

(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 8 月 24 日 (24.08.2023)



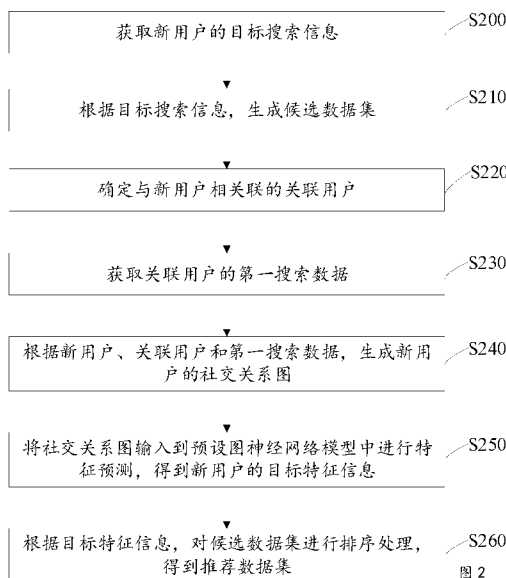
(10) 国际公布号
WO 2023/155306 A1

- (51) 国际专利分类号:
G06F 16/9536 (2019.01) **G06N 3/08** (2006.01)
G06N 3/04 (2006.01) **G06Q 50/00** (2012.01)
- (21) 国际申请号: PCT/CN2022/090754
- (22) 国际申请日: 2022 年 4 月 29 日 (29.04.2022)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202210143649.3 2022年2月16日 (16.02.2022) CN
- (71) 申请人: 平安科技(深圳)有限公司 (PING AN TECHNOLOGY(SHENZHEN)CO., LTD.) [CN/CN]; 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

- (72) 发明人: 吴粤敏(WU, Yuemin); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。 舒畅(SHU, Chang); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。 陈又新(CHEN, Youxin); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。
- (74) 代理人: 广州嘉权专利商标事务所有限公司 (JIAQUAN IP LAW); 中国广东省广州市天河区黄埔大道西 100 号富力盈泰广场 A 栋 910, Guangdong 510627 (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, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT,

(54) Title: DATA RECOMMENDATION METHOD AND APPARATUS BASED ON GRAPH NEURAL NETWORK AND ELECTRONIC DEVICE

(54) 发明名称: 基于图神经网络的数据推荐方法及装置、电子设备



S200 Obtain target search information of a new user
S210 Generate a candidate data set according to the target search information
S220 Determine an associated user associated with the new user
S230 Obtain first search data of the associated user
S240 Generate a social relation graph of the new user according to the new user, the associated user, and the first search data
S250 Input the social relation graph into a preset graph neural network model for feature prediction to obtain target feature information of the new user
S260 According to the target feature information, sort the candidate data set to obtain a recommended data set

(57) Abstract: The present application specifically relates to the technical field of artificial intelligence, and discloses a data recommendation method and apparatus based on a graph neural network and an electronic device. The method comprises: obtaining target search information of a new user, and generating a candidate data set according to the target search information; determining an associated user associated with the new user, and obtaining first search data of the associated user; generating a social relation graph of the new user according to the new user, the associated user, and the first search data; inputting the social relation graph into a preset graph neural network model for feature prediction to obtain target feature information of the new user; and according to the target feature information, sorting the candidate data set to obtain a recommended data set. In this way, the cold start problem on a user side can be solved, the accuracy of data recommendation for a new user is improved, and the use experience of the new user on a data search function is improved.

(57) 摘要: 本申请公开了一种基于图神经网络的数据推荐方法及装置、电子设备, 具体涉及人工智能技术领域。该方法包括: 获取新用户的目标搜索信息, 并根据目标搜索信息, 生成候选数据集。确定与新用户相关联的关联用户, 并获取关联用户的第一搜索数据。根据新用户、关联用户和第一搜索数据, 生成新用户的社交关系图。将社交关系图输入到预设图神经网络模型中进行特征预测, 得到新用户的目标特征信息。根据目标特征信息, 对候选数据集进行排序处理, 得到推荐数据集, 能够缓解用户侧冷启动问题, 提升对新用户进行数据推荐的准确性, 进而改善了新用户对数据搜索功能的使用体验。

JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, 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。

(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, 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, 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)。

根据细则4.17的声明:

— 发明人资格 (细则4.17(iv))

本国际公布:

— 包括国际检索报告 (条约第21条(3))。

基于图神经网络的数据推荐方法及装置、电子设备

本申请要求于 2022 年 2 月 16 日提交中国专利局、申请号为 202210143649.3，发明名称为“基于图神经网络的数据推荐方法及装置、电子设备”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，具体涉及一种基于图神经网络的数据推荐方法及装置、电子设备。

背景技术

现有技术中，当用户向搜索引擎输入关键词进行检索时，推荐系统通常会根据用户输入的关键词进行候选数据集的召回和排序，从而向用户推荐可能感兴趣的数据。然而，发明人发现，现有的数据推荐算法，比如基于机器学习的推荐算法，需要根据用户的历史行为数据进行推荐数据排序，因此，容易因新用户的历史行为数据不足而出现数据推荐准确性较低的问题，影响了新用户对搜索引擎的使用体验。

技术问题

以下是发明人意识到的现有技术的技术问题：现有的数据推荐算法，比如基于机器学习的推荐算法，需要根据用户的历史行为数据进行推荐数据排序，容易因新用户的历史行为数据不足而出现数据推荐准确性较低的问题，影响了新用户对搜索引擎的使用体验。

技术方案

第一方面，本申请实施例提供了一种基于图神经网络的数据推荐方法，所述方法包括：获取新用户的目标搜索信息，并根据所述目标搜索信息，生成候选数据集；确定与所述新用户相关联的关联用户；获取所述关联用户的第一搜索数据；根据所述新用户、所述关联用户和所述第一搜索数据，生成所述新用户的社交关系图；将所述社交关系图输入到预设图神经网络模型中进行特征预测，得到所述新用户的目标特征信息；根据所述目标特征信息，对所述候选数据集进行排序处理，得到推荐数据集。

第二方面，本申请实施例还提出了一种基于图神经网络的数据推荐装置，所述装置包括：获取模块，用于获取新用户的目标搜索信息；第一生成模块，用于根据所述目标搜索信息，生成候选数据集；确定模块，用于确定与所述新用户相关联的关联用户；所述获取模块，还用于获取所述关联用户的第一搜索数据；第二生成模块，用于根据所述新用户、所述关联用户和所述第一搜索数据，生成所述新用户的社交关系图；预测模块，用于将所述社交关系图

输入到预设图神经网络模型中进行特征预测，得到所述新用户的目标特征信息；排序模块，用于根据所述目标特征信息，对所述候选数据集进行排序处理，得到推荐数据集。

第三方面，本申请实施例还提出了一种电子设备，所述电子设备包括存储器和处理器，所述存储器上存储有程序，所述程序被所述处理器执行时实现一种基于图神经网络的数据推荐方法，所述方法包括：获取新用户的目标搜索信息，并根据所述目标搜索信息，生成候选数据集；确定与所述新用户相关联的关联用户；获取所述关联用户的第一搜索数据；根据所述新用户、所述关联用户和所述第一搜索数据，生成所述新用户的社交关系图；将所述社交关系图输入到预设图神经网络模型中进行特征预测，得到所述新用户的目标特征信息；根据所述目标特征信息，对所述候选数据集进行排序处理，得到推荐数据集。

第四方面，本申请提供了一种存储介质，用于计算机可读存储，存储介质存储有一个或者多个程序，一个或者多个程序可被一个或者多个处理器执行，以实现一种基于图神经网络的数据推荐方法，所述方法包括：获取新用户的目标搜索信息，并根据所述目标搜索信息，生成候选数据集；确定与所述新用户相关联的关联用户；获取所述关联用户的第一搜索数据；根据所述新用户、所述关联用户和所述第一搜索数据，生成所述新用户的社交关系图；将所述社交关系图输入到预设图神经网络模型中进行特征预测，得到所述新用户的目标特征信息；根据所述目标特征信息，对所述候选数据集进行排序处理，得到推荐数据集。

有益效果

本申请提出的基于图神经网络的数据推荐方法及装置、电子设备，可以确定与新用户相关联的关联用户，并获取关联用户的第一搜索数据，从而根据新用户、关联用户和第一搜索数据，生成新用户的社交关系图，故社交关系图同时融合了用户社交关系以及不同用户的第一搜索数据。基于此，将社交关系图输入到预设图神经网络模型中，使得预设图神经网络模型沿着社交关系链的深层传播识别关联用户之间的共同数据需求和评价模式特征，从而为 newUser 确定更具备可解释性的目标特征信息。之后，再基于新用户的目标特征信息，对新用户的候选数据集进行排序，得到为 newUser 推送的推荐数据集，能够缓解用户侧冷启动问题，提升对 newUser 进行数据推荐的准确性，进而改善了新用户对数据搜索功能的使用体验。

附图说明

附图用来提供对本申请技术方案的进一步理解，并且构成说明书的一部分，与本申请的实施例一起用于解释本申请的技术方案，并不构成对本申请技术方案的限制。

图 1 是本申请实施例所应用的一种电子设备的结构框图；

图 2 是本申请实施例一提供的基于图神经网络的数据推荐方法的流程图；

图 3 是本申请实施例中一种社交关系图的示意图；

图 4 是本申请实施例中一种预设图神经网络模型的结构示意图；

图 5 是本申请实施例二提供的基于图神经网络的数据推荐方法的流程图；

图 6 是本申请实施例所应用的一种基于图神经网络的数据推荐装置的结构框图。

本发明的实施方式

应当理解，此处所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。

在后续的描述中，使用用于表示元件的诸如“模块”、“部件”或“单元”的后缀仅为为了有利于本申请的说明，其本身没有特有的意义。因此，“模块”、“部件”或“单元”可以混合地使用。

本申请实施例可以基于人工智能技术对相关的数据进行获取和处理。人工智能 (artificial intelligence, AI) 是利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能，感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。其中，人工智能基础技术一般包括如传感器、专用人工智能芯片、云计算、分布式存储、大数据处理技术、操作/交互系统、机电一体化等技术，而人工智能软件技术主要包括计算机视觉技术、机器人技术、生物识别技术、语音处理技术、自然语言处理技术以及机器学习/深度学习等几大方向。

本申请提供一种基于图神经网络 (graph neural networks, GNN) 的数据推荐方法，应用于一种电子设备。参照图 1 所示，图 1 是本申请实施例所应用的一种电子设备的结构框图。

在本实施例中，电子设备可以是服务器、智能手机、平板电脑、便携计算机、桌上型计算机等具有运算功能的终端设备。

该电子设备包括：存储器 11、处理器 12、网络接口 13 及数据总线 14。

存储器 11 包括至少一种类型的可读存储介质，至少一种类型的可读存储介质可为如闪存、硬盘、多媒体卡、卡型存储器等的非易失性存储介质。在一些实施例中，可读存储介质可以是电子设备的内部存储单元，例如该电子设备的硬盘。在另一些实施例中，可读存储介质也可以是电子设备的外部存储器，例如电子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，闪存卡 (Flash Card) 等。

在本实施例中，存储器 11 的可读存储介质通常用于存储安装于电子设备的基于图神经网络的数据推荐程序、多种样本集及预先训练好的模型等。存储器 11 还可以用于暂时地存储已经输出或者将要输出的数据。

处理器 12 在一些实施例中可以是一中央处理器 (Central Processing Unit, CPU)，微处理器或其他数据处理芯片，用于运行存储器 11 中存储的程序代码或处理数据，例如执行数据推荐程序等。

网络接口 13 可选地可以包括标准的有线接口、无线接口 (如 WI-FI 接口)，通常用于在该电子设备与其他电子设备之间建立通信连接。

数据总线 14 用于实现这些组件之间的连接通信。

可选的，该电子设备还可以包括用户接口，用户接口可以包括输入单元比如键盘 (Keyboard)、语音输入装置比如麦克风 (microphone) 等具有语音识别功能的设备、语音输出装置比如音响、耳机等，可选地用户接口还可以包括标准的有线接口、无线接口。

可选的，该电子设备还可以包括显示器，显示器也可以称为显示屏或显示单元。在一些

实施例中可以是 LED 显示器、液晶显示器、触控式液晶显示器以及有机发光二极管 (Organic Light-Emitting Diode, OLED) 触摸器等。显示器用于显示在电子设备中处理的信息以及用于显示可视化的用户界面。

可选的, 该电子设备还包括触摸传感器。触摸传感器所提供的供用户进行触摸操作的区域称为触控区域。此外, 这里的触摸传感器可以为电阻式触摸传感器、电容式触摸传感器等。而且, 触摸传感器不仅包括接触式的触摸传感器, 也可包括接近式的触摸传感器等。此外, 触摸传感器可以为单个传感器, 也可以为例如阵列布置的多个传感器。

此外, 该电子设备的显示器的面积可以与触摸传感器的面积相同, 也可以不同。可选地, 将显示器与触摸传感器层叠设置, 以形成触摸显示屏。该装置基于触摸显示屏侦测用户触发的触控操作。

下面对本申请实施例公开的一种基于图神经网络的数据推荐方法进行具体说明。

实施例一

如图 2 所示, 图 2 是本申请实施例一提供的基于图神经网络的数据推荐方法的流程图。基于图 1 所示的电子设备实施例, 处理器 12 执行存储器 11 中存储的程序时实现如下步骤:

步骤 S200: 获取新用户的目标搜索信息。

在本申请实施例中, 新用户可以是第一次输入搜索信息的用户, 目标搜索信息包括但不限于: 新用户语音输入的搜索语句或关键词; 新用户通过搜索引擎输入的文字信息或图片信息等。具体的, 每次接收到搜索请求时, 可以先获取搜索请求对应的用户信息, 用户信息包括但不限于用户 ID、用户 IP 地址和用户设备码等用于标识用户身份的信息。之后, 确认数据库中是否已录入用户信息, 若未录入用户信息, 则根据用户信息新建新用户并录入数据库。相应的, 若已录入用户信息, 则从数据库中确定用户信息对应的目标用户, 并直接获取目标用户的目标特征信息, 执行步骤 S260。

步骤 S210: 根据目标搜索信息, 生成候选数据集。

在本申请实施例中, 可以预先通过网络爬虫搜集互联网的网页链接以及具体访问资源(比如某一文档、购物页的某一商品等)等大数据信息, 再对这些大数据信息进行提取并组织建立索引库。实际应用中, 可以由检索器将用户输入的目标搜索信息拆分成具有检索意义的字或词, 再根据拆分得到的字或词, 在索引库中快速检出候选数据以生成候选数据集。示例性的, 当新用户 in 搜索引擎中输入搜索关键词“XX 科技”时, 调取包含“XX 科技”的网页链接集合以加入候选数据集。

步骤 S220: 确定与新用户相关联的关联用户。

在本申请实施例中, 关联用户可以包括但不限于: 新用户使用社交软件时关联的好友用户、工作用户(包括客户、同级同事、上下级同事等)、临时对话用户(比如评论或转发用户)、关注用户、点赞用户、点踩用户和收藏用户等; 以及, 新用户的设备通讯录联系人。相应的, 可以通过获取新用户的账号信息访问权限或使用设备信息访问权限, 从新用户访问的社交应用软件、办公应用软件或多媒体应用软件等多种信息渠道确定与新用户相关联的关联用户。

步骤 S230：获取关联用户的第一搜索数据。

其中，第一搜索数据可以包括关联用户的历史搜索信息，比如关联用户利用搜索引擎搜索过的文本、图片等。进一步的，第一搜索数据具体还可以包括关联用户对历史搜索信息的评价数据，比如关联用户对任一网页链接或具体访问资源的评价内容（包括喜欢、不喜欢和评价类型等）；关联用户对历史搜索信息的推荐数据，比如关联用户使用软件分享功能时发送的数据。

步骤 S240：根据新用户、关联用户和第一搜索数据，生成新用户的社交关系图。

在本申请实施例中，根据新用户和关联用户，可以确定新用户与关联用户之间的社交关系信息，比如同事、好友、关注、点赞或不喜欢等。根据关联用户和第一搜索数据，可以确定关联用户和第一搜索数据之间的关联关系信息，比如喜欢、不喜欢、推荐、好评或差评等。基于此，可选的，分别以新用户、关联用户以及第一搜索数据为不同节点，再根据不同节点之间的关系信息，为不同节点之间建立连接关系，得到新用户的社交关系图。

请参阅图 3，图 3 是本申请实施例中一种社交关系图的示意图。如图 3 所示，该社交关系图中包括四个用户节点和两个数据节点，四个用户节点包括新用户节点 A、第一关联用户节点 B、第二关联用户节点 C 和第三关联用户节点 D，两个数据节点包括物品节点 1 和物品节点 2。其中，与新用户相关的社交关系信息包括：新用户节点 A 与第一关联用户节点 B 连接，其连接关系为“不喜欢”；新用户节点 A 与第二关联用户节点 C 连接，其连接关系为点赞关系；新用户节点 A 与第三关联用户节点 D 连接，其连接关系为关注关系。与第二关联用户节点 C 相关的关系信息包括：第二关联用户节点 C 与物品节点 1 连接，其连接关系为“不喜欢”；第二关联用户节点 C 与物品节点 2 连接，其连接关系为“喜欢”。可见，社交关系图采用图数据结构，便于梳理不同用户之间以及用户与资源之间的关系。

步骤 S250：将社交关系图输入到预设图神经网络模型中进行特征预测，得到新用户的目标特征信息。

作为一种可选的实施方式，预设图神经网络模型的训练步骤可以包括：

首先，获取训练样本以及训练样本的标注数据，训练样本包括第一用户的第一社交关系样本图以及第二用户的第二社交关系样本图，第一用户和第二用户为任意两位用户，标注数据用于表示第一用户与第二用户是否相似。其中，第二社交关系样本图和第一社交关系样本图可参照上述对社交关系图的说明，不再赘述。一种实现方式中，先取第一用户的第一社交关系样本图以及第二用户的第二社交关系样本图作为训练样本，再对第一社交关系样本图和第二社交关系样本图进行节点搜索。若在第一社交关系样本图搜索到第二用户对应的第二节点，和/或，在第二社交关系样本图中搜索到第一用户对应的第一节点，则将训练样本的标注数据确定为第一用户与第二用户相似。若未在第一社交关系样本图搜索到第二节点，并且未在第二社交关系样本图中搜索到第一节点，将训练样本的标注数据确定为第一用户与第二用户不相似。

之后，将第一社交关系样本图和第二社交关系样本图输入预设图神经网络模型。通过预设图神经网络模型生成第一社交关系样本图的第一特征信息以及第二社交关系样本图的第二

特征信息，并对第一特征信息和第二特征信息进行融合，得到融合结果。其中，融合结果可以表示第一特征信息与第二特征信息之间的相似值 S 。

最后，利用标注数据对融合结果进行验证，得到验证结果，若验证结果未达到预设准确率，则重新执行获取训练样本以及训练样本的标注数据的步骤，若验证结果达到预设准确率，训练结束。预设准确率可以由人为设定及调整，比如取 90% 或 95%，不做具体限定。

可以理解，每次用于训练预设图神经网络模型的训练样本数量可以是多个。一种实现方式中，可以计算每个训练样本的准确值 R ，再对多个训练样本的准确值 R 进行平均计算，得到验证结果，准确值 R 的计算方式为：

$$R = \begin{cases} \frac{S}{s_1}, & S < s_1, \text{ 标注数据为相似} \\ 1 - (\frac{S-s_2}{s_2}), & S > s_2, \text{ 标注数据为不相似} \\ 1, & S \geq s_1, \text{ 标注数据为相似} \\ 1, & S \leq s_2, \text{ 标注数据为不相似} \end{cases},$$

其中， s_1 是标注数据为相似时设定的标准值（比如 0.9）， s_2 为标注数据为不相似时设定的标准值（比如 0.1）。

另一种实现方式中，也可以在标注数据为相似的情况下，设定第一取值范围（比如 0.8 至 1），若训练样本对应的相似值 S 属于第一取值范围，则样本计数值 N 累计加一。以及，在标注数据为不相似的情况下，设定第二取值范围（比如 0 至 0.2），若训练样本对应的相似值 S 属于第二取值范围，则样本计数值 N 累计加一。基于此，遍历所有训练样本后，取最终确定的样本计数值 N 与训练样本数量之间的比值作为验证结果。

进一步的，作为一种可选的实施方式，预设图神经网络模型可以包括第一特征生成模块、第二特征生成模块以及融合模块，则在预设图神经网络模型的训练步骤中，可以将第一社交关系样本图输入第一特征生成模块，得到第一特征信息，以及将第二社交关系样本图输入第二特征生成模块，得到第二特征信息。再将第一特征信息和第二特征信息输入融合模块，得到融合结果。具体的，融合模块可以采用点积运算单元，则将第一特征信息和第二特征信息输入融合模块中进行点积运算，得到第一特征信息与第二特征信息之间的相似值 S 。

此外，在训练预设图神经网络模型的过程中，可以通过均方根误差损失函数 L ，对预设图神经网络模型输出的融合结果与标注数据计算损失值，并通过损失值的反向传播进行预设图神经网络模型的参数调整，直至训练结束。均方根误差损失函数 L 满足：

$$L = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2},$$

其中， m 为训练样本的数量， y_i 为第 i 个训练样本对应的融合结果， $\hat{y}_i$ 为第 i 个训练样本对应的标注数据。

相应的，将社交关系图输入到预设图神经网络模型中进行特征预测，得到新用户的目标特征信息，具体为：将社交关系图输入到第一特征生成模块或第二特征生成模块中进行特征预测，得到新用户的目标特征信息。可见，训练好的预设图神经网络模型能够智能分析用户社交关系及关联数据特征，为不同用户生成可靠的特征信息，并同时满足不同用户之间的相

似对比结果，因此在实际应用中，只需将新用户的社交关系图输入预设图神经网络模型中的特征生成模块，即可得到新用户的目标特征信息，具备更优的模型应用性能和可靠性。

请参与图 4，图 4 是本申请实施例中一种预设图神经网络模型的结构示意图。如图 4 所示，作为一种可选的实施方式，第一特征生成模块和第二特征生成模块均包括依次连接的双层图卷积层（double graph convolutional layer, DGCL）、池化层和全连接层，形成相似度-双向图卷积网络（similarity bi-directional graph convolutional networks, S-BIGCN）。具体的，双层图卷积层的参数共享，适用于无向图结构的社会关系图，而池化层可以采用平均池化层，既防止过拟合，又使得特征生成模块具备更优的训练速度以及推理速度，不需要耗费过多的计算机 GPU 资源，故预设图神经网络模型更易落地。

其他可选的实施方式中，预设图神经网络模型也可以采用图注意力网络模型、图自编码网络模型或图生成网络模型等。

步骤 S260：根据目标特征信息，对候选数据集进行排序处理，得到推荐数据集。

在本申请实施例中，候选数据集包括多个候选数据，一种可选的实施方式中，可以获取每个候选数据相关的社交关系图，并将候选数据相关的社交关系图输入到预设图神经网络模型中，得到候选数据的特征信息。将目标特征信息与每个候选数据的特征信息代入相似度计算公式中，得到每个候选数据对应的排序值。根据每个候选数据对应的排序值大小，即可实现对候选数据集的排序处理。其中，相似度计算公式包括但不限于余弦相似度计算方式、欧氏距离计算方式、曼哈顿距离计算公式或 Jaccard 相似系数计算公式等。

可见，实施上述方法实施例，将同时融合了用户社交关系以及不同用户的第一搜索数据的社交关系图输入到预设图神经网络模型中，使得预设图神经网络模型沿着社交关系链的深层传播识别关联用户之间的共同数据需求和评价模式特征，从而为新用户确定更具备可解释性的目标特征信息。之后，再基于新用户的目标特征信息，对新用户的候选数据集进行排序，得到为新用户推送的推荐数据集，能够缓解用户侧冷启动问题，提升对新用户进行数据推荐的准确性，进而改善了新用户对数据搜索功能的使用体验。

实施例二

如图 5 所示，图 5 是本申请实施例二提供的基于图神经网络的数据推荐方法的流程图。基于图 1 所示的电子设备实施例，处理器 12 执行存储器 11 中存储的程序时实现如下步骤：

步骤 S500：获取新用户的目标搜索信息。

步骤 S510：根据目标搜索信息，生成候选数据集。

步骤 S520：确定与新用户相关联的关联用户。

步骤 S530：获取关联用户的第一搜索数据。

步骤 S540：根据新用户、关联用户和第一搜索数据，生成新用户的社交关系图。

步骤 S550：将社交关系图输入到预设图神经网络模型中进行特征预测，得到新用户的目标特征信息。

可以理解的是，本实施例中步骤 S500 至 S550 的具体实现方式还可以参照上述实施例一中对步骤 S200 至 S250 的描述，在此不再赘述。

步骤 S560: 根据目标特征信息, 获取新用户的多位相似用户以及相似用户与新用户的相似度。

作为一种可选的实施方式, 步骤 S560 具体为: 获取多位已知用户的第三特征信息。根据目标特征信息以及每位已知用户的第三特征信息, 计算新用户与已知用户之间的相似度, 并取相似度在预设阈值范围内的已知用户作为相似用户。其中, 相似度的取值属于 $[0, 1]$, 相似度的取值越大, 表示新用户与已知用户的相似程度越高。预设阈值范围可以是人为指定, 比如 $[0.8, 1]$; 或者, 也可以对新用户与不同已知用户之间的相似度从大到小进行排序, 取排序第 k 位的相似度为目标阈值 d , k 为正整数并由人为指定, 再根据目标阈值 d 确定预设阈值范围 (比如预设阈值范围为 $[d, 1]$), 对此不做限定。

具体的, 可以从数据库中获取已录入的多位已知用户。一种实现方式中, 将目标特征信息和每位已知用户的第三特征信息代入预设的相似度计算公式中, 得到新用户与已知用户之间的相似度。比如, 相似度计算公式可以是余弦相似度计算方式, 即:

$$\text{sim}(u, v) = \frac{|u \cap v|}{\sqrt{|u| * |v|}},$$

其中, $\text{sim}(u, v)$ 为新用户与已知用户之间的相似度, u 为目标特征信息, v 为已知用户的第三特征信息。

可见, 通过挖掘新用户与已知用户的相似性, 筛选出新用户的相似用户, 以便根据相似用户的行为数据进行推荐参考。

进一步的, 作为一种可选的实施方式, 获取多位已知用户的第三特征信息, 具体可以包括以下步骤:

获取多位已知用户的历史搜索信息以及已知用户对历史搜索信息的操作数据, 已知用户的历史搜索信息包括多个第二搜索数据。第二搜索数据可以指响应于已知用户的搜索请求后, 向已知用户推荐过的历史数据。操作数据包括对每个第二搜索数据的操作记录, 操作记录对应的操作类型包括但不限于以下至少一种: 点击次数、收藏操作、分享次数、具体评分情况、点赞操作和点踩操作, 用于表示已知用户对不同第二搜索数据的反馈情况。

之后, 根据历史搜索信息和操作数据, 生成评分矩阵。从评分矩阵中获取每位已知用户对应的行向量, 行向量包括多个第一评分元素, 多个第一评分元素与多个第二搜索数据一一对应。根据已知用户对应的行向量, 即可确定已知用户的第三特征信息。

其中, 评分矩阵的行表示不同已知用户, 评分矩阵的列表示不同第二搜索数据, 则第 i 行第 j 列的矩阵元素表示第 i 位已知用户对第 j 个第二搜索数据的评分元素。

一种实现方式中, 如果采用的操作类型为一种, 则根据该操作类型的操作记录, 确定相应的评分元素。可以理解, 如果操作记录为数值, 比如点击次数、分享次数或具体评分情况等, 则可以直接取操作记录为评分元素。如果操作记录并非数值, 可以先对操作记录进行量化处理, 得到评分元素。比如, 针对收藏操作, 将已收藏操作量化为“1”, 将未收藏操作量化为“0”; 将点赞操作量化为“1”; 以及, 将点踩操作量化为“-1”等。

示例性的, 假设已知用户E对第二搜索数据1、第二搜索数据2和第二搜索数据3的点

击次数分别为 1、1 和 0，已知用户 F 对应的点击次数分别为 1、0 和 0，已知用户 G 对应的点击次数分别为 2、1 和 0，可以生成如下表 1 所示的评分矩阵。

表 1 评分矩阵示例表

	第二搜索数据 1	第二搜索数据 2	第二搜索数据 3
已知用户 E	1	1	0
已知用户 F	1	0	1
已知用户 G	2	1	0

基于表 1，取行向量[1, 1, 0]作为已知用户 E 的第三特征信息，取行向量[1, 0, 1]作为已知用户 F 的第三特征信息，取行向量[2, 1, 0]为已知用户 G 的第三特征信息。

另一种实现方式中，如果采用的操作类型为两种及以上，则可以根据不同操作类型的加权重，对所有操作类型的操作记录进行加权求和计算，得到相应的评分元素。比如，假设点击次数和分享次数各自的加权重为 0.5，已知用户 E 分别对第二搜索数据 1、第二搜索数据 2 和第二搜索数据 3 的分享次数为 1、0 和 0，则已知用户 E 的行向量= $[1*0.5+1*0.5, 1*0.5, 0]$ =[1, 0.5, 0]。

此外，还可选的，还可以将评分矩阵中每个行向量（或列向量）进行归一化处理，得到归一化后的行向量（或列向量），使得归一化后的行向量（或列向量）中每个评分元素分布在[0, 1]中。

基于此，根据已知用户对应的行向量，确定已知用户的第三特征信息之后，可以对第三特征信息进行存储，便于实际应用中直接调取已知用户的第三特征信息。进一步的，也可以在训练好（或每次更新完）预设图神经网络模型后，并获得已知用户的社交关系图的情况下，将已知用户的社交关系图输入到预设图神经网络模型中，得到新的第三特征信息，并利用新的特征信息替代原有的第三特征信息进行存储，实现特征更新作用。

步骤 S570：获取多位相似用户对候选数据集的第一评分信息。

在本申请实施例中，候选数据集可以包括多个候选数据。一种可选的实施方式中，根据历史搜索信息和操作数据，生成评分矩阵之后，还可以从评分矩阵中获取每个第二搜索数据对应的列向量，列向量包括多个第二评分元素，多个第二评分元素与多位已知用户一一对应。根据第二搜索数据对应的列向量，确定第二搜索数据的第四特征信息。

仍以表 1 为例，可以取列向量[1, 1, 2]为第二搜索数据 1 的第四特征信息，取列向量[1, 0, 1]为第二搜索数据 2 的第四特征信息，取列向量[0, 1, 0]为第二搜索数据 3 的第四特征信息。

相应的，步骤 S570 可以是：通过将候选数据与第二搜索数据进行匹配，从多个第二搜索数据的第四特征信息中，确定每个候选数据的第四特征信息。根据相似用户的第三特征信息以及候选数据的第四特征信息，计算相似用户对候选数据的第一评分值。

具体的，可以将相似用户的第三特征信息 w 以及候选数据的第四特征信息 c 代入相似度计算公式中，得到相似用户对候选数据的第一评分值 $\text{score}(u, c)$ ，比如，相似度计算公式采用

余弦相似度计算方式, 即:

$$\text{score}(w, c) = \frac{|wnc|}{\sqrt{|w|*|c|}}.$$

其他实现方式中, 也可以基于候选数据的标签生成候选数据的第四特征信息, 或者采用其他基于神经网络学习的方法生成候选数据的第四特征信息。

步骤 S580: 根据第一评分信息, 并结合每位相似用户与新用户的相似度, 计算新用户对候选数据集的第二评分信息。

在本申请实施例中, 第二评分信息包括新用户对每个候选数据的第二评分值。可选的, 计算第二评分信息的方式具体可以为:

$\text{Like}(u, c_j) = \sum_{i=1}^k (\text{sim}(u, w_i)) * \text{score}(w_i, c_j)$, 其中, $\text{Like}(u, c_j)$ 为新用户对第 j 个候选数据 c_j 的第二评分值, $\text{sim}(u, w_i)$ 为新用户与第 i 位相似用户的相似度, $\text{score}(w_i, c_j)$ 为第 i 位相似用户对第 j 个候选数据 c_j 的第一评分值。

步骤 S590: 根据第二评分信息, 对候选数据集进行排序处理, 得到推荐数据集。

在本申请实施例中, 具体的, 可以按照第二评分值从大到小的次序, 对候选数据集中多个候选数据进行降序处理, 得到推荐数据集。或者, 也可以在对候选数据集中多个候选数据进行降序处理之后, 将降序处理的结果分为 p 组推荐数据 (p 为正整数, 且 $p \geq 2$), 再对每组推荐数据进行乱序处理, 得到乱序后的 p 组推荐数据, 以形成推荐数据集。

可见, 实施上述步骤 S560 至步骤 S590, 通过筛选新用户的相似用户, 并结合相似用户的数据评分情况预测新用户对不同候选数据的兴趣偏好, 实现了对候选数据集的个性化排序, 从而优化新用户的搜索满意度。

可以理解的是, 本申请中涉及的数据库可以采用 Elastic Search、MongoDB 或 Redis 等, 不做具体限定。

可见, 实施上述方法实施例, 通过筛选新用户的相似用户, 并结合相似用户的数据评分情况预测新用户对不同候选数据的兴趣偏好, 实现了对候选数据集的协同过滤, 能够缓解用户侧冷启动问题, 提升对新用户进行数据推荐的准确性, 进而改善了新用户对数据搜索功能的使用体验。

本申请实施例还提供一种基于图神经网络的数据推荐装置。请参阅图 6, 图 6 是本申请实施例所应用的一种基于图神经网络的数据推荐装置的结构框图。如图 6 所示, 该基于图神经网络的数据推荐装置 600 包括:

获取模块 610, 用于获取新用户的目标搜索信息。

第一生成模块 620, 用于根据目标搜索信息, 生成候选数据集。

确定模块 630, 用于确定与新用户相关联的关联用户。

获取模块 610, 还用于获取关联用户的第一搜索数据。

第二生成模块 640, 用于根据新用户、关联用户和第一搜索数据, 生成新用户的社交关系图。

预测模块 650, 用于将社交关系图输入到预设图神经网络模型中进行特征预测, 得到新

用户的目标特征信息。

排序模块 660, 用于根据目标特征信息, 对候选数据集进行排序处理, 得到推荐数据集。

需要说明的是, 本实施例的具体实现过程可参见上述方法实施例的具体实现过程, 亦不再赘述。

本申请实施例还提供了一种电子设备, 该电子设备包括存储器和处理器, 存储器上存储有程序, 程序被处理器执行时实现一种基于图神经网络的数据推荐方法, 该方法包括: 获取新用户的目标搜索信息, 并根据目标搜索信息, 生成候选数据集; 确定与新用户相关联的关联用户; 获取关联用户的第一搜索数据; 根据新用户、关联用户和第一搜索数据, 生成新用户的社交关系图; 将社交关系图输入到预设图神经网络模型中进行特征预测, 得到新用户的目标特征信息; 根据目标特征信息, 对候选数据集进行排序处理, 得到推荐数据集。

本申请实施例还提供了一种存储介质, 用于计算机可读存储, 存储介质存储有一个或者多个程序, 一个或者多个程序可被一个或者多个处理器执行, 以实现一种基于图神经网络的数据推荐方法, 该方法包括: 获取新用户的目标搜索信息, 并根据目标搜索信息, 生成候选数据集; 确定与新用户相关联的关联用户; 获取关联用户的第一搜索数据; 根据新用户、关联用户和第一搜索数据, 生成新用户的社交关系图; 将社交关系图输入到预设图神经网络模型中进行特征预测, 得到新用户的目标特征信息; 根据目标特征信息, 对候选数据集进行排序处理, 得到推荐数据集。该存储介质可以是非易失性, 也可以是易失性, 不做具体限定。

本领域普通技术人员可以理解, 上文中所公开方法中的全部或某些步骤、系统、设备中的功能模块/单元可以被实施为软件、固件、硬件及其适当的组合。

在硬件实施方式中, 在以上描述中提及的功能模块/单元之间的划分不一定对应于物理组件的划分; 例如, 一个物理组件可以具有多个功能, 或者一个功能或步骤可以由若干物理组件合作执行。某些物理组件或所有物理组件可以被实施为由处理器, 如中央处理器、数字信号处理器或微处理器执行的软件, 或者被实施为硬件, 或者被实施为集成电路, 如专用集成电路。这样的软件可以分布在计算机可读介质上, 计算机可读介质可以包括计算机存储介质 (或非暂时性介质) 和通信介质 (或暂时性介质)。如本领域普通技术人员公知的, 术语计算机存储介质包括在用于存储信息 (诸如计算机可读指令、数据结构、程序模块或其他数据) 的任何方法或技术中实施的易失性和非易失性、可移除和不可移除介质。计算机存储介质包括但不限于 RAM、ROM、EEPROM、闪存或其他存储器技术、CD-ROM、数字多功能盘 (DVD) 或其他光盘存储、磁盒、磁带、磁盘存储或其他磁存储装置、或者可以用于存储期望的信息并且可以被计算机访问的任何其他的介质。此外, 本领域普通技术人员公知的是, 通信介质通常包含计算机可读指令、数据结构、程序模块或者诸如载波或其他传输机制之类的调制数据信号中的其他数据, 并且可包括任何信息递送介质。

以上参照附图说明了本申请的优选实施例, 并非因此局限本申请的权利范围。本领域技术人员不脱离本申请的范围和实质内所作的任何修改、等同替换和改进, 均应在本申请的权利范围之内。

权 利 要 求 书

1. 一种基于图神经网络的数据推荐方法，其中，所述方法包括：

获取新用户的目标搜索信息，并根据所述目标搜索信息，生成候选数据集；

确定与所述新用户相关联的关联用户；

获取所述关联用户的第一搜索数据；

根据所述新用户、所述关联用户和所述第一搜索数据，生成所述新用户的社交关系图；

将所述社交关系图输入到预设图神经网络模型中进行特征预测，得到所述新用户的目标特征信息；

根据所述目标特征信息，对所述候选数据集进行排序处理，得到推荐数据集。

2. 根据权利要求1所述的方法，其中，所述预设图神经网络模型的训练步骤包括：

获取训练样本以及所述训练样本的标注数据，所述训练样本包括第一用户的第一社交关系样本图以及第二用户的第二社交关系样本图，所述第一用户和所述第二用户为任意两位用户，所述标注数据用于表示所述第一用户与所述第二用户是否相似；

将所述第一社交关系样本图和所述第二社交关系样本图输入预设图神经网络模型；

通过所述预设图神经网络模型生成所述第一社交关系样本图的第一特征信息以及所述第二社交关系样本图的第二特征信息，并对所述第一特征信息和所述第二特征信息进行融合，得到融合结果；

利用所述标注数据对所述融合结果进行验证，得到验证结果，若所述验证结果未达到预设准确率，则重新执行所述获取训练样本以及所述训练样本的标注数据的步骤，若所述验证结果达到所述预设准确率，训练结束。

3. 根据权利要求2所述的方法，其中，所述获取训练样本以及所述训练样本的标注数据，包括：

取第一用户的第一社交关系样本图以及第二用户的第二社交关系样本图作为训练样本；

对所述第一社交关系样本图和所述第二社交关系样本图进行节点搜索；

若在所述第二社交关系样本图中搜索到所述第一用户对应的第一节点，和/或，在所述第一社交关系样本图中搜索到所述第二用户对应的第二节点，将所述训练样本的标注数据确定为所述第一用户与所述第二用户相似；

若未在所述第一社交关系样本图搜索到所述第二节点，并且未在所述第二社交关系样本图中搜索到所述第一节点，将所述训练样本的标注数据确定为所述第一用户与所述第二用户不相似。

4. 根据权利要求1至3任一项所述的方法，其中，所述根据所述目标特征信息，对所述候选数据集进行排序处理，得到推荐数据集，包括：

根据所述目标特征信息，获取所述新用户的多位相似用户以及所述相似用户与所述新用户的相似度；

获取多位所述相似用户对所述候选数据集的第一评分信息；

根据所述第一评分信息，并结合每位所述相似用户与所述新用户的相似度，计算所述新用户对所候选数据集的第二评分信息；

根据所述第二评分信息，对所述候选数据集进行排序处理，得到推荐数据集。

5. 根据权利要求4所述的方法，其中，所述根据所述目标特征信息，获取所述新用户的多位相似用户以及所述相似用户与所述新用户的相似度，包括：

获取多位已知用户的第三特征信息；

根据所述目标特征信息以及每位所述已知用户的第三特征信息，计算所述新用户与所述已知用户之间的相似度，并取相似度在预设阈值范围内的已知用户作为相似用户。

6. 根据权利要求5所述的方法，其中，所述获取多位已知用户的第三特征信息，包括：

获取多位已知用户的历史搜索信息以及所述已知用户对所述历史搜索信息的操作数据，所述历史搜索信息包括多个第二搜索数据；

根据所述历史搜索信息和所述操作数据，生成评分矩阵；

从所述评分矩阵中获取每位已知用户对应的行向量，所述行向量包括多个第一评分元素，多个所述第一评分元素与多个所述第二搜索数据一一对应；

根据所述已知用户对应的行向量，确定所述已知用户的第三特征信息。

7. 根据权利要求6所述的方法，其中，所述候选数据集包括多个候选数据；所述根据所述历史搜索信息和所述操作数据，生成评分矩阵之后，所述方法还包括：

从所述评分矩阵中获取每个第二搜索数据对应的列向量，所述列向量包括多个第二评分元素，多个所述第二评分元素与多位所述已知用户一一对应；

根据所述第二搜索数据对应的列向量，确定所述第二搜索数据的第四特征信息；

所述获取多位所述相似用户对所述候选数据集的第一评分信息，包括：

从多个所述第二搜索数据的第四特征信息中，确定每个所述候选数据的第四特征信息；

根据所述相似用户的第三特征信息以及所述候选数据的第四特征信息，计算所述相似用户对所述候选数据的第一评分值。

8. 一种基于图神经网络的数据推荐装置，其中，所述装置包括：

获取模块，用于获取新用户的目标搜索信息；

第一生成模块，用于根据所述目标搜索信息，生成候选数据集；

确定模块，用于确定与所述新用户相关联的关联用户；

所述获取模块，还用于获取所述关联用户的第一搜索数据；

第二生成模块，用于根据所述新用户、所述关联用户和所述第一搜索数据，生成所述新用户的社交关系图；

预测模块，用于将所述社交关系图输入到预设图神经网络模型中进行特征预测，得到所述新用户的目标特征信息；

排序模块，用于根据所述目标特征信息，对所述候选数据集进行排序处理，得到推荐数据集。

9. 一种电子设备，其中，所述电子设备包括存储器、处理器、存储在所述存储器上并可

在所述处理器上运行的程序以及用于实现所述处理器和所述存储器之间的连接通信的数据总线,所述程序被所述处理器执行时实现一种基于图神经网络的数据推荐方法,所述方法包括:

获取新用户的目标搜索信息,并根据所述目标搜索信息,生成候选数据集;

确定与所述新用户相关联的关联用户;

获取所述关联用户的第一搜索数据;

根据所述新用户、所述关联用户和所述第一搜索数据,生成所述新用户的社交关系图;

将所述社交关系图输入到预设图神经网络模型中进行特征预测,得到所述新用户的目标特征信息;

根据所述目标特征信息,对所述候选数据集进行排序处理,得到推荐数据集。

10. 根据权利要求 9 所述的电子设备,其中,所述预设图神经网络模型的训练步骤包括:

获取训练样本以及所述训练样本的标注数据,所述训练样本包括第一用户的第一社交关系样本图以及第二用户的第二社交关系样本图,所述第一用户和所述第二用户为任意两位用户,所述标注数据用于表示所述第一用户与所述第二用户是否相似;

将所述第一社交关系样本图和所述第二社交关系样本图输入预设图神经网络模型;

通过所述预设图神经网络模型生成所述第一社交关系样本图的第一特征信息以及所述第二社交关系样本图的第二特征信息,并对所述第一特征信息和所述第二特征信息进行融合,得到融合结果;

利用所述标注数据对所述融合结果进行验证,得到验证结果,若所述验证结果未达到预设准确率,则重新执行所述获取训练样本以及所述训练样本的标注数据的步骤,若所述验证结果达到所述预设准确率,训练结束。

11. 根据权利要求 10 所述的电子设备,其中,所述获取训练样本以及所述训练样本的标注数据,包括:

取第一用户的第一社交关系样本图以及第二用户的第二社交关系样本图作为训练样本;

对所述第一社交关系样本图和所述第二社交关系样本图进行节点搜索;

若在所述第二社交关系样本图中搜索到所述第一用户对应的第一节点,和/或,在所述第一社交关系样本图中搜索到所述第二用户对应的第二节点,将所述训练样本的标注数据确定为所述第一用户与所述第二用户相似;

若未在所述第一社交关系样本图搜索到所述第二节点,并且未在所述第二社交关系样本图中搜索到所述第一节点,将所述训练样本的标注数据确定为所述第一用户与所述第二用户不相似。

12. 根据权利要求 9 至 11 任一项所述的电子设备,其中,所述根据所述目标特征信息,对所述候选数据集进行排序处理,得到推荐数据集,包括:

根据所述目标特征信息,获取所述新用户的多位相似用户以及所述相似用户与所述新用户的相似度;

获取多位所述相似用户对所述候选数据集的第一评分信息;

根据所述第一评分信息,并结合每位所述相似用户与所述新用户的相似度,计算所述新

用户对所述候选数据集的第二评分信息；

根据所述第二评分信息，对所述候选数据集进行排序处理，得到推荐数据集。

13. 根据权利要求 12 所述的电子设备，其中，所述根据所述目标特征信息，获取所述新用户的多位相似用户以及所述相似用户与所述新用户的相似度，包括：

获取多位已知用户的第三特征信息；

根据所述目标特征信息以及每位所述已知用户的第三特征信息，计算所述新用户与所述已知用户之间的相似度，并取相似度在预设阈值范围内的已知用户作为相似用户。

14. 根据权利要求 13 所述的电子设备，其中，所述获取多位已知用户的第三特征信息，包括：

获取多位已知用户的历史搜索信息以及所述已知用户对所述历史搜索信息的操作数据，所述历史搜索信息包括多个第二搜索数据；

根据所述历史搜索信息和所述操作数据，生成评分矩阵；

从所述评分矩阵中获取每位已知用户对应的行向量，所述行向量包括多个第一评分元素，多个所述第一评分元素与多个所述第二搜索数据一一对应；

根据所述已知用户对应的行向量，确定所述已知用户的第三特征信息。

15. 一种存储介质，用于计算机可读存储，其中，所述存储介质存储有一个或者多个程序，所述一个或者多个程序可被一个或者多个处理器执行，以实现一种基于图神经网络的数据推荐方法，所述方法包括：

获取新用户的目标搜索信息，并根据所述目标搜索信息，生成候选数据集；

确定与所述新用户相关联的关联用户；

获取所述关联用户的第一搜索数据；

根据所述新用户、所述关联用户和所述第一搜索数据，生成所述新用户的社交关系图；

将所述社交关系图输入到预设图神经网络模型中进行特征预测，得到所述新用户的目标特征信息；

根据所述目标特征信息，对所述候选数据集进行排序处理，得到推荐数据集。

16. 根据权利要求 15 所述的存储介质，其中，所述预设图神经网络模型的训练步骤包括：

获取训练样本以及所述训练样本的标注数据，所述训练样本包括第一用户的第一社交关系样本图以及第二用户的第二社交关系样本图，所述第一用户和所述第二用户为任意两位用户，所述标注数据用于表示所述第一用户与所述第二用户是否相似；

将所述第一社交关系样本图和所述第二社交关系样本图输入预设图神经网络模型；

通过所述预设图神经网络模型生成所述第一社交关系样本图的第一特征信息以及所述第二社交关系样本图的第二特征信息，并对所述第一特征信息和所述第二特征信息进行融合，得到融合结果；

利用所述标注数据对所述融合结果进行验证，得到验证结果，若所述验证结果未达到预设准确率，则重新执行所述获取训练样本以及所述训练样本的标注数据的步骤，若所述验证结果达到所述预设准确率，训练结束。

17. 根据权利要求 16 所述的存储介质, 其中, 所述获取训练样本以及所述训练样本的标注数据, 包括:

取第一用户的第一社交关系样本图以及第二用户的第二社交关系样本图作为训练样本;

对所述第一社交关系样本图和所述第二社交关系样本图进行节点搜索;

若在所述第二社交关系样本图中搜索到所述第一用户对应的第一节点, 和/或, 在所述第一社交关系样本图中搜索到所述第二用户对应的第二节点, 将所述训练样本的标注数据确定为所述第一用户与所述第二用户相似;

若未在所述第一社交关系样本图搜索到所述第二节点, 并且未在所述第二社交关系样本图中搜索到所述第一节点, 将所述训练样本的标注数据确定为所述第一用户与所述第二用户不相似。

18. 根据权利要求 15 至 17 任一项所述的存储介质, 其中, 所述根据所述目标特征信息, 对所述候选数据集进行排序处理, 得到推荐数据集, 包括:

根据所述目标特征信息, 获取所述新用户的多位相似用户以及所述相似用户与所述新用户的相似度;

获取多位所述相似用户对所述候选数据集的第一评分信息;

根据所述第一评分信息, 并结合每位所述相似用户与所述新用户的相似度, 计算所述新用户对所述候选数据集的第二评分信息;

根据所述第二评分信息, 对所述候选数据集进行排序处理, 得到推荐数据集。

19. 根据权利要求 18 所述的存储介质, 其中, 所述根据所述目标特征信息, 获取所述新用户的多位相似用户以及所述相似用户与所述新用户的相似度, 包括:

获取多位已知用户的第三特征信息;

根据所述目标特征信息以及每位所述已知用户的第三特征信息, 计算所述新用户与所述已知用户之间的相似度, 并取相似度在预设阈值范围内的已知用户作为相似用户。

20. 根据权利要求 19 所述的存储介质, 其中, 所述获取多位已知用户的第三特征信息, 包括:

获取多位已知用户的历史搜索信息以及所述已知用户对所述历史搜索信息的操作数据, 所述历史搜索信息包括多个第二搜索数据;

根据所述历史搜索信息和所述操作数据, 生成评分矩阵;

从所述评分矩阵中获取每位已知用户对应的行向量, 所述行向量包括多个第一评分元素, 多个所述第一评分元素与多个所述第二搜索数据一一对应;

根据所述已知用户对应的行向量, 确定所述已知用户的第三特征信息。

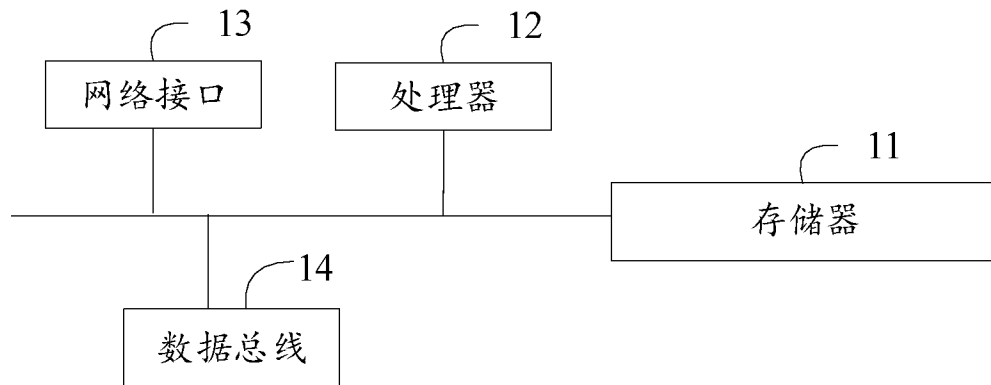


图 1

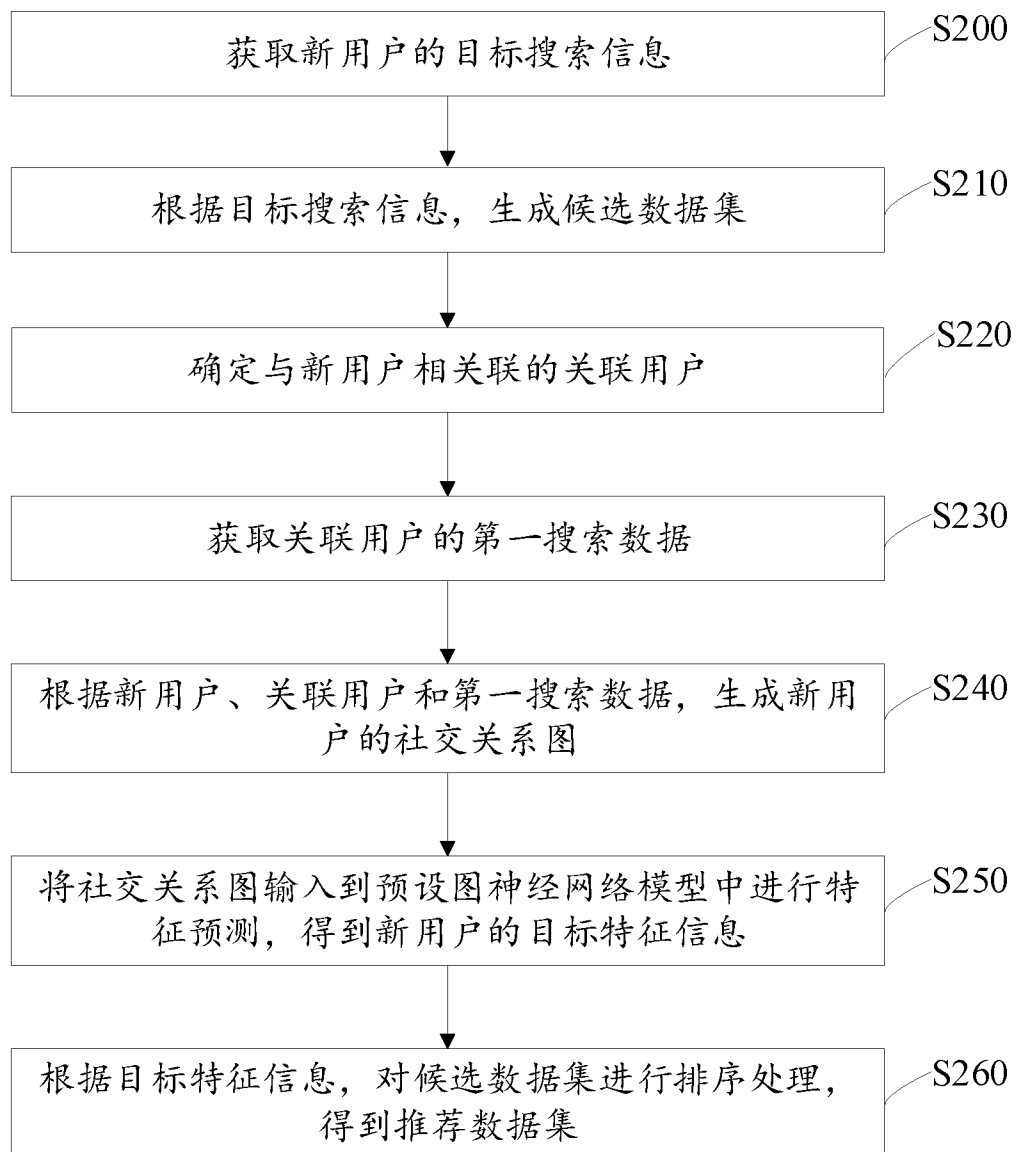


图 2

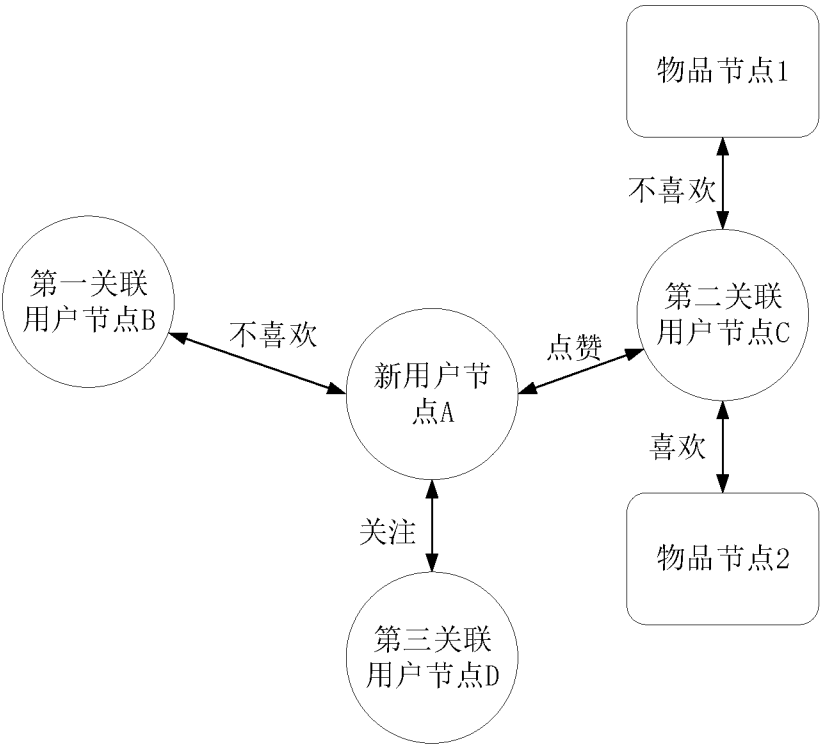


图 3

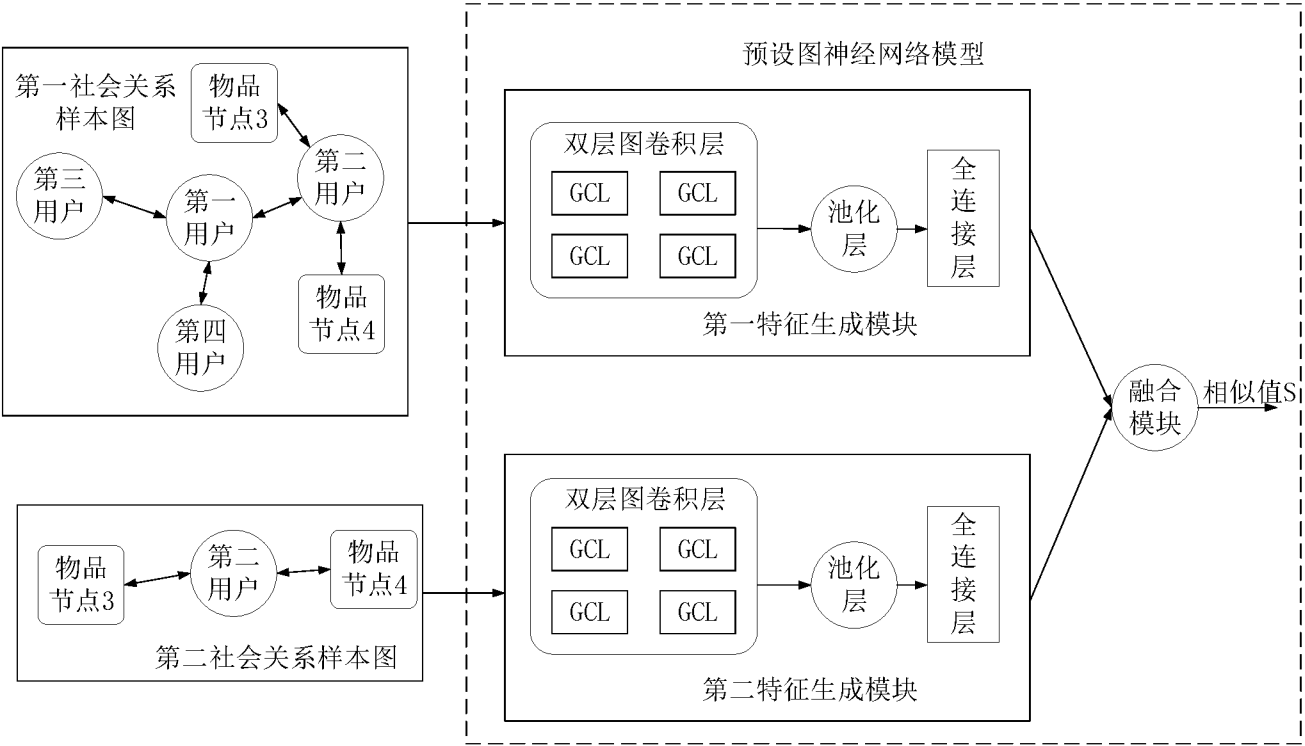


图 4

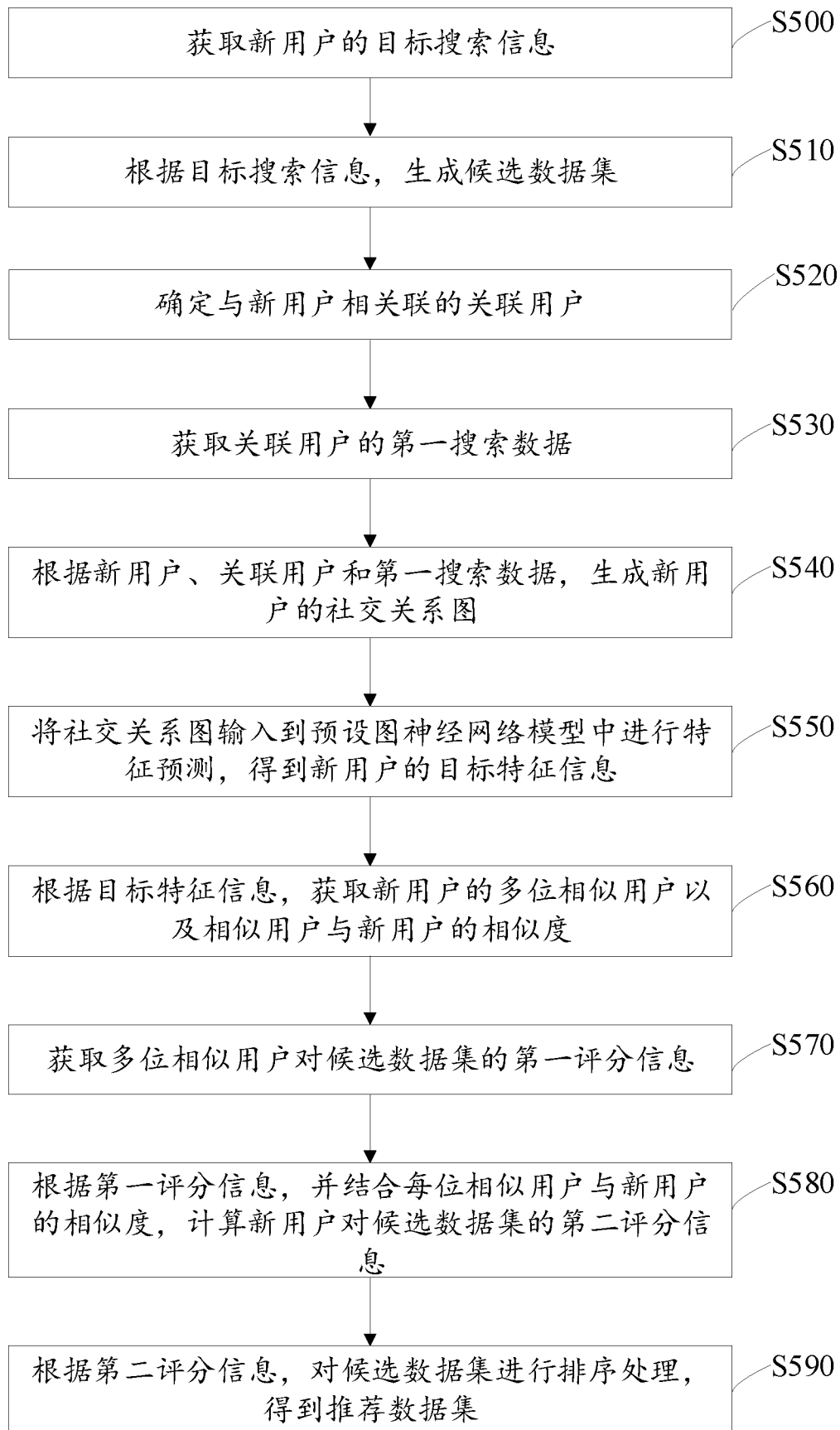


图 5

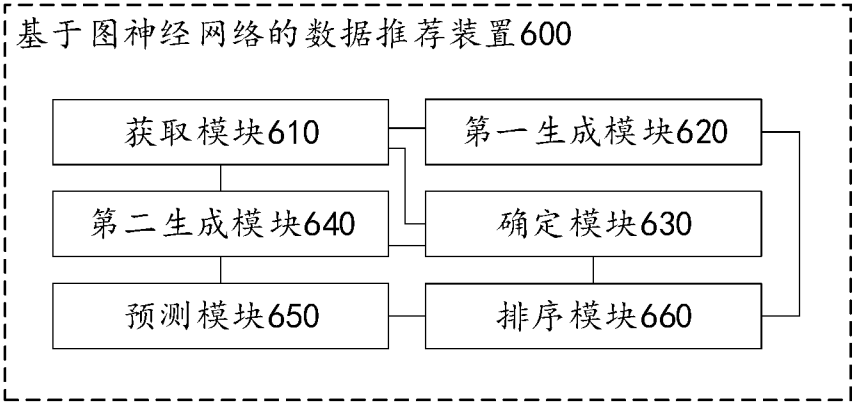


图 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/090754

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/9536(2019.01)i; G06N 3/04(2006.01)i; G06N 3/08(2006.01)i; G06Q 50/00(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)

G06F G06N G06Q

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)

CNPAT, CNKI, WPI, EPODOC, IEEE: 图神经网络, 推荐, 新用户, 搜索, 数据集, 关联用户, 社交, 关系图, 候选, 排序, 训练, 模型, graph neural network, GNN, recommend, new user, search, data set, associated user, relational graph, candidate, rank, train, model

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 112115367 A (BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO., LTD.) 22 December 2020 (2020-12-22) description, paragraphs [0044]-[0056] and [0076]-[0084]	1-20
A	CN 113392289 A (BEIJING SANKUAI ONLINE TECHNOLOGY CO., LTD.) 14 September 2021 (2021-09-14) entire document	1-20
A	CN 112966165 A (PEKING UNIVERSITY) 15 June 2021 (2021-06-15) entire document	1-20
A	CN 113282831 A (BEIJING IQIYI SCIENCE & TECHNOLOGY CO., LTD.) 20 August 2021 (2021-08-20) entire document	1-20
A	WO 2021233199 A1 (HUAWEI TECHNOLOGIES CO., LTD.) 25 November 2021 (2021-11-25) entire document	1-20
A	WO 2022024663 A1 (OMRON CORP.) 03 February 2022 (2022-02-03) entire document	1-20

☐ 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

“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

27 October 2022

Date of mailing of the international search report

16 November 2022

Name and mailing address of the ISA/CN

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

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2022/090754

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	112115367	A	22 December 2020	None			
CN	113392289	A	14 September 2021	None			
CN	112966165	A	15 June 2021	None			
CN	113282831	A	20 August 2021	None			
WO	2021233199	A1	25 November 2021	CN	113688304	A	23 November 2021
WO	2022024663	A1	03 February 2022	JP	2022026106	A	10 February 2022

国际检索报告

国际申请号

PCT/CN2022/090754

A. 主题的分类

G06F 16/9536(2019.01)i; G06N 3/04(2006.01)i; G06N 3/08(2006.01)i; G06Q 50/00(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)

G06F G06N G06Q

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))

CNPAT, CNKI, WPI, EP0DOC, IEEE: 图神经网络, 推荐, 新用户, 搜索, 数据集, 关联用户, 社交, 关系图, 候选, 排序, 训练, 模型, graph neutral network, GNN, recommend, new user, search, data set, associated user, relational graph, candidate, rank, train, model

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
A	CN 112115367 A (北京百度网讯科技有限公司) 2020年12月22日 (2020 - 12 - 22) 说明书第[0044]-[0056], [0076]-[0084]段	1-20
A	CN 113392289 A (北京三快在线科技有限公司) 2021年9月14日 (2021 - 09 - 14) 全文	1-20
A	CN 112966165 A (北京大学) 2021年6月15日 (2021 - 06 - 15) 全文	1-20
A	CN 113282831 A (北京爱奇艺科技有限公司) 2021年8月20日 (2021 - 08 - 20) 全文	1-20
A	WO 2021233199 A1 (华为技术有限公司) 2021年11月25日 (2021 - 11 - 25) 全文	1-20
A	WO 2022024663 A1 (OMRON CORPORATION) 2022年2月3日 (2022 - 02 - 03) 全文	1-20

☐ 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

“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

2022年10月27日

国际检索报告邮寄日期

2022年11月16日

ISA/CN的名称和邮寄地址

中国国家知识产权局(ISA/CN)

中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

郭婉莹

电话号码 86-(10)-53961361

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2022/090754

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	112115367	A	2020年12月22日	无	
CN	113392289	A	2021年9月14日	无	
CN	112966165	A	2021年6月15日	无	
CN	113282831	A	2021年8月20日	无	
WO	2021233199	A1	2021年11月25日	CN 113688304 A	2021年11月23日
WO	2022024663	A1	2022年2月3日	JP 2022026106 A	2022年2月10日