

会话场景下基于特征增强的图神经推荐方法

黄震华¹⁾林小龙¹⁾孙圣力²⁾汤庸¹⁾陈运文³⁾

¹⁾(华南师范大学计算机学院广州 510631)

²⁾(北京大学软件与微电子学院北京 102600)

³⁾(达而观智能(深圳)有限公司研发部广东深圳 518063)

摘要 基于图神经网络的会话推荐(简称图神经会话推荐)是近年来推荐系统领域的一个研究重点和热点,这主要是因为它们引入了会话图拓扑结构信息来提高物品和会话特征表示的准确性,因而,在一定程度上提升了会话推荐的性能。然而,现有图神经会话推荐方法仍然存在两方面的不足,从而影响其性能:1)它们所构建的会话图中物品间的相关性权重均是在模型训练之前就预先指定并保持固定不变,导致不能准确捕捉物品间的相关性;2)它们只从单个会话的物品序列中产生物品和会话的局部特征,而缺乏从整个会话数据集出发,全局考虑不同物品之间以及不同会话之间的相关性,并由此来生成物品和会话的全局特征,从而充分表示物品和会话的语义特征。为此,本文提出了一种新颖的会话场景下基于特征增强的图神经推荐方法 FA-GNR (Feature Augmentation based Graph Neural Recommendation)。FA-GNR 方法首先基于单个会话构建物品间相关性权重可学习优化的会话图,并借鉴 GRU (Gated Recurrent Unit) 神经网络来产生物品局部特征,同时基于会话数据集,通过 GloVe (Global Vectors) 词嵌入方法产生物品全局特征,从而融合物品的局部和全局特征来生成其语义特征。然后,FA-GNR 方法基于物品语义特征,利用局部注意力机制来产生会话的局部特征,同时基于物品的全局特征,并通过全局注意力机制来产生会话的全局特征,从而融合会话的局部和全局特征来生成其语义特征。最后,在物品和会话语义特征的基础上,FA-GNR 方法通过交叉熵损失来学习给定会话下不同物品的点击概率分布。在多个公开数据集上的实验结果表明,FA-GNR 方法的推荐性能优于目前主流的方法。

关键词 会话推荐;图神经网络;特征增强;注意力;深度学习

中图法分类号 TP311

Feature Augmentation based Graph Neural Recommendation Method in Session Scenarios

HUANG Zhen-Hua¹⁾LIN Xiao-Long¹⁾SUN Sheng-Li²⁾TANG Yong¹⁾CHEN Yun-Wen³⁾

¹⁾(School of Computer Science, South China Normal University, Guangzhou, Guangdong 510631)

²⁾(School of Software & Microelectronics, Peking University, Beijing 102600)

³⁾(Research and Development Department, DataGrand Inc., Shenzhen, Guangdong 518063)

Abstract Recommender systems are an effective way to solve the information overload problem. They can filter information from an overwhelming volume of data based on a user's historical preference to find the content a user is really interested in, and help the user to efficiently obtain the information they want. They have been widely used in various fields, such as news, video, shopping, travel recommendations, etc. Recently, graph neural network based session recommendation (referred to as graph neural session recommendation) has become a re-

本课题得到国家自然科学基金(62172166, 61772366, U1811263)、上海市自然科学基金(17ZR1445900)资助。黄震华(通信作者), 博士, 教授, 博士生导师, 主要研究领域为机器学习、数据挖掘、推荐系统。E-mail: jukiehuang@163.com。林小龙, 硕士研究生, 主要研究领域为推荐系统、神经网络。E-mail: m18826480489@163.com。孙圣力, 博士, 副教授, 主要研究领域为机器学习、数据挖掘、数据库。E-mail: slsun@ss.pku.edu.cn。汤庸, 博士, 教授, 博士生导师, 中国计算机学会(CCF)杰出会员, 主要研究领域为教育大数据、数据挖掘、数据库。E-mail: ytang@scnu.edu.cn。陈运文, 博士, 高级工程师, 主要研究领域为机器学习、数据挖掘、自然语言处理。E-mail: chenyunwen@datagrand.com。

search focus and hot in the recommender system community. This is mainly since they introduce the topological structure information of session graph to improve the accuracy of item and session feature representation, and therefore, to a certain extent, can improve the performance of session recommendation. However, existing graph neural session recommendation methods still have two main drawbacks, which affects the further improvement of performance of session recommendation. Firstly, the correlation weights between items in the session graph they construct are all pre-specified before the model training and remain fixed, which leads to the inability to capture the correlations between items accurately. Secondly, they only generate the local features of items and sessions from the item sequence of a single session. And they lack of the perspective of an entire session dataset and the global consideration of the correlations between different items and between different sessions, which can be used to produce global features of items and sessions. This leads to inability to represent the semantic features of items and sessions adequately. To address the above two drawbacks, we innovatively propose FA-GNR (Feature Augmentation based Graph Neural Recommendation method in session scenarios) in this paper. The FA-GNR method first constructs a session graph with learning and optimizing correlation weights between items based on a single session, which is used to generate item local features via GRU (Gated Recurrent Unit) neural network. At the same time, based on the session dataset, item global features are generated by the GloVe (Global Vectors) word embedding method. In this way, item local and global features are fused to produce item semantic features. Then, the FA-GNR method utilizes the local attention mechanism to generate session local features based on item semantic features. Meanwhile, based on item global features, the global attention mechanism is used to generate session global features. In this way, session local and global features are fused to produce session semantic features. Finally, based on item and session semantic features, the FA-GNR method uses cross-entropy loss to learn the click probability distribution of different items under a given session. Experimental results on several public datasets (e.g. Yoochoose1_64, Yoochoose1_4, and Diginetica) indicate that the recommendation performance of FA-GNR is better than that of existing mainstream methods in the graph neural session recommendation task. For example, on the Yoochoose1_4 and Diginetica datasets, the P@20 metric of FA-GNR exceeds the state-of-the-art graph-based neural network methods by 2.36% and 3.21% on average, respectively. At the same time, ablation, t-SNE visualization and hyper-parameter experiments further demonstrate the effectiveness of FA-GNR.

Key words session recommendation; graph neural network; feature augmentation; attention; deep learning

1 引言

在网络服务中，为了预测用户将来的偏好行为，个性化推荐系统需要从用户的历史行为数据中学习用户的偏好。基于会话的推荐系统（简称会话推荐系统）^[1]是其中的一个重要研究分支。与一般的个性化推荐系统不同，会话推荐系统利用用户的点击行为序列构建用户的行为模式，进而预测下一项用户可能感兴趣的物品。通常，在会话推荐系统中，用户可以产生多个会话，而单个会话由用户在一段时间内连续点击的一个物品序列组成。用户产生的会话数据则被用来训练学习用户对物品的偏好，进而生成推荐模型，来预测用户在当前会话下点击下一项物品的概率。

近年来，由于会话推荐系统具有较高的实际应用价值，受到了学术界和工业界的关注，并提出了许多有效的解决方案。据我们调研，早期的方法大多是基于近邻和协同过滤^[2]的思想，将物品的相似度定义为出现在同一会话中的频率，并基于相似度的计算预测下一项物品。我们把这类方法称之为传统会话推荐方法。通常，传统会话推荐方法只关注物品间的相似度量，而忽略对物品深层特征的挖掘，从而严重影响了最终的推荐性能。为了解决该问题，研究人员近些年提出了基于深度学习^[3]的方法，即深度学习会话推荐方法。它们大都使用深度神经网络，比传统方法更能有效捕获物品的深层特征，例如基于循环神经网络 RNN (Recurrent Neural Network)^[4]的 GRU4Rec (Gated Recurrent Unit for

Recommendation) 方法^[5]以及基于编码器-解码器 (Encoder-decoder) 结构^[6]的 NARM (Neural Attentive Recommendation Machine) 方法^[7]。

在此基础上, 研究人员发现完整的会话不仅包含点击序列, 还涉及其拓扑结构的信息, 而融入拓扑结构信息的特征, 将能更充分地表示会话的特征。在此思想的指导下, 研究人员提出了基于图神经网络 GNN (Graph Neural Network)^[8]的会话推荐方法。特别, 基于图神经网络的推荐方法 (简称图神经推荐方法) 充分考虑了会话序列的拓扑结构信息, 根据会话序列构建会话图, 并通过图神经网络进行物品特征的表示学习, 然后基于学习到的物品特征来生成会话特征。特别, 在会话图中, 物品被视为图节点, 同时, 会话序列内被连续点击的物品由有向边连接, 边的权重代表着不同物品间的相关性。目前, 主流的方法主要有 SR-GNN (Session-based Recommendation with GNNs)^[9]、GCE-GNN (Global Context Enhanced GNNs)^[10]以及 FGNN (Full GNN)^[11]等。我们发现, 虽然现有图神推荐方法在一定程度上能够提升会话推荐的性能, 但是仍然存在两个方面的不足: (1) 现有方法在构建会话图时, 图中节点 (即物品) 之间的相关性权重均是在模型训练之前就预先指定, 并在模型训练过程中保持固定不变, 这将导致不能准确地捕捉物品之间的相关性; (2) 现有方法只从个会话的物品序列中产生物品和会话的局部特征, 而缺乏全局视觉, 即围绕整个会话数据集, 全局考虑不同物品之间以及不同会话之间的相关性, 并由此来进一步生成物品和会话的全局特征, 从而不能充分表示物品和会话的语义特征。现有图神经推荐方法的这两个不足将直接影响了会话推荐性能的进一步提升。

为了解决现有图神经会话推荐方法的上述两个不足, 本文提出了一种基于特征增强的图神经推荐方法 FA-GNR (Feature Augmentation based Graph Neural Recommendation)。FA-GNR 方法首先根据会话序列构建会话图, 图中节点 (即物品) 间的相关性权重在模型训练中进行学习优化, 并将会话图输入到图神经网络部件中, 来产生物品局部特征; 同时, 基于整个会话数据集, 通过全局特征抽取部件, 计算物品的全局共现矩阵, 进而获取物品的全局特征; 然后, 通过物品线性变换部件将物品的局部和全局特征进行融合, 从而得到其最终的语义特征。在此基础上, FA-GNR 首先基于物品语义特征, 借

鉴 NARM 方法^[7]的思想, 通过局部注意力部件获得会话的局部特征; 同时, 考虑到会话的全局特性, 根据会话相似度, 获取与当前会话最相似的 Top- k 个会话, 并将包含自身在内的 $k+1$ 个会话通过全局注意力部件得到会话的全局特征; 然后, 将会话的局部和全局特征经过会话线性变换部件融合得到其最终的语义特征。当得到物品和会话语义特征后, FA-GNR 将它们输入到相关性预测部件中, 来计算给定会话下不同物品的点击概率分布。

归纳起来, 本文的主要贡献如下:

(1) 通过对会话图中有向边赋予可学习的注意力权重, 来捕获物品间的相关性, 该注意力权重随着模型的训练学习而动态优化。

(2) 为了提高物品和会话特征表示的充分性和准确性, 本文分别对它们的特征进行增强, 即在局部特征的基础上融入全局特征, 从而通过融合两种特征来生成各自最终的语义特征。

(3) 在公开数据集 Yoochoose (包括两个子集: Yoochoose1_4 和 Yoochoose1_64) 以及 Diginetica 上进行了大量的实验。实验结果表明, FA-GNR 方法的推荐效果优于目前主流的方法。

本文接下来组织如下: 第 2 节介绍国内外相关工作; 第 3 节给出本文 FA-GNR 方法, 并对其详细描述; 第 4 节给出实验结果与对比分析; 第 5 节总结全文并对未来的研究工作进行展望。

2 相关工作

2.1 传统会话推荐方法

在传统会话推荐方法中, 物品相似度定义为同一会话中出现的频率。其中较为经典的是文献[12]提出的物品协同过滤方法。随后, 研究人员进一步提出了序列化推荐方法^[13]。该类方法大都是基于马尔可夫链^[14], 使用序列化数据和给定用户的最后点击行为来预测用户下一步可能点击哪些物品。文献[15]提出了一种基于马尔可夫链的序列化推荐方法, 并探讨了如何使用概率决策树模型^[16]来提取序列化模式以学习用户的下一个行为状态。

矩阵分解 MF (Matrix Factorization)^[17]是推荐系统的常用方法, 基本目标是将用户物品评分矩阵简化为两个低维矩阵, 每个矩阵代表用户的潜在因素或评价的物品。但是这种方法很容易在训练过程中造成模型过拟合, 因此不适合于会话推荐。为此, 文献[18]提出了一种基于贝叶斯理论^[19]的方法, 在

先验知识下最大化后验概率，通过学习获得两个参数矩阵，进而计算多个物品的偏序关系。

2.2 深度学习会话推荐方法

近年来，基于深度学习的方法广泛应用于会话推荐中。其中，文献[5]提出的 GRU4Rec 方法，首次在会话推荐中使用 RNN 网络^[20]。在此基础上，文献[21]通过引入数据扩充技术来改进 GRU4Rec。除了 RNN 网络，文献[22]提出了使用基于邻域的方法来捕获共现信号。文献[23]充分考虑物品的内容特征，使用 3D 卷积神经网络^[24]来学习更为准确的物品特征表示。文献[7]提出的 NARM 方法，将注意力机制融入到堆栈 GRU 编码器中，以捕获更具代表性的物品转换信息。为了进一步降低由时间序列引起的偏差，文献^[25]提出了一种有效的方法 STAMP (Short-Term Attention/Memory Priority)，完全使用注意力层代替了循环编码器。

受 Transformer^[26]启发，文献[27]使用了多层注意力来捕获物品之间的相关性。文献[28]提出了 CSRM (Collaborative Session-based Recommendation Machine) 方法，它一种将协作邻域信息应用于会话推荐的新型混合框架。文献[29]使用了变分自编码器，通过关注用户的兴趣转移来捕获用户的序列行为特征。此外，文献[30]提出了一种新颖的网络 MCPN (Mixture-Channel Purpose Routing Network) 来模拟会话推荐中用户的多种购买意图。

2.3 图神经会话推荐方法

随着图神经网络在计算机视觉和自然语言处理领域的成功^[31-34]，近几年，研究人员考虑如何将图神经网络融入到会话推荐中来有效提升推荐性能。其中，文献[9]提出的 SR-GNN 方法，用于学习会话图上的物品特征，然后通过关注每个学习的物品特征来获得具有代表性的会话特征。

文献[35]提出了一种基于图神经网络的新颖方法 RAGNN (Repeat Aware Graph Neural Network)。RAGNN 方法有效结合图神经网络与重复探索机制，以便能够更好地实现会话推荐。文献[36]将基于自回归移动平均滤波器的图卷积层集成到图神经网络中。这样，可以通过一系列会话来捕获物品之间的复杂转换，在此基础上，使用注意网络将每个会话表示为该会话当前兴趣和全局偏好的组成。文献[37]提出了 PA-GGAN (Position-Aware Gated Graph Attention Network) 方法。该方法设计一种有效的

反向位置机制，并在节点特征聚集时引入自注意机制，从而增强了门控图神经网络。

文献[10]提出了 GCE-GNN 方法，从会话图和全局图中学习两个层次的物品特征。该方法通过构建全局图来获取物品的全局特征需要巨大的时间和空间开销。文献[11]提出的 FGNN 方法，能够学习物品间复杂的依赖关系。为了在单个会话的特征表示学习中利用跨会话信息，FGNN 方法进一步构造了一个连接不同会话的宽连接会话图和一个新颖的掩码-读出函数来改进会话的特征表示。另外，文献[38]提出 LESSR (Lossless Edge-order preserving aggregation and Short-cut graph attention for Session-based Recommendation) 方法，来解决有损会话编码和低效远程依赖捕获这两个问题。

3 FA-GNR 方法

3.1 问题定义

假定推荐系统中共有 m 个物品和 n 个会话序列，它们构成的集合分别为 $V = \{v_1, v_2, \dots, v_m\}$ 和 $S = \{s_1, s_2, \dots, s_n\}$ 。第 i 个会话序列定义为 $SQ_i = \langle v_{i,1}, v_{i,2}, \dots, v_{i,t_i} \rangle \subseteq V$ ，其中 t_i 为 SQ_i 中最后一次点击行为发生的时刻。那么，会话推荐任务是预测 SQ_i 在下一时刻 t_{i+1} 所有物品的点击概率分布： $p(v_{i,t_{i+1}} | SQ_i), v_{i,t_{i+1}} \in V$ ，并选取 p 值最大的物品 $v_{i,max}$ 作为 t_{i+1} 时刻的推荐物品。

3.2 FA-GNR 方法总体架构

本文 FA-GNR 方法的总体架构如图 1 所示，主体部分由两个分支组成。其中一个分支接收输入的会话(例如 $\langle v_1, v_3, v_1, v_2, v_3, v_4 \rangle$)，并构建会话图(详见 3.3 小节)，然后将会话图结构信息输入到图神经网络部件 (Graph Nerual Network) 中，图的出、入度权重矩阵^[9]以及注意力权重矩阵经图神经网络部件后，生成物品局部特征(详见 3.4 小节)。另一个分支将整个会话数据集输入到全局特征抽取部件中。受 GloVe (Global Vectors) 词嵌入方法^[39]启发，该部件基于会话数据集构建物品全局共现矩阵，进而获得物品全局特征(详见 3.5 小节)。在获得物品局部和全局特征后，FA-GNR 方法通过物品线性变换部件 (ItemLinear-Transformation) 融合物品的这两类特征，并生成其最终的语义特征。

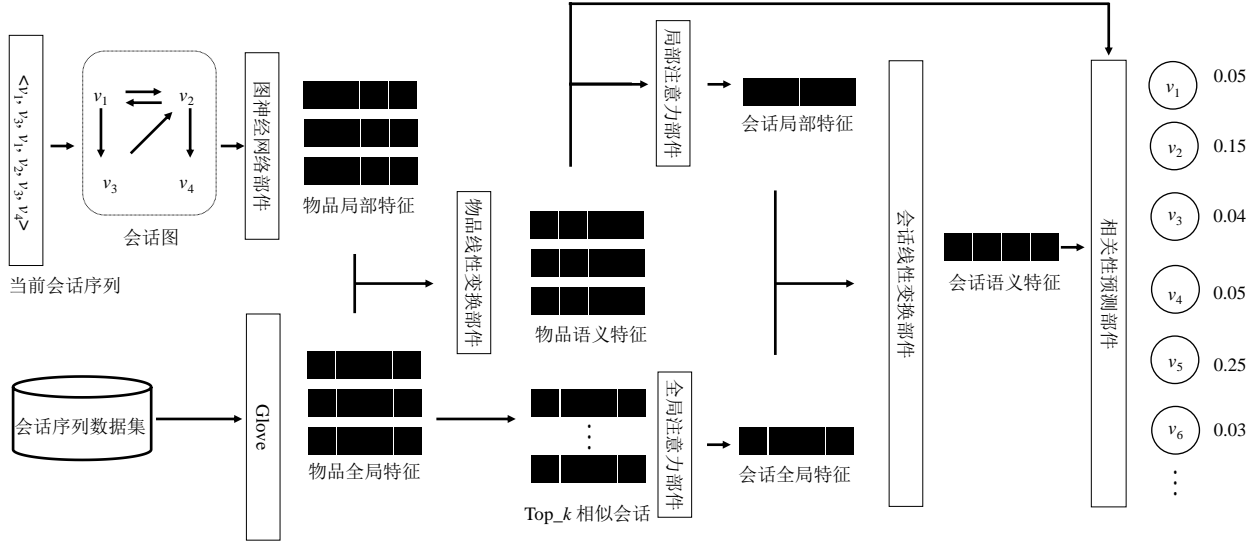


图 1FA-GNR 方法总体架构图

在此基础上，本文从全局和局部两个角度考虑会话的特征表示。FA-GNR 方法首先将物品局部特征输入局部注意力部件（Local Attention）中，根据用户的当前兴趣偏好，分配不同的注意力权重，从而融合会话中不同物品的局部特征来生成会话的局部特征（详见 3.6 小节）。另一方面，FA-GNR 方法将物品的全局特征替代会话序列中的物品，生成会话序列矩阵。由于生成整个数据集所有会话序列矩阵耗时巨大，因此，我们仅对当前同一最小批的会话建立会话序列矩阵。接着，FA-GNR 方法基于会话序列矩阵，获取与当前会话最相似的 Top- k 个会话，并将它们与当前会话输入到全局注意力部件（Global Attention）中，来生成会话的全局特征（详见 3.7 小节）。在获得会话全局和局部特征后，FA-GNR 方法通过会话线性变换部件（Session Linear Transformation）融合会话的这两类特征，并生成其最终的语义特征。

当获得所有物品和会话的语义特征之后，我们可以将它们输入到相关性预测部件（Correlation Prediction）中，来产生所有物品的概率分布。

3.3 会话图构建

给定会话序列 $SQ = \langle v_1, v_2, \dots, v_t \rangle$ ，基于 SQ 的会话图定义为一个有向图 $G_s = (V_s, E_s)$ ，其中 $V_s \subseteq V$ 为 SQ 中物品构成的顶点集合，每条有向边 $e_{ij} = (v_i \rightarrow v_j) \in E_s$ 代表用户在会话 SQ 中先后点击物品 v_i 和 v_j 。例如，对于序列 $SQ = \langle v_1, v_3, v_1, v_2, v_3, v_4 \rangle$ ，我们可以得到如图 2(a) 所示的会话图 G_s 。

同时， G_s 中的每条边 $e_{ij} = (v_i \rightarrow v_j)$ 关联三类权重信息，即出、入度权重 a_{ij}^{out} 和 a_{ij}^{in} ，以及注意力权重 a_{ij}^{att} 。其中 a_{ij}^{out} 和 a_{ij}^{in} 是模型训练之前就事先确定的，分别等于 e_{ij} 出现的次数与 v_i 出度和入度的比值。而 a_{ij}^{att} 是在模型训练的过程中学习得到的，其计算公式如下^[40]：

$$e_{ij} = \mathbf{q}^T((\mathbf{W}_1 \mathbf{v}_i) \oplus (\mathbf{W}_1 \mathbf{v}_j)), \quad (1)$$

$$a_{ij}^{att} = \frac{\exp(\text{LeakyReLU}(e_{ij}))}{\sum_{z \in N_i} \exp(\text{LeakyReLU}(e_{iz}))}, \quad (2)$$

其中， $\mathbf{v}_i, \mathbf{v}_j \in \mathbb{R}^d$ 分别为物品 v_i 和 v_j 的 d 维局部特征， $\mathbf{q} \in \mathbb{R}^{2d}$ 和 $\mathbf{W}_1 \in \mathbb{R}^{d \times d}$ 为两个可学习的模型参数， N_i 代表物品 v_i 在 G_s 中邻接物品的集合， \oplus 为向量拼接操作符， $\exp(\cdot)$ 为指数函数，而 $\text{LeakyReLU}(\cdot)$ 为非线性激活函数^[41]。

在此基础上，我们可以得到 G_s 的三个权重矩阵，即出度权重矩阵 \mathbf{A}_s^{out} 、入度权重矩阵 \mathbf{A}_s^{in} 以及注意力权重矩阵 \mathbf{A}_s^{att} 。不失一般性，我们将这三个矩阵总称为会话图结构信息矩阵 \mathbf{A}_s ，即 \mathbf{A}_s 由 \mathbf{A}_s^{out} 、 \mathbf{A}_s^{in} 和 \mathbf{A}_s^{att} 水平拼接而成。例如，对于图 2(a) 中的会话图 G_s ，我们可以得到如图 2(b) 所示的三个权重矩阵。值得注意的是， \mathbf{A}_s^{att} 不是事先确定的，其最优值是通过模型训练学习得到。

3.4 物品局部特征生成

图神经网络部件以会话图结构信息矩阵 \mathbf{A}_s 为输入来生成物品局部特征。在本文工作中，受文献^[9]的启发，图神经网络部件借鉴 GRU 的神经网络结

构，如图 3 所示。

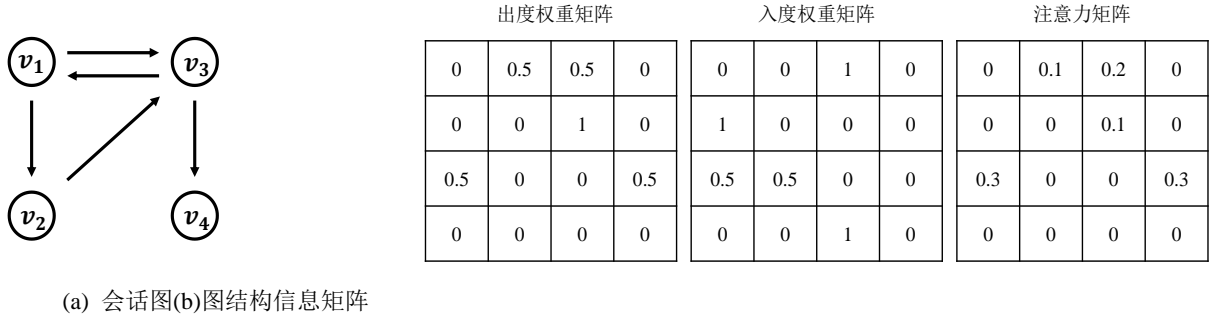


图 2 基于会话序列的会话图与图结构信息矩阵

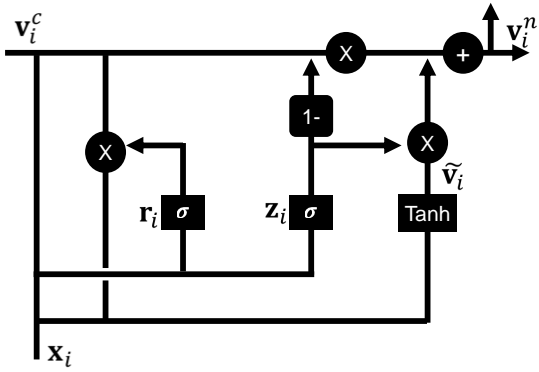


图 3 基于 GRU 的神经网络部件

下面，具体给出 SQ 中物品 v_i 的局部特征生成过程。本文首先获取 v_i 对应的输入信息 \mathbf{x}_i 。假设 SQ 包含 ρ 个不同的物品： v_1, v_2, \dots, v_ρ ，它们当前局部特征构成的矩阵记为 $\mathbf{V}^c = [\mathbf{v}_1^c, \mathbf{v}_2^c, \dots, \mathbf{v}_\rho^c]^T$ ，那么有：

$$\mathbf{x}_i = (\mathbf{A}_{s,i}^{out} \mathbf{V}^c \oplus \mathbf{A}_{s,i}^{in} \mathbf{V}^c \oplus \mathbf{A}_{s,i}^{att} \mathbf{V}^c) \mathbf{W}_2 + \mathbf{b}, \quad (3)$$

其中， $\mathbf{A}_{s,i}^{out}$ 、 $\mathbf{A}_{s,i}^{in}$ 和 $\mathbf{A}_{s,i}^{att} \in \mathbb{R}^{1 \times \rho}$ ，代表相应矩阵第 i 行， $\mathbf{W}_2 \in \mathbb{R}^{3d \times d}$ 和 $\mathbf{b} \in \mathbb{R}^{1 \times d}$ 为两个可学习的模型参数。不难看出： $\mathbf{x}_i \in \mathbb{R}^{1 \times d}$ 。

接着，基于 \mathbf{x}_i 和 \mathbf{v}_i^c ，我们可以分别得到重置门 (Reset gate) 和更新门 (Update gate) 特征：

$$\mathbf{r}_i = \sigma(\mathbf{W}_r [\mathbf{v}_i^c + (\mathbf{x}_i)^T]), \quad (4)$$

$$\mathbf{z}_i = \sigma(\mathbf{W}_z [\mathbf{v}_i^c + (\mathbf{x}_i)^T]), \quad (5)$$

其中， \mathbf{W}_r 和 $\mathbf{W}_z \in \mathbb{R}^{d \times d}$ 为两个可学习的模型参数， $\sigma(\cdot)$ 为 Sigmoid 非线性激活函数^[42]。不难看出： \mathbf{r}_i 和 $\mathbf{z}_i \in \mathbb{R}^{d \times 1}$ 。

在此基础上，我们将融合 \mathbf{r}_i 、 \mathbf{v}_i^c 和 \mathbf{x}_i 来生成候选状态特征：

$$\tilde{\mathbf{v}}_i = \tanh(\mathbf{W}_h [\mathbf{r}_i \otimes \mathbf{v}_i^c + (\mathbf{x}_i)^T]), \quad (6)$$

其中， $\mathbf{W}_h \in \mathbb{R}^{d \times d}$ 为可学习的模型参数， $\tanh(\cdot)$ 为非线性激活函数^[43]， \otimes 为向量位乘 (Element-wise multiplication) 操作符。不难看出： $\tilde{\mathbf{v}}_i \in \mathbb{R}^{d \times 1}$ 。

最后，基于 \mathbf{z}_i 、 \mathbf{v}_i^c 和 $\tilde{\mathbf{v}}_i$ 来得到 v_i 新的局部特征：

$$\mathbf{v}_i^n = (\mathbf{1} - \mathbf{z}_i) \otimes \mathbf{v}_i^c + \mathbf{z}_i \otimes \tilde{\mathbf{v}}_i. \quad (7)$$

3.5 物品全局特征与语义特征生成

物品全局特征抽取部件基于 Glove 词嵌入方法^[39]。本文首先利用会话数据集 SD 构建物品的全局共现矩阵 $\mathbf{C} \in \mathbb{R}^{m \times m}$ ，其中 m 为会话数据集中的物品数量，而 \mathbf{C} 中每一个元素的取值定义：

$$\mathbf{C}[i, j] = \sum_{SQ \in SD} \mathbb{1}(v_i \in SQ \wedge v_j \in SQ), \quad (8)$$

其中， $\mathbb{1}(\cdot)$ 为指示函数，若括号中内容为真则函数返回 1，否则为 0。不难看出， $\mathbf{C}[i, j]$ 取值越大，则 v_i 和 v_j 同时出现的频率越高，全局越相关。

本文记物品全局特征为 \mathbf{g} ，则基于文献[39]， \mathbf{g} 与共现矩阵 \mathbf{C} 的关系可表示为：

$$\mathbf{g}_i \mathbf{g}_j^T + b_i + b_j = \log(\mathbf{C}[i, j]), \quad (9)$$

其中， \mathbf{g}_i 和 $\mathbf{g}_j \in \mathbb{R}^{d \times 1}$ 分别是物品 v_i 和 v_j 的全局特征，而 b_i 和 $b_j \in \mathbb{R}^1$ 为两个可学习的参数。在此基础上，我们用如下损失函数来优化物品的全局特征：

$$\mathcal{L}_{item-g} = \sum_{i,j=1}^m \mathcal{J}(\mathbf{C}[i, j]) \cdot \mathcal{H}(i, j)^2. \quad (10)$$

$\mathcal{J}(\mathbf{C}[i, j])$ 为非递减函数，表示如下：

$$J(\mathbf{C}[i,j]) = \begin{cases} \left(\frac{\mathbf{C}[i,j]}{c_{max}}\right)^\beta & \text{if } \mathbf{C}[i,j] < c_{max}, \\ 1 & \text{otherwise} \end{cases} \quad (11)$$

其中, $c_{max}=100$, $\beta=0.75$ 。 $\mathcal{h}(i,j)=\mathbf{g}_i \mathbf{g}_j^\top + b_i + b_j - \log(\mathbf{C}[i,j])$ 。

当获得每个物品的局部和全局特征表示后, 将它们通过物品线性变换部件融合生成其最终的语义特征:

$$\mathbf{v}_i^{sem} = \mathbf{W}_3(\mathbf{v}_i \oplus \mathbf{g}_i), \quad (12)$$

其中, $\mathbf{W}_3 \in \mathbb{R}^{d \times 2d}$ 为可学习的模型参数。不难看出: $\mathbf{v}_i^{sem} \in \mathbb{R}^{d \times 1}$ 。

3.6 会话局部特征生成

受文献[7]启发, 局部注意力部件将整合会话 SQ 内部所有物品的语义特征, 从而生成会话的局部特征。具体地, 本文首先计算不同物品 v_i 在会话内部的关注度 (即注意力权重) a_i^{loc} :

$$\varepsilon_i = \mathbf{a}^\top \sigma(\mathbf{W}_4 \mathbf{v}_i^{sem} + \mathbf{W}_5 \mathbf{v}_{last}^{sem}), \quad (13)$$

$$a_i^{loc} = \frac{\exp(\varepsilon_i)}{\sum_{v_j \in S} \exp(\varepsilon_j)}, \quad (14)$$

其中, $\mathbf{a} \in \mathbb{R}^{d \times 1}$ 、 \mathbf{W}_4 和 $\mathbf{W}_5 \in \mathbb{R}^{d \times d}$ 为三个可学习的模型参数, \mathbf{v}_i^{sem} 为 v_i 的语义特征, 而 \mathbf{v}_{last}^{sem} 为 SQ 中最后一个物品的语义特征。

在此基础上, 可以得到会话 SQ 的局部特征:

$$\mathbf{s}^{loc} = \sum_{i=1}^{|SQ|} a_i^{loc} \mathbf{v}_i^{sem}, \quad (15)$$

其中 $|SQ|$ 为会话序列长度。不难看出: $\mathbf{s}^{loc} \in \mathbb{R}^{d \times 1}$ 。

3.7 会话全局特征与语义特征生成

全局注意力部件基于物品全局特征, 构建会话序列特征, 并根据序列特征相似度选取 Top- k 个相似会话融进当前会话, 来生成该会话全局特征。

具体来说, 首先对于最小批中的每一个会话 $SQ = \langle v_1, v_2, \dots, v_t \rangle$, 构造它的序列特征 \mathbf{s}^{seq} :

$$\varpi_i = \mathbf{e}^\top \sigma(\mathbf{W}_6 \mathbf{g}_i + \mathbf{W}_7 \mathbf{g}_t), \quad (16)$$

$$a_i^{seq} = \frac{\exp(\varpi_i)}{\sum_{v_j \in SQ} \exp(\varpi_j)}, \quad (17)$$

$$\mathbf{s}^{seq} = \sum_{i=1}^t a_i^{seq} \mathbf{g}_i, \quad (18)$$

其中, $\mathbf{e} \in \mathbb{R}^{d \times 1}$ 、 \mathbf{W}_6 和 $\mathbf{W}_7 \in \mathbb{R}^{d \times d}$ 为三个可学习的模型参数, \mathbf{g}_i 和 \mathbf{g}_t 分别为 v_i 和 v_t 的全局特征。不难看出: $\mathbf{s}^{seq} \in \mathbb{R}^{d \times 1}$ 。

基于此, 计算当前会话 SQ_{cur} 和最小批中其它每个会话 SQ_j 序列特征间的相似度:

$$sim_{cur,j} = (\mathbf{s}^{seq})^\top \cdot \mathbf{s}_j^{seq}. \quad (19)$$

基于上述序列特征相似度, 选取出 Top- k 个相似度最大的会话, 对这 k 个会话的序列特征分配可学习的注意力权重参数, 并得到当前会话 SQ_{cur} 的全局辅助特征:

$$\mathbf{s}_{cur}^{ass} = \sum_{j=1}^k \theta_j \mathbf{s}_j^{seq}, \text{ s.t. } \sum_{j=1}^k \theta_j = 1, \quad (20)$$

其中, $\theta_1 \sim \theta_k$ 为 k 个可学习的注意力权重参数, 其和等于 1。不难看出: $\mathbf{s}_{cur}^{ass} \in \mathbb{R}^{d \times 1}$ 。

然后, 将 \mathbf{s}_{cur}^{ass} 与 SQ_{cur} 的序列特征融合生成最终的全局特征:

$$\mathbf{s}_{cur}^{glo} = \mathbf{s}_{cur}^{seq} + \mathbf{s}_{cur}^{ass}. \quad (21)$$

不难看出: $\mathbf{s}_{cur}^{glo} \in \mathbb{R}^{d \times 1}$ 。

当获得每个会话 SQ 的局部和全局特征表示之后, 将它们通过会话线性变换部件融合生成其最终的语义特征:

$$\mathbf{s}^{sem} = \mathbf{W}_8(\mathbf{s}^{loc} \oplus \mathbf{s}^{glo}), \quad (22)$$

其中, $\mathbf{W}_8 \in \mathbb{R}^{d \times 2d}$ 为可学习的模型参数。不难看出: $\mathbf{s}^{sem} \in \mathbb{R}^{d \times 1}$ 。

3.8 模型学习与推荐

当生成会话和物品的语义特征之后, 本文计算当前会话 SQ_{cur} 语义特征与每个物品 v_i 语义特征的乘积, 得到它们间的推荐分数:

$$r(S_{cur}, v_i) = (\mathbf{s}_{cur}^{sem})^\top \cdot \mathbf{v}_i^{sem}. \quad (23)$$

在此基础上, 获取给定 SQ_{cur} 的情况下, 每个物品 v_i 被推荐的概率值:

$$\hat{p}(v_i | SQ_{cur}) = \frac{\exp(\mathcal{Q}(SQ_{cur}, v_i))}{\sum_{v_j \in V} \exp(\mathcal{Q}(SQ_{cur}, v_j))}. \quad (24)$$

最后, 本文构造交叉熵损失函数^[44]来训练和学习 FA-GNR 方法:

$$\mathcal{L}_{FA-GNR} = \sum_{i=1}^{|V|} p_i \log(\hat{p}_i) + (1 - p_i) \log(1 - \hat{p}_i), \quad (25)$$

其中, p_i 和 \hat{p}_i 分别为 v_i 被推荐的真实值 (one-hot 编码) 和预测概率值。

下面, 我们给出 FA-GNR 方法训练的伪代码 (如算法 1 所示) 及其时间、空间复杂度分析。

算法 1. FA-GNR 方法训练伪代码

输入: 会话数据集 SD ; 物品集合 V ; 最小批大小 b ;
相似会话数据量 k ;

输出: 训练完毕后的物品语义特征 $\{\mathbf{v}^{sem}\}$ 和会话语义特征 $\{\mathbf{s}^{sem}\}$;

BEGIN

1. 基于 SD 使用公式(8)构建共现矩阵 C ;

2. 随机初始化物品全局特征 $\{\mathbf{g}_i\}$;

3. FOR $i=1 \sim c_1$ DO

 /* c_1 为物品全局特征抽取部件训练周期 */

4. FOR $i=1$ to $|V|$, $j=1$ to $|V|$ DO

5. 根据公式(9)和(10)计算 \mathcal{L}_{item-g} ;

6. 使用优化器 AdaGrad^[45]更新 \mathbf{g}_i 和 \mathbf{g}_j ;

7. END FOR

8. END FOR

9. 随机初始化物品局部特征 $\{\mathbf{v}_i\}$;

10. FOR $i=1 \sim c_2$ DO

 /* c_2 为 FA-GNR 方法训练周期 */

11. FOR $j=1$ to $|SD|/b$ DO

12. 基于序列 SQ_j 构建会话图 G_S ;

13. 基于 G_S 产生会话图结构信息矩阵 \mathbf{A}_S :

$\mathbf{A}_S = \mathbf{A}_S^{out} \oplus \mathbf{A}_S^{in} \oplus \mathbf{A}_S^{att}$;

14. 根据公式(3)-(7)产生物品局部特征 $\{\mathbf{v}_z\}_{z=1}^{|SQ_j|}$;

15. 根据公式(12)产生物品语义特征 $\{\mathbf{v}_z^{sem}\}_{z=1}^{|SQ_j|}$;

16. 根据公式(13)-(15)产生会话局部特征 \mathbf{s}_j^{loc} ;

17. 根据公式(16)-(18)产生会话序列特征 \mathbf{s}_j^{seq} ;

18. 根据公式(19)获取 Top- k 个相似会话;

19. 根据公式(20)-(21)产生会话全局特征 \mathbf{s}_j^{glo} ;

20. 根据公式(22)产生会话语义特征 \mathbf{s}_j^{sem} ;

21. 根据公式(23)-(25)计算 \mathcal{L}_{FA-GNR} ;

22. 使用优化器 ADAM^[46]更新模型参数并得到更新后的 $\{\mathbf{v}_z^{sem}\}_{z=1}^{|SQ_j|}$ 和 \mathbf{s}_j^{sem} ;

23. END FOR

24. END FOR

25. 返回所有物品和会话的语义特征 $\{\mathbf{v}^{sem}\}, \{\mathbf{s}^{sem}\}$;

END

为了便于描述, 本文令 $\hbar = |SD|/b$ 。在算法 1 中, 步骤 1 的时间开销为 $O(|V|^2)$; 步骤 2 的时间开销为 $O(|V|)$; 步骤 5 和 6 的时间开销均为 $O(c_1|V|^2)$; 步骤 9 的时间开销为 $O(|V|)$; 而步骤 12-22 的时间开销均为 $O(c_2\hbar)$ 。因此, FA-GNR 方法的时间复杂度为 $O((2c_1 + 1)|V|^2 + 2|V| + 12c_2\hbar) = O((2c_1 + 1)|V|^2 + 12c_2\hbar)$ 。

在算法 1 中, 步骤 1 的空间开销为 $O(|V|^2)$; 步骤 2 的空间开销为 $O(d|V|)$, 其中 d 为特征维度; 步骤 5 的空间开销为 $O(2d)$; 步骤 6 的空间开销为 $O(d|V|)$; 步骤 9 的空间开销为 $O(|V|)$; 步骤 12 的空间开销为 $O(3n^2\hbar)$, 其中 n 为最小批中最大的会话图的节点数; 步骤 13 中产生 \mathbf{A}_S^{out} 和 \mathbf{A}_S^{in} 的空间开销为 $(2n^2\hbar)$, 而产生 \mathbf{A}_S^{att} 的空间开销为 $O(n^2 + d^2 + 4d)$; 步骤 14 的空间开销为 $O((3n^2 + 6d^2 + nd)\hbar)$; 步骤 15 的空间开销为 $O((2d^2 + 2d)\hbar)$; 步骤 16 和 17 的空间开销均为 $O((2d^2 + 3d)n\hbar)$; 步骤 18 的空间开销为 $(2dn\hbar)$; 步骤 19 的空间开销为 $O((k + 1)d\hbar)$; 步骤 20 的空间开销为 $O((2d^2 + 2d)\hbar)$; 步骤 21 的空间开销为 $O((2d + |V|)\hbar)$; 最后, 步骤 21 的空间开销为 $O((13d^2 + (n + 4)d + n^2 + k)\hbar)$ 。由于 $|V|$ 一般远远大于 d , n 以及 \hbar , 故 FA-GNR 方法的空间复杂度为 $O(|V|^2)$ 。

在得到训练好的物品和会话语义特征 $\{\mathbf{v}^{sem}\}, \{\mathbf{s}^{sem}\}$ 之后, 给定当前会话 SQ_{cur} , FA-GNR 方法将 SQ_{cur} 的语义特征 \mathbf{s}_{cur}^{sem} 与每个物品 v_i 的语义特征 \mathbf{v}_i^{sem} 输入到相关性预测部件中, 并通过公式(23)计算它们间推荐分数 $r(SQ_{cur}, v_i)$ 。在此基础上, 选取推荐分数最高的 N 项物品作为被推荐的候选项。

4 实验评估

本节将通过多组实验来验证 FA-GNR 方法的有效性, 首先对实验数据集、比较方法、评估指标和超参设置进行描述, 在此基础上, 实现对比实验、消融 (Ablation) 实验、t-SNE (t-distributed Stochastic Neighbor Embedding) 可视化实验以及超参实验, 并对实验结果进行详细的分析。

4.1 数据集

实验主要使用两个流行的公开数据集：第一个是 Yoochoose^[47]，它包含 6 个月内电子商务网站上的点击流数据；第二个为 Diginetica^[48]，它包含另一个电子商务网站上 5 个月内的点击流数据。

本文首先对这两个数据集进行了预处理。按照文献 [7] 中的处理方法，对于 Yoochoose 数据集，使用第二天的会话数据进行测试，并从测试集中滤除点击次数仅为一次的会话。对于 Diginetica 数据集，采取同样的预处理方法，唯一的区别是使用下一周的会话数据进行测试。考虑到 Yoochoose 数据集过于庞大，基于目前相关工作的做法，本文在总体规模上按 1/4 与 1/64 的比例构建子数据集 Yoochoose1_4 和 Yoochoose1_64。具体构建方式为：首先将 Yoochoose 中的会话按照时间戳由大到小排序，然后分别顺序选取 1/4 和 1/64 比例的会话作为 Yoochoose1_4 和 Yoochoose1_64。对于每个数据集，我们随机选择 80% 的会话数据作为训练集，并将其余的视为测试集。表 1 给出了上述实验数据集的具体统计信息。

表 1 实验中使用数据集的统计信息

统计数据	Yoochoose1_64	Yoochoose1_4	Diginetica
点击	557248	8326407	982961
训练会话	369859	5917746	719470
测试会话	55898	55898	60858
物品种类	16766	29618	43097
平均长度	6.16	5.71	5.12

4.2 比较方法

本文将 FA-GNR 方法与现有 13 个主流方法进行比较（现有方法的超参设置与原文保持一致）：

- Pop^[9]：推荐当前数据集中点击频率最高的 K 个物品；

- S-Pop^[9]：推荐所有会话最后一项中点击频率最高的 K 个物品；

- Item-KNN^[2]：将物品同时出现的频率作为相似度，并推荐相似度最高的 K 个物品；

- BPR-MF^[18]：基于贝叶斯后验优化的个性化排序方法；

- GRU4Rec^[5]：使用 RNN 神经网络处理会话的长期特征，其中，超参设置为{特征维度 100，最小批大小 50，优化器 AdaGrad，学习率 0.01}；

- NARM^[7]：首次将注意力机制用来捕捉会话特征，其中，超参设置为{特征维度 100，最小批大

小 512，学习率 0.001}；

- CSRM^[28]：将协作邻域信息应用于会话推荐的混合框架，其中，超参设置为{特征维度 150，最小批大小 512，优化器 ADAM，学习率 0.0005}；

- BERT4Rec^[49]：引入了序列推荐变压器的双向编码器表示，其中超参设置为{特征维度 256，最小批大小 128，优化器 ADAM，学习率 0.001}；

- SR-GNN^[9]：构建会话图，并使用图神经网络提取会话拓扑结构特征，其中，超参设置为{特征维度 100，最小批大小 100，优化器 ADAM，学习率 0.001}；

- GCE-GNN^[10]会话图扩展为全局图，在全局图的基础上聚合高阶邻居节点特征，其中，超参设置为{特征维度 128，最小批大小 100，优化器 ADAM，学习率 0.001}；

- FGNN(19)^[50]构建会话图来研究基于会话推荐系统的物品转换模式，并提出了一种考虑会话图中序列顺序和潜在顺序的新模型，其中，超参设置为{特征维度 128，最小批大小 100，优化器 ADAM，学习率 0.001}；

- FGNN^[11]：泛连接会话图处理会话特征，其中，超参设置为{特征维度 100，最小批大小 100，学习率 0.001}；

- LESSR^[38]：一种无损嵌入方案和快捷图注意层，其中超参设置为{特征维度 32，最小批大小 512，学习率 0.001}。

上述 13 个对比方法中，前 4 个方法为传统会话推荐方法，第 5-8 个方法为基于深度学习的会话推荐方法，而后 5 个方法为基于图神经网络的会话推荐方法。

4.3 评估指标与超参设置

本文所有实验均在如下环境中完成：Python3.7、Pytorch1.6.0、GeForce RTX2080Ti GPU。为了保持与现有工作相同的评估设置，使用前 20 个推荐物品来测试每一个对比方法。另外，本文采用与现有工作一致的两个评估指标 P@20（Precision）^[49]和 MRR@20（Mean Reciprocal Rank）^[51]：

$$P@20 = \frac{n_{hit}}{|ST|}, \quad (26)$$

其中， ST 为测试的会话集， n_{hit} 为真实点击的物品出现在推荐列表前 20 位的次数。

$$\text{MRR@20} = \frac{1}{|ST|} \sum_{SQ \in ST} \frac{1}{\text{Rank}(SQ)} \quad (27)$$

其中, ST 与公式(26)含义相同, 为测试的会话集, $\text{Rank}(SQ)$ 表示在会话 SQ 中, 真实点击物品出现在推荐列表前 20 位的位次。 $P@20$ 和 MRR@20 指标越高, 说明推荐性能越好。

本文 FA-GNR 方法的超参设置如下: (1) 使用基于最小批的 ADAM 优化器^[46]优化模型参数, 最小批大小为 128; (2) 物品和会话特征 (包括局部特征、全局特征和语义特征) 维度为 128; (3) 在 $\{0.01, 0.008, 0.006, 0.005\}$ 之间调整学习率; (4) 在 $\{5e-5, 3e-5, \dots, 1e-5\}$ 中设置归一化系数; (5) 学习率衰减系数在 $\{0.05, 0.1, 0.15, 0.2\}$ 中选择; (6) 相似会话的 k 值在 $\{0, 2, 3, 4, 6\}$ 中取值。此外, 本文执行提前停止策略^[52], 即如果在连续 10 个时间段内测试数据上的 $P@20$ 和 MRR@20 指标没有增加, 则提前终止训练。

4.4 与现有方法的推荐性能对比

为了检验 FA-GNR 方法的整体推荐性能, 本文将它与现有 13 个具有代表性的方法进行比较。表 2 给出了它们的推荐性能。从中可以看出:

(1) 传统会话推荐方法的准确度普遍不高, 例如 Pop、S-Pop、Item-KNN 和 BPR-MF 这 4 种传统方法在 Diginetica 数据集上的 MRR@20 指标分别只达到 0.01%、0.02%、3.27% 以及 0.37%。这主要是因为它们只基于物品的相似性进行推荐, 而不考虑会话整体信息, 如会话之间单击顺序的信息。

(2) 由于基于深度学习的会话推荐方法能够捕捉到物品与会话的深层特征, 因此在推荐准确度上要明显优于传统方法。例如, 早期的 GRU4Rec 方法充分考虑到会话之间的序列信息, 并使用 GRU 神经网络来处理序列特征, 从而在 Yoochoose1_64 数据集上的 $P@20$ 指标比性能最好的传统方法 Item-KNN 提升了 22.63%。然而, 该方法缺乏考虑用户当前的偏好兴趣。为此, NARM 方法提出对用户点击行为进行建模, 并利用注意系数来捕捉用户当前的偏好, 从而在 Yoochoose1_64 数据集上的 $P@20$ 指标比 GRU4Rec 方法提升了 11.51%。而 CSRMM 方法将协作邻域信息应用到新的基于会话的推荐的混合框架中, 使得在 Yoochoose1_64 数据集上的 $P@20$ 指标比 NARM 方法再提升了 2.23%。

(3) 与基于深度学习的会话推荐方法相比, 基于图神经网络的方法表现出更好的性能。这主要

是因为基于图神经网络的方法不仅仅使用用户点击序列, 还利用了图拓扑结构信息, 因此更能充分表示物品和会话的特征。例如, 在 Yoochoose1_4 数据集上, 基于图神经网络方法的平均 MRR@20 指标超过基于深度学习方法 2.49%。特别, 在基于图神经网络的方法中, 尽管 FGNN 和 GCE-GNN 方法都使用了全局会话图结构, 但推荐性能比 SR-GNN 方法差, 这表明它们没有正确捕获会话之间的依赖关系。而我们发现另一个方法 LESSR 适用于会话序列平均长度较长的数据集, 如 Yoochoose1_64; 对于平均长度较短的数据集, 如 Diginetica (平均长度为 5.12), 无法显著提高推荐性能。

(4) 本文 FA-GNR 方法的推荐性能在绝大多数情况下均超过现有基于图神经网络的会话推荐方法。例如, 在 Yoochoose1_4 数据集上, FA-GNR 方法的平均 $P@20$ 指标超过基于图神经网络方法 2.36%。这主要有两个原因: 1) FA-GNR 方法所构建会话图中边的权重不是预先指定的, 而是随着模型的训练进行动态更新和优化, 这样能够更准确地捕获会话内部物品之间的相关性; 2) FA-GNR 方法在物品和会话局部特征的基础上, 通过特征增强的方式融入了全局特征, 从而同时考虑物品和会话的全局和局部的依赖关系, 这样, 物品和会话的特征表示更为充分和准确。另外, 我们发现, 与现有方法相比, 本文 FA-GNR 方法在物品数量更多的数据集 (如 Diginetica) 上更具优势。这主要是因为 FA-GNR 方法是通过 Glove 词嵌入方法计算物品的全局共现矩阵, 而物品数越多意味着矩阵的元素就越多, 即可供训练的样本更多, 从而能够有效提高物品全局特征表示的准确性。

(5) 在 Yoochoose1_4 数据集上 SR-GNN 的 MRR@20 指标比 FA-GNR 方法领先 0.11%, 这可能是因为 Yoochoose1_4 数据集的平均长度 5.71 小于 Yoochoose1_64 数据集的 6.16。会话平均长度的减小将会导致共现次数的减少, 从而影响物品全局特征的学习, 导致推荐时真实点击物品位次靠后。

4.5 消融实验

本小节将通过消融实验来检验 FA-GNR 方法各主要组成部分的有效性。本文主要构造如下 5 个变种方法 (Variant):

- M_1 : 在 FA-GNR 方法中, 只使用物品的局部特征, 而不使用其全局特征;
- M_2 : 在 FA-GNR 方法中, 只使用会话的局部

特征，而不使用其全局特征，即不使用集成 Top- k 相似会话部分；

• M_3 ：在 FA-GNR 方法中，会话图不使用物品间的注意力权重矩阵；

• M_4 ：在 FA-GNR 方法中，会话图不使用物品间的注意力权重矩阵，同时，会话只是用局部特征

而不使用全局特征；

• M_5 ：在 FA-GNR 方法中，会话图不使用物品间的注意力权重矩阵，同时，物品只是用局部特征而不使用全局特征。

表 3 给出了 FA-GNR 方法与 5 个变种方法的推荐性能。由于 M_1 变种方法不使用物品的全局特征，

表 2 本文 FA-GNR 方法与现有 13 个方法的推荐性能 (%)

方法	Yoochoose1_64		Yoochoose1_4		Diginetica	
	P@20	MRR@20	P@20	MRR@20	P@20	MRR@20
Pop	3.36	0.61	0.6	0.22	0.09	0.01
S-pop	0.97	0.15	0.63	0.26	0.23	0.02
Item-KNN	27.25	6.94	32.68	9.05	14.81	3.21
BPR-MF	1.71	1.79	1.48	0.16	1.92	0.37
GRU4Rec	49.88	14.08	42.16	11.45	19.38	4.15
NARM	67.62	28.71	67.45	27.45	46.27	16.54
CSRM	69.13	29.10	69.87	28.99	49.47	16.30
BERT4Rec	65.72	28.35	66.11	28.09	43.20	12.95
SR-GNN	70.32	30.85	71.01	31.42	50.5	17.36
FGNN(19)	66.52	26.95	67.24	27.46	48.18	15.91
FGNN	67.74	27.90	68.77	28.66	49.11	16.69
GCE-GNN	66.80	26.31	67.68	27.01	45.61	13.11
LESSR	65.60	24.52	68.14	28.25	51.44	18.02
FA-GNR	70.54	31.43	71.26	31.31	52.18	18.32

表 3FA-GNR 方法及其 5 个变种方法的推荐性能 (%)

方法	Yoochoose1_64		Yoochoose1_4		Diginetica	
	P@20	MRR@20	P@20	MRR@20	P@20	MRR@20
M_1	69.34	29.9	70.14	30.73	49.69	16.48
M_2	69.69	29.86	70.23	30.81	51.76	17.88
M_3	70.50	30.52	70.80	31.16	52.01	17.97
M_4	68.74	29.34	70.44	30.97	50.95	17.30
M_5	68.26	29.08	69.73	29.85	49.50	16.33
FA-GNR	70.54	31.43	71.26	31.31	52.18	18.32

而 Top- k 相似会话的获取是基于物品的全局特征，所以实际上 M_1 变种方法同时也不会集成 Top- k 相似会话，即不会使用会话全局特征。因此， M_1 变种方法在 Yoochoose1_64 数据集上的 P@20 指标比 FA-GNR 方法降低了 1.20%，这说明物品和会话的全局特征能够提高推荐性能。 M_2 变种方法不使用会话全局特征，此时，它在 Yoochoose1_64 数据集上的 P@20 指标比 FA-GNR 方法降低了 0.85%，这验

证了会话全局特征的有效性。 M_3 变种方法不使用物品间的注意力权重矩阵，其性能略低于 FA-GNR 方法。 M_4 变种方法在 M_3 的基础上同时不使用会话全局特征，此时，它在 Yoochoose1_64 数据集上的 P@20 指标比 FA-GNR 方法降低了 1.80%，这说明物品间的注意力权重矩阵和会话全局特征结合使用能够有效提高推荐性能。最后， M_5 变种方法在 M_3 的基础上同时不使用物品的全局特征，此时，它

在 Yoochoose1_64 数据集上的 P@20 指标比 FA-GNR 方法降低了 3.23%，这说明物品间的注意力权重矩阵与物品全局特征结合使用同样能够有效提高推荐性能。综上可知，FA-GNR 方法的各个组成部分对于提高推荐性能均有作用。

由于在消融实验中集成 Top-k 相似会话带来较大的性能提升，因此，下面进一步重点检验 2 种不

表 4FA-GNR 方法及其 2 个集成变种方法的推荐性能 (%)

方法	Yoochoose1_64		Yoochoose1_4		Diginetica	
	P@20	MRR@20	P@20	MRR@20	P@20	MRR@20
S ₁	69.97	29.96	70.66	30.98	51.18	17.70
S ₂	67.24	28.68	70.95	31.06	50.68	17.66
FA-GNR	70.54	31.43	71.26	31.31	52.18	18.32

能够有效捕捉不同会话的重要性。S₁ 集成变种方法对 Top-k 相似会话进行均值操作，即均衡不同会话间的内部属性，导致生成的会话全局特征准确度不高。例如，S₁ 集成变种方法在 Yoochoose1_64 数据集上的 P@20 指标比 FA-GNR 方法降低了 0.57%。另外，S₂ 集成变种方法采用池化策略，而该策略通常用于图片数据，抓取图片像素矩阵上具有代表性的信息点。不同于图片，会话序列矩阵不是一个整体，它是由不同物品全局特征组成，如果会话序列矩阵通过池化操作进行集成，那么不但提取不到关键信息，反而丢失了诸多物品与会话的内部关联。例如，S₂ 集成变种方法在 Diginetica 数据集上的 MRR@20 指标比 FA-GNR 方法降低了 0.62%。

4.6 t-SNE可视化

本小节将通过 t-SNE 降维技术^[54]来可视化显示本文 FA-GNR 方法特征表示的有效性。由于 SR-GNN 方法在现有方法中的推荐性能最好，因此，为了简单而不失一般性，本文与该方法对比。

本文从 Yoochoose1_64 数据集中选取 6 个物品以及分别以这 6 个物品为最后点击项的 50 个会话，共 300 个会话。图 4(a)和(b)分别显示了 SR-GNN 和 FA-GNR 方法通过 t-SNE 降维技术在二维空间上的会话语义特征表示，不同标签的会话用不同的形状区分。从中可以发现，相同形状的会话通常趋向于靠拢，但靠拢程度在区域密度上有所差别。我们可以明显看出图 4(b)中的会话分布更紧凑，并且层次更为分明。这表明在 FA-GNR 方法中，同类标签会话的语义特征更加接近，从而在计算物品的推荐分数时，也会获得更准确的结果。

同的集成变种方法对于推荐性能的影响：

- S₁: 集成会话序列矩阵的平均值；
- S₂^[53]: 借鉴池化思想在会话矩阵上操作。

表 4 给出了 FA-GNR 方法及其 2 个集成变种方法的推荐性能。从中可以看出，FA-GNR 方法取得最好的推荐性能，这是因为不同的会话对于当前会话的依赖程度不同，因此分配可学习的注意力权重

4.7 超参数性能影响

本小节将检验 FA-GNR 方法中两个重要超参数对推荐性能的影响，即相似会话数量 k 与特征维度 d ，其中 $k \in \{0, 2, 3, 4, 6\}$ ， $d \in \{16, 32, 64, 96, 128\}$ 。为了简单，本文只给出 P@20 指标的实验结果，如图 5 所示。由于 MRR@20 指标实验结果的变化趋势与 P@20 指标类似，故此省略。

从图 5 可以看出，增大特征维度 d 并不能持续带来更好的推荐性能。例如，对于 Yoochoose1_64 数据集，当 $k=6$ 时，随着 d 取值的增大，推荐性能将逐渐上升，当 $d=96$ 时，推荐性能达到最优，之后当 $d=128$ 时，推荐性能反而下降不如 $d=96$ 时的性能。这主要是因为随着 d 的增大，FA-GNR 方法的学习能力逐步增强，但过强的学习能力则会造成过拟合现象，从而导致推荐性能的下降，因此，特征维度不是越大越好。另一方面，增大相似会话数量 k 也不能持续带来更好的推荐性能。例如，对于 Yoochoose1_4 数据集，当 $d=128$ 时，随着 k 取值的增大，推荐性能将逐渐上升，当 $k=4$ 时，推荐性能达到最优，之后当 $k=6$ 时，推荐性能反而下降不如 $k=4$ 时的性能。这是因为当 k 增大时，FA-GNR 方法会整合更多的相似对话，从而能够更为充分地表示会话的全局特征，然而，整合太多的相似会话会带来额外的噪声，从而导致推荐性能的下降。因此，相似会话数量不是越多越好。

5 结论与将来工作

为了解决现有会话推荐方法的不足，本文提出了一种基于特征增强的图神经推荐方法 FA-GNR。FA-GNR 方法在构建会话图的过程中，对不同的有向边赋予不同的可学习注意力权重，从而能够准确捕获会话图中物品之间的相关性。同时，对物品和会话的特征进行增强，在局部特征的基础上融入全局特征。特别，对于物品，基于 GloVe 词嵌入方法

计算物品全局共现矩阵，进而获取物品全局特征；而对于会话，首先获取与当前会话最相似的 Top-k 个会话，并通过全局注意力机制来获得会话全局特征。最后，在多个公开数据集上进行对比实验、消融实验、t-SNE 可视化实验以及超参实验，实验结果表明，FA-GNR 方法具有有效性。

本文将来的工作主要包括：1) FA-GNR 方法目前只使用图神经网络来提取物品和会话特征而非改进图神经网络结构，因此，为了进一步提升推

荐

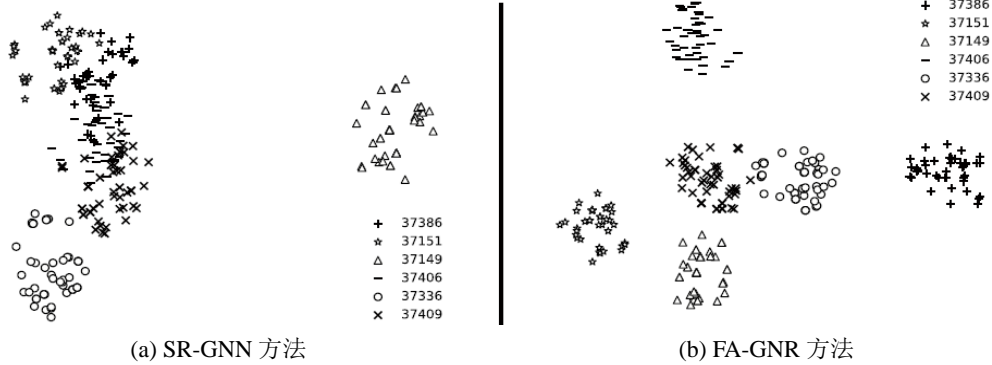


图 4 通过 t-SNE 降维技术在二维空间上的特征表示

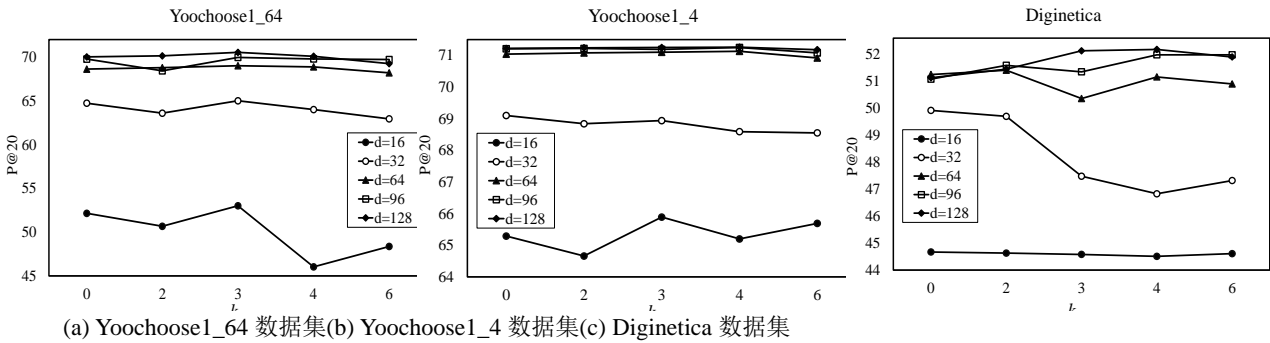


图 5 超参数性能影响

性能，我们考虑设计更为有效的图神经网络来增强物品和会话的特征表示；2) 物品的全局特征仅由其共现频率来刻画，这将很大程度上影响物品特征表示的准确性，我们考虑融入物品在全局会话中出现的上下文或次序等更多维度的全局辅助信息；3) FA-GNR 方法主要考虑推荐准确度，而没有太多关注推荐效率，从落地角度出发，我们考虑将知识蒸馏技术^[55]融入 FA-GNR 方法中以完成高效的在线推理，从而提升其推荐的实时效率。

参考文献

- [1] Zhong T, Wen Z, Zhou F, Trajcevski G, Zhang K. Session-based recommendation via flow-based deep generative networks and bayesian inference. *Neurocomputing*, 2020, 391: 129-141
- [2] Davidson J, Liebald B, Liu J, et al. The youtube video recommendation system//Proceedings of the fourth ACM conference on Recommender systems. Barcelona, Spain, 2010: 293-296
- [3] Isensee F, Jaeger P F, Kohl S AA, et al. Nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 2021, 18(2): 203-211
- [4] Lin J C W, Shao Y, Djenouri Y, et al. Asrnn: a recurrent neural network with an attention model for sequence labeling. *Knowledge-Based Systems*, 2021, 212: 106548
- [5] Hidasi B, Karatzoglou A, Baltrunas L, Tikk D. Session-based recommendations with recurrent neural networks. *arXiv preprint, arXiv:1511.06939*, 2015

- [6] Badrinarayanan V, Kendall A, Cipolla R. Segnet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 2017, 39(12): 2481-2495
- [7] Li J, Ren P, Chen Z, Ren Z, Lian T, Ma J. Neural attentive session-based recommendation//*Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. Singapore, 2017: 1419-1428
- [8] Guo Z, Tang L, Guo T, et al. Deep graph neural network-based spammer detection under the perspective of heterogeneous cyberspace. *Future Generation Computer Systems*, 2021, 117: 205-218
- [9] Wu S, Tang Y, Zhu Y, et al. Session-based recommendation with graph neural networks//*Proceedings of the AAAI Conference on Artificial Intelligence*. Hawaii, USA, 2019, 33(01): 346-353
- [10] Wang Z, Wei W, Cong G, et al. Global context enhanced graph neural networks for session-based recommendation//*Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. Xi'an, China, 2020: 169-178
- [11] Qiu R, Huang Z, Li J, et al. Exploiting cross-session information for session-based recommendation with graph neural networks. *ACM Transactions on Information Systems*, 2020, 38(3): 1-23
- [12] Sarwar B, Karypis G, Konstan J, et al. Item-based collaborative filtering recommendation algorithms//*Proceedings of the 10th International Conference on World Wide Web*. HongKong, China, 2001: 285-295
- [13] Gantner Z, Rendle S, Freudenthaler C, et al. Mymedialite: a free recommender system library//*Proceedings of the fifth ACM Conference on Recommender systems*. Chicago, USA, 2011: 305-308
- [14] Rendle S, Freudenthaler C, Schmidt-Thieme L. Factorizing personalized markov chains for next basket recommendation//*Proceedings of the 19th international conference on World Wide Web*. New York, USA, 2010: 811-820
- [15] Zimdars A, Chickering D M, Meek C. Using temporal data for making recommendations. *arXiv preprint arXiv:1301.2320*, 2013
- [16] Cheng J, Wu X, Zhou M, et al. A novel method for detecting new overlapping community in complex evolving networks. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2018, 49(9): 1832-1844
- [17] Chandrasekaran V, Sanghavi S, Parrilo P A, et al. Rank-sparsity incoherence for matrix decomposition. *SIAM Journal on Optimization*, 2011, 21(2): 572-596
- [18] Rendle S, Freudenthaler C, Gantner Z, et al. Bpr: bayesian personalized ranking from implicit feedback. *arXiv preprint arXiv:1205.2618*, 2012
- [19] Wang S, Du R, Liu Y. The learning and optimization of full bayes classifiers with continuous attributes. *Chinese Journal of Computers*, 2012, 35(10): 2129-2138 (in Chinese)
(王双成, 杜瑞杰, 刘颖. 连续属性完全贝叶斯分类器的学习与优化. *计算机学报*, 2012, 35(10): 2129-2138)
- [20] Li Y C, Xiong D Y, Zhang M. A survey of neural machine translation. *Chinese Journal of Computers*, 2018, 41(12): 2734-2755 (in Chinese)
(李亚超, 熊德意, 张民. 神经机器翻译综述. *计算机学报*, 2018, 41(12): 2734-2755)
- [21] Tan Y K, Xu X, Liu Y. Improved recurrent neural networks for session-based recommendations//*Proceedings of the 1st workshop on deep learning for recommender systems*. Boston, USA, 2016: 17-22
- [22] Ludewig M, Jannach D. Evaluation of session-based recommendation algorithms. *User Modeling and User-Adapted Interaction*, 2018, 28(4-5): 331-390
- [23] Tuan T X, Phuong T M. 3D convolutional networks for session-based recommendation with content features//*Proceedings of the eleventh ACM conference on recommender systems*. Como, Italy, 2017: 138-146.
- [24] 24O'Shea K, Nash R. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015
- [25] Liu Q, Zeng Y, Mokhosi R, et al. Stamp: short-term attention/memory priority model for session-based recommendation//*Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. London, UK, 2018: 1831-1839
- [26] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *arXiv preprint arXiv:1706.03762*, 2017
- [27] Kang W C, McAuley J. Self-attentive sequential recommendation//*Proceedings of the 2018 IEEE International Conference on Data Mining*. Singapore, 2018: 197-206
- [28] Wang M, Ren P, Mei L, et al. A collaborative session-based recommendation approach with parallel memory modules//*Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. Paris, French, 2019: 345-354
- [29] Jin Y, Liu M, Li Y, et al. Variational auto-encoder based bayesian-poisson tensor factorization for sparse and imbalanced count data. *Data Mining and Knowledge Discovery*, 2021, 35(2): 505-532
- [30] Wang S, Hu L, Wang Y, et al. Modeling multi-purpose sessions for next item recommendations via mixture-channel purpose routing networks// *Proceedings of the International Joint Conference on Artificial Intelligence*. Macao, China, 2019: 3771-3777
- [31] Xu B, Cen K, Huang J, Shen H, Cheng X. A survey on graph convolutional neural network. *Chinese Journal of Computers*, 2020, 43(5): 755-780 (in Chinese)
(徐冰冰, 岑科廷, 黄俊杰, 沈华伟, 程学旗. 图卷积神经网络综述. *计算机学报*, 2020, 43(5): 755-780)
- [32] Shi W, Rajkumar R. Point-gnn: Graph neural network for 3d object detection in a point cloud//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Virtual Conference, 2020: 1711-1719
- [33] Huang Z, Shan G, Cheng J, et al. TRec: an efficient recommendation system for hunting passengers with deep neural networks. *Neural Computing and Applications*, 2019, 31(1): 209-222
- [34] Yin Y, Lai S, Song L, et al. An external knowledge enhanced graph-based neural network for sentence ordering. *Journal of Artificial Intelligence Research*, 2021, 70: 545-566
- [35] Xian X, Fang L, Sun S. Regnn: a repeat aware graph neural network for session-based recommendations. *IEEE Access*, 2020, 8: 98518-98525
- [36] Wang H, Xiao G, Han N, et al. Session-based graph convolutional arma filter recommendation model. *IEEE Access*, 2020, 8: 6205362064
- [37] Wang J, Xu Q, Lei J, et al. Pa-ggan: session-based recommendation

- with position-aware gated graph attention network//Proceedings of the 2020 IEEE International Conference on Multimedia and Expo. London, UK, 2020:1-6
- [38] Chen T, Wong R C W. Handling information loss of graph neural networks for session-based recommendation//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. California, USA, 2020: 1172-1180
- [39] Pennington J, Socher R, Manning C D. Glove: global vectors for word representation//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha, Qatar, 2014: 1532-1543
- [40] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks. arXiv preprint arXiv:1710.10903, 2017
- [41] Che X, Kang W, Deng B, Yang K, Li J. A prediction model of SDN routing performance based on graph neural network. Acta Electronica Sinica, 2021, 49(3): 484-491 (in Chinese)
(车向北, 康文倩, 邓彬, 杨柯涵, 李剑. 一种基于图神经网络的 SDN 路由性能预测模型. 电子学报, 2021, 49(3): 484-491)
- [42] Zhong Y, Deng W, Hu J, et al. Sface: sigmoid-constrained hypersphere loss for robust face recognition. IEEE Transactions on Image Processing, 2021, 30: 2587-2598
- [43] Xie Y, Raj A N J, Hu Z, et al. A twofold lookup table architecture for efficient approximation of activation functions. IEEE Transactions on Very Large Scale Integration Systems, 2020, 28(12): 2540-2550
- [44] Ghazvininejad M, Karpukhin V, Zettlemoyer L, et al. Aligned cross entropy for non-autoregressive machine translation//Proceedings of the International Conference on Machine Learning. Bangkok, Thailand, 2020: 3515-3523
- [45] D'áosses, Bottou L, Bach F, et al. On the convergence of adam and adagrad. arXiv preprint arXiv:2003.02395, 2020
- [46] Mada M, Farmadi A, Budiman I, et al. Gru, adagrad, Rmsprop, adam implementation of gru and adam optimization method for stock price prediction. Journal of Data Science and Software Engineering, 2021, 2(01): 36-45
- [47] Pan Z, Cai F, Ling Y, et al. Rethinking item importance in session-based recommendation//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. Xi'an, China, 2020: 1837-1840
- [48] Lei T, Xiong Y, Tian P, et al. Minsr: multi-level interests network for session-based recommendation//International Conference on Database Systems for Advanced Applications. Jeju, Korea, 2020: 650-657
- [49] Sun F, Liu J, Wu J, et al. Sequential recommendation with bidirectional encoder representations from transformer//Proceedings of the 28th ACM international conference on information and knowledge management. Beijing, China, 2019: 1441-145
- [50] Qiu, Ruihong, et al. Rethinking the item order in session-based recommendation with graph neural networks//Proceedings of the 28th ACM International Conference on Information and Knowledge Management. Beijing, China, 2019:579-588
- [51] Huang Z, Xu X, Ni J, et al. Multimodal representation learning for recommendation in internet of things. IEEE Internet of Things Journal, 2019, 6(6): 10675-10685
- [52] Ghanbari E, Shakeri A. ERR. Rank: an algorithm based on learning to rank for direct optimization of expected reciprocal rank. Applied Intelligence, 2019, 49(3): 1185-1199
- [53] Huang Z, Xu X, Zhu H, et al. An efficient group recommendation model with multiattention-based neural networks. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(11): 4461-4474
- [54] Wu X, Gori M, Monfardini G, Scarselli F. A new model for learning in graph domains//Proceedings of the 2005 IEEE International Joint Conference on Neural Networks. Montreal QC, Canada, 2005(2): 729-734
- [55] He R, Hu Y, et al. Learning an evolutionary embedding via massive knowledge distillation. International Journal of Computer Vision, 2020, 128(8): 2089-2106

HUANG Zhen-Hua, Ph.D., professor, Ph.D. supervisor. His research interests mainly include deep learning, recommendation system, and data mining.

LIN Xiao-Long, M.S. candidate. His research mainly include recommender system and neural network.

SUN Sheng-Li, Ph.D., associate professor. His research interests mainly include machine learning, data mining, and database.

Background

The exponential growth of information in the Internet

TANG Yong, Ph.D., professor, Ph.D. supervisor. His research interests mainly include educational big data, data mining, and database.

CHEN Yun-Wen, Ph.D., senior engineer. His research interests mainly include machine learning, data mining, and NLP.

causes the information overload problem, making it impossible

for users to obtain their content of interest quickly and accurately. Recommender systems are an effective way for solving this problem. Session recommender system is an important research branch of recommender system. Different from the general recommender system, the session recommender system uses the user's click behavior sequence to construct the user's behavior pattern, and then predicts the next item that the user may be interested in. Generally, in a session recommender system, a user can generate multiple sessions, and a single session is composed of a sequence of items that the user clicks continuously over a period of time. The session data generated by the user is employed to train and learn the user's preferences for items, and then to generate a recommendation model to predict the probability of the user clicking the next item in the current session.

Graph neural session recommendation has become a research focus and hot in the session recommender system community in recent years. However, existing methods still has two drawbacks, which affects the further improvement of performance of session recommendation. First, the correlation weights between items in the session graph they construct are all pre-specified before the model training and remain fixed, which leads to the inability to capture the correlations between items accurately. Second, they only generate the local features of items and sessions from the item sequence of a single session. And they lack of the perspective of an entire session dataset and the global consideration of the correlations between different items and between different sessions, which can be used to produce global features of items and sessions. This leads to

inability to represent the semantic features of items and sessions adequately.

To solve the above two drawbacks, we propose FA-GNR in this paper. The FA-GNR method first constructs a session graph with learning and optimizing correlation weights between items based on a single session, which is employed to generate item local features via GRU neural network. At the same time, based on the session dataset, item global features are generated by the GloVe word embedding method. In this way, item local and global features are fused to produce item semantic features. Then, the FA-GNR method utilizes the local attention mechanism to generate session local features based on item semantic features. At the same time, based on item global features, the global attention mechanism is used to generate session global features. In this way, session local and global features are fused to produce session semantic features. Finally, based on item and session semantic features, the FA-GNR method uses cross-entropy loss to learn the click probability distribution of different items under a given session. Experimental results over several public datasets show that the recommendation performance of FA-GNR is better than that of existing mainstream methods.

This work is partially supported by the National Natural Science Foundation of China (No. 62172166, No. 61772366, and No. U1811263), which aims to research on deep learning-based recommender system, and the Natural Science Foundation of Shanghai (No. 17ZR1445900), which aims to research on graph neural network.