

基于 AGRU-GNN 的图网络社交推荐算法^①



卓佳宁, 雷景生, 周雪雪

(上海电力大学 计算机科学与技术学院, 上海 200090)

通讯作者: 卓佳宁, E-mail: jeslyn576@gmail.com

摘要: 在推荐系统中, 用户对物品的兴趣是动态变化的, 会受用户自身历史行为、朋友历史行为甚至短时热点等多方面因素影响. 而如何在推荐系统中对用户的时序兴趣进行描述并提取有效信息, 一直以来是推荐算法的一大挑战之一. 本文在图神经网络 (GNN) 推荐算法的基础上, 提出一种基于注意力门控循环单元 (Attention-GRU) 的改进图网络算法, 对用户、物品的交互时序历史进行特征建模, 于此同时结合社交网络将此时序特征在用户、物品之间传播. 算法在 Ciao 与 Epionions 数据集上进行了验证, 并与其他相关工作进行对比, 证明了该模型有效地提取了用户、物品的时序特征, 提升了推荐系统的有效性.

关键词: 图网络推荐; 社交网络推荐; 门控循环单元; 注意力机制

引用格式: 卓佳宁, 雷景生, 周雪雪. 基于 AGRU-GNN 的图网络社交推荐算法. 计算机系统应用, 2021, 30(5): 219-227. <http://www.c-s-a.org.cn/1003-3254/7926.html>

AGRU-GNN Graph Network for Social Recommendation

ZHUO Jia-Ning, LEI Jing-Sheng, ZHOU Xue-Xue

(College of Computer and Science, Shanghai University of Electric Power, Shanghai 200090, China)

Abstract: In a recommendation system, users' interest in items changes dynamically and is affected by various factors such as users' own historical behaviors, their friends' historical behaviors, and even short-term hot spots. How to describe users' temporal interests in a recommendation system and extract effective information has always been one of the challenges for the recommendation algorithms. On the basis of the Graph Neural Network (GNN) recommendation algorithm, we propose an improved graph network algorithm based on the Attention Gated Recurrent Unit (Attention-GRU) in this study. Furthermore, feature modeling is performed on the temporal interactive history of users and items, and in combination with social network, the temporal characteristics are transmitted between users and items. In addition, the proposed algorithm is verified on the Ciao and Epionions data sets and compared with other related work, proving that the model proposed in this study can effectively extract the temporal characteristics of users and items and improve the effectiveness of the recommendation systems.

Key words: graph network recommendation; social network recommendation; Gated Recurrent Unit (GRU); attention mechanism

这是一个信息爆炸的时代, 大数据这把双刃剑在为我们带来丰富的知识讯息的同时, 也产生了不少“数据泡沫”, 用户常常被这些无效信息包围, 无从下手, 而

用户真正感兴趣的信息被阻绝在外. 为了解决这一问题, 推荐系统应运而生, 旨在为每个特定用户从海量数据中筛选出其各自感兴趣的内容并推荐给用户. 推荐

① 基金项目: 国家自然科学基金 (61672337)

Foundation item: National Natural Science Foundation of China (61672337)

收稿时间: 2020-09-21; 修改时间: 2020-10-21; 采用时间: 2020-11-04; csa 在线出版时间: 2021-04-28

系统的关键之一在于对用户兴趣以及物品特征的建模表示,从而能够进一步通过评分模型对用户、物品预测相关程度,并以此为依据进行推荐。另一方面,社交网络中好友、大V等社交关系也可能会影响用户的潜在兴趣,社交网络也越来越多地被结合到推荐系统中来^[1-3],在推荐时利用通过社交网络的传播的信息,提高了推荐的针对性与有效性。

基于图神经网络(GNNs)的推荐算法^[4,5]是推荐系统应用的热门算法之一。GNN在图结构拓扑上应用深度神经网络的技术对局部图的邻域信息进行迭代,使得信息能够在非结构化的节点关系中进行交互计算,最终得到各个节点在图拓扑结构下的特征域表示。社交推荐本质上是在非规则的用户-物品关系、用户-用户关系中聚合并表示各个节点的本征信息与交互关系,因此GNN在社交推荐中得到了广泛的应用^[6,7]。

图神经网络通常描述的是静态拓扑关系,一方面,节点之间的连接关系不会随时间发生改变。对于推荐系统而言,时时刻刻都有海量新数据产生,新用户、新物品以及新交互事件都会改变已有图网络的拓扑结构。另一方面,对于特定节点,其邻居节点的时序性通常被GNN所忽略。例如,对于推荐系统中的用户-物品图结构而言,以用户为中心,则其邻居节点即为该用户交互过的物品,而用户与不同物品的交互自然存在时间上的先后,而交互时间的时序性则可能暗含了用户兴趣随时间的变化情况。

本文主要针对第二点提出改进GNN算法,利用门控循环单元(GRU)对时间序列信息进行选择性遗忘与选择性记忆,以此对用户-物品交互历史以及物品-用户交互历史进行建模,增强了GNN算法在局部图邻域迭代过程中对时序信息的抽象能力,使得推荐系统对用户节点以及物品节点的表示结合了其特点随时间的变化,提升了推荐的效果。具体而言,我们分别建立了以用户节点为中心的用户-物品图,每个用户节点的相邻节点为其交互过的物品,这些相邻节点通过交互时间戳属性排序,在GNN进行局部图邻域迭代时通过GRU学习物品交互历史潜在关联,并通过注意力机制与用户表征进行聚合;类似的,同时建立以物品节点为中心的物品-用户图,每个物品节点的相邻节点为与其交互过的用户,同样按照交互事件的时间戳排序,并通过注意力GRU模型进行聚合,以此来表征对该物品感兴趣的群体特征变化;另外,还根据用户间的社交关系

建立了用户-用户图网络,通过注意力机制来传递社交网络中相邻用户之间的相互作用。

1 相关工作

1.1 基于非时序特征的GNN推荐算法

深度学习主要关注例如文字的序列结构、例如图片的平面结构,现在处理这些数据的做法也比较成熟,关注序列任务的NLP领域多用RNN、Transformer、CNN对数据进行Encoder,而关注平面结构的CV领域更多使用CNN及其各种变体对数据进行Encoder。在现实世界中更多的数据表示并不是序列或者平面这种简单的排列,而是表现为更为复杂的图结构。

Kipf和Welling^[8]提出了用于半监督图分类的图卷积网络(GCNs)。模型通过利用节点属性和图结构来学习节点表示,它由多个图卷积层组成,每个图卷积层使用当前节点的表示及其相邻节点的表示的组合来更新节点表示。通过这个过程,可以捕获节点之间的依赖关系,然而在原来的公式中,当更新节点表示时,所有的邻居被赋予静态权值。Fan等^[9]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。Ying等^[10]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^[11]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。上述文献证明了图网络在推荐上的有效性,但都没有考虑时序性对用户兴趣以及物品的影响。

1.2 基于时序特征的推荐算法

对于随时间变化的用户兴趣进行建模在推荐系统领域也引起了大量关注,这些模型大多基于(高斯)矩阵分解^[12]。例如,Xiong等^[13]通过分解(用户、物品、时间)张量来学习时间表示。Koren^[14]开发了一个类似的模型叫做timeSVD++。Gopalan等进行了类似的建模,但是使用了泊松分解^[15]。但是,这些方法假定用户的兴趣在长期范围内缓慢而平稳地变化,通常以月或年为序。

为了有效地捕捉用户的短期兴趣,最近的研究将RNN引入到用户最近的(有序的)行为模型中。例如,Hidasi等^[16]首先提出了SessionRNN来模拟用户在会

话中的兴趣. Wu 等^[17]使用两个独立的 RNNs 根据新的观察结果更新用户和项的表示. Beutel 等^[18]构建了一个基于 *rmnde* 的推荐器, 可以包含辅助上下文信息. 本文基于这种思路将 GRU 引入图网络, 并且辅之以注意力机制, 来捕捉用户的时变兴趣和物品的时变特征, 进而提升了推荐系统在时序评分数据集上的有效性.

2 基于 AGRU 的图网络社交推荐算法

本文以图网络为骨架建立推荐系统, 定义用户集合 $\{\mathcal{U}|u_1, u_2, \dots, u_N\}$ 与物品集合 $\{\mathcal{V}|v_1, v_2, \dots, v_M\}$, 则图顶点集合为 $\mathcal{U} \cup \mathcal{V}$. 图网络中存在两种非同质边缘, 第一种边表示用户与物品之间的交互关系, 即 $\{\mathcal{R}|\mathcal{U} \times \mathcal{V}\} = \{\mathcal{R}|r_{i,j} := (u_i, v_j), u_i \in \mathcal{U}, v_j \in \mathcal{V}\}$, 描述了用户 u_i 对物品 v_j 进行过评价, 边 $r_{i,j}$ 的权重记为 $w_{i,j}^R$, 表示用户 u_i 对物品 v_j 的评分量化分值, 边 $r_{i,j}$ 的时间戳记为 $t_{i,j}^R$, 表示此次评分事件发生的时刻; 第二种边表示用户与用户之间的有向交互, 即 $\{\mathcal{S}|\mathcal{U} \times \mathcal{U}\} = \{\mathcal{S}|s_{i,j} := (u_i \rightarrow u_j), u_i, u_j \in \mathcal{U}\}$, 表示用户 u_i 在社交平台上关注了用户 u_j , 边 $s_{i,j}$ 的权重记为 $w_{i,j}^S$, 表示用户 u_i 与 u_j 社交关系重要性的量化值, 本文不考虑社交网络的时序特征, 因此边 $s_{i,j}$ 无时间戳属性.

因此, 本文推荐系统可描述为: 给定用户集合 \mathcal{U} 、物品集合 \mathcal{V} , 以及评分边集合 \mathcal{R} 、社交边集合 \mathcal{S} , 经过一定的训练算法得到优化预测函数 \mathcal{F} , 使得对于任意 $u_i \in \mathcal{U}, v_j \in \mathcal{V}, r_{i,j} \notin \mathcal{R}$, 可预测用户 u_i 对物品 v_j 的评分 $\hat{w}_{i,j}^R = \mathcal{F}(u_i, v_j)$, 作为是否推荐的依据. 以下将分别以用户节点与物品节点为中心, 展开介绍图网络局部图邻域聚合的具体算法细节.

2.1 图网络局部图邻域聚合

GNN 中常用局部图邻域聚合技术^[8,19]来传递邻近节点之间的信息, 并迭代训练各节点的抽象表示. 对于给定图 $\mathcal{G}(\mathcal{V}, \mathcal{E})$, 定义 k 阶邻域函数 $Neigh^k(v_i) = \{v_k | (v_i, v_{i,1}), \dots, (v_{i,k-1}, v_k) \in \mathcal{E}\}$, 即 v_i 与 v_k 之间存在长度为 k 的路径. 本文采用一阶邻域函数, 简记为 \mathcal{N} . 则 GNN 局部图邻域聚合算法可见算法 1.

算法 1. GNN 局部图邻域聚合算法的第 n 次迭代

- 1) 遍历图 \mathcal{G} 的顶点 \mathcal{V} , 对每个顶点 v_i , 执行第 2)~4) 步操作;
- 2) 计算 v_i 的邻域节点集合 $\mathcal{N}(v_i) = Neigh(v_i)$;
- 3) 利用聚合函数 Agg 计算 $\mathcal{N}(v_i)$ 的聚合表示结果 $v_i^{agg} = Agg(\mathcal{N}(v_i))$;
- 4) 将聚合结果 v_i^{agg} 与顶点自身 v_i 拼接, 并进行线性仿射与非线性激活, 得到该顶点本次迭代的表示 $v_i^{(n)}$.

上述算法中聚合函数 Agg 有多种选择, 常见的如拼接、注意力、卷积等. 本文根据边 $r_{i,j}$ 存在时序性的特点, 选择基于注意力机制的门控循环单元作为邻域聚合函数, 以提取各节点邻域集合中的时序信息, 提高推荐系统对时序数据的学习能力.

本文推荐系统图模型中 (如图 1 所示), 顶点集合由用户集合 \mathcal{U} 、物品集合 \mathcal{V} 两部分组成, 边集合由评分关系集合 \mathcal{R} 、社交关系集合 \mathcal{S} 构成, 其中不包含关于物品集合 \mathcal{V} 的闭集边集, 因此整个图可划分为 3 个局部图: 用户-物品图、用户-用户图以及物品-用户图, 以下将分别讨论其邻域聚合算法. 另外, 用户 u_i 、物品 v_j 分别通过嵌入层获得其向量表示 e_i^U, e_j^V ; 考虑到数据集中评分的取值为离散整数 0~5, 权重 $w_{i,j}^R$ 也通过对评分分值的嵌入向量表示, 即 $e_{i,j}^R$.

2.2 用户模型: AGRU 聚合+社交影响力聚合

考虑图模型中的用户节点 u_i 以及以其为端点的评分边 \mathcal{R}_{u_i} 构成的局部图 (图 1 左上角子图), 即用户-物品图. 则 u_i 的邻域节点为用户 u_i 所评价过的物品集合 $\mathcal{N}^V(u_i)$, 根据各评分边的时间戳 t , 可得到用户评分事件的时序排列. 为提取用户评价物品的时序特征, 本文在用户-物品图邻域聚合采用基于注意力的门控循环单元 (Attention-GRU) 来提取评分序列特征, 算法如算法 2 所示.

算法 2. 用户-物品图 AGRU 聚合算法, 第 n 次迭代

- 1) 遍历用户集合 \mathcal{U} , 对每个顶点 u_i , 执行第 2)~8) 步操作;
- 2) 计算 u_i 在物品集合 \mathcal{V} 上的邻域节点集合 $\mathcal{N}^V(u_i) = Neigh^V(u_i)$, 邻域节点数量 $L = |\mathcal{N}^V(u_i)|$;
- 3) 将 $\mathcal{N}^V(u_i)$ 中的物品节点 v_j 按时间戳 $t_{i,j}$ 升序排列, 得到长度为 L 的物品序列 $Seq(u_i) = [v_1, v_2, \dots, v_L]$, 其中 $t_{i,1} \leq t_{i,2} \leq \dots \leq t_{i,L}$;
- 4) 对序列 $Seq(u_i)$ 中的每个物品元素 v_j , 查询其嵌入表示 e_j^V , 以及对评分 $w_{i,j}^R$ 的嵌入表示 $e_{i,j}^R$; 将 e_j^V 与 $e_{i,j}^R$ 拼接输入多层感知机降维得到包含评分分值信息的物品表示 e_j' ; 更新 $Seq(u_i) = [e_1', e_2', \dots, e_L']$;
- 5) 将序列 $Seq(u_i)$ 作为 GRU 的输入, 同时输入初始隐藏状态 h_0 , 得到 GRU 在每个时刻末层输出构成的长度为 L 的序列 $Outseq = [h_1^{(k)}, h_2^{(k)}, \dots, h_L^{(k)}]$, 其中 k 为 GRU 的纵向堆叠层数;
- 6) 以 GRU 末层最后时刻的隐藏状态 $h_L^{(k)}$ 为参考, 用注意力机制对 $Outseq$ 中的 L 个元素进行加权平均得到 h' ;
- 7) 拼接 h' 与 $h_L^{(k)}$, 并经过线性层以及非线性 $ReLU$ 激活, 得到 u_i^{agg} , 作为邻域集合 $\mathcal{N}^V(u_i)$ 的聚合结果, 即 $Agg(\mathcal{N}^V(u_i)) = u_i^{agg}$;
- 8) 将聚合结果 u_i^{agg} 与顶点自身嵌入表示 e_i^U 拼接, 并进行线性仿射与非线性激活, 得到该用户节点第 n 次迭代的表示 $u_i^{(n)}$.

上述算法中, 第 4) 步采用双层感知机降维:

$$x_j^{VR} = CONCAT(e_j^V, e_{i,j}^R) \quad (1)$$

$$y_j = ReLU(W_1 x_j^{VR} + b_1) \quad (2)$$

$$e_j' = ReLU(W_2 y_j + b_2) \quad (3)$$

第5)步中第 $m = 1, 2, \dots, k$ 层GRU^[20]在时刻 $t = 1, 2, \dots, L$ 执行下列计算:

$$x_t = \begin{cases} e_t', & m = 1 \\ \delta_t^{(m-1)} * h_t^{(m-1)}, & m > 1 \end{cases} \quad (4)$$

$$r_t = \sigma(W_{ir} x_t + b_{ir} + W_{hr} h_{(t-1)} + b_{hr}) \quad (5)$$

$$z_t = \sigma(W_{iz} x_t + b_{iz} + W_{hz} h_{(t-1)} + b_{hz}) \quad (6)$$

$$n_t = \tanh(W_{in} x_t + b_{in} + r_t * (W_{hn} h_{(t-1)} + b_{hn})) \quad (7)$$

$$h_t = (1 - z_t) * n_t + z_t * h_{(t-1)} \quad (8)$$

其中, δ 表示Dropout函数, *表示Hadamard乘积, σ 表

示Sigmoid激活函数:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (9)$$

每个GRU单元的内部结构如图2所示, x_t 表示当前时刻的输入, $h_{(t-1)}$ 则表示该GRU单元上一时刻的隐藏状态. 两者拼接并分别通过独立全连接层得到重置因子 r_t 和遗忘因子 z_t , 由于经过了Sigmoid激活, 两种因子的取值范围均在0到1之间. 重置因子对上一时刻的隐藏状态进行重置, 并将结果与当前输入拼接后传递给tanh激活的全连接层, 得到以当前输入为主、隐藏状态为辅的表示 n_t . 另一方面, 遗忘因子 z_t 与记忆因子 $1 - z_t$ 分别对 $h_{(t-1)}$ 和 n_t 调制, 实现了对序列历史隐含信息的选择遗忘以及当前输入的选择记忆.

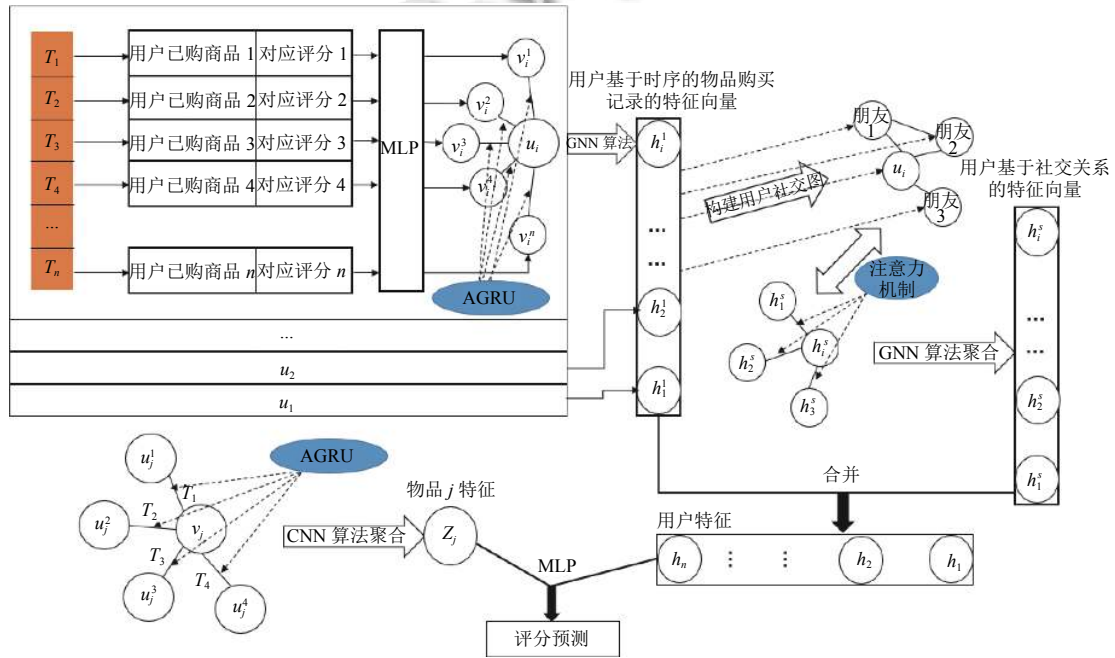


图1 基于AGRU-GNN的图网络社交推荐算法架构

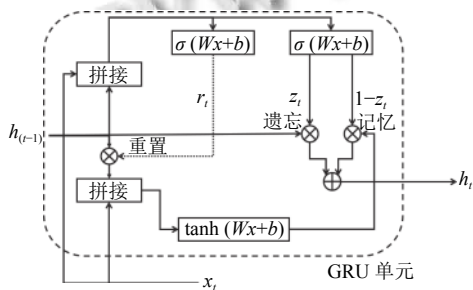


图2 GRU单元内部结构

取GRU第 k 层的隐藏状态 $Outseq = [h_1^{(k)}, h_2^{(k)}, \dots, h_L^{(k)}]$ 作为输出, 并基于输出单元数量为1的多层感知机+

Softmax实现注意力机制(如图3所示), 获得 $h_i^{(k)}$ 的注意力权重系数 p_i :

$$x_i = CONCAT(h_i^{(k)}, h_L^{(k)}) \quad (10)$$

$$y_i = ReLU(W_{a1} x_i + b_{a1}) \quad (11)$$

$$z_i = ReLU(W_{a2} y_i + b_{a2}) \quad (12)$$

$$a_i = ReLU(W_{a3} z_i + b_{a3}) \quad (13)$$

$$p_i = \frac{\exp(a_i)}{\sum_{i=1}^L \exp(a_i)} \quad (14)$$

则邻域 $\mathcal{N}^V(u_i)$ 聚合特征:

$$\text{Agg}(\mathcal{N}^{\mathcal{V}}(u_i)) = u_i^{\text{agg}} = \sum_{i=1}^L p_i h_i^{(k)} \quad (15)$$

最终将邻域聚合特征与节点自身特征融合得到 $u_i^{(n)}$:

$$x_i = \text{CONCAT}(u_i^{\text{agg}}, e_i^{\mathcal{U}}) \quad (16)$$

$$y_i = \text{ReLU}(W_1 x_i + b_1) \quad (17)$$

$$u_i^{(n)} = \text{ReLU}(W_2 y_i + b_2) \quad (18)$$

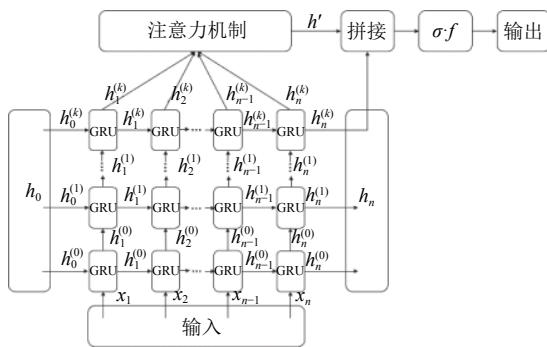


图3 基于注意力机制的GRU聚合算法架构

另一方面,考虑用户节点 u_i 以及以其为起始端点的社交边 S_{u_i} 构成的局部图,即用户-用户图(图1右上角子图)。本算法中社交边集为有向边集,每条边表示用户的单向关注关系,则 u_i 的邻域节点为用户 u_i 所关注的用户集合 $\mathcal{N}^{\mathcal{U}}(u_i)$ 。社交边 $s_{i,j}$ 的权重 $w_{i,j}^S$ 量化为被关注者 u_j 的影响力因子:

$$w_{i,j}^S = \left| \left\{ u_k \mid u_k \in \mathcal{U}, s_{k,j} \in \mathcal{S} \right\} \right| \quad (19)$$

在用户-用户局部图的邻域聚合的第 n 次迭代中,使用用户-物品局部图第 n 次迭代聚合的结果 $u_i^{(n)}$ 作为节点 u_i 的向量表示,使得AGRU捕获到的时序特征也能够通过社交边进行传播。具体聚合过程如下:

首先,查询用户-物品图在用户集合 $\mathcal{N}^{\mathcal{U}}(u_i)$ 上的迭代结果得到:

$$\{u_j^{(n)} \mid u_j \in \mathcal{N}^{\mathcal{U}}(u_i)\} \quad (20)$$

随后通过如式(10)–(14)所示的注意力网络得到 $u_i^{(n)}$ 关于其每个邻域节点的边权加权表示 $w_{i,j}^S u_j^{(n)}$ 的注意力权重 p_j 。则社交局部图的邻域聚合特征

$$\text{Agg}(\mathcal{N}^{\mathcal{U}}(u_i)) = u_i^{\text{agg},S} = \sum_j p_j w_{i,j}^S u_j^{(n)} \quad (21)$$

最终经过式(16)–式(18),将 $u_i^{\text{agg},S}$ 与节点自身特征 $u_i^{(n)}$ 融合降维后得到 $u_i^{(n)*}$,即为用户 u_i 在第 n 次迭代后的最终向量表示。

2.3 物品模型:对偶AGRU聚合

类似地,考虑以物品节点为中心情形下的局部图邻域聚合算法。由于评分边集 \mathcal{R} 是无向边集,因此物品-用户图(图1左下角子图)与用户-物品图完全对偶,对于物品 v_i ,其邻域用户集合 $\mathcal{N}^{\mathcal{U}}(v_i)$ 也可通过评分边的时间戳排序得到用户时间序列,表示对物品 v_i 评价过的用户在时间上的先后次序。因此对于物品的特征迭代算法与算法2关于 $(\mathcal{U}, \mathcal{V})$ 对偶,可对称地得到物品 v_i 在第 n 次迭代后的表示 $v_i^{(n)}$ 。显然地,物品社交图没有明显的物理意义,因此物品模型中不包含自身社交影响力聚合,用户 u_i 在第 n 次迭代后的最终向量表示 $v_i^{(n)*} = v_i^{(n)}$ 。当然,用户社交的影响因素会在物品进行邻域聚合时传播给物品模型。

2.4 评分预测模型

根据第2.2、2.3节对用户 u_i 、物品 v_i 的AGRU聚合算法,分别得到第 n 次迭代后的向量表示 $u_i^{(n)*}, v_i^{(n)*}$ 。将其拼接后经过单输出多层感知机(如图1右下角子图),得到对评分关系 $r_{i,j}$ 权重(即评分分值)的预测值 $\hat{w}_{i,j}^R$ 。

在训练过程中,以一个批次(batch)数据的均方差MSE作为目标函数对模型中的参数进行优化:

$$\text{MSE}(\text{batch}) = \frac{\sum_{k=1}^{\text{bsize}} (\hat{w}_{i_k, j_k}^R - w_{i_k, j_k}^R)^2}{\text{bsize}} \quad (22)$$

其中,bsize为批次规模。

在推理过程中,则 $\hat{w}_{i,j}^R$ 即为推荐系统给出的用户 u_i 对物品 v_j 的评分分值预测。

3 实验结果与分析

3.1 数据集选取与预处理

实验采用Ciao数据集和Epinions数据集对算法进行验证,两者均为包含社交关系信息的在线购物数据^[21]。数据集包含两部分数据:(1)评分信息,每条评分数据包括用户索引、物品索引、评分分值(0–5)以及时间戳4个维度;(2)社交信息,每条社交数据包括关注者用户索引以及被关注者用户索引,表示用户单向关注关系。Ciao数据集与Epinions数据集的特点如表1所示。

对数据集的预处理包括数据清洗、索引重映射、数据集切分、图建模、时间戳排序5个步骤。数据清洗主要完成数据去重,评分数据中包含少量重复评分边

$r_{i,j}$, 根据其时间戳 $t_{i,j}^R$ 保留较新的数据; 随后对数据集用户集合 \mathcal{U} 、物品集合 \mathcal{V} 分别进行索引重映射, 将原始索引映射至 $[1, 2, \dots, |\mathcal{U}|], [1, 2, \dots, |\mathcal{V}|]$ 范围内, 方便嵌入层映射; 之后将数据集中的评分边 \mathcal{R} 集随机划分为训练评分边集 $\mathcal{R}^{\text{train}}$ 和测试评分边集 $\mathcal{R}^{\text{test}}$, 分别比较训练集比重为 60% 和 80% 的实验结果; 对训练集和测试集, 分别建立第 2 节中描述的推荐图模型, 并对每个用户节点的邻域物品节点按时间戳提前排序, 则图网络迭代过程中无需重复排序, 以降低计算量, 对于物品节点的邻域用户节点, 也按同样方法提前排序。

表 1 Ciao 数据集与 Epinions 数据集特征

数据集	Ciao	Epinions
用户数量	7375	22164
物品数量	105114	296277
评分数量	284086	922267
评分密度(%)	0.0366	0.0140
社交联系	111781	355813
社交密度(%)	0.2055	0.0724
时间跨度	3969	4326
日均评分数量	72	213

3.2 评价指标

本文采用衡量推荐系统推荐效果的两个常用指标作为评判标准: 均方根误差 $RMSE$ 和平均绝对误差 MAE :

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (23)$$

$$MAE(y, \hat{y}) = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i| \quad (24)$$

其中, y 表示测试集所有评分边的权重 (真实值) $\{w_{i,j}^{\text{test}}\}$, \hat{y} 则表示推荐系统对测试集各评分边的预测权重 (预测值) $\{\hat{w}_{i,j}^{\text{test}}\}$. $RMSE$ 与 MAE 越小, 表明推荐系统预测的评分与真实值差距越小, 推荐效果越好。

3.3 对比模型

我们选取下列模型作为本算法对比的基准模型:

(1) 概率矩阵分解 (PMF)^[12]: 对用户潜在因子、物品潜在因子建模, 利用用户-物品评分概率矩阵进行推荐;

(2) 社交推荐 (SoRec)^[22]: 基于用户-物品评分矩阵提取联合潜在因子特征, 并结合社交矩阵进行推荐;

(3) 社交矩阵分解 (SocialMF)^[3]: 考虑了社交信任度在用户-物品评分矩阵分解空间中的传播;

(4) 神经网络矩阵分解 (NCF)^[23]: 将神经网络与矩阵分解结合进行推荐;

(5) 深度社交网络推荐 (Deep-SoR)^[24]: 利用深度神经网络对用户社交关系中的特征进行训练, 并通过概率矩阵分解实现推荐;

(6) 图网络推荐 (GraphRec)^[9]: 根据社交关系和评分关系建立图网络, 利用注意力机制迭代图节点, 训练用户、物品的嵌入表示来实现推荐。

3.4 实验环境与超参数设置

本文采用 Python 3.8, PyTorch 1.6, Cuda 10.2 作为搭建算法的环境, 工作站 CPU 为 4 核 Intel(R) Core(TM) i5-7500@3.40 GHz, GPU 型号为 NVIDIA GP102 [TITAN Xp].

本算法以 batch 为单位进行训练, 为了避免过拟合, 在各非线性激活函数输出端增加 dropout 层, 丢弃概率全局统一设置; 同时在图网络的每一批次迭代中, 对同一 batch 的用户、物品向量表示进行批次归一化. 算法中的超参数经过验证集 (从训练集中抽取 20% 得到) 反复调整得到: 嵌入维度 (用户、物品、评分采用统一嵌入维度) 为 64, GRU 纵向堆叠层数 $k=2$, dropout 丢弃概率为 0.5, 批次大小为 64, 学习率为 0.0005.

3.5 结果分析

本文在 Ciao、Epinions 数据集分别选取 60%、80% 作为训练集, 其余作为测试集对提出的算法进行了验证, 并与第 3.3 节列出的几种基准模型进行了对比, $RMSE$ 和 MAE 比较结果如表 2.

从表 2 的结果可以看出, 本文所提出的基于 AGRU-GNN 算法在 Ciao 数据集和 Epinions 数据集上比表中所述的基准模型误差更小. 其中 PMF 和 NCF 只利用了用户与物品的评分数据, 而 SoRec、SocialMF、Deep-SoR 和 GraphRec 均使用了数据集中的社交关系数据, 因此平均比同类模型具有更小的误差. 本文所采用的 AGRU 聚合算法, 通过 GRU 对评分事件的时序信息提取, 并加以注意力机制对不同时刻的输出进行了有选择的加权平均, 结合社交局部图使得用户对物品的时变兴趣信息能够基于社交影响关系在用户之间传播, 从而达到了更好的推荐效果. 这证明了本文提出的算法在时序推荐数据集上具有一定的优越性.

我们还探究了不同超参数设置下, 对本文推荐算法效果的影响. 以 Ciao(80%) 数据集为例, 基准超参数的设置为 GRU 纵向堆叠层数 $k=2$, dropout 概率为 0.5, 迭代周期数为 5. 图 4 展示了在基准超参数设置下, 模

型预测误差随迭代周期数 (epoch) 的变化, 一次迭代周期表示图网络中所有顶点 $U \cup V$ 均被聚合迭代了一次。可以看出模型的收敛速度较快, 5 个周期后便达到了最优值, 随着迭代次数的增加, 模型出现了一定程度的过

拟合。收敛速度快得益于 GRU 对时序信息具有较强的表示能力, 快速地提取了用户、物品的时变特征; 另一方面, 由于 GRU 内部结构较为复杂, 也增大了模型的参数量, 因此随即出现了一定程度的过拟合现象。

表 2 本文算法 AGRU-GNN 与其他推荐模型对比结果

训练集	评价指标	PMF	SoRec	SocialMF	NCF	Deep-SoR	GraphRec	AGRUGNN
Ciao(60%)	MAE	0.9498	0.8491	0.8349	0.8250	0.7798	0.7540	0.7410
	RMSE	1.2017	1.0730	1.0601	1.0799	1.0435	1.0093	0.9892
Ciao(80%)	MAE	0.9019	0.8399	0.8268	0.7998	0.7741	0.7387	0.7029
	RMSE	1.1240	1.0652	1.0499	1.0615	1.0298	0.9794	0.9659
Epinions(60%)	MAE	1.0211	0.9101	0.9001	0.9097	0.8518	0.8441	0.8332
	RMSE	1.2740	1.1565	0.1419	1.1645	1.1134	1.0878	1.0719
Epinions(80%)	MAE	0.9952	0.8958	0.8836	0.9070	0.8410	0.8168	0.8072
	RMSE	1.2128	1.1435	0.1317	1.1475	1.1001	1.0631	1.0454

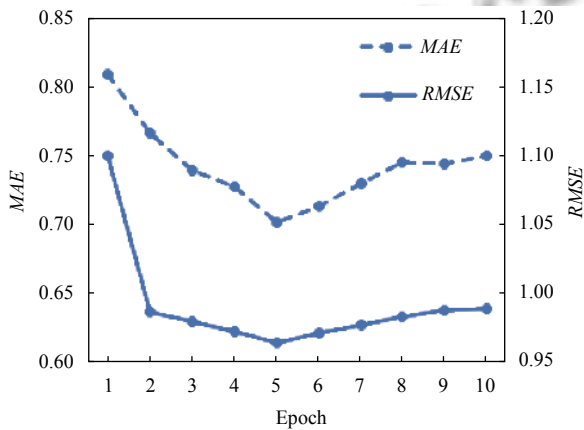


图 4 模型预测误差随迭代周期数的变化

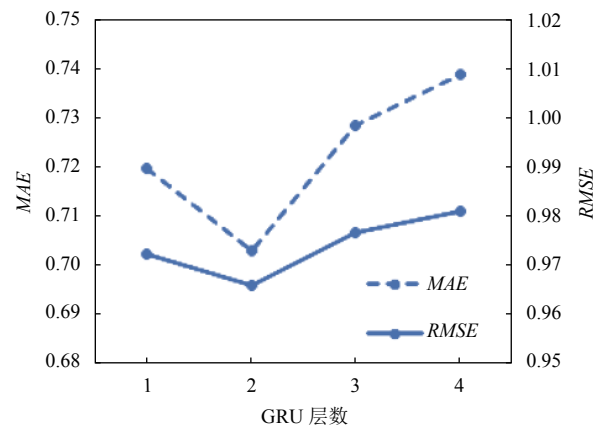


图 5 模型最佳预测误差随 GRU 层数的变化

GRU 的层数一定程度上决定了模型的复杂程度, 为了避免严重的过拟合, 我们研究了不同 GRU 层数设置对模型误差的影响, 如图 5。可以看出, 双层 GRU 在此数据集上具有较好的推荐效果, 而层数过大则存在较严重的过拟合, 层数过小则削弱了模型对时序信息的学习能力。

为了更好地抑制过拟合, 模型中在大部分非线性激活单元的输出端以及 GRU 层间加入了 dropout 层, 并通过全局统一的 dropout 概率加以调节。图 6 展示了模型最佳预测误差随 dropout 概率的变化, 以 50% 概率进行丢弃时, 模型在预测误差和过拟合之间达到了较好的平衡。即使以较大的概率 75% 进行丢弃, 模型的功能依然没有受损, 但预测误差开始陡增。

另外, 我们还对比了本文模型的若干变种模型, 分别为:

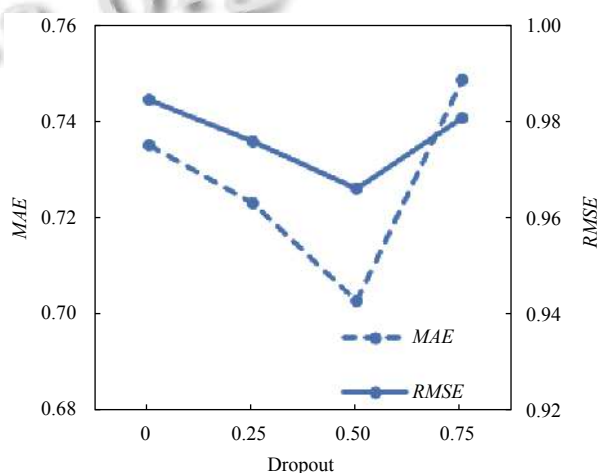


图 6 模型最佳预测误差随 dropout 概率的变化

(1) AGRU-GNN: 采用 GRU 作为图网络时序聚合模型, 并在物品-用户图、社交图的邻域聚合均采用了

注意力机制,即本文提出的最终模型;

(2) GRU-GNN: 禁用了模型中的注意力机制,其余参数与(1)相同;

(3) ARNN-GNN: 采用 Vanilla RNN 作为图网络时序聚合模型,也同时采用了注意力机制;

(4) GNN(即 GraphRec^[9]): 基准图网络模型,主要采用注意力聚合。

四种变种模型分别在 Ciao(80%) 和 Epinions(80%) 数据集上进行了对比,结果如图 7 所示。GRU 对时序信息的提取筛选能力明显优于普通 RNN,普通 RNN 对时序信息的选择记忆、遗忘能力不足,导致 ARNN-GNN 的推荐有效性甚至弱于基准模型 GraphRec;另外,注意力机制的存在也对模型整体性能有一定提升,这说明注意力机制有助于在用户的物品交互时间序列关注重点事件,同样地,在社交图中也对不同的好友存在不同的关注权重。

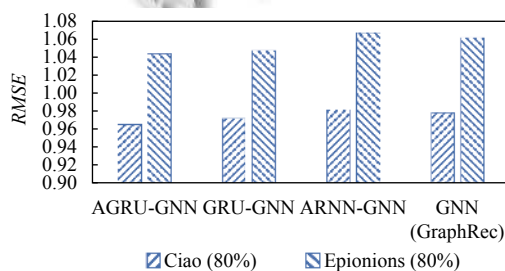


图 7 变种模型最佳预测误差 RMSE 对比

综上所述,本文提出的基于 AGRU 的图网络聚合算法,在带有时间戳及社交信息的推荐数据集 Ciao、Epinions 上取得了良好的性能,证明了本算法中采用的 AGRU 聚合从用户时序评价历史中提取了有效信息,提升了推荐系统的准确度。

4 结论与展望

本文在神经网络 GNN 架构的基础上,基于推荐系统中评价事件的时序性以及用户特征、物品特征的时变特点,提出了一种基于门控循环单元的注意力聚合算法。通过 Ciao、Epinions 数据集,证明了本算法在时序评价数据集上具有更低的预测误差,说明用户的时变兴趣、物品的时变特征对于推荐系统也至关重要。GRU 对于时间序列有较强的学习能力,后续工作可考虑基于此算法实现推荐系统的在线训练,即输入数据是随时间流动的流式数据,图结构也随数据的增长动

态变化,使 GRU 从流式数据中捕获动态的时序特征,以此实现推荐系统的在线演变和动态推荐。在线的动态推荐更加接近实际生活中的推荐系统产品,当然,还需要解决冷启动、计算开销大甚至是分布式架构设计等问题。

参考文献

- 1 Ma H, Zhou DY, Liu C, *et al.* Recommender systems with social regularization. Proceedings of the fourth ACM International Conference on Web Search and Data Mining. Hong Kong, China. 2011. 287–296.
- 2 Tang JL, Hu X, Gao HJ, *et al.* Exploiting local and global social context for recommendation. Proceedings of the Twenty-Third international joint conference on Artificial Intelligence. Beijing, China. 2013. 2712–2718.
- 3 Jamali M, Ester M. A matrix factorization technique with trust propagation for recommendation in social networks. Proceedings of the Fourth ACM Conference on Recommender Systems. Barcelona, Spain. 2010. 135–142.
- 4 Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering. Proceedings of the 30th Conference on Neural Information Processing Systems. Barcelona, Spain. 2016. 3844–3852.
- 5 Derr T, Ma Y, Tang JL. Signed graph convolutional networks. Proceedings of 2018 IEEE International Conference on Data Mining (ICDM). Singapore. 2018. 929–934.
- 6 Wu L, Sun PJ, Hong RC, *et al.* SocialGCN: An efficient graph convolutional network based model for social recommendation. arXiv preprint arXiv: 1811.02815, 2019.
- 7 van der Berg R, Kipf T N, Welling M. Graph convolutional matrix completion. arXiv preprint arXiv: 1706.02263, 2017.
- 8 Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv: 1609.02907, 2017.
- 9 Fan WQ, Ma Y, Li Q, *et al.* Graph neural networks for social recommendation. The World Wide Web Conference. San Francisco, CA, USA. 2019. 417–426.
- 10 Ying R, He RN, Chen KF, *et al.* Graph convolutional neural networks for web-scale recommender systems. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London, UK. 2018. 974–983.
- 11 Wang X, He X, Wang M, *et al.* Neural graph collaborative filtering. arXiv: 1905.08108, 2019.

- 12 Salakhutdinov R, Mnih A. Probabilistic matrix factorization. Proceedings of the 20th International Conference on Neural Information Processing Systems. Vancouver, BC, Canada. 2007. 1257–1264.
- 13 Xiong L, Chen X, Huang TK, *et al.* Temporal collaborative filtering with bayesian probabilistic tensor factorization. Proceedings of the 2010 SIAM International Conference on Data Mining. Columbus, OH, USA. 2010. 211–222.
- 14 Koren Y. Collaborative filtering with temporal dynamics. Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Paris, France. 2009. 447–456.
- 15 Gopalan P, Hofman JM, Blei DM. Scalable recommendation with hierarchical poisson factorization. Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence. Amsterdam, the Netherlands. 2015. 326–335.
- 16 Hidasi B, Karatzoglou A, Baltrunas L, *et al.* Session-based recommendations with recurrent neural networks. arXiv preprint arXiv: 1511.06939, 2016.
- 17 Wu CY, Ahmed A, Beutel A, *et al.* Recurrent recommender networks. Proceedings of the Tenth ACM International Conference on Web Search and Data Mining. Cambridge, UK. 2017. 495–503.
- 18 Beutel A, Covington P, Jain S, *et al.* Latent cross: Making use of context in recurrent recommender systems. Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. Marina Del Rey, CA, USA. 2018. 46–54.
- 19 Hamilton WL, Ying R, Leskovec J. Inductive representation learning on large graphs. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, CA, USA. 2017. 1024–1034.
- 20 Cho K, van Merriënboer B, Gulcehre C, *et al.* Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv: 1406.1078, 2014.
- 21 Tang JL, Gao HJ, Liu H, *et al.* eTrust: Understanding trust evolution in an online world. Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Beijing, China. 2012. 253–261.
- 22 Ma H, Yang HX, Lyu MR, *et al.* SoRec: Social recommendation using probabilistic matrix factorization. Proceedings of the 17th ACM Conference on Information and Knowledge Management. Napa Valley, CA, USA. 2008. 931–940.
- 23 He XN, Liao LZ, Zhang HW, *et al.* Neural collaborative filtering. Proceedings of the 26th International Conference on World Wide Web. Perth, Australia. 2017. 173–182.
- 24 Fan WQ, Li Q, Cheng M. Deep modeling of social relations for recommendation. Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18). New Orleans, LA, USA. 2018. 8075–8076.