

DOI: 10.11992/tis.201908034

图神经网络推荐研究进展

吴国栋^{1,2}, 查志康², 涂立静², 陶鸿², 宋福根¹

(1. 东华大学 管理学院, 上海 200051; 2. 安徽农业大学 信息与计算机学院, 安徽 合肥 230036)

摘 要: 图神经网络 (graph neural network, GNN) 具有从图的领域对数据进行特征提取和表示的优势, 近年来成为人工智能研究的热点, 图神经网络推荐也是推荐系统研究的一个新方向。本文对 GNN 模型进行深入研究的基础上, 分析了 GNN 推荐过程, 并从无向单元图推荐、无向二元图推荐、无向多元图推荐 3 个方面详细讨论了现有 GNN 推荐研究取得的主要进展及不足, 阐明了现有 GNN 推荐研究中存在的主要难点, 最后提出了 GNN 上下文推荐、GNN 跨领域推荐、GNN 群组推荐、GNN 推荐的可解释性等未来 GNN 推荐的研究方向。

关键词: 图神经网络; 推荐系统; 深度学习; 实体联系; 社交关系; 协同过滤; 无向图; 有向图

中图分类号: TP301 **文献标志码:** A **文章编号:** 1673-4785(2020)01-0014-11

中文引用格式: 吴国栋, 查志康, 涂立静, 等. 图神经网络推荐研究进展 [J]. 智能系统学报, 2020, 15(1): 14-24.

英文引用格式: WU Guodong, ZHA Zhikang, TU Lijing, et al. Research advances in graph neural network recommendation[J]. CAAI transactions on intelligent systems, 2020, 15(1): 14-24.

Research advances in graph neural network recommendation

WU Guodong^{1,2}, ZHA Zhikang², TU Lijing², TAO Hong², SONG Fugeng¹

(1. School of Business and Management, Donghua University, Shanghai 200051, China; 2. School of Information and Computer, Anhui Agricultural University, Hefei 230036, China)

Abstract: Graph neural network (GNN) has the advantage of feature extraction and representation of data from the field of Graph. In recent years, it has become a hotspot of artificial intelligence research, and the recommendation of Graph Neural Network is also a new direction of recommendation system research. Based on the in-depth study of GNN model, this paper analyzes the GNN recommendation process, and discusses in detail the main progress and deficiencies of GNN recommendation studies from three aspects: undirected unit graph recommendation, undirected binary graph recommendation and undirected multivariate graph recommendation. The main difficulties in existing GNN recommendation studies are clarified, and the research directions of GNN recommendation in the future, including GNN contextual recommendation, GNN cross-domain recommendation, GNN group recommendation, and GNN recommendation's interpretability, and so on, are pointed out in the end.

Keywords: graph neural network; recommendation system; deep learning; entity relations; society relation; collaborative filtering; undirected graph; directed graph

图神经网络 (graph neural network, GNN) 作为一种从卷积神经网络 (convolutional neural networks, CNN) 和图嵌入思想启发而来的新型拓展神经网络^[1], 可以在图领域对数据进行特征提取

和表示, 是一种高效、易扩展的新型的神经网络结构, 在学习图数据方面表现出了强大的功能。与传统的深度学习方法相比, GNN 可以通过构建的图模型来反映实体及其之间的联系。目前, 国内外对 GNN 的研究都取得了一定的进展, 如自然语言处理^[2]、文本分类^[3]、特征关系提取^[4]、图片分类^[5]、疾病预测^[6]等。在推荐系统中, 实体间的关系有用户与用户、用户与物品、物品与物品之

收稿日期: 2019-08-30.

基金项目: 国家自然科学基金资助项目 (31671589); 安徽省自然科学基金重点项目 (KJ2017A152, KJ2019A0211).

通信作者: 吴国栋. E-mail: gdwu1120@qq.com.

间关系, 传统的推荐方法主要关注用户与物品间的关系, 很少关注用户相互间及物品相互间的关系。图神经网络 GNN 的发展, 为人们进一步分析推荐系统实体及其相互间的关系提供了更好的表示方式, 近年来, 基于 GNN 推荐的相关研究越来越受到广大学者关注, 并取得了一定的成果。本文分析了 GNN 推荐的主要过程及其研究进展, 提出了 GNN 推荐当前存在的难点, 并指出未来 GNN 推荐的主要研究方向。

1 图神经网络

图神经网络是利用一定的方法对节点进行描述, 并经过不断的节点状态更新, 得到具有包含邻居节点信息和图形拓扑结构特点的状态^[7], 最终将这些节点通过特定方法进行输出, 得到需要的结果。式 (1) 表示节点状态的更新方式, 式 (2) 表示对更新完的节点状态进行输出:

$$\mathbf{h}_v = f(\mathbf{X}_v, \mathbf{X}_{\text{CO}[v]}, \mathbf{h}_{\text{ne}[v]}, \mathbf{X}_{\text{ne}[v]}) \quad (1)$$

$$\mathbf{o}_v = g(\mathbf{h}_v, \mathbf{X}_v) \quad (2)$$

式中: f 是局部转换函数; g 为局部输出函数; \mathbf{o} 表示输出结果; \mathbf{h}_v 表示节点当前的嵌入化结果; \mathbf{X}_v 代表节点 v 的特征; $\mathbf{X}_{\text{CO}[v]}$ 表示对应节点 v 的边的属性; $\mathbf{h}_{\text{ne}[v]}$ 表示节点 v 更新前的状态; $\mathbf{X}_{\text{ne}[v]}$ 表示节点 v 的邻居节点的特征。

式 (1)、(2) 均表示以一个节点为单位进行操作, 当对所有的节点进行上述运算的时候, 可以更加简洁的用式 (3)、式 (4) 进行描述:

$$\mathbf{H} = F(\mathbf{H}, \mathbf{X}) \quad (3)$$

$$\mathbf{O} = G(\mathbf{H}, \mathbf{X}_N) \quad (4)$$

式中: \mathbf{H} 表示所有节点的状态; \mathbf{O} 表示对所有的节点进行输出之后的结果; \mathbf{X} 表示边的特征; \mathbf{X}_N 分别表示所有节点的特征; $F(\cdot)$ 、 $G(\cdot)$ 分别表示全局转换函数与全局输出函数。由此可知, 当对所有节点的状态进行从 t 至 $t+1$ 轮更新时, 可以表示为式 (5):

$$\mathbf{H}^{t+1} = F(\mathbf{H}^t, \mathbf{X}) \quad (5)$$

GNN 有多种变体, 如根据训练方法的差异有 GraphSAGE^[8]、FastGCN^[9]、ControlVariate^[10]、Co-training GCN^[11] 和 Self-training GCN^[11] 等; 根据信息传播计算方式的不同有 Spectral Network^[12]、MoNet^[13]、DCNN^[14]、Gate Graph Neural Networks^[15]、Sentence LSTM^[16] 等。

图神经网络 GNN 按照其构成拓扑结构边的类型, 分为有向图 GNN 和无向图 GNN, 对于不同的边类型, 对应的局部转换函数 f 的计算方法也有所差异。

1) 有向图 GNN

实际上有向图比无向图可以反映更多的信息, 比如某种次序或者逻辑关系, 在文献 [17] 中, Kampffmeyer 等通过使用两个权值矩阵 \mathbf{W}_p 、 \mathbf{W}_c 来进行协同获得更加精确的处理结果, 如式 (6):

$$\mathbf{H}' = \sigma(\mathbf{D}_p^{-1} \mathbf{A}_p \cdot \sigma(\mathbf{D}_c^{-1} \mathbf{A}_c \mathbf{H}^{-1} \mathbf{W}_c) \mathbf{W}_p) \quad (6)$$

式中: p 代表父节点 (指向它的节点); c 代表子节点 (它指向的节点); $\mathbf{D}_p^{-1} \mathbf{A}_p$ 、 $\mathbf{D}_c^{-1} \mathbf{A}_c$ 分别代表对父节点和子节点邻接矩阵的标准化矩阵; σ 是非线性激活函数。

2) 无向图 GNN

没有次序关系的节点就构成无向图, 图神经网络 GNN 对于无向图的处理只需要将其视为两个有向图的叠加即可。实际上, 无向图的计算相比有向图更为简单, 对其进行更新时, 只需一个权重矩阵 \mathbf{W} 正常经过加权, 通过激活函数更新即可。

本文将图神经网络 GNN 按照构成它的节点是否表示相同的实体, 分为单元图 GNN, 二元图 GNN 与多元图 GNN。单元图 GNN 中节点仅表示 1 种实体类型, 即节点的类型唯一, 在推荐系统中常常体现在用户社交关系图, 如根据用户的朋友关系构建对应的 GNN 模型; 二元图 GNN 中的节点所表示的实体类型为 2 种, 诸如用户-物品购买关系图, 用户-项目标签图等; 图中节点所表示实体类型超过 2 种的为多元图 GNN, 在实体种类及其关系较复杂的时候可以使用这种类型的 GNN 模型。

2 推荐系统

推荐系统就是通过分析用户的历史行为对用户的兴趣建模, 从而主动给用户推荐能够满足其兴趣和需求的物品 (或信息)。常见的推荐模型有基于用户行为的推荐, 如协同过滤 (collaborative filtering, CF)^[18]、隐语义模型 (latent factor model, LFM)^[19]; 基于内容的推荐, 如标签推荐^[20]; 混合推荐^[21], 即通过多种不同的推荐模型相互协作, 如在推荐系统的不同时期分别使用不同的模型或者使用多个模型的推荐结果再进行决策。协同过滤分为基于用户的协同过滤^[22]和基于物品的协同过滤^[23], 基于用户行为的推荐算法往往存在着冷启动问题^[24]和数据稀疏性问题^[25]。

随着深度学习的发展, 基于深度学习的推荐研究也很多, Elkahky 等^[26]提出了一种多视图深度神经网络模型 (Multi-View deep neural network, Multi-ViewDNN), 通过用户和物品 (项目) 两种信

息实体的语义匹配来实现用户的物品推荐; Zheng 等^[27]提出了一种深度协作神经网络模型(deep cooperative neural network, DeepCoNN), 利用两个并行的神经网络模型学习用户和项目的隐特征, 然后在两个神经网络上构建一个交互层来预测用户对项目的评分; Liu 等^[28]研究了位置社交网络中的行为预测问题, 通过利用循环神经网络抓住序列行为之间的依赖关系, 从而基于用户的历史行为序列, 帮助预测下一时刻的行为。

3 GNN 推荐主要过程

一般说来, GNN 推荐主要过程分为 3 个步骤: 1) 根据推荐系统实体及其相互关系构建对应的 GNN 模型。主要考虑如何将推荐系统中不同实体映射为 GNN 中图节点, 不同实体间的联系映射为 GNN 图中对应的边, 如何用的神经网络函数去拟合 GNN 中不同的节点。在 GNN 推荐模型中, 不论是用户节点还是物品节点, 都可以根据相关信息做嵌入化处理, 得到对应节点的向量化描述。如对一个物品节点, 可以根据不同用户对其评论或评分信息做嵌入化处理; 对一个用户节点, 可以根据其社交关系或物品购买记录做嵌入化处理。边作为反映 GNN 推荐模型中实体之间的联系, 通常有两种处理方式: 一是根据联系的类型及强弱程度不同, 进行嵌入化操作; 二是对边不进行任何处理, 仅仅当成 GNN 推荐模型中, 用于信息传播算法作用的媒介。2) 决定 GNN 模型的信息传播与更新方法。GNN 模型的变体很多, 不同变体之间, 信息的聚集和更新方式也不同。文献[29]指出了在不同的 GNN 变体中, 出现的相关聚集过程。如 GCN 模型, 通常选择 mean 聚集函数; GraphSAGE 模型, 通常采用 max 函数; GIN 模型, 通常使用的 sum 函数等。不同的聚集函数适合于不同的图结构, mean 函数可以反映图中节点的分布情况, max 函数可以反映主要特征, sum 可以考虑到比较全面的图形结构特征等。信息更新过程, 主要是将信息聚集之后的结果, 与中心节点进行特定的运算, 并作为下一层节点的初始状态。这方面, 不同的模型差异也很大, 如 GCN 模型进行信息更新时, 不会考虑中心节点的信息; GraphSAGE 则会将聚集结果和中心节点的向量进行连接操作; GIN 模型会将两者进行直接相加等。充分考虑推荐系统实体间关系, 及对应 GNN 模型的结构特征, 选择最合适的信息传播算法, 进行信息更新, 往往可以取得更

好的推荐效果。3) 提取出更新后的节点(边、子图)特征, 并用相关算法实现推荐。将 GNN 模型完成更新之后的节点取出, 作为该节点对应的实体特征, 采用相关算法进行推荐。如文献[30]中, 使用协同过滤推荐的方法; 文献[31]中, 将 GNN 提取的物品特征和节点特征, 通过 MLP 进行评分预测推荐; 文献[32]中, 通过两种不同的神经网络, 进行试验结果预测与推荐。

4 GNN 推荐相关研究

由于有向图 GNN 推荐时, 需要两个权值矩阵来实现聚集与更新操作, 相对无向图 GNN 推荐, 只需一个权重矩阵来说, 就变得更为复杂。因此, 在现有图神经网络 GNN 推荐的相关研究中, 主要集中在无向图 GNN 推荐方面。以下重点从无向单元图 GNN 推荐、无向二元图 GNN 推荐以及无向多元图 GNN 推荐 3 个方面探讨 GNN 推荐相关研究。

4.1 无向单元图 GNN 推荐

这种 GNN 推荐模型节点类型单一, 操作比较简单。主要是将推荐系统中的实体映射到一个无向单元图 GNN, 图中所有节点都表示为一种实体类型。如 Cui^[33]等将 GNN 用于服装推荐, 服装的类别作为节点, 各个类别之间的关系作为边, 各类服装放入对应的节点之中形成子图, 构成了一个无向单元图 GNN 推荐模型, 推荐与指定服装匹配度较高的目标服装实现套装搭配, 取得了较好的效果。

文献[30]中, 考虑到用户的行为往往受到自己当前兴趣、朋友兴趣等多种因素影响, 通过构建用户—朋友的无向单元图 GNN, 分别对用户节点、朋友节点进行描述, 并使用相关传播算法更新用户信息节点, 最后和物品嵌入化得到的结果进行运算, 输出用户对物品可能的喜爱程度, 如图 1 所示。具体步骤如下:

1) 将用户近期的浏览内容, 通过循环神经网络(RNN)得到表示用户兴趣的 h_n , 并用 h_n 描述用户的节点信息;

2) 将朋友的近期行为记录, 通过 RNN 得到 s_k^s , 用来表示该朋友的短期兴趣;

3) 将朋友长期行为向量化, 得 s_k^l , 见式(7):

$$s_k^l = W_u[k, :] \quad (7)$$

式中: $W_u[k, :]$ 表示对用户 u 从第 k 次记录一直取到最新的记录, 进行向量化处理。

4) 将上述两个向量联合起来, 构成代表目标

节点用户的嵌入化向量 s_k , 见式 (8):

$$s_k = \text{ReLU}(W_1[s_k^s, s_k^l]) \quad (8)$$

式中: ReLU 为非线性激活函数, $W_1[s_k^s, s_k^l]$ 表示对向量 s_k^s, s_k^l 进行连接处理后使用 W_1 进行偏置。

5) 目标节点的初始状态选取为 h_u , 邻居节点的初始状态选取为 s_k , 利用构建完成的用户与好友之间关系的拓扑图, 通过计算每层的目标节点和邻居节点的相似度关系, 提取出注意力系数 α_{uk} , 见式 (9):

$$\alpha_{uk}^{(l)} = \frac{\exp(f(h_u^{(l)}, h_k^{(l)}))}{\sum_{j \in N(u) \cup \{u\}} \alpha_{uj}^{(l)} h_k^{(l)}} \quad (9)$$

式中: l 表示当前的更新次数, $f(h_u^l, h_k^l)$ 表示目标

节点与他的第 k 个邻居的相似度计算结果。

6) 将所有的邻居节点信息通过注意力系数相加后, 得到表示目标节点所有邻居节点的信息向量 \hat{h}_u , 见式 (10), 再将 \hat{h}_u 经过偏置加权之后作为节点的下一轮状态进行更新, 见式 (11):

$$\hat{h}_u = \sum_{k \in N(u) \cup \{u\}} \alpha_{uk}^{(l)} h_k^{(l)} \quad (10)$$

$$h_u^{(l+1)} = \text{ReLU}(W^{(l)} \hat{h}_u^{(l)}) \quad (11)$$

式中: W 是可训练的权值矩阵, $l, l+1$ 表示当前更新的次数。

7) 将节点的特征与嵌入化得到的物品特征进行相似度运算, 通过 softmax 进行输出, 此时输出的结果就是某用户对物品可能的感兴趣程度。

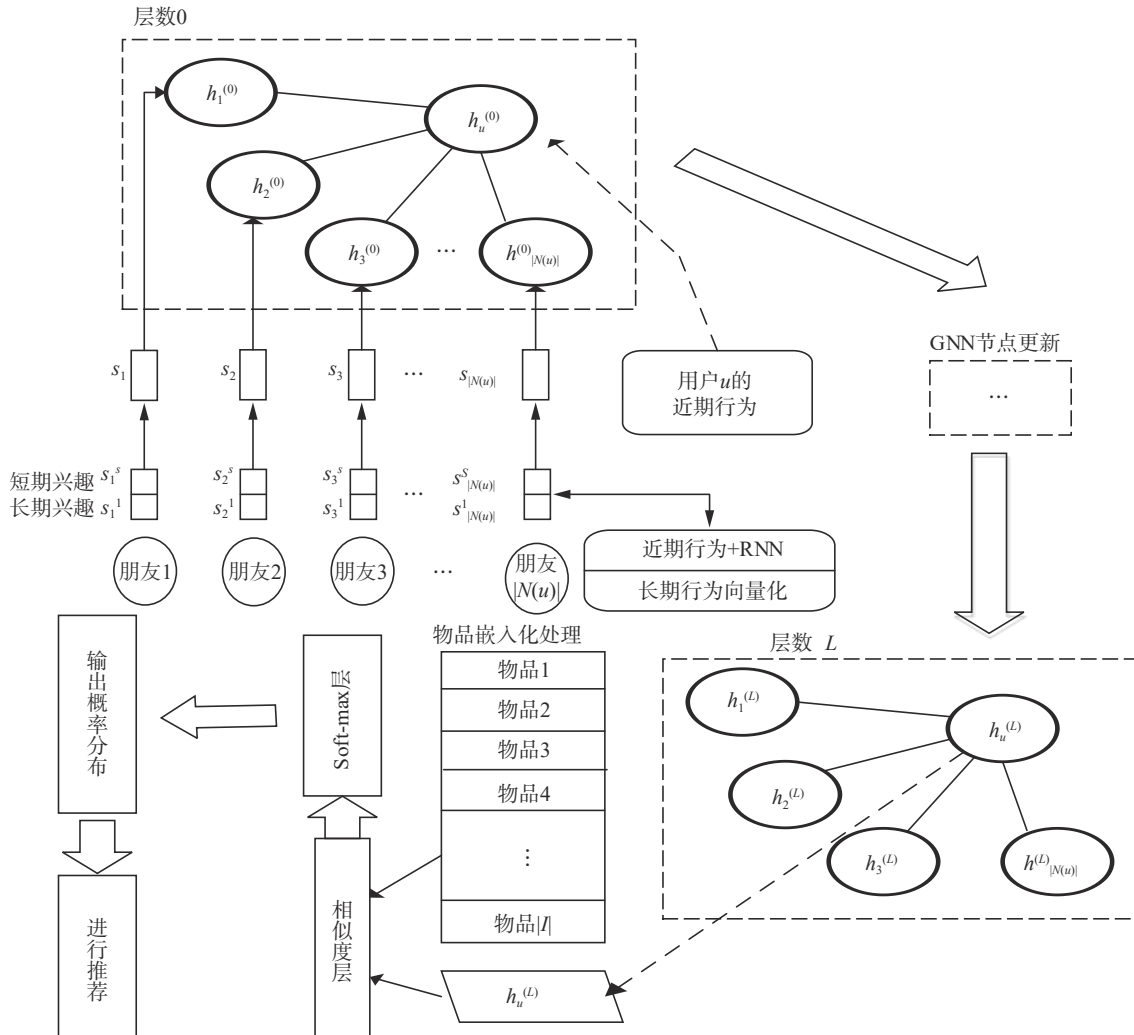


图 1 无向单元图 GNN 推荐模型

Fig. 1 Single-unit undirected graph GNN model

该模型充分考虑到社会关系与用户本人的当前兴趣, 但对物品特征的提取比较粗糙, 没有考虑到用户与物品之间的互动关系, 同时不同的朋友对用户的影响程度也是不一样的, 此模型也没

有加以区分。

4.2 无向二元图 GNN 推荐

在实际推荐中, 最常见的关系就是用户与项目(物品)的交互记录, 这两种实体关系就可以构

成一个完整的二元拓扑图。Fan等^[31]通过构建用户-物品关系二元图(用户购买的所有物品)、用户-朋友关系单元图(用户的社交关系)、物品-用户关系二元图(物品被用户购买的记录)这3个无向图

用于对物品进行评分预测。同样考虑到对用户可能的行为,不仅与他自己曾经购买的物品有关,也和他的社交关系有关,该文献提出了一种新型图神经网络 GNN 推荐系统框架 GraphRec, 如图 2。

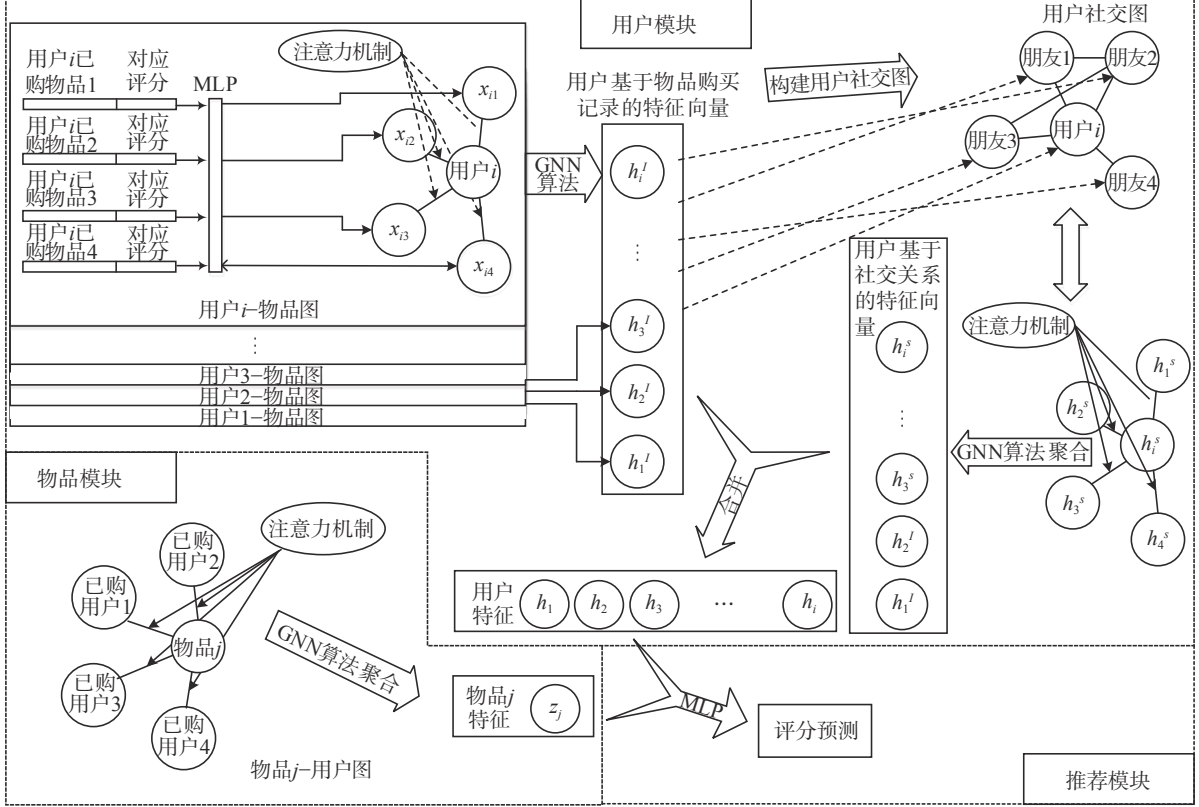


图 2 基于社交关系的 GraphRec 模型

Fig. 2 Social-based recommendation model-GraphRec

该框架主要由 3 个模块组成,分别是用户-物品关系和用户-朋友关系构成用户模块,物品-用户关系构成物品模块和由一个多层感知器(MLP)所构成的推荐模块。通过信息传播更新算法,对这 3 个图中的节点进行计算,根据结果进行评分预测。具体过程如下:

首先根据用户购买不同物品行为以及用户之间的社交关系构建相应的户-物品关系图和用户-朋友关系图,再在图上分别对用户 i 和物品 j 进行嵌入化处理,得到代表用户信息的 p_i 向量,代表物品信息的 q_j 向量。同时每一个特定的用户对于他所购买的物品对应的评分,用独热编码向量化成 e_r 。

用户模块分为两部分,第 1 部分是对某用户的购买行为进行特征刻画,第 2 部分是对该用户的社交关系进行刻画,将两个特征结合起来通过一个多层感知器,就得到代表该用户的特征。用户模块的实现方法如下:

1) 对所提取的物品 q_a 向量和对应的评分

e_r 向量连接处理,通过一个多层感知器得到 x_{ia} 向量(其中 a 表示用户购买的相关物品, i 表示某一个用户),见式 (12):

$$x_{ia} = g_r([q_a \oplus e_r]) \quad (12)$$

2) 将计算出的 x_{ia} 和用户 p_i 通过一个两层神经网络得到该用户对物品的注意力系数 α^* , 见式 (13):

$$\alpha_{ia}^* = W_2^T \cdot \sigma(W_1 \cdot [x_{ia} \oplus p_i] + b_1) + b_2 \quad (13)$$

其中, w_1, w_2 是可训练的参数, b_1, b_2 是偏置项, $\sigma(\cdot)$ 是非线性函数。

3) 对 α^* 进行 softmax 处理,得到 α_{ia} 作为最终的注意力系数,见式 (14):

$$\alpha_{ia} = \frac{\exp(\alpha_{ia}^*)}{\sum_{a \in C(i)} \exp(\alpha_{ia}^*)} \quad (14)$$

4) 用 α_{ia} 对之前得到的 x_{ia} 向量进行加权,并且通过一个多层感知器(GNN 的简单聚合函数)得到用户-物品特征 h_i , 见式 (15):

$$h_i^l = \sigma(w \cdot \left\{ \sum_{a \in C(i)} \alpha_{ia} x_{ia} \right\} + b) \quad (15)$$

5) 用同样的方法对该用户的好友进行处理, 得到对应 h_{io} (o 表示用户好友集 i 的一个好友);

6) 使用 2) 中的方法对 h_{io} 和用户向量 p_i 通过一个两层神经网络提取注意力系数 β^* , 见式 (16), 并且通过 softmax 进行处理可以得到 β , 见式 (17):

$$\beta_{io}^* = W_2^T \cdot \sigma(W_1 \cdot [h_{io}^l \oplus p_i] + b_1) + b_2 \quad (16)$$

$$\beta_{io} = \frac{\exp(\beta_{io}^*)}{\sum_{o \in N(i)} \exp(\beta_{io}^*)} \quad (17)$$

7) 将得到的 β 对 h_i 进行加权并通过一个多层感知器, 最后得到具有刻画用户 i 社交关系的向量 h_i^s , 见式 (18):

$$h_i^s = \sigma(w \cdot \left\{ \sum_{o \in N(i)} \beta_{io} h_o^l \right\} + b) \quad (18)$$

其中, b 是偏置项, $\sigma(\cdot)$ 是非线性函数。将 h_s 和 h_i 连接起来通过多层感知器得到代表用户的向量 h 。

物品模块使用类似的方法对用户向量 p_i 和评分向量 e_r 进行连接, 获得 z_{jt} 向量 (j 表示物品, t 表示某用户), 然后对 z 向量使用上述步骤 2) 中的方法提取出注意力系数, 并且对周围的用户向量使用该系数加权后, 结果通过一个多层感知器就得到代表物品 j 特征的向量 z_j 。

预测模块是将上述两个模块得到的 h 向量, 通过一个多层感知器, 最终输出预测评分。

损失函数定义为均方误差, 见式 (19):

$$\text{Loss} = \frac{1}{2|O|} \sum_{i,j \in O} (r'_{ij} - r_{ij})^2 \quad (19)$$

其中 i, j 表示用户 i 、物品 j , r 表示实际评分, r' 表示预测评分, $|O|$ 表示所有评分记录的个数。

该方法将用户的社交关系融入神经网络, 同时对于用户和物品的特征刻画都相互关联起来, 并且引入了注意力机制, 取得了不错的效果, 但是在用户模块中两个 GNN 模型无法并行计算, 增加了模型的时间复杂度。

在文献 [32] 中, Rex 构建了用户-标签无向二元图, 通过 PinSage 算法 (随机游走和图卷积结合) 成功应用于超大规模的网页内容推荐。Wang 等 [34] 构建用户-物品图, 不同之处在于 GNN 算法中使用目标节点及其邻居节点的点积进行更新, 并且通过协同过滤的方法进行推荐。文献 [35] 为了解决推荐系统的冷启动问题, 利用知识图谱为推荐提供更多的辅助信息, 针对不同用户将知识图谱转换为多个用户的个性加权图, 并且通过 GCN 学习出每个物品的表示向量, 然后根据用户特征向量计算出用户与物品有交互的概率。

4.3 无向多元图 GNN 推荐

多元图 GNN 会大大增加模型的复杂度, 同时计算复杂性和刻画节点的难度也随之增加, 但多元图 GNN 往往更加贴近现实, 可以得到更加客观的结果。文献 [36] 中, 考虑到在一家医院的病患进行就诊时, 需要进行项目测试以及根据其具体情况开出不同的药方, 一般基于经验进行诊断, 可能会出现误差, 而且数据难以全面收集。Mao 将图神经网络 GNN 推荐运用在诊疗研究中, 如图 3 所示。通过一个无向多元图 GNN 构建病患、医疗科室、药方以及测试结果之间的关系, 其中医院的各个科室作为核心节点, 进行几轮信息的迭代计算之后, 根据计算出的节点信息通过两种不同的神经网络, 推荐可能的药方, 以及某项诊疗测试的可能结果。

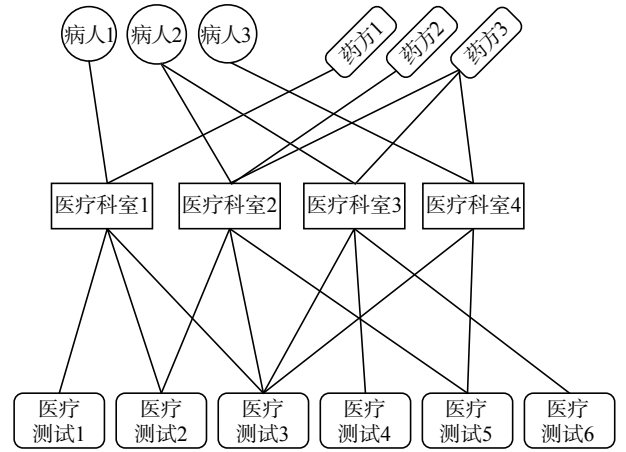


图 3 患者医疗诊断关系和医疗测试结果

Fig. 3 The graph of patients' consults relation and laboratory testing results

该研究中, 节点之间的关系通过 3 个邻接矩阵进行表示, 如图 4, 分别是 $A_{E \times P}$ 、 $A_{E \times M}$ 、 $A_{E \times L}$, 其中 E 为医疗科室, P 为患者, M 为药方, L 为医疗测试结果; $A_{E \times P}$ 表示医疗科室与病人之间的关系, 如果病人在该医疗科室进行过就诊, 则为 1, 否则为 0。假设一个病人可以对应多个医疗科室, 但是一个医疗科室只能对应一个病人, 也就是说, 一行最多只有一个 1。 $A_{E \times M}$ 表示医疗科室与药方之间的关系, 如果该科室可以开出对应的药方, 则该值为 1, 反之为 0。 $A_{E \times L}$ 表示医疗科室和医疗测试结果之间的关系, 如果某个医疗科室有该实验测试则该两点之间有权重, 选择的值是平时测试的平均值, 为了解决一些测试结果是 0 的项目进行区别, 作者额外增加了一个 $M_{E \times L}$ 矩阵, 对于存在的项目则对应数字为 1, 否则为 0。最后一个矩阵表示医疗科室和药方之间的关系, 如果该医疗科室有对应的药方则该点为 1, 否则为 0。

	P_1	P_2	P_3		M_1	M_2	M_3		L_1	L_2	L_3	L_1	L_2	L_3		L_1	L_2	L_3
E_1	1	0	0	E_1	1	0	0	E_1	0.3	0.0	0.1	0	0	0	E_1	1	1	1
E_2	0	1	0	E_2	0	1	1	E_2	0	0.1	0	0	0.5	0	E_2	0	1	0
E_3	0	1	0	E_3	0	0	0	E_3	0	0	0.1	0.8	0	0	E_3	0	0	1
E_4	0	0	1	E_4	0	0	0	E_4	0	0	0.0	0	0	0.6	E_4	0	0	1
$A_{E \times P}$				$A_{E \times M}$				$A_{E \times L}$				$M_{E \times L}$						

图4 病患、医疗科室、药方以及医疗测试结果关系矩阵

Fig. 4 Relation matrix between patients encounters medications and laboratory test data

在式(20)中, f_1, f_2 分别表示两种激活函数, $N_{(v)}$ 表示 v 节点的所有邻居, k 表示第 k 轮运算。

$$H^{(k+1)}(v) = f_1^{W_1^{(k)}} \left(H^{(k)}(v), f_2^{W_2^{(k)}} \left(\{H^{(k)}(w) | w \in N(v)\} \right) \right) \quad (20)$$

Mao 对激活函数 ϕ 使用了最简单的求和函数^[36], 如式(21)所示, 对节点类型为 i 的节点的第 k 轮计算, 其中 h_j 表示他的连接节点。

$$H_i^{(k+1)} = \phi \left(\sum_{j=0}^n A_{ij} \cdot H_j^{(k)} \cdot W_j^{(k)} \right) \quad (21)$$

经过 k 轮计算之后, 将计算的节点结果输入两种神经网络, 分别进行药方和医疗测试结果的推荐。

在式(22)中, 损失函数表示为两项之和, 第1项 L_m 为药方预测的误差, 见式(23), 第2项 L_L 表示使用交叉熵的方式表示医疗测试结果的预测误差, 见式(24), λ 是系数通过这种引入的混合误差, 很有效地避免了过拟合问题。

$$L = L_m(P, A_{E \times M}) + \lambda L_L(V, A_{E \times L}) \quad (22)$$

$$L_m(P, A) = -\frac{1}{N_E} \sum_i \sum_j \frac{N_n}{N_p} a_{ij} \log p_{ij} + (1 - a_{ij}) \log (1 - p_{ij}) \quad (23)$$

$$L_L(V, A) = -\frac{1}{N_E} \sum_i \sum_j m_{ij} (v_{ij} - a_{ij})^2 \quad (24)$$

该模型较单元图与二元图神经网络, 结构更加复杂, 但能更有效地对不同类型的数据信息进行充分使用。同时, 通过将模型中的对象进行节点再划分, 成功处理了一些实验过程中的数据缺失值问题; 使用新的正则化规则, 避免了过拟合情况, 取得了不错效果。不足之处在于实验相关数据仅采用一家医院, 对于规模更大的数据集, GNN 图形结构将更复杂, 实际应用有一定困难。

5 现有 GNN 推荐研究难点

5.1 多源多模态信息的 GNN 推荐问题

在一个系统中, 数据来源与数据结构往往呈现多源异构性。以电子商务系统为例, 数据来源

有商家提供的商品描述与分类信息, 消费者与系统交互的点击、浏览、收藏、交易信息, 消费者的评分及评论信息等; 数据形式有文本信息, 图像信息、数值评分与标签信息等, 甚至还有语音及视频信息等。现有推荐模型对于这些多源多模态数据很难同时进行有效处理, 导致数据利用不充分或数据对象之间的关系表示不明显。同样的, 现有 GNN 推荐模型, 对节点向量化的方式比较单一, 如一般只能对用户留下的评论数据或者评分加以处理, 难以将多源多模态信息进行融合表示以达到最佳效果。如何充分集成使用多源多模态数据信息是 GNN 推荐有待解决的一个问题。

5.2 GNN 的有向图推荐问题

目前 GNN 推荐模型主要是无向图 GNN 推荐, 其中无向图的边用来反映节点之间的互动关系。但是有向图比无向图可以反映更多的信息, 比如某种次序或者逻辑关系, 有向图的推荐结果往往可以取得更好的效果。推荐系统中也存在着对应的有向关系, 如信任关系、关注关系等。与无向图 GNN 推荐相比, 基于有向图 GNN 的推荐更具有挑战性, 也更为复杂。

5.3 GNN 的动态推荐问题

对于数据规模往往很大的图形结构, 很难进行实时的图模型更新, 而且图神经网络 GNN 迭代过程中往往需要根据整个图的信息向量化才能完成计算, 对于不断更新的数据每次进行重新计算, 需要花费很高的成本, 如何在较低成本下完成动态的 GNN 更新, 现有 GNN 推荐研究并没有考虑。如文献[30]中, 用户的社交情况可能会实时发生变化, 新朋友可能会对用户的兴趣造成很大影响, 如果可以实时更新用户的社交图谱, 就可以完成对 GNN 动态推荐。

5.4 GNN 网络更新层数问题

标准的深度神经网络推荐研究中, 网络层数可以达到成百上千层, 由于更深的结构, 可以显著提高网络的表达能力。在 GNN 推荐过程中, 一般层数仅仅为个位数。因为堆叠多个 GNN 层

将会导致过度平滑,所有顶点将收敛到相同的值。如何选择 GNN 更新的层数以及其他相关的参数,以缓解堆叠多层 GNN 推荐造成的平滑问题,是 GNN 推荐难点所在。

5.5 数据稀疏问题

GNN 推荐中,通过构建节点和边,将各个对象用边“连接”起来,在缓解数据稀疏性方面相比传统的推荐模型会有提升。但 GNN 推荐模型在面临数据稀疏的时候,往往会出现两个问题:1) 如果节点对应的对象没有任何相关记录时,相关推荐算法将难以实现,如文献[36]中,如果有新来的病人或者病人的就诊记录很少,如仅为一条记录, GNN 推荐模型会难以得到可以代表用户特征的最终结果;2) 节点属性维度缺失问题。一个向量化的多维节点,每一个维度都反映一个对应的属性,如果该属性的值是未知的,在向量化的过程中就无法写入相应的数值,将会影响到推荐的效果。在文献[36]中就出现了病人未参加部分医疗测试的情况,作者通过重新构建测试结果,解决这个问题,但模型的复杂度也随之增加。

6 GNN 推荐未来研究方向

1) 基于上下文的 GNN 推荐

用户的兴趣和行为往往是随着时间、空间等因素动态发生变化的,在目前的推荐系统中,很多学者对基于时间上下文的个性化推荐^[37],基于情景上下文的推荐^[38]都有一定的研究,通过联系上下文往往会获得更加精确的推荐效果。现有 GNN 推荐研究不论是构建用户-用户关系图还是用户-物品(项目)关系图,都只是单纯的将用户(物品)相关信息进行处理。时间与空间上下文信息,都没有考虑在内。如在文献[31]中,作者没有考虑到用户兴趣会随着时间变化对浏览的记录产生的影响,文献[36]中,没有考虑到患者就诊记录的时间问题,但在不同时期同一个患者情况会存在很大差异。如何在 GNN 推荐模型中引入时间、内容以及位置等不同的上下文信息,以提高推荐性能是值得研究的重要课题之一。

2) 基于 GNN 的跨领域推荐

跨领域推荐^[39]对解决推荐系统的冷启动和数据稀疏性问题有很大的帮助,通过从辅助域到目标域的推荐可以取得良好的推荐效果。在传统的推荐研究中,跨领域推荐取得了一定成效,如基于标签的跨领域推荐^[40],基于共享知识的跨领域

推荐^[41]等。现有 GNN 推荐研究中,往往只是对“一张”图进行“加工”,停留在一个单一的层面上,如何从一张图过渡到另一张图,实现不同 GNN 图模型间信息的有效迁移,是一个值得研究的课题。如文献[36]中,作者仅仅考虑到病人在一家医院的就诊情况,如果考虑到病人的更多信息,如生活习惯、职业、年龄等将这些作为辅助域的信息充分利用,就可以获得更加精确的推荐效果。

3) 基于 GNN 的群组推荐

推荐系统的应用越来越广泛,推荐的对象从单个用户的个性化推荐发展为面向多个用户的群组推荐。现有 GNN 推荐研究中,主要是个性化推荐,很少涉及 GNN 群组推荐,如何通过 GNN 进行群组推荐未来值得研究的问题。

4) 基于 GNN 的套餐推荐

当前的推荐系统大多推荐内容单一,仅能满足用户的部分需求,现实生活中,用户需求往往是多方面的。如对旅游的用户进行推荐时,如果能通过 GNN 的高度归纳性将路线规划、住宿环境、交通方式、特产商铺、消费价格等多个因素进行综合考虑,为用户推荐最适合的套餐服务,这将会给用户带来更好的旅游体验。

5) GNN 推荐与其他推荐方法融合

深度学习发展迅猛,在不同领域的研究都取得了突破性进展,而 GNN 推荐实质上只是获取节点的状态之后再用函数进行输出,可以考虑融合其他推荐算法或者传统深度学习技术,减少模型的时间复杂度,同时提高推荐的效果。

6) GNN 推荐的可解释性研究

GNN 的应用范围很广泛,比如文献[42]利用 GNN 模拟各个物理系统间的关系,进行演绎来预测整个系统的下一个状态,这说明 GNN 具有一定的因果推理能力。现有 GNN 推荐研究中,尽管取得了不错的推荐效果,但是推荐结果往往很难给出一个令用户信服的科学解释。如在文献[31]中,作者使用多个 GNN 模型得出特征后,通过 MLP 进行评分预测,仍然缺乏一定的可解释性。

7 结束语

图神经网络(GNN)推荐方法可以通过节点、边及对应的拓扑结构直接反映推荐系统中实体及其相互间关系,直接在图上对相关信息进行不断更新计算,具有很强的归纳性,有效避免了传统推荐方法中信息利用不完全带来诸多问题,越来

越受到学术界和产业界的关注。本文深入分析了 GNN 推荐及其过程,对现有 GNN 推荐相关研究从无向单元图推荐、无向二元图推荐、无向多元图推荐 3 个方面进行详细探讨,对各自的优点与不足进行了分析,指出了 GNN 推荐面临的难点及未来的研究方向,对基于图的深度推荐系统进一步研究具有一定的借鉴意义。

参考文献:

- [1] ZHOU J, CUI G, ZHANG Z, et al. Graph neural networks: a review of methods and applications[J]. arXiv: 1812.08434, 2018.
- [2] BASTINGS J, TITOV I, AZIZ W, et al. Graph convolutional encoders for syntax-aware neural machine translation[C]//Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen, Denmark, 2017: 1957–1967.
- [3] HENAFF M, BRUNA J, LECUN Y. Deep convolutional networks on graph-structured data[J]. arXiv:1506.05163, 2015.
- [4] ZHANG Yuhao, QI Peng, MANNING C D. Graph convolution over pruned dependency trees improves relation extraction[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium, 2018: 2205–2215.
- [5] WANG Xiaolong, YE Yufei, GUPTA A. Zero-shot recognition via semantic embeddings and knowledge graphs[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 6857–6866.
- [6] RHEE S, SEO S, KIM S. Hybrid approach of relation network and localized graph convolutional filtering for breast cancer subtype classification[J]. arXiv: 1711.05859, 2017.
- [7] KAWAMOTO T, TSUBAKI M, OBUCHI T. Mean-field theory of graph neural networks in graph partitioning[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Red Hook, USA, 2018: 4361–4371.
- [8] HAMILTON W, YING Zhotao, LESKOVEC J. Inductive representation learning on large graphs[C]//Advances in Neural Information Processing Systems. Long Beach, US, 2017: 1024–1034.
- [9] CHEN J, MA T, XIAO C. Fastgcn: fast learning with graph convolutional networks via importance sampling[J]. arXiv:1801.10247, 2018.
- [10] CHEN Jianfei, ZHU Jun, SONG Le. Stochastic training of graph convolutional networks with variance reduction[C]//Proceedings of the 35th International Conference on Machine Learning. Stockholm, Sweden, 2018: 941–949.
- [11] LI Q, HAN Z, WU X M. Deeper insights into graph convolutional networks for semi-supervised learning[C]//Thirty-Second AAAI Conference on Artificial Intelligence. New Orleans, USA, 2018.
- [12] BRUNA J, ZAREMBA W, SZLAM A, et al. Spectral networks and locally connected networks on graphs[C]//International Conference on Learning Representations (ICLR2014). Banff, Canada, 2014: 1–14.
- [13] MONTI F, BOSCAINI D, MASCI J, et al. Geometric deep learning on graphs and manifolds using mixture model CNNs[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 5115–5124.
- [14] ATWOOD J, TOWSLEY D. Diffusion-convolutional neural networks[C]//Advances in Neural Information Processing Systems. Barcelona, Spain, 2016: 1993–2001.
- [15] LI Y, ZEMEL R, BROCKSCHMIDT M, et al. Gated graph sequence neural networks [J]. arXiv:1511.05493, 2015.
- [16] ZHANG Yue, LIU Qi, SONG Linfeng. Sentence-state LSTM for text representation[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne, Australia, 2018: 317–327.
- [17] KAMPFFMEYER M, CHEN Yinbo, LIANG Xiaodan, et al. Rethinking knowledge graph propagation for zero-shot learning[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 11487–11496.
- [18] 黄璐, 林川杰, 何军, 等. 融合主题模型和协同过滤的多样化移动应用推荐 [J]. 软件学报, 2017, 28(3): 708–720.
- [19] HUANG Lu, LIN Chuanjie, HE Jun, et al. Diversified mobile app recommendation combining topic model and collaborative filtering[J]. Journal of software, 2017, 28(3): 708–720.
- [19] 胡堰, 彭启民, 胡晓惠. 一种基于隐语义概率模型的个性化 Web 服务推荐方法 [J]. 计算机研究与发展, 2014, 51(8): 1781–1793.
- [20] HU Yan, PENG Qimin, HU Xiaohui. A personalized Web service recommendation method based on latent semantic probabilistic model[J]. Journal of computer research and development, 2014, 51(8): 1781–1793.
- [20] 刘建勋, 石敏, 周栋, 等. 基于主题模型的 Mashup 标签推荐方法 [J]. 计算机学报, 2017, 40(2): 520–534.

- LIU Jianxun, SHI Min, ZHOU Dong, et al. Topic model based tag recommendation method for Mashups[J]. *Chinese journal of computers*, 2017, 40(2): 520–534.
- [21] 曹俊豪, 李泽河, 江龙, 等. 一种融合协同过滤和用户属性过滤的混合推荐算法[J]. *电子设计工程*, 2018, 26(9): 60–63.
- CAO Junhao, LI Zehe, JIANG Long, et al. A hybrid recommendation algorithm based on collaborative filtering and user attribute filtering[J]. *Electronic design engineering*, 2018, 26(9): 60–63.
- [22] 张双庆. 一种基于用户的协同过滤推荐算法[J]. *电脑知识与技术*, 2019, 15(1): 19–21.
- ZHANG Shuangqing. User-based collaborative filtering recommendation algorithm[J]. *Computer knowledge and technology*, 2019, 15(1): 19–21.
- [23] 邓园园, 吴美香, 潘家辉. 基于物品的改进协同过滤算法及应用[J]. *计算机系统应用*, 2019, 28(1): 182–187.
- DENG Yuanyuan, WU Meixiang, PAN Jiahui. Improved item-based collaborative filtering algorithm and its application[J]. *Computer systems & applications*, 2019, 28(1): 182–187.
- [24] 高玉凯, 王新华, 郭磊, 等. 一种基于协同矩阵分解的用户冷启动推荐算法[J]. *计算机研究与发展*, 2017, 54(8): 1813–1823.
- GAO Yukai, WANG Xinhua, GUO Lei, et al. Learning to recommend with collaborative matrix factorization for new users[J]. *Journal of computer research and development*, 2017, 54(8): 1813–1823.
- [25] 王伟, 陈伟, 祝效国, 等. 众筹项目的个性化推荐: 面向稀疏数据的二分图模型[J]. *系统工程理论与实践*, 2017, 37(4): 1011–1023.
- WANG Wei, CHEN Wei, ZHU K, et al. Personalized recommendation of crowd-funding campaigns: a bipartite graph approach for sparse data[J]. *Systems engineering-theory & practice*, 2017, 37(4): 1011–1023.
- [26] ELKAHKY A M, SONG Yang, HE Xiaodong. A multi-view deep learning approach for cross domain user modeling in recommendation systems[C]//Proceedings of the 24th International Conference on World Wide Web. Florence, Italy, 2015: 278–288.
- [27] ZHENG Lei, NOROOZI V, YU P S. Joint deep modeling of users and items using reviews for recommendation[C]//Proceedings of the Tenth ACM International Conference on Web Search and Data Mining. New York, USA, 2017: 425–434.
- [28] LIU Q, WU S, WANG L, et al. Predicting the next location: a recurrent model with spatial and temporal contexts[C]//Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. Arizona, USA, 2016: 194–200.
- [29] XU,KEYULU,et al.How powerful are graph neural networks[J]. arXiv:1810.00826, 2018.
- [30] SONG Weiping, XIAO Zhiping, WANG Yifan, et al. Session-based social recommendation via dynamic graph attention networks[C]//Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining. New York, United States, 2019: 555–563.
- [31] FAN Wenqi, MA Yao, LI Qing, et al. Graph neural networks for social recommendation[C]//The World Wide Web Conference. New York, USA, 2019: 417–426.
- [32] YING R, HE Ruining, CHEN Kaifeng, et al. Graph convolutional neural networks for web-scale recommender systems[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York, USA, 2018: 974–983.
- [33] CUI Zeyu, LI Zekun, WU Shu, et al. Dressing as a whole: outfit compatibility learning based on node-wise graph neural networks[C]//The World Wide Web Conference. New York, USA, 2019: 307–317.
- [34] WANG X, HE X, WANG M, et al. Neural graph collaborative filtering[J]. arXiv:1905.08108, 2019.
- [35] WANG Hongwei, ZHAO Miao, XIE Xing, et al. Knowledge graph convolutional networks for recommender systems[C]//The World Wide Web Conference. San Francisco, USA, 2019: 3307–3313.
- [36] MAO C, YAO L, LUO Y. MedGCN: Graph convolutional networks for multiple medical tasks[J]. arXiv: 1904.00326, 2019.
- [37] 刘云, 王颖, 亓国涛, 等. 时间上下文的协同过滤 Top-N 推荐算法[J]. *计算机技术与发展*, 2017, 27(7): 79–82.
- LIU Yun, WANG Ying, QI Guotao, et al. Collaborative filtering top-N recommendation algorithm based on time context[J]. *Computer technology and development*, 2017, 27(7): 79–82.
- [38] 沈记全, 王磊, 侯占伟, 等. 基于情景上下文与信任关系的旅游景点推荐算法[J]. *计算机应用研究*, 2018, 35(12): 3640–3643.
- SHEN Jiquan, WANG Lei, HOU Zhanwei, et al. Attractions recommendation algorithm based on situational context and trust relationship[J]. *Application research of computers*, 2018, 35(12): 3640–3643.
- [39] 李林峰, 刘真, 魏港明, 等. 基于共享知识模型的跨领域推荐算法[J]. *电子学报*, 2018, 46(8): 1947–1953.

LI Linfeng, LIU Zhen, WEI Gangming, et al. Cross-domain recommendation algorithm based on sharing knowledge pattern[J]. *Acta electronica sinica*, 2018, 46(8): 1947–1953.

- [40] 邢长征, 杨晓婷. 基于 SVD++ 与标签的跨域推荐模型[J]. *计算机工程*, 2018, 44(4): 225–230.

XING Changzheng, YANG Xiaoting. Cross-domain recommendation model based on SVD++ and tag[J]. *Computer engineering*, 2018, 44(4): 225–230.

- [41] 李林峰, 刘真, 魏港明, 等. 基于共享知识模型的跨领域推荐算法[J]. *电子学报*, 2018, 46(8): 1947–1953.

LI Linfeng, LIU Zhen, WEI Gangming, et al. Cross-domain recommendation algorithm based on sharing knowledge pattern[J]. *Acta electronica sinica*, 2018, 46(8): 1947–1953.

- [42] BATTAGLIA P, PASCANU R, LAI M, et al. Interaction networks for learning about objects, relations and physics[C]// *Proceedings of the 30th International Conference on Neural Information Processing Systems*. Red Hook, USA, 2016: 4502–4510.

作者简介:



吴国栋, 副教授, 中国计算机学会会员, 主要研究方向为深度学习、推荐系统。主持安徽省自然科学研究重点项目 1 项、一般项目 1 项、安徽省科技攻关重点项目 1 项。发表学术论文 30 余篇。



查志康, 硕士研究生, 主要研究方向为推荐系统。



涂立静, 讲师, 主要研究方向为人工智能、机器学习。主持安徽省自然科学研究一般项目 1 项、安徽农业大学青年基金项目 1 项。发表学术论文 10 余篇。

第四届亚洲人工智能技术大会

The 4th Asian Conference on Artificial Intelligence Technology

由中国人工智能学会、重庆市大数据应用发展管理局、重庆理工大学、重庆市巴南区人民政府联合主办, 重庆理工大学期刊社、重庆市巴南区科学技术局、重庆市巴南区大数据应用发展管理局、重庆两江人工智能学院联合承办, 重庆市人工智能学会协办, 重庆市科学技术协会指导的“第四届亚洲人工智能技术大会 (ACAIT 2020)”将作为 2020 年中国智能产业博览会期间的唯一国际学术会议在重庆召开。

征稿范围 (但不局限于):

1) 人工智能理论基础; 2) 人工智能应用; 3) 模式识别; 4) 机器感知与虚拟现实; 5) 自然语言处理和机器翻译; 6) 图像和语音处理; 7) 计算机视觉; 8) 神经网络与计算智能; 9) 知识科学与知识工程; 10) 生物信息学与人工生命; 11) 机器学习; 12) 深度学习及其应用; 13) 数据挖掘; 14) 面向大数据的人工智能技术; 15) 智能控制与智能管理; 16) 粗糙集与软计算; 17) 智能搜索; 18) 智能推理; 19) 智能规划; 20) 智能信息处理; 21) 智能制造; 22) 智能机器人; 23) 物联网; 24) 工业互联网; 25) 智能通信与网络; 26) 人机交互/普适计算; 27) 智慧能源; 28) 自动程序设计。

联系方式:

联系人: 贺柳、徐佳忆

电话: 023-62561406

邮箱: cqznjs@126.com; xb@cqut.edu.cn