛条件时，停止训练过程。

9、根据权利要求 8 所述的方法，其特征在于，所述方法还包括：

获取文献引用关系图；

通过所述第一网络嵌入模型提取所述文献引用关系图的第一嵌入向量；

通过所述分类器对所述第一嵌入向量进行分类处理，得到各文献的主题或所属领域。

10、根据权利要求 8 所述的方法，其特征在于，所述方法还包括：

获取媒体交互图；

通过所述第一网络嵌入模型提取所述媒体交互图的第二嵌入特征；

通过所述分类器对所述第二嵌入特征进行分类处理，得到所述对象节点对应的兴趣类型；

按照所述兴趣类型向所述对象节点对应的媒体账号推荐目标媒体。

11、根据权利要求 8 所述的方法，其特征在于，所述方法还包括：

获取社交关系图；

通过所述第一网络嵌入模型提取所述社交关系图的第三嵌入特征；

通过所述分类器对所述第三嵌入特征进行分类处理，得到社交对象感兴趣的通信群组；

向所述社交对象推送所述感兴趣的通信群组。

12、根据权利要求 1 至 7 任一项所述的方法，其特征在于，所述基于调整后的所述第一网络嵌入模型对所述数据网络图进行节点特征提取，得到用于对所述数据网络图中各节点分类的嵌入向量之后，所述方法还包括：

通过第二网络嵌入模型对所述数据网络图进行节点特征提取，并基于提取的节点特征重构出目标邻接矩阵；

依据所述目标邻接矩阵和矩阵标签之间的损失值，对所述第二网络嵌入模型的参数进行调整；

当调整后的所述第二网络嵌入模型达到收敛条件时，通过调整后的所述第二网络嵌入模型获得所述数据网络图中各节点的结构信息；

将所述嵌入向量和所述结构信息之间的拼接向量，作为用于对所述数据网络图中各节点分类的目标嵌入向量。

13、根据权利要求 12 所述的方法，其特征在于，所述方法还包括：
通过分类器对所述目标嵌入向量进行分类处理，得到预测结果；
基于所述预测结果与分类标签之间的损失值，对所述分类器进行参数调整；
当调整后的所述分类器达到所述收敛条件时，停止训练过程。

14、根据权利要求 12 所述的方法，其特征在于，所述方法还包括：
获取文献引用关系图；
通过所述第一网络嵌入模型提取所述文献引用关系图的第一嵌入向量；
通过所述第二网络嵌入模型提取所述文献引用关系图的第一结构数据；
通过所述分类器，对拼接所述第一嵌入向量和所述第一结构数据所得的目标嵌入向量进行分类处理，得到各文献的主题或所属领域。

15、根据权利要求 12 所述的方法，其特征在于，所述方法还包括：
获取媒体交互图；
通过所述第一网络嵌入模型提取所述媒体交互图的第二嵌入特征；
通过所述第二网络嵌入模型提取所述媒体交互图的第二结构数据；
通过所述分类器，对拼接所述第二嵌入特征和所述第二结构数据所得的目标嵌入向量进行分类处理，得到所述对象节点对应的兴趣类型；
按照所述兴趣类型向所述对象节点对应的媒体账号推荐目标媒体。

16、根据权利要求 12 所述的方法，其特征在于，所述方法还包括：
获取社交关系图；
通过所述第一网络嵌入模型提取所述社交关系图的第三嵌入特征；
通过所述第二网络嵌入模型提取所述社交关系图的第三结构数据；
通过所述分类器，对拼接所述第三嵌入特征和所述第三结构数据所得的目标嵌入向量进行分类处理，得到社交对象感兴趣的通信群组；
向所述社交对象推送所述感兴趣的通信群组。

17、一种数据网络图的嵌入装置，其特征在于，所述装置包括：
第一提取模块，用于通过第一网络嵌入模型对数据网络图和负样本网络图进行节点特征提取，得到正样本嵌入向量和负样本嵌入向量；所述数据网络图为正样本网络图，是基于不平衡的对象数据集构建所得的不平衡网络图；

第二提取模块，用于通过所述第一网络嵌入模型对所述数据网络图的第一增强图和第二增强图进行节点特征提取，得到第一全局嵌入向量和第二全局嵌入向量；

确定模块，用于确定所述正样本嵌入向量与所述第一全局嵌入向量、所述第二全局嵌入向量之间的第一匹配度，以及确定所述负样本嵌入向量与所述第一全局嵌入向量、所述第二全局嵌入向量之间的第二匹配度；

调整模块，用于依据所述第一匹配度和所述第二匹配度确定损失值，并基于所述损失值调整所述第一网络嵌入模型的参数；

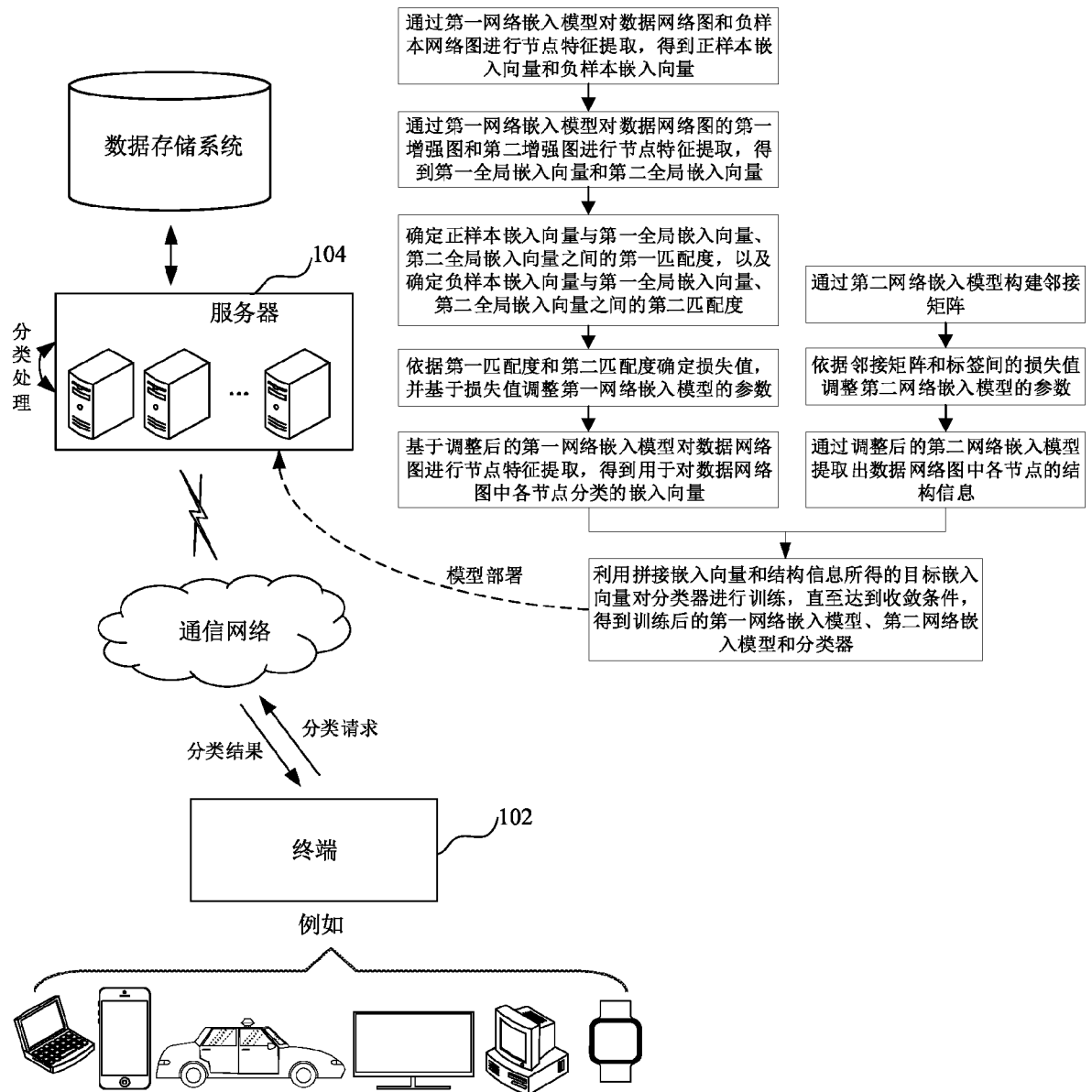
第三提取模块，用于基于调整后的所述第一网络嵌入模型对所述数据网络图进行节点特征提取，得到用于对所述数据网络图中各节点分类的嵌入向量。

18、一种计算机设备，包括存储器和处理器，所述存储器存储有计算机程序，其特征在于，所述处理器执行所述计算机程序时实现权利要求 1 至 16 任一项所述的方法的步骤。

19、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，所述计算机程序

被处理器执行时实现权利要求 1 至 16 任一项所述的方法的步骤。

20、一种计算机程序产品，包括计算机程序，其特征在于，该计算机程序被处理器执行时实现权利要求 1 至 16 任一项所述的方法的步骤。



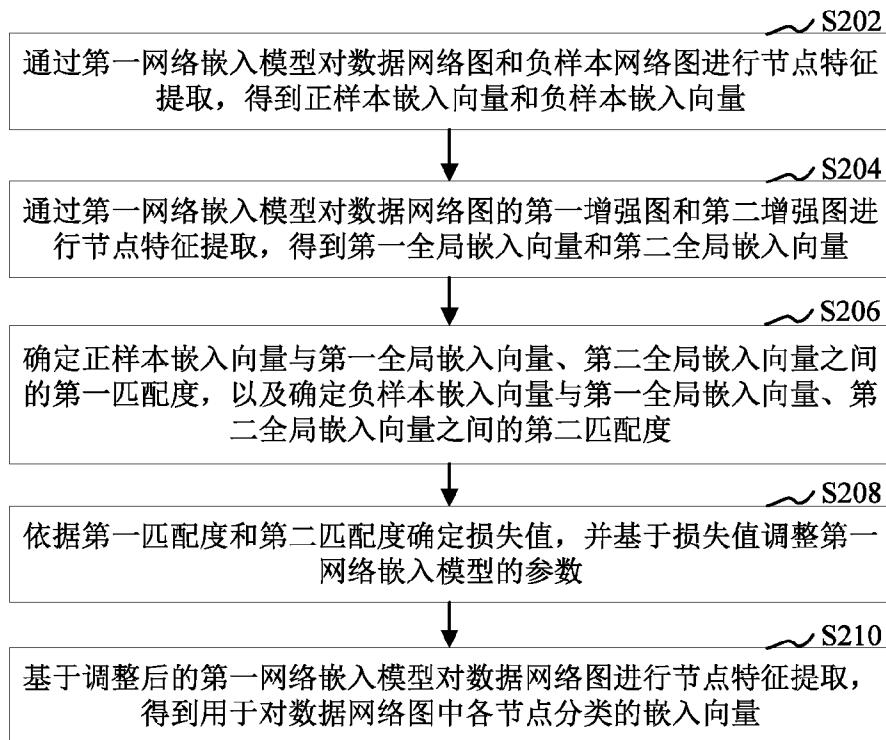


图 2

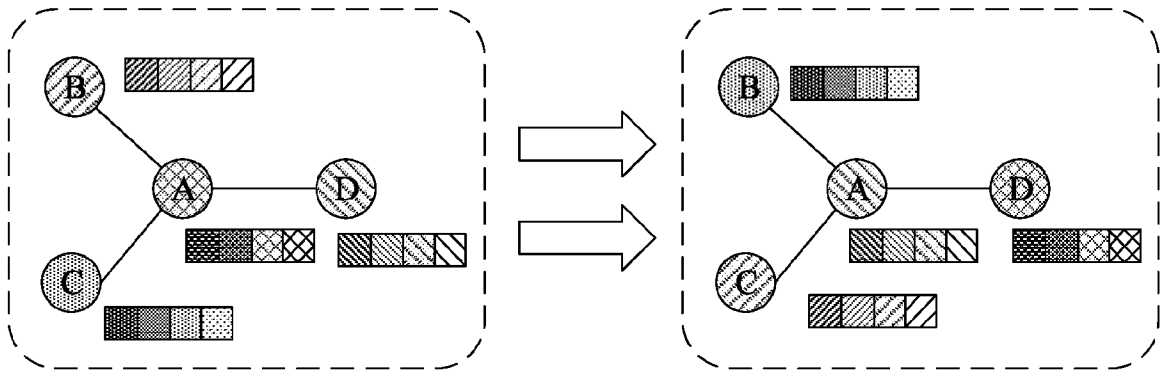


图 3

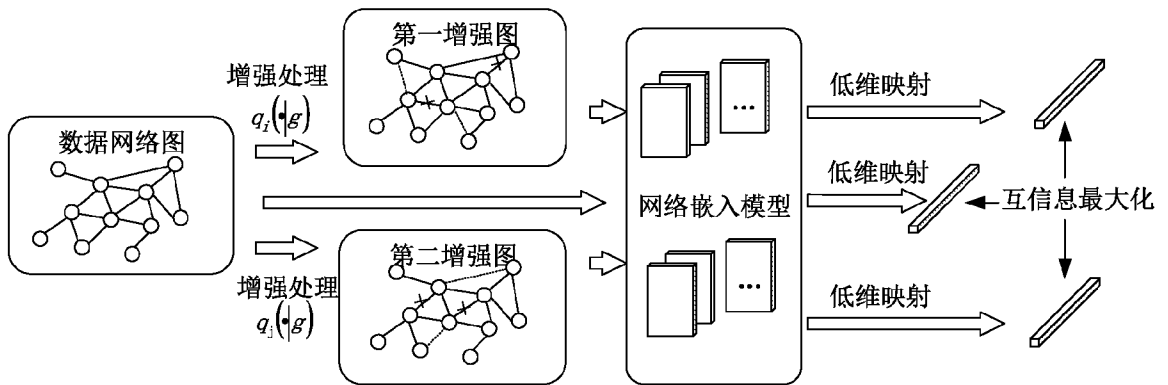


图 4

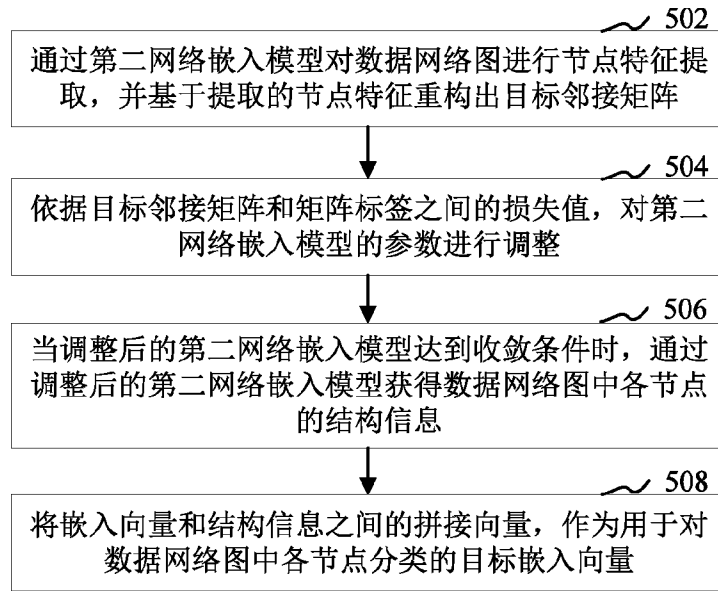


图 5

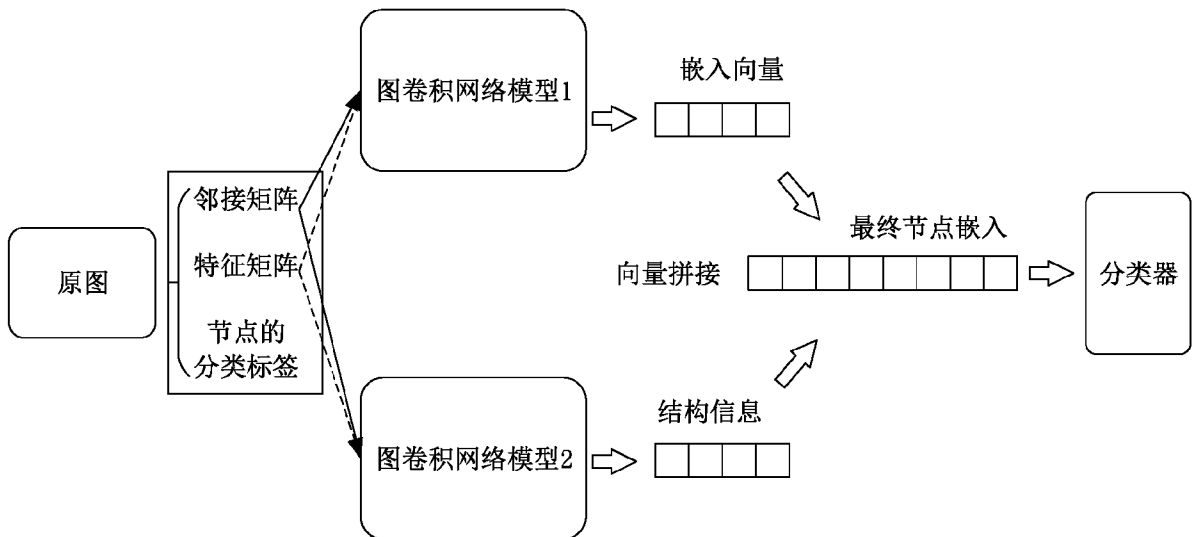


图 6

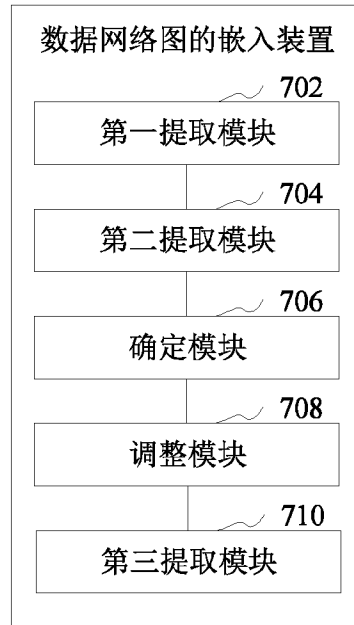


图 7

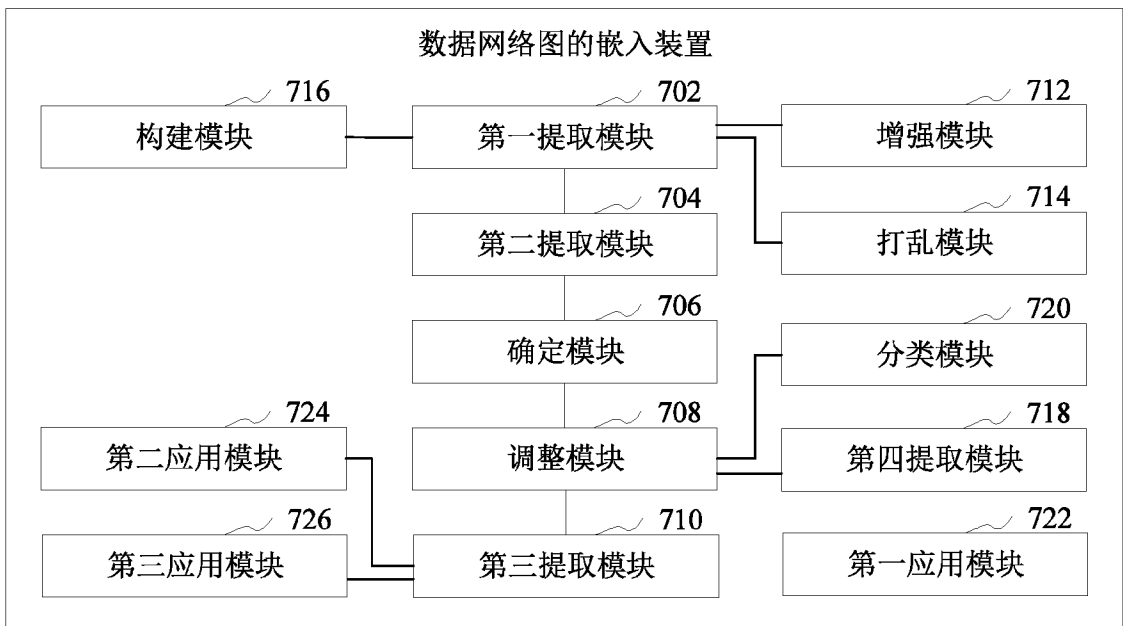


图 8

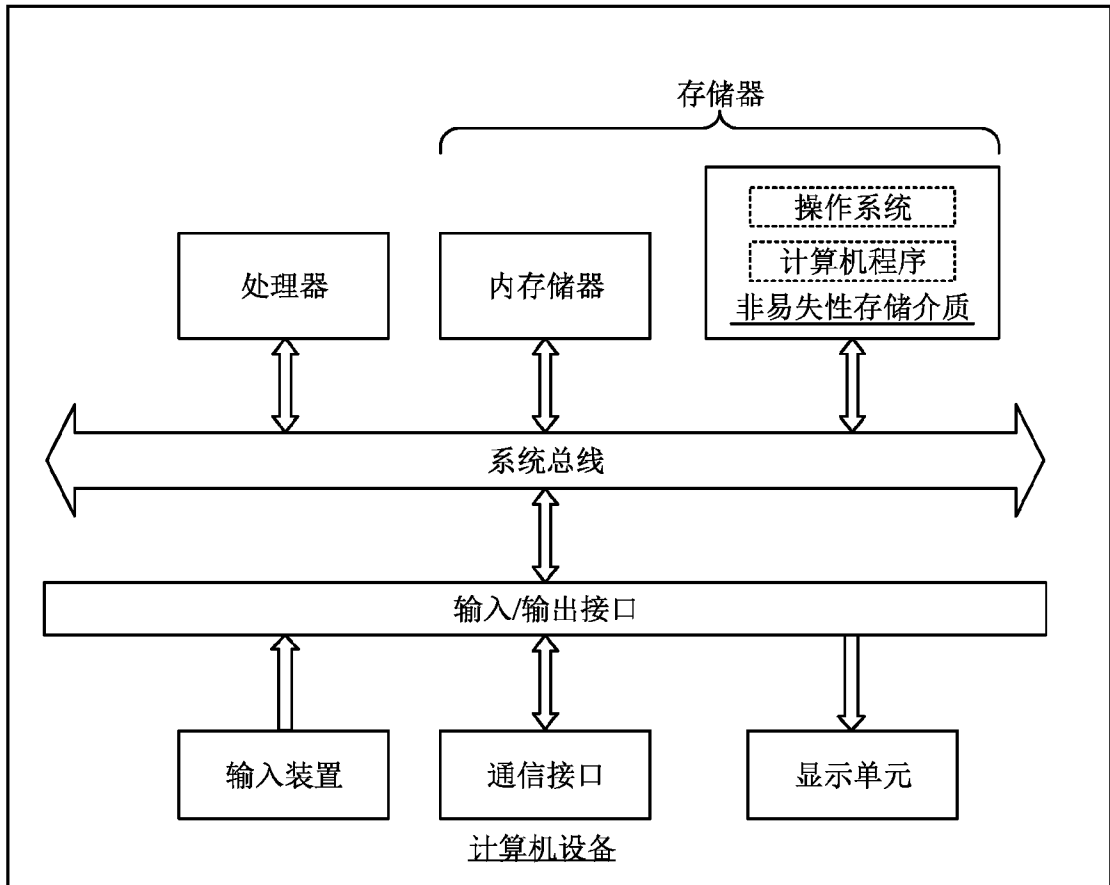


图 9

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/092130

A. CLASSIFICATION OF SUBJECT MATTER G06V10/86(2022.01)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: G06V, G06K 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, CNABS, CNKI, DWPI, ENTXT, ENTXTC: 不平衡, 节点, 嵌入向量, 数据集, 图网络, GNN, node, embedding, vector, data, set, unbalance		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 111723292 A (TRIP COMPUTER TECHNOLOGY (SHANGHAI) CO., LTD.) 29 September 2020 (2020-09-29) entire document	1-20
A	CN 112766500 A (ALIPAY (HANGZHOU) INFORMATION TECHNOLOGY CO., LTD.) 07 May 2021 (2021-05-07) entire document	1-20
A	WO 2020114122 A1 (ALIBABA GROUP HOLDING LTD.) 11 June 2020 (2020-06-11) entire document	1-20
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
<p>* Special categories of cited documents:</p> <p>“A” document defining the general state of the art which is not considered to be of particular relevance</p> <p>“D” document cited by the applicant in the international application</p> <p>“E” earlier application or patent but published on or after the international filing date</p> <p>“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)</p> <p>“O” document referring to an oral disclosure, use, exhibition or other means</p> <p>“P” document published prior to the international filing date but later than the priority date claimed</p> <p>“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</p> <p>“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</p> <p>“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</p> <p>“&” document member of the same patent family</p>		
Date of the actual completion of the international search 11 August 2023		Date of mailing of the international search report 21 August 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
Information on patent family members

International application No.

PCT/CN2023/092130

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	111723292	A	29 September 2020	None			
CN	112766500	A	07 May 2021	None			
WO	2020114122	A1	11 June 2020	TW	202032422	A	01 September 2020
				TWI	709086	B	01 November 2020

国际检索报告

国际申请号

PCT/CN2023/092130

<p>A. 主题的分类</p> <p>G06V10/86(2022.01) i</p> <p>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</p>														
<p>B. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>IPC: G06V, G06K</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>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))</p> <p>CNTXT, CNABS, CNKI, DWPI, ENTXT, ENTXTC:不平衡, 节点, 嵌入向量, 数据集, 图网络, GNN, node, embedding, vector, data, set, unbalance</p>														
<p>C. 相关文件</p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>CN 111723292 A (携程计算机技术(上海)有限公司) 2020年9月29日 (2020 - 09 - 29) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 112766500 A (支付宝(杭州)信息技术有限公司) 2021年5月7日 (2021 - 05 - 07) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>WO 2020114122 A1 (ALIBABA GROUP HOLDING LTD) 2020年6月11日 (2020 - 06 - 11) 全文</td> <td>1-20</td> </tr> </tbody> </table> <p><input type="checkbox"/> 其余文件在C栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。</p> <p>* 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</p>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	A	CN 111723292 A (携程计算机技术(上海)有限公司) 2020年9月29日 (2020 - 09 - 29) 全文	1-20	A	CN 112766500 A (支付宝(杭州)信息技术有限公司) 2021年5月7日 (2021 - 05 - 07) 全文	1-20	A	WO 2020114122 A1 (ALIBABA GROUP HOLDING LTD) 2020年6月11日 (2020 - 06 - 11) 全文	1-20
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求												
A	CN 111723292 A (携程计算机技术(上海)有限公司) 2020年9月29日 (2020 - 09 - 29) 全文	1-20												
A	CN 112766500 A (支付宝(杭州)信息技术有限公司) 2021年5月7日 (2021 - 05 - 07) 全文	1-20												
A	WO 2020114122 A1 (ALIBABA GROUP HOLDING LTD) 2020年6月11日 (2020 - 06 - 11) 全文	1-20												
国际检索实际完成的日期	2023年8月11日	国际检索报告邮寄日期												
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 白雪涛 电话号码 (+86) 62412069												

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2023/092130

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	111723292	A	2020年9月29日	无			
CN	112766500	A	2021年5月7日	无			
WO	2020114122	A1	2020年6月11日	TW	202032422	A	2020年9月1日
				TWI	709086	B	2020年11月1日