·人工智能与模式识别·　　文章编号：1000-3428(2022)05-0082-09　　文献标志码：A　　中图分类号：TP399

# 基于复杂结构信息的图神经网络序列推荐算法

胡承佐[1,2]，王庆梅[1,2]，李迪超[3]，王　铮[4]

（1.北京科技大学 国家材料服役安全科学中心，北京 100083；2.南方海洋科学与工程广东省实验室，广东 珠海 519080；
3.澳门大学 计算机与信息科学系，澳门 999078；4.北京科技大学 计算机与通信工程学院，北京 100083）

**摘　要：**图结构因其在序列推荐场景中的自然适应性而备受关注，而现有的基于图神经网络的会话序列推荐算法虽然能够利用图结构信息达到较好的推荐效果，但是没有考虑用户在会话序列中的重复点击行为和项目之间的复杂转换，且未很好地利用图中复杂的结构信息，导致推荐的效果受到一定程度的限制。提出有向与无向信息同注意力相融合的图神经网络序列推荐算法，并基于推荐算法给出项目隐含向量建模算法，结合会话序列图中的有向结构信息与无向结构信息，通过考虑用户的重复点击行为和引入注意力机制建立会话中点击项目的复杂转换模型。图节点在特征传播的过程中平衡邻居节点信息与自身信息的比例，以更准确地预测推荐过程中生成的会话向量。在Diginetica、Yoochoose 1/64、Yoochoose 1/4 3个数据集上的实验结果表明，与SR-GNN、TAGNN算法相比，该算法精度最高提升4.34%，能够更好地预测用户在会话中的下一次点击精度。

**关键词：**图结构；图神经网络；会话序列；推荐算法；注意力机制

**开放科学（资源服务）标志码（OSID）：**

**中文引用格式：**胡承佐，王庆梅，李迪超，等.基于复杂结构信息的图神经网络序列推荐算法[J].计算机工程，2022，48(5)：82-90，97.

**英文引用格式：**HU C Z，WANG Q M，LI D C，et al.Sequence recommendation algorithm of graph neural networks based on complex structure information[J].Computer Engineering，2022，48(5)：82-90，97.

# Sequence Recommendation Algorithm of Graph Neural Networks Based on Complex Structure Information

HU Chengzuo[1,2]，WANG Qingmei[1,2]，LI Dichao[3]，WANG Zheng[4]

（1.National Center for Materials Service Safety，University of Science and Technology Beijing，Beijing 100083，China；
2.Southern Marine Science and Engineering Guangdong Laboratory，Zhuhai，Guangdong 519080，China；
3.Department of Computer and Information Science，University of Macau，Macao 999078，China；
4.School of Computer and Communication Engineering，University of Science and Technology Beijing，Beijing 100083，China）

**【Abstract】**Graph structures have received significant attention，owing to their natural adaptability for sessions.Thus，many researchers have investigated Graph Neural Networks（GNN）-based recommending algorithms and achieved state-of-the-art performances.Existing session-based recommendations based on GNN can yield relatively accurate recommendations，utilizing structural graph information.However，they neither consider repetitive submissions from users and complex transition between items nor fully utilize complex graph structural information.Consequently，they result in prediction losses.This paper proposes a GNN sequence recommendation algorithm based on the fusion of directed and undirected information with attention.The proposed algorithm combines directed and undirected structural information of session graphs into new hidden embeddings of items.By using repetitive behavioral information and attention mechanisms，the model incorporates complex transitions of items to form better session embeddings.During feature propagation，each node strikes a balance between preserving its information and absorbing its neighbors' information，improving the accuracy of recommendation predictions.The experimental results for Diginetica，Yoochoose 1/64，and Yoochoose 1/4 data sets show that compared with the best existing algorithms，that is，Session-based Recommendation with GNN（SR-GNN）and Target Attentive GNN（TAGNN），the accuracy of the algorithm can be improved by up to 4.34%.The proposed algorithm can predict the accuracy of the user's next click better in a session.

**【Key words】**graph structure；Graph Neural Networks（GNN）；session sequence；recommending algorithm；attention mechanism

**DOI：**10.19678/j.issn.1000-3428.0061308

**基金项目：**南方海洋科学与工程广东省实验室（珠海）创新团队建设项目（311020012）。

**作者简介：**胡承佐（1996—），男，硕士研究生，主研方向为推荐系统；王庆梅（通信作者），副研究员；李迪超，博士研究生；王　铮，助理教授。

**收稿日期：**2021-03-29　　**修回日期：**2021-05-22　　**E-mail：**qmwang@ustb.edu.cn

## 0　概述

推荐系统是数据挖掘和机器学习领域最重要的应用之一，它能够帮助平台用户缓解信息过载的问题，并在电商平台、音乐网站等许多网页应用中挑选出有价值的信息。在大部分推荐系统中用户的行为序列是按照时间排列的，并且呈现出匿名性和大数据量的特征。为了预测用户在下一时刻的行为信息，基于会话序列的推荐通过挖掘用户历史行为中的序列顺序特征信息，从而学习用户的喜好[1]。会话序列是指在一段时间间隔内的由用户点击而产生的项目序列，而基于会话序列的推荐能捕捉到序列内部的依赖关系对序列预测的重要性[2]。用户在某一个会话序列中通常有着一个共同的目的，如购买下装衣物；而用户在不同序列之间的行为特性可能关联性不大，如在别的会话中用户的目的是购买手机配件等。

鉴于其较高的实际价值，基于会话序列的推荐在近年来得到了研究人员很大的关注，并且出现了许多具有良好效果的研究成果。早期的算法主要基于马尔科夫链和循环神经网络（Recurrent Neural Networks，RNN）。随着近期图神经网络（Graph Neural Networks，GNN）的兴起并且在许多下游任务中有着较好的表现[3-4]，有研究人员将GNN应用到了基于会话序列推荐中[5-6]。尽管这些基于GNN的算法有着较好的表现，但是这些算法也存在以下问题：忽略了点击序列中重复出现的项目，多次出现的项目与其他项目的重要程度是不同的，这些项目在一定程度上能够体现用户偏好信息；在生成项目的向量表示时没有较好地利用会话序列图中的结构信息，只考虑项目之间的方向性还有所欠缺，引入项目之间的无向关系能够更好地学习用户的行为信息。

为解决以上问题，本文提出一种基于复杂结构信息的图神经网络序列推荐算法。利用带注意力的图卷积网络和门控图神经网络，分别提取序列图中项目之间的无向结构信息与有向结构信息，并引入注意力机制建模项目之间的复杂转换，从而得到项目隐含向量，根据隐含向量利用注意力网络结合会话的全局信息与局部信息，生成准确的会话向量表示。

## 1　相关工作

### 1.1　传统推荐算法

早期的基于邻域建模算法在用户不匿名的推荐场景中使用最邻近算法进行会话序列预测[7-9]，需要测量项目之间或者会话之间的相似度。DAVIDSON等[7]提出一种通过项目共同出现的模式来计算项目之间相似度的算法，并根据待预测会话序列中的项目推荐最有可能跟其共同出现的项目。PARK等[9]提出一种将会话序列转换成向量的算法，然后计算会话向量之间的余弦相似度。DIAS等[8]基于PARK的研究提出用聚类算法将稀疏会话向量转换为稠密向量，然后再计算稠密向量之间的余弦相似度，但它受到数据稀疏性的影响，且没有考虑到会话向量内部项目之间的复杂转换关系。

而基于马尔科夫链的算法可以更好地获得序列中顺序信息。最简单的基于马尔科夫链的算法利用训练集中项目的转换频率来计算得到转换矩阵[10]，但不能应对那些在训练集中没出现过的转换关系。FPMC算法[11]通过一种张量分解的算法将转换矩阵进行分解，从而解决了该问题。另外一种解决算法是马尔科夫隐嵌入[12]，首先把项目映射到欧式空间中，然后通过计算项目之间的欧式距离从而估算项目之间的转换概率。因为状态空间存在着数据爆炸的问题，基于马尔科夫链的算法多数都只考虑了用户点击序列对连续项目之间的单向转换，导致会话中的其他项目被忽略。

### 1.2　基于深度学习的算法

循环神经网络（RNN）对会话序列有着强大的建模能力，能很好地解决基于马尔科夫链算法的不足。GRU4Rec[13]是一个基于RNN的会话序列推荐算法，它的原理是将多个GRU层堆叠在一起。受计算机视觉和自然语言处理领域中非常流行的注意力机制启发，LI等[14]采用带注意力的混合编码算法对用户的序列行为和目标进行建模，并且实验证明了学习到的序列表现结果十分有效。因此，后续基于RNN的工作都融入了注意力的机制[15-17]。

近几年图神经网络在许多任务中都有着较好的表现[18]，也有一些研究人员将图神经网络引入到基于会话序列的推荐中。SR-GNN[4]将会话序列建模为不带权重的有向会话图，图中的边代表项目之间的转换关系，然后利用门控图神经网络（Gated Graph Neural Networks，GGNN）在有边相连的节点间进行信息的传播。XU等[6]在SR-GNN的基础上利用GGNN来提取局部信息，并且用自注意力网络来捕获远距离项目之间的全局依赖关系。上述算法证明，对于基于会话序列的推荐，GNN是一个值得研究的方向。

本文的主要贡献如下：

1）利用会话序列建立会话图。根据会话图的邻接关系，利用图卷积网络提取图中的无向结构信息，通过门控图神经网络提取图中的有向结构信息，最后对中间项目隐含向量通过线性变换得到最终的项目隐含向量。

2）在提取会话图中的结构信息时，给会话序列中出现的重复点击项目分配更高的注意力，并在生成项目隐含向量时引入注意力机制，根据项目间依

赖的程度修改相应项目的权重系数。

3)进行大量的实验并在3个公开数据集上证明了该算法比现有的算法表现更优。

## 2 算法描述与模型框架

### 2.1 公式化描述

基于会话序列的推荐旨在根据用户当前的会话序列数据给出用户下个时刻点击项目的预测,而在该过程中完全不依赖用户的长期偏好信息。

在基于会话序列的推荐中,令 $V=\{v_i\}_{i=1}^{m}$ 代表所有会话序列中出现过的 $m$ 个项目的集合。那么一条长度为 $n$ 的匿名会话序列能够用列表 $s=[v_i]_{i=1}^{n}$ 来表示,且会话 $s$ 中的项目是按时间先后顺序排列的,每个 $v_i\in V$ 代表了用户在会话 $s$ 中点击的项目。基于会话序列的推荐就是要预测用户的下一个点击,即会话 $s$ 中的序列标签 $v_{n+1}$。利用基于会话序列的推荐模型,对每个会话 $s$ 都可以得到所有可能项目的概率 $\hat{\boldsymbol{y}}$,其中 $\hat{\boldsymbol{y}}=\{\hat{y}_1,\hat{y}_2,\cdots,\hat{y}_m\}$。概率向量 $\hat{\boldsymbol{y}}$ 中包括了出现在当前会话后下一个点击项目的所有可能情况,且每个元素的值都代表了对应项目的推荐得分,$\hat{\boldsymbol{y}}$ 中排名最高的前 $K$ 个的项目即为将要推荐的候选项目。

### 2.2 总体框架

对基于会话序列的推荐,首先从历史会话序列的信息中构建有向会话序列图。GCN和GNN分别能提取会话图中项目转换的无向结构信息和有向结构信息,并相应地生成精确的项目隐含向量。而后将得到的项目隐含向量输入到注意力网络中,同时考虑会话的全局信息与局部信息,从而构造出更可靠的会话表示,并以此推断下一次的点击项目。

模型的整体框架如图1所示。首先将每个会话序列 $s$ 转换为有向会话序列图 $G_s=(V_s,\varepsilon_s,\boldsymbol{A}_s)$,其中:$V_s$ 代表点集;$\varepsilon_s$ 代表边集;$\boldsymbol{A}_s$ 代表邻接矩阵的集合。在会话图 $G_s$ 中每个节点都代表一个项目 $v_i\in V$,而且每条边 $(v_{i-1},v_i)\in\varepsilon_s$ 都代表了用户先后点击了项目 $v_{i-1}$ 和项目 $v_i$。将 $\boldsymbol{A}_s$ 定义为3个邻接矩阵 $\boldsymbol{A}_s^{(\mathrm{in})}$、$\boldsymbol{A}_s^{(\mathrm{out})}$ 和 $\boldsymbol{A}_s^{(\mathrm{und})}$ 的拼接,其中:$\boldsymbol{A}_s^{(\mathrm{und})}$ 表示无向图的带权重邻接矩阵;$\boldsymbol{A}_s^{(\mathrm{in})}$ 和 $\boldsymbol{A}_s^{(\mathrm{out})}$ 分别表示带权重的入度邻接矩阵和出度邻接矩阵。然后依次对每个会话图 $G_s$ 进行处理,根据会话图 $G_s$ 生成每个项目 $v_i\in s$ 对应的初始项目隐含向量 $\boldsymbol{x}_i$,依次通过带无向注意力网络的图卷积网络、带有向注意力网络的门控图神经网络,分别得到每个图中涉及的所有节点的中间隐含向量,再通过一个线性层得到精确的最终的项目隐含向量,将得到的项目隐含向量输入目标注意力网络,从而得到每条会话序列所对应的会话隐含向量。最后通过线性变换和一个softmax层对每个会话预测所有可能项目被点击的概率。

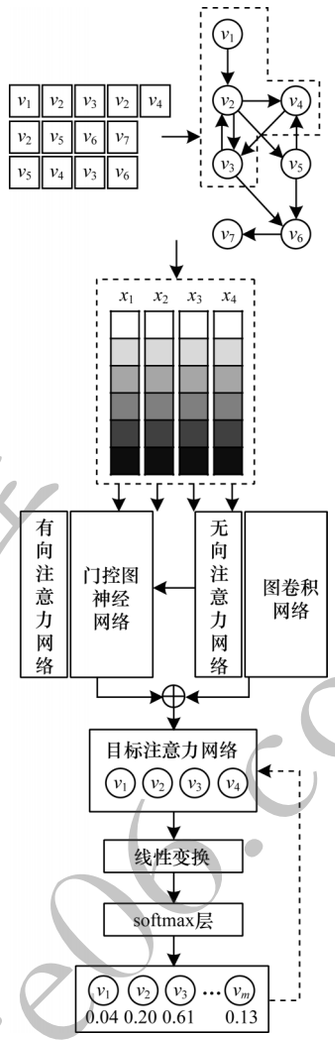



图1 本文模型总体框架

Fig.1 Overall framework of the proposed method

### 2.3 项目隐含向量

在建立好会话图 $G_s$ 之后,首先将每个节点 $v_i\in V$ 映射到随机嵌入向量空间中得到 $d$ 维向量表示 $\boldsymbol{x}_i\in\mathbb{R}^d$,再通过对应的神经网络模型和线性层得到 $\boldsymbol{h}_i\in\mathbb{R}^d$。图神经网络对基于会话序列的推荐有天然的适应性,因为它可以在考虑到丰富的节点连接关系的前提下自动提取会话图的特征。下面将介绍生成最终节点向量过程中使用的两个网络模型。

#### 2.3.1 带注意力的无向结构信息

本文采用文献[19]算法构建图卷积网络(GCN)。会话图 $G_s$ 中的带权重无向邻接矩阵 $\boldsymbol{A}_s^{\mathrm{und}}\in\mathbb{R}^{n\times n}$ 是一个稀疏且对称的邻接矩阵,其中,$a_{ij}$ 代表了节点 $v_i$ 和 $v_j$ 之间的边权重,节点间无相连关系则表示为 $a_{ij}=0$。将度矩阵 $\boldsymbol{D}$ 定义为对角矩阵 $\boldsymbol{D}=\mathrm{diag}(d_1,d_2,\cdots,d_n)$,且对角线上的值等于邻接矩阵的行元素之和 $d_i=\sum_j a_{ij}$。图中的每个节点 $v_i$ 都有对应的 $d$ 维特征向量 $\boldsymbol{x}_i\in\mathbb{R}^d$,所以总的特征矩阵 $\boldsymbol{X}\in\mathbb{R}^{n\times d}$ 就是图中每个特征向量的堆叠,即 $\boldsymbol{X}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n]^{\mathrm{T}}$。

与卷积神经网络(CNN)和多层感知机(MLP)类

似，GCN在多层结构中对于每个节点的特征$v_i$进行学习并得到新的特征表示，然后再输入对应的线性分类器。对于第$k$层的图卷积层，矩阵$\boldsymbol{H}^{(k-1)}$表示所有节点的输入向量，$\boldsymbol{H}^{(k)}$表示节点的输出向量。最初的$d$维节点向量即初始输入的特征，并输入到首层GCN中：

$$\boldsymbol{H}^{(0)}=\boldsymbol{X} \tag{1}$$

层数为$K$的GCN相当于对图中所有节点的特征向量$\boldsymbol{x}_i$应用一个$K$层的MLP模型，每个节点的隐含向量表示在每层的一开始都和其邻居节点进行均值化。在每个图卷积层中，节点的向量表示有3个更新阶段：特征传播，线性转换和逐点非线性激活。本文用于学习项目隐含向量所到的只有特征传播阶段。

特征传播是GCN和MLP之间的本质区别。在每层的最开始，每个节点$v_i$的特征都与它的局部邻居的特征向量进行均值化：

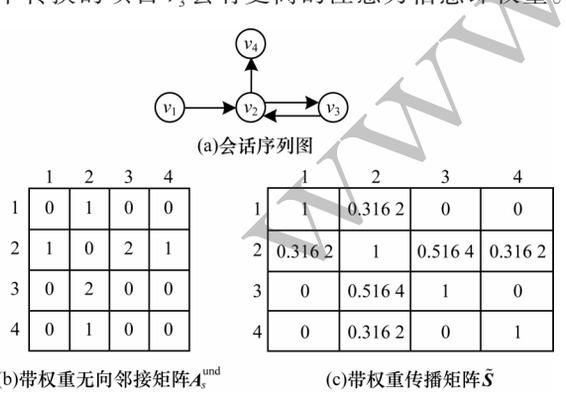
$$\bar{\boldsymbol{h}}_i^{(k)}\leftarrow\frac{1}{d_i+1}\boldsymbol{h}_i^{(k)}+\sum_{j=1}^{n}\frac{a_{ij}}{\sqrt{(d_i+1)(d_j+1)}}\boldsymbol{h}_j^{(k-1)} \tag{2}$$

上述的更新公式可以用整个图的简单矩阵操作来简化。$\boldsymbol{S}$代表“对称归一化”后带自环的邻接矩阵：

$$\boldsymbol{S}=\tilde{\boldsymbol{D}}^{-\frac{1}{2}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{D}}^{-\frac{1}{2}} \tag{3}$$

$$\tilde{\boldsymbol{S}}=\alpha(\boldsymbol{S}+\boldsymbol{I}-\boldsymbol{S}\odot\boldsymbol{I})+\beta\boldsymbol{I} \tag{4}$$

其中：$\tilde{\boldsymbol{A}}=\boldsymbol{A}_s^{\text{und}}+\boldsymbol{I}$并且$\tilde{\boldsymbol{D}}$是$\tilde{\boldsymbol{A}}$的度矩阵；$\odot$是点乘运算符。由于邻接矩阵$\boldsymbol{A}_s^{\text{und}}$中考虑了重复项目的权重，因此矩阵$\tilde{\boldsymbol{A}}$中带有对重复点击物品额外的注意力。同时由于$\boldsymbol{S}$带自环，对称归一化的过程导致多条边相连的项目权重相比单边相连或无边相连的项目权重更少。为了提高有边项目的权重并减少其他项目的噪声对其进行干扰，传播矩阵$\tilde{\boldsymbol{S}}$中对于有边相连的项目通过式(4)的左半部分提高权重，即提高矩阵中自我信息的注意力。最后用超参$\alpha$和超参$\beta$控制传播矩阵信息和单位阵信息的比例，从而控制传播过程中带注意力的节点信息的吸收比例。根据图2中给出的具体示例，可以看到邻接矩阵$\boldsymbol{A}_s^{\text{und}}$和传播矩阵$\tilde{\boldsymbol{S}}$中有重复点击的项目$v_2$，以及在重复点击过程中转换的项目$v_3$会有更高的注意力信息即权重。




(a)会话序列图

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 0 | 1 | 0 | 0 |
| 2 | 1 | 0 | 2 | 1 |
| 3 | 0 | 2 | 0 | 0 |
| 4 | 0 | 1 | 0 | 0 |

(b)带权重无向邻接矩阵$\boldsymbol{A}_s^{\text{und}}$

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1 | 0.316 2 | 0 | 0 |
| 2 | 0.316 2 | 1 | 0.516 4 | 0.316 2 |
| 3 | 0 | 0.516 4 | 1 | 0 |
| 4 | 0 | 0.316 2 | 0 | 1 |

(c)带权重传播矩阵$\tilde{\boldsymbol{S}}$

**图2　会话序列图与其对应的带权重无向邻接矩阵$\boldsymbol{A}_s^{\text{und}}$和带权重传播矩阵$\tilde{\boldsymbol{S}}$**

**Fig.2　Session sequence graph and its corresponding weighted undirected adjacency matrix $\boldsymbol{A}_s^{\text{und}}$ and weighted propagation matrix $\tilde{\boldsymbol{S}}$**

因此，式(2)的等价更新形式就可以变成一个对于所有节点的简单稀疏矩阵相乘：

$$\tilde{\boldsymbol{H}}^{(k)}=\tilde{\boldsymbol{S}}\boldsymbol{H}^{(k-1)} \tag{5}$$

上述步骤沿着图上边对节点的隐含向量表示进行局部平滑，将GCN作为特征预处理算法对特征进行传播后，使得节点能够吸收相邻节点的带注意力信息，并最终使得局部相连的节点能够有相似的预测表现[20]。

### 2.3.2　带注意力的有向结构信息

本文根据文献[21]构建模型GGNN。对于会话图$G_s$中的节点$v_i$，节点向量的更新公式如下：

$$\boldsymbol{a}_i^{(t)}=\hat{\boldsymbol{A}}_{i:}[\boldsymbol{h}_{v_1}^{(t-1)},\boldsymbol{h}_{v_2}^{(t-1)},\cdots,\boldsymbol{h}_{v_n}^{(t-1)}]^{\mathrm{T}}\boldsymbol{T}+\boldsymbol{b} \tag{6}$$

$$\boldsymbol{z}_i^t=\sigma(\boldsymbol{W}_z\boldsymbol{a}_i^{(t)}+\boldsymbol{U}_z\boldsymbol{h}_{v_i}^{(t-1)}) \tag{7}$$

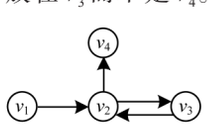
$$\boldsymbol{r}_i^t=\sigma(\boldsymbol{W}_r\boldsymbol{a}_i^{(t)}+\boldsymbol{U}_r\boldsymbol{h}_{v_i}^{(t-1)}) \tag{8}$$

$$\boldsymbol{h}_{v_i}^{(t)}=\tanh(\boldsymbol{W}_o\boldsymbol{a}_i^{(t)}+\boldsymbol{U}_o(\boldsymbol{r}_i^t\odot\boldsymbol{h}_{v_i}^{(t-1)})) \tag{9}$$

$$\boldsymbol{h}_{v_i}^{(t)}=(1-\boldsymbol{z}_i^t)\odot\boldsymbol{h}_{v_i}^{(t-1)}+\boldsymbol{z}_i^t\odot\boldsymbol{h}_{v_i}^{(t)} \tag{10}$$

其中：$\boldsymbol{T}\in\mathbb{R}^{d\times 2d}$和$\boldsymbol{b}\in\mathbb{R}^d$控制着权重和偏置项的大小；$\boldsymbol{z}_i^t\in\mathbb{R}^{d\times d}$和$\boldsymbol{r}_i^t\in\mathbb{R}^{d\times d}$分别是重置门和更新门；权重矩阵$\boldsymbol{W}_z$、$\boldsymbol{U}_z$，$\boldsymbol{W}_r$、$\boldsymbol{U}_r$和$\boldsymbol{W}_o$、$\boldsymbol{U}_o$分别代表了重置门、更新门和输出门中可学习的网络参数；$\boldsymbol{h}_{v_i}\in\mathbb{R}^d$代表节点$v_i$的隐含向量；$[\boldsymbol{h}_{v_1}^{(t-1)},\boldsymbol{h}_{v_2}^{(t-1)},\cdots,\boldsymbol{h}_{v_n}^{(t-1)}]$是会话中的节点向量序列，且$[\boldsymbol{h}_{v_1}^{(0)},\boldsymbol{h}_{v_2}^{(0)},\cdots,\boldsymbol{h}_{v_n}^{(0)}]=[\bar{\boldsymbol{h}}_{v_1},\bar{\boldsymbol{h}}_{v_2},\cdots,\bar{\boldsymbol{h}}_{v_n}]$，即将GCN模型的最终输出作为GGNN模型的初始输入；$\sigma(\cdot)$是sigmoid函数；$\odot$是点乘运算符；邻接矩阵$\hat{\boldsymbol{A}}\in\mathbb{R}^{n\times 2n}$代表图中节点的交流；$\hat{\boldsymbol{A}}_{i:}\in\mathbb{R}^{1\times 2n}$代表节点$v_i$在$\hat{\boldsymbol{A}}$中的两列矩阵块。

矩阵$\hat{\boldsymbol{A}}$定义为入度矩阵$\boldsymbol{A}_s^{(\text{in})}$和出度矩阵$\boldsymbol{A}_s^{(\text{out})}$的拼接，它们分别代表会话图中输入边和输出边的加权连接。例如，给定会话序列$s=[v_1,v_2,v_3,v_2,v_4]$，对应的会话图$G_s$和邻接矩阵$\hat{\boldsymbol{A}}$如图3所示。可以看到有向邻接矩阵中的加权是根据节点之间的紧密联系程度设置的，例如$v_2$从$v_3$到$v_4$各有一条边，但是两者权重不同的原因是$v_2$与$v_3$之间相连的边更多，代表两者之间相似度更高。为达到更好的预测效果，从$v_2$的角度来讲应该更多地吸收$v_3$的信息，所以模型要把更多的注意力放在$v_3$而不是$v_4$。




(a)会话序列图

| | 出度 | | | | 入度 | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 1/2 | 1/2 | 1/2 | 0 | 1/2 | 0 |
| 3 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |

(b)邻接矩阵$\hat{\boldsymbol{A}}$

**图3　会话序列图与其对应的邻接矩阵$\hat{\boldsymbol{A}}$**

**Fig.3　Session sequence graph and its corresponding adjacency matrix $\hat{\boldsymbol{A}}$**

所以，对于每个会话图$G_s$，GGNN模型在相邻节点之间传播带注意力的节点信息，而重置门控和更新门控分别决定需要进行舍弃或者保留的信息。

2.3.3 项目隐含向量的生成

在经过图卷积网络（GCN）和门控图神经网络（GGNN）的信息处理后，分别得到$\tilde{\boldsymbol{H}}=\mathrm{GCN}(\boldsymbol{X})$，$\hat{\boldsymbol{H}}=\mathrm{GGNN}(\tilde{\boldsymbol{H}})$。前者是对初始嵌入向量进行了带注意力的无向结构信息处理；后者在前者的基础上更加精细地提取图结构中带注意力的有向结构信息。

为了平衡带注意力的无向结构信息与有向结构信息的比例，采用式（11）进行控制：

$$\boldsymbol{H}=\gamma\hat{\boldsymbol{H}}+(1-\gamma)\tilde{\boldsymbol{H}} \tag{11}$$

其中：$\gamma$为超参数。

### 2.4 会话向量表示的生成

2.4.1 基于目标注意力的向量

在得到每个项目的向量表示后，进一步构建目标向量，从而能在考虑到目标项目的前提下对历史行为的相关性进行分析，目标项目是指所有待预测的候选项目。因为在实际应用场景中，用户得到的推荐项目只匹配其一小部分的兴趣，所以利用文献[22]提出的目标注意力模型来计算目标会话中所有项目对目标项目的注意力得分。

本文利用局部目标注意力模型来计算会话$s$中所有项目$v_i$对每个目标项目$v_t\in V$的注意力得分$\beta_{i,t}$：

$$\beta_{i,t}=\mathrm{softmax}(e_{i,t})=\frac{\exp(\boldsymbol{h}_{v_i}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{h}_{v_t})}{\sum_{j=1}^{m}\exp(\boldsymbol{h}_{v_i}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{h}_{v_t})} \tag{12}$$

其中：$\boldsymbol{h}_{v_i}\in\mathbb{R}^d$和$\boldsymbol{h}_{v_t}\in\mathbb{R}^d$分别为项目$v_i$与$v_t$的隐含向量表示。

在式（12）中，会话中的项目与候选目标分别匹配，并且用带权重矩阵$\boldsymbol{W}\in\mathbb{R}^{d\times d}$来进行成对的非线性转换。然后再用softmax函数对得到的自注意力分数进行归一化，并得到最后的注意力分数。

对于每个会话序列$s$，用户对目标项目$v_t$的兴趣可以表示为：

$$\boldsymbol{s}_{\mathrm{target}}^{t}=\sum_{i=1}^{s_n}\beta_{i,t}\boldsymbol{h}_{v_i} \tag{13}$$

最终得到基于目标注意力的向量$\boldsymbol{s}_{\mathrm{target}}^{t}$，$\boldsymbol{s}_{\mathrm{target}}^{t}\in\mathbb{R}^d$，它代表了用户对不同目标项目之间产生的兴趣程度。

2.4.2 会话向量的生成

本文利用会话$\boldsymbol{s}$中涉及到的项目向量进一步地探索用户的短期和长期喜好，从而得到会话中的局部向量与全局向量，并综合2.4.1节中计算得出的基于目标注意力的向量生成最终的会话向量。

首先是局部向量，在一个会话序列$\boldsymbol{s}$中，用户最终的行为通常是由当前序列中最后一个交互的项目决定的。所以，将用户的短期兴趣表现为局部向量$\boldsymbol{s}_{\mathrm{local}}\in\mathbb{R}^d$，且该局部向量即为会话序列中最后一个项目$v_{s_n}$的向量表示。

$$\boldsymbol{s}_{\mathrm{local}}=\boldsymbol{h}_{v_{s_n}} \tag{14}$$

对于全局向量，将用户的长期偏好定义为全局向量$\boldsymbol{s}_{\mathrm{global}}\in\mathbb{R}^d$，其聚合了会话$\boldsymbol{s}$中所有出现的项目向量。同时，利用注意力机制来引入最后交互的项目$v_{s_n}$与整个会话中出现的项目$[v_1,v_2,\cdots,v_n]$之间的依赖关系。

$$\boldsymbol{\alpha}_i=\boldsymbol{q}^{\mathrm{T}}\sigma(\boldsymbol{W}_1\boldsymbol{h}_{v_{s_n}}+\boldsymbol{W}_2\boldsymbol{h}_{v_i}+\boldsymbol{c}) \tag{15}$$

$$\boldsymbol{s}_{\mathrm{global}}=\sum_{i=1}^{s_n}\boldsymbol{\alpha}_i\boldsymbol{h}_{v_i} \tag{16}$$

其中：$\boldsymbol{q}$，$\boldsymbol{c}\in\mathbb{R}^d$且$\boldsymbol{W}_1$，$\boldsymbol{W}_2\in\mathbb{R}^{d\times d}$是相应的权重参数。

最后对于前面得到局部向量、全局向量和基于目标注意力的向量，将三者进行拼接并利用线性转换得到会话序列$\boldsymbol{s}$所对应的会话向量。

$$\boldsymbol{s}_h=\boldsymbol{W}_3[\boldsymbol{s}_{\mathrm{target}}^{t};\ \boldsymbol{s}_{\mathrm{local}};\ \boldsymbol{s}_{\mathrm{global}}] \tag{17}$$

其中：权重参数$\boldsymbol{W}_3\in\mathbb{R}^{d\times 3d}$，将3个向量拼接的结果投射到向量空间$\mathbb{R}^d$中。值得注意的是，对不同的目标项目会对应地生成不同的会话向量。

### 2.5 推荐生成

在得到每个会话序列$\boldsymbol{s}$对应的会话向量$\boldsymbol{s}_h$之后，对于所有的候选目标项目$v_i\in V$的得分$\hat{\boldsymbol{z}}_i$进行计算，即将候选项目向量$\boldsymbol{h}_{v_i}$与会话向量$\boldsymbol{s}_h$进行相乘，得到：

$$\hat{\boldsymbol{z}}_i=\boldsymbol{s}_h^{\mathrm{T}}\boldsymbol{h}_{v_i} \tag{18}$$

然后通过softmax函数，获得模型的输出向量$\hat{\boldsymbol{y}}$：

$$\hat{\boldsymbol{y}}=\mathrm{softmax}(\hat{\boldsymbol{z}}_i) \tag{19}$$

其中：$\hat{\boldsymbol{z}}\in\mathbb{R}^m$代表所有候选目标项目的预测推荐得分；$\hat{\boldsymbol{y}}\in\mathbb{R}^m$代表目标项目在会话序列$\boldsymbol{s}$下一时刻被点击的概率。

对于每个会话图$G_s$，将损失函数定义为预测值与实际值的交叉熵：

$$L(\hat{\boldsymbol{y}})=-\sum_{i=1}^{m}\boldsymbol{y}_i\log_a(\hat{\boldsymbol{y}}_i)+(1-\boldsymbol{y}_i)\log_a(1-\hat{\boldsymbol{y}}_i) \tag{20}$$

其中：$\boldsymbol{y}$代表会话序列下一时刻真实点击项目的独热编码向量。

最后使用基于时间的反向传播（BPTT）算法来训练提出的模型。值得注意的是，在基于会话序列的推荐场景中，多数会话都是相对较短的序列。为了防止过拟合的出现，采用较小的训练次数是比较适宜的。

## 3 实验结果与分析

本节介绍使用的数据集、数据预处理策略和评价指标，将提出算法与其他算法进行比较，最后在不同的实验设置下给出模型的详细分析。

### 3.1 数据集和数据预处理

本文选择在实际应用中两个有代表性的数据集来评估所提出算法，数据集分别是RecSys Challenge 2015发布的公开数据集Yoochoose和CIKM Cup 2016发布的公开数据集Diginetica。Yoochoose数据集包含了电子购物平台上6个月内的用户点击流，而Diginetica数

据集中只包含了交易成功的数据，即用户的购买流。

为了公平比较，本文遵循文献[14，23]的预处理算法，在两个数据集中将长度为1的会话序列和总出现次数小于5次的项目滤去，同时遵循文献[1]的预处理算法，将长度大于20的会话序列滤去。因为一个会话序列如果长度过长，那么用户的主要目的在项目转换间很有可能已经发生了改变，后续的推荐如果基于若干目的中的一个进行推荐，则很难达到精准预测的效果。经过上述处理，最终在Yoochoose数据集中得到了7 897 532条会话和37 470个项目，在Diginetica数据集中得到了185 517条会话和43 093个项目。对于训练集和测试集的划分，在Yoochoose数据集中选择最后一天的数据作为测试集，在Diginetica数据集中则选择最后一个星期的数据进行测试，剩余的数据则作为训练集输入模型。由于Yoochoose训练集规模过于庞大，遵循文献[14，23]的做法，将Yoochoose训练数据中距离测试集时间最近的1/64和1/4部分的数据划分出来作为训练数据。

和文献[24]的策略相同，本文进一步通过切分输入序列数据来生成对应的序列和标签。对于输入会话序列$\boldsymbol{s}=[v_1,v_2,\cdots,v_n]$，作为一种数据增强策略，生成一系列的序列和标签$([v_1],v_2),([v_1,v_2],v_3),\cdots,([v_1,v_2,\cdots,v_{n-1}],v_n)$，其中$[v_1,v_2,\cdots,v_{n-1}]$是生成的序列，而$v_n$代表了下一时刻点击的项目，即序列的标签。

数据集具体情况如表1所示。

**表1　实验数据集统计结果**

**Table 1　Statistical results of experimental datasets**

| 数据集 | 点击数 | 训练集 | 测试集 | 项目数 | 平均长度 |
|---|---|---|---|---|---|
| Diginetica | 876 847 | 691 330 | 58 903 | 43 093 | 4.73 |
| Yoochoose 1/64 | 475 331 | 337 829 | 48 337 | 17 026 | 3.85 |
| Yoochoose 1/4 | 7 565 283 | 5 405 277 | 48 337 | 30 282 | 3.83 |

### 3.2　评价指标

本文采用在基于会话序列的推荐中常用的两个度量标准作为算法的评价指标：

1）P@20（Precision）。是一种被广泛使用的预测精度的度量标准，它代表了算法推荐结果的前20项中正确推荐的比例。

2）MRR@20（Mean Reciprocal Rank）。是算法推荐结果中正确推荐项目的倒数排名均值。当真实结果在算法的推荐排位中超过20时，对应的倒数排名为0，MMR度量标准是一种考虑了推荐顺位的算法，较大的MRR值代表了在推荐列表中真实结果位于排名列表的顶部，这也证明了推荐系统的有效性。

### 3.3　参数设置

根据文献[4，14，22-23]，本文在两个数据集中均将隐含向量的维度设置为$d=100$。所有超参设置都利用均值为0、标准差为0.1的高斯分布函数进行初始化。同时还采用小批量Adam优化器对这些参数进行优化，并且将初始的学习率$\eta$设置为0.001，且每3个训练周期衰减0.1。此外，批处理大小设置为100，L2正则化参数设置为$10^{-5}$。

### 3.4　与相关算法的比较

为了评估所提算法的性能，将其与会话序列推荐问题的现有算法中有代表性的算法进行比较，即POP、S-POP、Item-KNN、BPR-MF、FPMC、GRU4Rec、RepeatNet、NARM、STAMP、SR-GNN、TAGNN。

1）POP和S-POP算法的策略是分别在训练集和当前会话序列中推荐前$K$个出现频率最高的项目。

2）Item-KNN[25]推荐与当前会话序列项目相似的项目，其中相似度定义为会话向量之间的余弦相似度。

3）BPR-MF[26]是一种基于贝叶斯后验优化的个性化排序算法，它通过随机梯度下降优化成对项目排序的目标函数。

4）FPMC[11]是一种基于马尔科夫链的序列推荐算法。

5）GRU4Rec[13]使用循环神经网络来对用户序列进行建模。

6）RepeatNet[16]在重复消费的场景下将常规神经推荐算法与新的重复推荐机制集成在一起。

7）NARM[14]使用带有注意力机制的循环神经网络来捕捉用户的主要意图和序列行为特征。

8）STAMP[23]利用自注意力机制捕捉用户在当前会话序列中的大致意图和最后一次点击行为的兴趣点所在。

9）SR-GNN[4]通过使用门控图神经网络来捕捉项目转换之间的关系。

10）TAGNN[20]在使用图神经网络模型的基础上利用注意力网络捕捉会话序列中的项目与目标项目的相似程度。

各算法在P@20和MRR@20这2个指标上的性能表现如表2所示，其中加粗数字为最佳结果。本文提出算法能够灵活地在会话图上构建项目之间的联系，并且提取其中带注意力的有向结构信息和无向结构信息，使得后续目标注意力的学习能够更加准确，并且综合用户在会话中的全局兴趣和局部兴趣给出最后的推荐。由表2中的实验数据可以看

出，本文算法在3个数据集上的2个指标上都获得了最好的表现结果，这也证明了所提算法的有效性。

**表2 不同算法实验结果对比**

**Table 2 Comparison of experimental results of different algorithms**

| 算法 | Diginetica | | Yoochoose 1/64 | | Yoochoose 1/4 | |
|---|---|---|---|---|---|---|
| | P@20 | MRR@20 | P@20 | MRR@20 | P@20 | MRR@20 |
| POP | 1.18 | 0.28 | 7.31 | 1.69 | 1.37 | 0.31 |
| S-POP | 21.06 | 13.68 | 30.44 | 18.35 | 27.08 | 17.75 |
| Item-KNN | 35.75 | 11.57 | 51.60 | 21.81 | 52.31 | 21.70 |
| BPR-MF | 5.24 | 1.98 | 31.31 | 12.08 | 3.40 | 1.57 |
| FPMC | 22.14 | 6.66 | 45.62 | 15.01 | 51.86 | 17.50 |
| GRU4Rec | 30.79 | 8.22 | 60.64 | 22.89 | 59.53 | 22.60 |
| RepeatNet | 48.49 | 17.13 | 70.06 | 30.55 | 70.71 | 31.03 |
| NARM | 48.32 | 16.00 | 68.37 | 28.87 | 69.73 | 29.23 |
| STAMP | 46.62 | 15.13 | 68.74 | 28.67 | 70.44 | 30.00 |
| SR-GNN | 50.73 | 17.59 | 70.57 | 30.94 | 71.36 | 31.89 |
| TAGNN | 51.31 | 18.03 | 71.02 | 31.12 | 69.33 | 29.48 |
| 本文算法 | **53.35** | **18.45** | **72.00** | **32.47** | **72.34** | **32.73** |

从表2可以看出：传统推荐算法如POP和S-POP在基于会话序列的问题上表现不尽如人意，因为其忽略了用户在当前会话中的偏好，仅考虑了前$K$个最受欢迎的项目。BPR-MF说明利用会话中的语义信息具有一定意义，而表现更好的FPMC则说明利用一阶马尔科夫链来建模会话序列是相对有效的算法。同样作为传统推荐算法，Item-KNN比前两者更优。值得注意的是，Item-KNN仅依赖计算物品之间的相似度，这说明了物品的同时出现也是一种比较重要的信息。而Item-KNN没有考虑到会话中的时序信息，不能捕获到物品之间转换的信息。

与传统算法不同，基于深度学习的算法在所有数据集上的结果都有更好的表现。GRU4Rec是一种基于循环神经网络的算法，它的表现结果优于大部分传统算法，与一部分传统算法达到相近的程度。这说明了循环神经网络对于序列数据有着一定的建模能力。然而GRU4Rec主要聚焦在对会话序列进行建模，无法捕获会话中的用户偏好，其后出现的算法如NARM和STAMP都对GRU4Rec有着显著的提升。NARM显式地捕获用户在会话中的主要偏好，而STAMP利用注意力机制考虑用户的短期兴趣，也优于GRU4Rec。RepeatNet通过考虑用户的重复点击行为达到了较好的预测效果，这说明对用户的行为习惯进行建模具有一定的重要性。RepeatNet相比NARM与STAMP提升有限，可能是因为仅通过项目特征来对用户的重复点击习惯进行建模是不充分的，且基于RNN的结构无法捕获会话内的一些共同的依赖关系。

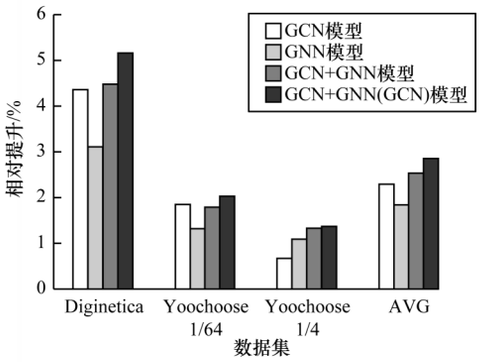
基于图神经网络的算法将每个会话序列都构建成一张子图，并且通过图神经网络对会话中的所有项目进行编码。SR-GNN和TAGNN比所有基于RNN的模型有着更好的结果。SR-GNN利用门控图神经网络来学习会话序列内项目之间的依赖关系，TAGNN进一步利用注意力机制挖掘会话内的项目与目标项目之间的依赖关系。然而，这些算法都完全根据会话图的有向关系进行学习，而没有综合考虑会话图中的无向关系，因为会话序列中物品和物品之间的关系有时往往不是单向的关系而是双向的，利用无向的结构信息能够捕获到物品之间更加全面的关系，且它们都忽视了会话序列中出现的重复点击特征，从直觉上来讲一个序列中重复出现项目的重要性应当是更大的。此外，在实际推荐场景中项目和项目之间的关联程度是多变的，而这些算法对于会话内项目和项目之间的依赖关系采用的是平均化算法，不能通过权重或者注意力的算法体现出某一项目对其他项目的依赖程度。

本文提出算法相比其他算法表现都要更优。具体而言，在3个数据集上，对于P@20，相对表现最佳的相关算法有3.55%、1.38%、1.37%的相对提升，对于MRR@20，对表现最佳的相关算法有1.92%、4.34%、2.63%的相对提升。本文算法能够很好地提取会话图中的结构信息，先后利用图卷积网络和门控图神经网络对图中的无向结构信息和有向结构信息进行提取，并将两者进行线性组合从而达到向量的精准表示。考虑到会话序列中的重复点击项目，通过注意力网络提高重复信息的权重，同时通过增加自环和矩阵操作提高会话图中节点的自我信息比例，使得节点不容易受到其他节点的噪声干扰。根据项目与项目之间不同的依赖关系，利用注意力网络分配不同的权重，从而使得网络能够生成精确的向量表示。

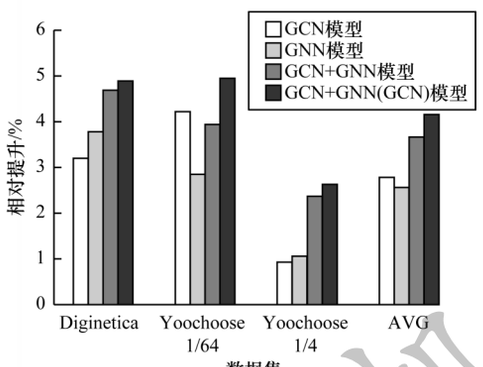
### 3.5 消融实验

本文提出的算法能够灵活地捕捉会话图中的结构信息和项目之间的关系。为了验证模型中各组成成分的实际作用，设置了几种模型变体进行消融实验。在实验环节中选择SR-GNN[4]作为对比的基准算法，实验中的数据以对比SR-GNN的相对提升百分比的形式展示。

本文进行以下有向结构信息和无向结构信息的组合分析：1）GCN，只提取会话图中的无向结构信息；2）GNN，只提取会话图中的有向结构信息；3）GCN+GNN，将随机初始向量同时输入到两个神经网络中，然后把模型输出结果进行线性组合；4）GCN+GNN（GCN），首先将随机初始向量输入到GCN中，然后把GCN模型的输出向量作为GNN模型的输入，最后将2个模型的输出结果进行线性组合。实验对比结果如图4所示，其中AVG代表4种组合条件在3个数据集上的平均表现。




(a)P@20指标上不同组成成分的比较结果



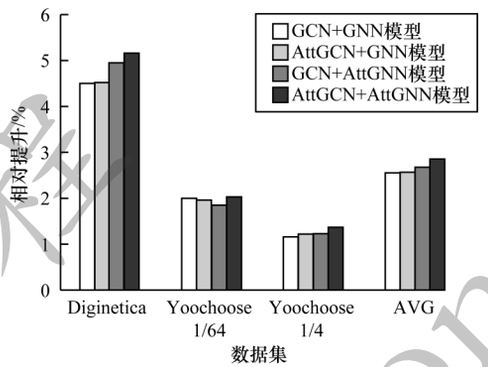

(b)MRR@20指标上不同组成成分的比较结果

**图4　不同结构信息的实验结果**

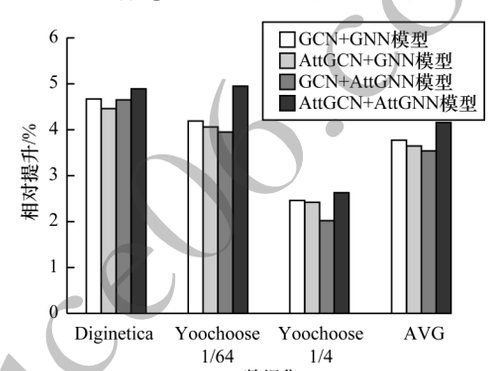
**Fig.4　Experimental results of different structural information**

从图4可以看出，综合有向结构信息与无向结构信息的GCN+GNN（GCN）模型在3个数据集的2个指标P@20、MRR@20上都取得了最佳的结果，这证明了综合考虑有向结构信息和无向结构信息的重要性。图4中平均数据AVG还显示了单独考虑无向结构信息相比单独考虑有向结构信息表现更加优异。从单个数据集的表现上可以看出，在Yoochoose 1/4数据集和Diginetica数据集的MRR@20指标上，考虑有向结构信息表现稍好于无向结构信息。这一定程度上反映了在基于会话序列的推荐中用户的偏好与项目之间的联系、项目之间的转换方向在不同的场景下有着不同的重要程度，但平均而言还是无向结构信息更重要。这说明在基于会话序列推荐的场景下用户在项目之间的转换方向虽然值得考虑，但还是需要重点考虑用户浏览的项目之间的联系，才能更好地学习到用户的偏好。而比前两者更好的做法是综合考虑有向结构信息和无向结构信息，图4中GCN结合GNN的方法在3个数据集的综合表现及平均表现AVG上，基本都要比单独使用GCN或GNN的表现要好。而其中GCN+GNN的方法中2个网络模型的输入数据都是随机嵌入向量，而GCN+GNN（GCN）的方法中GNN模型的输入是经过GCN模型提取无向结构信息的向量，这说明了相比直接使用随机向量，先对无向结构信息进行提取而后再进行有向结构信息的提取能够得到更精准表示的嵌入向量。

本文进行重复点击注意力信息和项目间依赖关系的组合分析如下：1）GCN+GNN，不考虑重复点击注意力信息和项目间不同的依赖关系；2）AttGCN+GNN，只在GCN中考虑重复点击项目的注意力信息；3）GCN+AttGNN，只在GNN中考虑项目之间不同程度的依赖关系；4）AttGCN+AttGNN，同时在GCN中融合重复点击的注意力信息和在GNN中融合带注意力的项目依赖关系。实验结果如图5所示。




(a)P@20指标上不同组成成分的表现




(b)MRR@20指标上不同组成成分的表现

**图5　不同注意力信息组成成分的实验结果**

**Fig.5　Experimental results of different attention information composition**

从图5可以看出，综合考虑重复点击注意力信息和项目间依赖关系的AttGCN+AttGNN在3个数据集的2个指标上均取得了最佳的实验结果，这说明重复点击行为和项目间的依赖关系在基于会话序列的推荐中有一定的重要性。根据图5（a），单独考虑重复点击注意力和项目间关系的注意力比不考虑任何注意力信息的GCN+GNN要有更好的表现，综合考虑两者的AttGCN+AttGNN效果最佳，说明了注意力信息确实能够使得重要的信息尽可能地保留并得到更加精准表现的嵌入向量。根据图5（b），虽然综合考虑两种注意力的AttGCN+AttGNN仍能获得最佳的实验表现，但是单独考虑两者注意力其中之一的表现会略差于不考虑任何注意力信息的GCN+GNN的表现，即单独使用注意力信息虽然能够提高推荐结果的精确率，但是对推荐排名的预测效果并不好。可能是向量在输入GCN和GNN模型时，如果一个模型考虑了注意力而另一个模

型没有考虑,那么2个模型中邻接矩阵的表现模式也就没有统一,导致向量在两个模型之间输入输出时不能同时利用注意力保留结构信息,反而使得结构信息因为注意力模式不一致受到干扰,所以最后没有生成精准表示的嵌入向量,即不能在预测阶段计算出每个项目准确的预测得分。

## 4 结束语

在基于会话序列的推荐场景中,用户的重复点击行为和图结构信息都能在不知用户历史偏好的情况下很好地预测出用户的行为。本文提出一种基于有向与无向信息同注意力相融合的图神经网络序列推荐算法。利用GCN与GNN模型提取会话序列图中的有向结构信息与无向结构信息并进行线性组合,在2个模型的内部引入注意力机制,对用户的重复点击和项目之间的复杂转换信息进行有效提取,使得生成的会话向量在推荐过程中预测更准确。在3个数据集上的实验结果表明,该算法精度优于目前现有的算法,并验证了注意力机制和复杂的结构信息的有效性。下一步将在研究复杂图结构的基础上,寻找更好的隐含向量表现方式和向量组合方式,提高会话序列推荐的准确率。

## 参考文献


[1] CHEN T W, WONG R C W. Handling information loss of graph neural networks for session-based recommendation[C]// Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA: ACM Press, 2020: 1172-1180.

[2] CHENG Z Y, SHEN J L, ZHU L, et al. Exploiting music play sequence for music recommendation[C]//Proceedings of IEEE IJCAI'17. Washington D. C., USA: IEEE Press, 2017: 3654-3660.

[3] QIU R H, LI J J, HUANG Z, et al. Rethinking the item order in session-based recommendation with graph neural networks[C]//Proceedings of the 28th ACM International Conference on Information and Knowledge Management. New York, USA: ACM Press, 2019: 579-588.

[4] WU S, TANG Y Y, ZHU Y Q, et al. Session-based recommendation with graph neural networks[EB/OL]. [2021-02-20]. https://arxiv.org/abs/1811.00855.

[5] 王健宗,孔令炜,黄章成,等. 图神经网络综述[J]. 计算机工程,2021,47(4):1-12.
WANG J Z, KONG L W, HUANG Z C, et al. Survey of graph neural network[J]. Computer Engineering, 2021, 47(4): 1-12. (in Chinese)

[6] XU C F, ZHAO P P, LIU Y C, et al. Graph contextualized self-attention network for session-based recommendation[C]//Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao, China: [s. n.], 2019: 3940-3946.

[7] DAVIDSON J, LIEBALD B, LIU J N, et al. The YouTube video recommendation system[C]//Proceedings of the 4th ACM Conference on Recommender Systems. New York, USA: ACM Press, 2010: 293-296.

[8] DIAS R, FONSECA M J. Improving music recommendation in session-based collaborative filtering by using temporal context[C]//Proceedings of the 25th IEEE International Conference on Tools with Artificial Intelligence. Washington D. C., USA: IEEE Press, 2013: 783-788.

[9] PARK S E, LEE S, LEE S G. Session-based collaborative filtering for predicting the next song[C]//Proceedings of the 1st ACIS/JNU International Conference on Computers, Networks, Systems and Industrial Engineering. Washington D. C., USA: IEEE Press, 2011: 353-358.

[10] SHANI G, BRAFMAN R I, HECKERMAN D. An MDP-based recommender system[EB/OL]. [2021-02-20]. https://arxiv.org/ftp/arxiv/papers/1301/1301.0600.pdf.

[11] RENDLE S, FREUDENTHALER C, SCHMIDT-THIEME L. Factorizing personalized Markov chains for next-basket recommendation[C]//Proceedings of the 19th IEEE International Conference on World Wide Web. Washington D. C., USA: IEEE Press, 2010: 811-820.

[12] CHEN S, MOORE J L, TURNBULL D, et al. Playlist prediction via metric embedding[C]//Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA: ACM Press, 2012: 714-722.

[13] HIDASI B, KARATZOGLOU A, BALTRUNAS L, et al. Session-based recommendations with recurrent neural networks[EB/OL]. [2021-02-20]. https://arxiv.org/abs/1511.06939.

[14] LI J, REN P J, CHEN Z M, et al. Neural attentive session-based recommendation[C]//Proceedings of 2017 ACM Conference on Information and Knowledge Management. New York, USA: ACM Press, 2017: 1419-1428.

[15] CHEN T W, WONG R C W. Session-based recommendation with local invariance[C]//Proceedings of 2019 IEEE International Conference on Data Mining. Washington D. C., USA: IEEE Press, 2019: 994-999.

[16] REN P J, CHEN Z M, LI J, et al. RepeatNet: a repeat aware neural recommendation machine for session-based recommendation[C]//Proceedings of AAAI Conference on Artificial Intelligence. [S. 1.]: AAAI Press, 2019: 4806-4813.

[17] SONG J, SHEN H, OU Z J, et al. ISLF: interest shift and latent factors combination model for session-based recommendation[C]//Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao, China: [s. n.], 2019: 5765-5771.

[18] 呼延康,樊鑫,余乐天,等. 图神经网络回归的人脸超分辨率重建[J]. 软件学报,2018,29(4):914-925.
HUYAN K, FAN X, YU L T, et al. Graph based neural network regression strategy for facial image super-resolution[J]. Journal of Software, 2018, 29(4): 914-925. (in Chinese)

[19] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks[EB/OL]. [2021-02-20]. https://arxiv.org/abs/1609.02907.

[20] WU F, ZHANG T Y, SOUZA A H J, et al. Simplifying graph convolutional networks[EB/OL]. [2021-02-20]. https://arxiv.org/abs/1902.07153.

[21] LI Y J, TARLOW D, BROCKSCHMIDT M, et al. Gated graph sequence neural networks[EB/OL]. [2021-02-20]. https://arxiv.org/abs/1511.05493.


(下转第97页)

（上接第90页）

[22] YU F，ZHU Y Q，LIU Q，et al. TAGNN：target attentive graph neural networks for session-based recommendation [C]//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York，USA：ACM Press，2020：1921-1924.

[23] LIU Q，ZENG Y F，MOKHOSI R，et al. STAMP：short-term attention/memory priority model for session-based recommendation [C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York，USA：ACM Press，2018：1831-1839.

[24] TAN Y K，XU X X，LIU Y. Improved recurrent neural networks for session-based recommendations [C]//Proceedings of the 1st IEEE Workshop on Deep Learning for Recommender Systems. Washington D. C.，USA：IEEE Press，2016：17-22.

[25] SARWAR B，KARYPIS G，KONSTAN J，et al. Item-based collaborative filtering recommendation algorithms [C]// Proceedings of the 10th International Conference on World Wide Web. Washington D. C.，USA：IEEE Press，2001：285-295.

[26] RENDLE S，FREUDENTHALER C，GANTNER Z，et al. BPR：Bayesian personalized ranking from implicit feedback [EB/OL]. [2021-02-20]. https://arxiv. org/abs/1205. 2618.

编辑　索书志