

基于霍克斯过程和图神经网络的会话推荐

杨真真^{*1} 闫孟儒¹ 杨永鹏² 陈亚杰¹

(1. 南京邮电大学理学院, 江苏南京 210023; 2. 南京信息职业技术学院网络与通信学院, 江苏南京 210023)

摘要: 针对传统会话推荐系统(Session-Based Recommendation System, SBRS)往往忽略了项目点击量之间的交互,以及遗漏了会话内项目之间的相对顺序的问题,本文提出了一种基于霍克斯过程和图神经网络(Hawkes Process and Graph Neural Network, HPGNN)的会话推荐方法。该方法提出了包含图神经位置感知层和图神经霍克斯层的双流结构,分别学习用户的长期和短期偏好。图神经位置感知层通过门控图神经网络(Gated Graph Neural Network, GGNN)来捕捉各个节点之间的交互关系,得到会话中每个项目的隐向量表示,并引入逐次递减的残差网络,有效地将之前的编码信息与当前网络融合,然后通过位置感知注意力网络来捕捉项目节点在会话中的位置信息,用于学习用户的长期偏好表示。图神经霍克斯层通过将霍克斯过程和GGNN相结合来捕捉连续时间的项目点击量之间的关系,用于更准确的表示用户的短期偏好。最后将两者进行线性组合,来更好地描述用户意图。实验结果表明,提出的HPGNN在Diginetica和Yoochoose1/64两个基准会话推荐数据集上的推荐性能均优于其他会话推荐模型。

关键词: 会话推荐; 推荐系统; 图神经网络; 霍克斯过程; 位置感知注意力网络

中图分类号: TP391;TP183 **文献标识码:** A **DOI:** 10.16798/j.issn.1003-0530.2024.04.013

引用格式: 杨真真,闫孟儒,杨永鹏,等. 基于霍克斯过程和图神经网络的会话推荐[J]. 信号处理,2024,40(4): 757-765. DOI: 10.16798/j.issn.1003-0530.2024.04.013.

Reference format: YANG Zhenzhen, YAN Mengru, YANG Yongpeng, et al. Hawkes process and graph neural network for session-based recommendation[J]. Journal of Signal Processing, 2024, 40(4): 757-765. DOI: 10.16798/j.issn.1003-0530.2024.04.013.

Hawkes Process and Graph Neural Network for Session-Based Recommendation

YANG Zhenzhen^{*1} YAN Mengru¹ YANG Yongpeng² CHEN Yajie¹

(1. College of Science, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu 210023, China;
2. School of Network and Communication, Nanjing Vocational College of Information Technology, Nanjing, Jiangsu 210023, China)

Abstract: To overcome the drawbacks of traditional session-based recommendation system (SBRS) models, which often ignore the interaction between item clicks and omit the relative position between items within a session, this study investigated a Hawkes process and graph neural network (HPGNN) method for session-based recommendation. This novel method incorporated a dual-stream structure consisting of a graph neural position-aware layer and graph neural

收稿日期: 2023-04-07; 修回日期: 2023-07-20

*通信作者: 杨真真 yangzz@njupt.edu.cn *Corresponding Author: YANG Zhenzhen, yangzz@njupt.edu.cn

基金项目: 国家自然科学基金(62071242, 62171232);江苏省研究生科研与实践创新计划项目(KYCX22_0955, SJCX23_0251);江苏省研究生教育学改革项目(JGKT23_C019);南京邮电大学科研项目(NY220207);南京邮电大学研究生教育学改革项目(JGKT23_XJ02)
Foundation Items: The National Natural Science Foundation of China (62071242, 62171232); The Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX22_0955; SJCX23_0251); Jiangsu Province Graduate Education Reform Project (JGKT23_C019); NUPTSF (NY220207); Graduate Education Reform Project of Nanjing University of Posts and Telecommunications (JGKT23_XJ02)

Hawkes layer to learn the long- and short-term user preferences, respectively. The graph neural position-aware layer captured the interaction between nodes by using a gated graph neural network (GGNN) to obtain the hidden vector representations of each item in a session. Furthermore, we introduced a gradually decreasing residual network that combined previous encoding information with the current network. In addition, we used a position-aware attention network to capture the position information of item nodes in the session to learn the long-term preferences of users. The graph neural Hawkes layer combined the GGNN and Hawkes process to represent the users' short-term preferences more accurately by capturing the relationship between item click counts over time. Finally, we linearly combined the two layers to better describe users' intent. Experimental results showed that our proposed HPGNN outperformed other existing state-of-the-art SBRS models on the Diginetica and Yoochoose1/64 datasets.

Key words: session-based recommendation; recommender system; graph neural network; Hawkes process; position-aware attention network

1 引言

随着移动设备的普及,互联网产生的信息和数据量呈现出了爆发式增长,推荐系统(Recommendation System, RS)在数字消费、服务和决策方面为用户缓解了信息过载的压力,成为人工智能最重要的应用之一。RS不仅可以为用户筛选出更多有效信息,还能帮助供应商提高平台使用体验,保留更多用户。

当前的大多数RS都是假设用户的档案和过去的交互记录是不断被记录的,然而,由于个性化推荐和个人隐私之间的冲突不断升级,许多种服务用户的身份和信息可能是未知的,只有正在进行会话的用户的历史记录可用。在这种情况下,传统依赖充分的用户交互的推荐系统的性能不佳^[1-3]。基于会话的推荐系统(Session-Based Recommendation System, SBRS)利用用户消费过程中生成的会话学习用户的偏好,每个会话由一个连续时间内发生的用户与项目间的交互构成,使得SBRS能够捕捉到用户准确的偏好^[3]。现有的基于深度学习的会话推荐系统大多使用循环神经网络(Recurrent Neural Network, RNN)及其变体来对项目序列的信息进行建模,例如,GRU4Rec^[4]利用门控循环单元(Gated Recurrent Unit, GRU)^[5],通过点击项目的特征对会话进行建模;NARM^[6]首次将注意力机制整合到深度学习网络中,利用注意力机制计算两个RNN子模块的输出之间的注意力权重,以便自适应的选择更重要的项目,从而更有效的表示用户的偏好;STAMP^[7]认为用户的最后一次点击尤为重要,并把最后一个点击的项目表示当作用户的短期偏好。

然而,基于深度学习的会话推荐只能模拟连续项目之间的单向转换,忽略了同一会话中其他上下文项目之间的单向转换。为了解决这个问题,SR-GNN^[8]通过图结构对目标会话进行建模,并利用图

神经网络(Graph Neural Network, GNN)对图上项目节点之间的复杂转换进行建模。虽然基于GNN的会话推荐取得了不错的效果,但现有的这类会话推荐大多仅关注用户的长期偏好表示,虽有方法对用户的短期偏好表示进行了研究,但仅仅将用户最后一项点击的项目表示作为用户的短期偏好,这种简单的表示方式往往并不能准确地描述用户的短期偏好。本文聚焦用户短期偏好的更准确表示方式,受霍克斯过程^[9]的启发,认为项目点击量之间存在的关系可以辅助模型更准确地表示用户的短期偏好。此外,现有的模型通常忽略了会话内项目之间的相对位置信息,从而影响了模型的推荐性能。

为了解决上述问题,本文将霍克斯过程和图神经网络相结合,提出了一种基于霍克斯过程和图神经网络的会话推荐系统。该模型提出了包含图神经网络位置感知层和图神经霍克斯层的双流结构来分别学习用户的长期和短期偏好。一方面,在图神经位置感知层,首先利用门控图神经网络(Gated Graph Neural Network, GGNN)来捕捉各个节点之间的交互关系以用于后续处理,并为了更好地利用过去的编码信息,提出了逐次递减的残差网络来与当前网络融合。其次,为了捕捉不同项目之间的位置关系,提出了位置感知注意力网络来将位置信息嵌入会话信息。通过会话的位置矩阵查询到项目的位置向量嵌入,并与最终的项目表示相加,得到位置感知后的项目表示,以表示用户的长期偏好。另一方面,受霍克斯过程(Hawkes Process, HP)启发,在图神经霍克斯层,采用霍克斯强度函数来考虑项目事件点击量之间的交互信息,利用霍克斯过程结合图神经网络来捕捉点击量之间的交互影响,并将结果作为用户的短期偏好,以弥补现有基于GNN的会话模型大多仅仅将最后一项的项目表示作为用户短期偏好的缺陷,以此来提高模型的推荐性能。

2 相关工作

会话推荐系统旨在通过学习会话内和会话间的依赖关系,通过给定的已知部分预测未知部分,或者给定历史会话预测当前会话,它可以是连续的,也可以是非连续的^[10]。SBRS是基于隐式反馈的推荐系统的典型应用,通常用户的身份未知,且没有明确的偏好^[11]。由于SBRS更加关注用户的当前偏好,且用户匿名,所以传统推荐系统中经典的协同过滤(Collaborative Filtering, CF)方法,例如矩阵分解(Matrix Factorization, MF)^[12-13]在这种情景下并不适用。目前主流的SBRS方法主要有三种类型:传统SBRS方法、基于RNN及其变体等深度神经网络的方法以及基于图神经网络的方法。

2.1 传统SBRS方法

传统SBRS一般利用经典的数据挖掘和机器学习方法来捕获嵌入会话数据中的依赖项进行会话推荐。Sarwar等人提出了基于项目邻域的方法^[14],项目相似度是根据同一会话中的共现情况计算的。之后,基于马尔可夫链的方法被提出,该类方法在难易程度上适中,并具有高效的表征学习能力。Shani等人^[15]将SBRS视为一个序列优化问题,并采用马尔可夫决策过程作为解决方案。Rendle等人^[16]提出的FPMC通过对用户的个性化概率转移矩阵进行因子分解,对每两次相邻点击之间的顺序行为进行建模,并为每个序列提供更准确的预测。BPR-MF^[17]将推荐视为一个排序问题,利用矩阵分解处理隐反馈数据,其中新会话的特征向量被定义为会话中项目的特征向量的平均值,并将其作为用户的最终特征向量。采用传统的SBRS方法相对简单,并且易于实现,但是它们只能捕获相邻项中的信息,很难捕获到项目转换表示出的偏好变化。

2.2 基于深度神经网络的SBRS

基于深度神经网络的方法近年来受到了广泛关注,这类方法主要利用RNN或其变体等深度神经网络强大序列建模能力对会话内和会话间的复杂的依赖关系进行建模^[18]。Hidasi等人提出的GRU4Rec^[4]是第一个利用RNN的变体GRU^[5]进行推荐的模型,该模型通过对多个GRU层进行堆叠,学习点击和点击项目的特征来对会话进行建模。后续出现了很多GRU的改进方法,例如Quadrana等人^[7]利用会话级和用户级两层GRU分别对会话中的序列和跨会话信息进行建模,考虑了用户的个性化信息。Li等人提出了NARM^[6],通过使用注意力机制来强调用户的主要目的,以此提高推荐的准确性。Donkers等

人^[19]设计了一个独特的基于用户的GRU模型,结合了用户特征来生成个性化的下一项建议。

2.3 基于图神经网络的SBRS

图神经网络的兴起影响了会话推荐的发展,许多工作将GNN应用到推荐系统^[20-24],用以对项目之间复杂的交互关系进行建模。SR-GNN^[8]开创性地将GNN引入推荐系统,首先对会话进行预处理,将会话序列建模为图结构数据,然后利用门控图神经网络(Gated Graph Neural Network, GGNN)^[25]将图结构的交互信息编码,学习节点交互的信息嵌入,并用注意力机制学习用户的全局偏好和当前偏好,以此生成下一个推荐项目。GC-SAN^[20]在用户全局偏好建模方面对SR-GNN进行改进,利用Transformer^[26]对会话上下文信息进行建模,用以学习会话推荐的长期依赖性。TAGNN^[27]在通过GNN学习项目嵌入后,利用不同的目标项目激活不同的用户兴趣,以生成个性化推荐。DGS^[28]使用动态图神经网络有效地捕捉用户兴趣随时间变化的动态变化,并能够在多个时间步生成推荐。IHGCN^[29]利用图卷积网络学习用户的特征,利用特征维度级的细粒度注意力机制挖掘用户的全局偏好,并与短期偏好融合得到最终的推荐结果。此外,还有一些模型利用社交图谱、商品信息等边信息进行补充,以增强个性化表征^[30-32]。但是上述模型都仅将会话内最后一个项目的嵌入表示作为用户的短期偏好,并且忽略了会话内项目之间的相对位置信息。

3 基于HPGNN的会话推荐

3.1 总体框架

基于HPGNN的会话推荐系统的总体结构如图1所示。

首先,将会话序列建模成图结构,然后,将得到的图结构通过图神经位置感知层和图神经霍克斯层分别学习用户的长期和短期偏好。首先,通过GGNN学习项目之间的交互信息,获得每个项目的节点表示,再引入逐次递减的残差网络,将过去的编码信息融入当前网络,并利用位置感知注意力网络将项目在会话中的位置信息进行嵌入,得到用户的长期偏好表示;同时,通过图神经霍克斯层来捕捉项目事件点击量之间的交互影响,得到用户的短期偏好表示;然后,将得到的用户长期和短期偏好通过一个线性变换层进行融合,得到会话的最终表示;最后,通过预测层计算每个候选项目的推荐得分,给出推荐结果。

3.2 问题定义

会话序列推荐旨在通过分析会话内或会话间

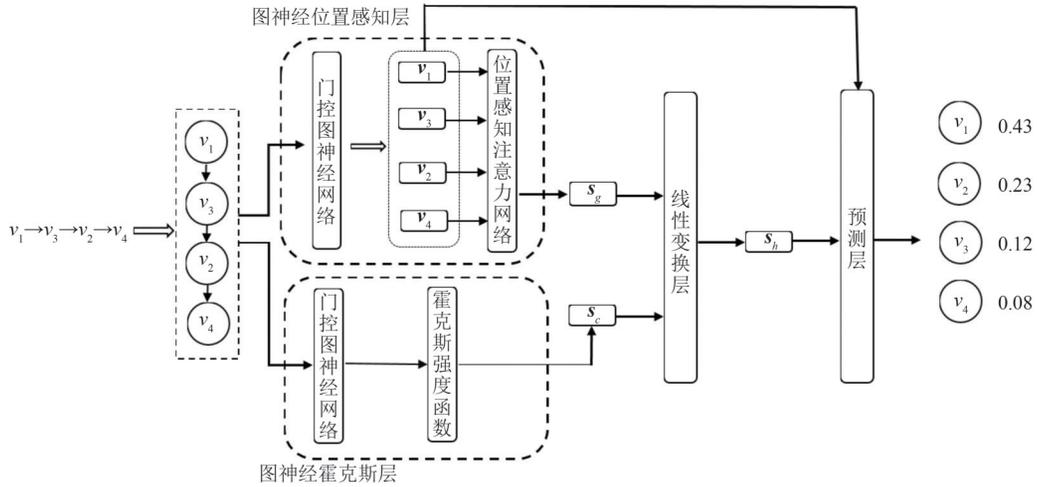


图1 HPGNN总体框架

Fig. 1 Overall framework of HPGNN

的复杂关系来预测用户可能点击或访问的下一个物品。设 $V = \{v_1, v_2, \dots, v_m\}$ 表示会话中所有项目组成的集合, 匿名用户的会话序列表示为 $S = \{v_{s,1}, v_{s,2}, \dots, v_{s,n}\}$, 其中 $v_{s,i} \in V$ 表示会话 S 中用户的点击项目在会话 S 中按时序排列。基于会话推荐的任务是预测会话 S 的下次点击的项目 $v_{s,n+1}$ 。推荐模型计算项目集合 V 中的每个候选项目的推荐分数, 输出分数向量 $\hat{y} = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_m\}$, 其中 \hat{y}_m 是项目 v_m 的推荐得分, \hat{y} 中分数最高的若干个项目将成为推荐的候选项目集, 并作为最终推荐结果。

3.3 提出的 HPGNN 方法

为了解决现有基于 GNN 的会话推荐模型仅仅将用户最后一项点击的项目表示作为用户短期偏好, 以及忽略了项目相对位置信息的问题, 提出了基于图神经位置感知层和图神经霍克斯层的双流结构。该结构利用图神经位置感知层学习项目节点表示, 以捕获用户的长期偏好; 利用图神经霍克斯层学习用户的短期偏好, 以解决仅仅将用户点击的最后一项项目表示作为用户短期偏好的问题。提出的基于 HPGNN 的会话推荐系统主要由会话图的构建、图神经位置感知层、图神经霍克斯层和预测层四部分组成。

3.3.1 会话图的构建

由于每条会话 S 都是根据点击时间进行排序的项目序列, 因此, 对于每个会话, 根据项目的点击时间顺序, 可以建模为有向图 $G = (V_s, E_s)$ 。图中的每个节点表示会话中的一个项目 $v_{s,i} \in V$, 每个有向边 $(v_{s,i}, v_{s,i+1})$ 表示用户在点击项目 $v_{s,i}$ 后点击了 $v_{s,i+1}$ 。所有节点集合表示为 V_s , 所有的边集合表示成 E_s 。根据会话图 G 中项目节点的出边和入边个数, 可以

得到图的邻接矩阵 A , A 包含出度矩阵 A_{out} 和入度矩阵 A_{in} 。需要注意的是, 由于每个节点的交互都不是唯一的, 需要将每个边的权重做归一化处理。例如节点 $v_{s,i}$ 有两条入边, 则其入度矩阵对应位置的值需要进行归一化处理, 将其赋值为 0.5。

3.3.2 图神经位置感知层

图神经位置感知层包括 GGNN 和位置感知注意力网络两部分, 其中 GGNN 用于学习用户的项目节点表示, 位置感知注意力机制利用位置信息辅助注意力机制来捕获会话序列的长距离依赖。

首先, GGNN 根据图的入度矩阵 A_{in} 、出度矩阵 A_{out} 和最初输入到 GGNN 模块的会话 $\mathbf{v}_s^{(0)} = [v_{s,1}^{(0)}, v_{s,2}^{(0)}, \dots, v_{s,m}^{(0)}]$, 采用 N 次迭代学习来获得每个项目节点的最终表示 $\mathbf{v}_s^{(N)} = [v_{s,1}^{(N)}, v_{s,2}^{(N)}, \dots, v_{s,m}^{(N)}]$ 。以 GGNN 第 t 步迭代为例描述其具体流程:

(1) 根据图的入度矩阵 A_{in} 、出度矩阵 A_{out} 和网络训练 $t-1$ 步迭代时的项目节点表示 $\mathbf{v}_s^{(t-1)} = [v_{s,1}^{(t-1)}, v_{s,2}^{(t-1)}, \dots, v_{s,m}^{(t-1)}]$ 来计算会话中项目节点对应的传播信息 $\mathbf{a}_s^{(t)}$, 具体如下所示:

$$\begin{aligned} \mathbf{a}_{s,in}^{(t)} &= \mathbf{A}_{in} [v_{s,1}^{(t-1)}, v_{s,2}^{(t-1)}, \dots, v_{s,m}^{(t-1)}] \mathbf{W}_{in} + \mathbf{b}_{in} \\ \mathbf{a}_{s,out}^{(t)} &= \mathbf{A}_{out} [v_{s,1}^{(t-1)}, v_{s,2}^{(t-1)}, \dots, v_{s,m}^{(t-1)}] \mathbf{W}_{out} + \mathbf{b}_{out} \\ \mathbf{a}_s^{(t)} &= \mathbf{a}_{s,in}^{(t)} \parallel \mathbf{a}_{s,out}^{(t)} \end{aligned} \quad (1)$$

其中 $\mathbf{v}_s^{(t-1)} = [v_{s,1}^{(t-1)}, v_{s,2}^{(t-1)}, \dots, v_{s,m}^{(t-1)}]$ 是通过 GGNN 的第 $t-1$ 步迭代得到的会话表示, m 表示会话内项目的个数, \mathbf{W}_{in} 和 \mathbf{W}_{out} 是可学习参数矩阵, \mathbf{A}_{in} 和 \mathbf{A}_{out} 分别表示会话图 G 的入度和出度矩阵, \mathbf{b}_{in} 和 \mathbf{b}_{out} 是偏置量, \parallel 表示拼接操作。

(2) GGNN 采用更新门和重置门机制确定每一次迭代训练需要丢弃或保留的信息。其中更新门

函数为:

$$\mathbf{z}_s^{(t)} = \sigma(\mathbf{W}_z \mathbf{a}_s^{(t)} + \mathbf{H}_z \mathbf{v}_s^{(t-1)}) \quad (2)$$

得到的 $\mathbf{z}_s^{(t)}$ 表示控制遗忘信息, \mathbf{W}_z 和 \mathbf{H}_z 是可学习的参数矩阵, $\sigma(\cdot)$ 表示 Sigmoid 函数。重置门函数为:

$$\mathbf{r}_s^{(t)} = \sigma(\mathbf{W}_r \mathbf{a}_s^{(t)} + \mathbf{H}_r \mathbf{v}_s^{(t-1)}) \quad (3)$$

得到的 $\mathbf{r}_s^{(t)}$ 表示更新信息, 其中 \mathbf{W}_r 和 \mathbf{H}_r 是可学习的参数矩阵。同时, 根据得到的更新信息 $\mathbf{r}_s^{(t)}$ 计算更新后的会话对应的用户信息 $\tilde{\mathbf{v}}_s^{(t)}$ 为:

$$\tilde{\mathbf{v}}_s^{(t)} = \tanh(\mathbf{W}_h \mathbf{a}_s^{(t)} + \mathbf{H}_h (\mathbf{r}_s^{(t)} \odot \mathbf{v}_s^{(t-1)})) \quad (4)$$

其中 $\tanh(\cdot)$ 是双曲正切函数, \mathbf{W}_h 和 \mathbf{H}_h 是可学习的参数矩阵, \odot 表示元素哈达玛积运算符。

(3) 求得第 t 步迭代的项目节点表示 $\mathbf{v}_s^{(t)}$ 为:

$$\mathbf{v}_s^{(t)} = (1 - \mathbf{z}_s^{(t)}) \odot \mathbf{v}_s^{(t-1)} + \mathbf{z}_s^{(t)} \odot \tilde{\mathbf{v}}_s^{(t)} + \frac{\mathbf{v}_s^{(t-1)}}{i_{\text{steps}}} \quad (5)$$

其中, $(1 - \mathbf{z}_s^{(t)}) \odot \mathbf{v}_s^{(t-1)}$ 表示选择遗忘哪些信息, $\mathbf{z}_s^{(t)} \odot \tilde{\mathbf{v}}_s^{(t)}$ 控制模型记住哪些新产生的信息, i_{steps} 为递减参数。同时, 为了将过去的信息与当前的信息进行融合, 还提出了逐次递减的残差网络应用于网络的更新迭代, 随着训练层数增多, 达到逐次递减的效果。

其次, 为了挖掘会话中项目之间的位置信息及对不同项目分配不同的注意力, 在图神经位置感知层的 GGNN 后还引入位置感知注意力网络, 其具体过程描述如下所示:

(1) 对 GGNN 得到的 $\mathbf{v}_s^{(N)}$ 加入位置信息, 其格式为:

$$\mathbf{v}_s^p = \mathbf{v}_s^{(N)} + \mathbf{p} \quad (6)$$

其中 \mathbf{p} 是会话中项目 $\mathbf{v}_s^{(N)}$ 的位置嵌入矩阵, 其格式如下所示:

$$\mathbf{p}_i^{(t)} = \begin{cases} \sin(t/10000^{2i/d}), & \text{if } t = 2k \\ \cos(t/10000^{2i/d}), & \text{if } t = 2k + 1 \end{cases} \quad (7)$$

其中 $\mathbf{p}_i^{(t)}$ 是位置嵌入矩阵 \mathbf{p} 的第 i 个会话项目的第 t 个位置的值, d 是会话中项目向量的维数, k 表示 t 位置奇偶位置。于是得到嵌入位置信息后的节点表示 $\mathbf{v}_{s,i}^p$ 。

(2) 为了捕获用户的长期偏好, 更好的预测用户下一个点击项目, 提出的 HPGNN 通过注意力机制聚集会话内所有节点向量的特征, 来捕获用户的全局兴趣偏好。考虑到这些嵌入的信息可能具有不同的优先级, 采用软注意力机制来更好地表示用户的长期偏好, 格式如下:

$$\alpha_i = \sigma(\mathbf{W}_1 \mathbf{v}_{s,m}^{(N)} + \mathbf{W}_2 \mathbf{v}_{s,i}^p + \mathbf{b})$$

$$\mathbf{s}_g = \sum_{i=1}^m \alpha_i \mathbf{v}_{s,i}^p \quad (8)$$

其中, $\mathbf{v}_{s,m}^{(N)}$ 表示目标会话中最后一个点击项目的嵌

入向量, m 表示会话内项目的总数, $\mathbf{v}_{s,i}^p$ 表示加入位置嵌入信息的项目矩阵对应的第 i 个项目向量, 参数 \mathbf{W}_1 和 \mathbf{W}_2 是项目嵌入向量的权重, \mathbf{b} 是偏置向量。

3.3.3 图神经霍克斯层

利用图神经霍克斯层学习用户的短期偏好, 以解决现有的基于 GNN 的会话推荐模型大多仅仅将用户点击的最后一项项目表示作为用户短期偏好的问题。霍克斯过程可以学习到时间序列中不同事件之间的影响程度, 并根据这些影响来预测未来事件发生的概率。会话序列在时间轴上离散排列, 本文默认两个事件的交互为一个确定的时间单位, 当会话在一段短时间内发生时, 这个时间差可以被忽略, 将整个会话的交互过程看作是发生在一个连续的时间域中, 其过程更加符合霍克斯过程的使用场景。受神经霍克斯过程将长短期记忆网络与霍克斯强度函数相结合的思想启发, 本文提出了图神经霍克斯层, 将霍克斯强度函数与图神经网络相结合, 学习项目点击量之间的交互影响, 并将其转化为用户的短期偏好。不仅考虑了用户的历史行为对他们当前行为的影响, 还更好地模拟了真实世界的交互过程。

在图神经霍克斯层中, 利用霍克斯强度函数捕获会话内某一事件类型在连续时间轴上的交互量信息, 其具体过程为:

首先, 图神经霍克斯层根据图的入度矩阵 \mathbf{A}_{in} 、出度矩阵 \mathbf{A}_{out} 和最初输入到该层的会话 $\mathbf{v}_s^{(0)} = [\mathbf{v}_{s,1}^{(0)}, \mathbf{v}_{s,2}^{(0)}, \dots, \mathbf{v}_{s,m}^{(0)}]$, 采用 M 次迭代学习来获得每个项目节点的最终表示 $\mathbf{v}_s^{(M)} = [\mathbf{v}_{s,1}^{(M)}, \mathbf{v}_{s,2}^{(M)}, \dots, \mathbf{v}_{s,m}^{(M)}]$ 以及会话中项目的强度信息 $\lambda_k^{(M)}$ 。以第 j 次迭代为例, 其格式为:

$$\begin{aligned} (\mathbf{v}_s^{(j)}, \mathbf{a}_s^{(j)}) &= \text{GGNN}(\mathbf{A}_{\text{in}}, \mathbf{A}_{\text{out}}, \mathbf{v}_s^{(j-1)}) \\ \mathbf{v}_{\text{hp}}^{(j)} &= \mathbf{v}_s^{(j)} \odot \sigma(\mathbf{W}_{\text{hp}} \mathbf{a}_s^{(j)} + \mathbf{H}_{\text{hp}} \mathbf{v}_s^{(j-1)}) \\ \lambda_k^{(j)} &= f_k(\mathbf{W}_k^T \mathbf{v}_{\text{hp}}^{(j)}) \end{aligned} \quad (9)$$

其中 $\mathbf{v}_s^{(j-1)}$ 和 $\mathbf{v}_s^{(j)}$ 分别表示会话中项目对应的第 $j-1$ 步和第 j 步迭代时学习得到的嵌入矩阵, GGNN 为门控图神经网络函数, $\mathbf{a}_s^{(j)}$ 为经 GGNN 处理后得到的会话中项目对应的传播信息, $\mathbf{v}_{\text{hp}}^{(j)}$ 表示第 j 步迭代时的中间矩阵, $\sigma(\cdot)$ 是 Sigmoid 函数, \mathbf{W}_{hp} 、 \mathbf{H}_{hp} 和 \mathbf{W}_k 是可学习参数矩阵, $\lambda_k^{(j)}$ 是网络迭代训练第 j 步时的霍克斯过程中的强度函数, $f_k(\cdot)$ 为形式任意的非线性函数, 本文采用 Sigmoid 函数。然后, 图神经霍克斯层最终输出会话中项目的强度函数 $\lambda_k^{(M)}$ 以及项目嵌入向量 $\mathbf{v}_s^{(M)}$, 最后, 将两者的哈达玛积表示为用户的短期偏好:

$$\mathbf{s}_c = \lambda_k^{(M)} \odot \mathbf{v}_s^{(M)} \quad (10)$$

3.3.4 预测层

为了准确描述用户的偏好,将用户的长期偏好与短期偏好相结合。具体而言,首先,利用注意力层得到的会话全局偏好表示 s_g 和由霍克斯过程计算得到的用户短期偏好 s_c 进行拼接操作,得到用户的最终会话表示,格式为:

$$s_{\text{final}} = s_g \parallel s_c \quad (11)$$

然后,使用用户的最终会话表示来计算每个候选项目 v_i 的推荐分数 \hat{z}_i , 格式为:

$$\hat{z}_i = s_{\text{final}}^T v_i \quad (12)$$

最后,使用 softmax 函数计算每个候选项目 v_i 被点击的概率 \hat{y} :

$$\hat{y} = \text{softmax}(\hat{z}) \quad (13)$$

其中, $\hat{y} = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_i]$ 表示候选项目 $v = [v_1, v_2, \dots, v_i]$ 将成为下一个点击项目的概率。并将预测值和真实值的交叉熵作为损失函数,其格式为:

$$L(\hat{y}) = - \sum_{i=1}^{|V|} y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \quad (14)$$

其中, y_i 表示真实值的一个热编码向量^[8]。

4 实验分析

本节描述了实验中使用的数据集、对比模型、评价指标和参数设置,并通过与当前先进的会话推荐模型进行对比,验证 HPGNN 的有效性和优越性。

4.1 实验环境和评价指标

4.1.1 数据集

本文使用 Diginetica 和 Yoochoose1/64 两个真实数据集来验证 HPGNN 的有效性和优越性。Diginetica 数据集来自 2016 年的 CIKM 杯,本文只使用其点击序列,其序列平均长度为 19.7。Yoochoose1/64 数据集来自 Recsys Challenge 2015,包含 6 个月内用户到电子商务网站的交互流,其序列平均长度为 6.9,本文只使用其前 1/64 点击流。为了实验的公平性,实验使用了与 Wu 等人^[8]相同的预处理方法。首先消除长度为 1 的会话和数据集中出现少于 5 次的项目;然后将这两个数据集划分为测试集和训练集,对于 Diginetica,前一天的数据为训练集,其余为测试集;对于 Yoochoose1/64,前 8 周的数据为训练集,其余为测试集。经过处理, Diginetica 包含 664609 个会话和 39056 个项目, Yoochoose1/64 包含 430448 个会话和 17319 个项目。

4.1.2 对比模型

将提出的 HPGNN 模型与如下的会话推荐模型进行对比:FPMC^[6]是使用马尔可夫链的经典推荐方法,通过为每个用户建立马尔可夫转移矩阵来整合序列和个性化信息;BPR-MF^[17]是一种基于矩阵

分解的推荐方法,新会话的特征向量被定义为会话中项目的特征向量的平均值,并将其作为用户的最终特征向量。GRU4Rec^[4]是一种基于 RNN 的推荐方法,通过多层 GRU 对用户进行建模;NARM^[6]是一种基于 RNN 的推荐方法,使用两个 RNN 分别捕获用户的全局偏好和当前偏好,并在 RNN 层后面添加交互式注意力网络;STAMP^[7]是一种基于注意力的推荐方法,使用最后点击的项目嵌入作为当前兴趣的表示,并通过注意力获得全局偏好;SR-GNN^[8]是一种基于 GNN 的推荐方法,将会话建模为图结构化数据,并使用 GNN 学习项目之间的复杂交互;GC-SAN^[20]是一种可以捕获全局上下文的 GNN 推荐方法,利用 Transformer 对会话上下文信息进行建模,以学习会话推荐的长期依赖性;DGSR^[28]是一种基于 GNN 的推荐方法,使用动态图神经网络有效地捕捉用户兴趣随时间变化的动态变化,并能够在多个时间步生成推荐;IHGCN^[29]也是一种基于 GNN 的推荐方法,利用图卷积网络学习用户的特征,利用特征维度级的细粒度注意力机制挖掘用户的全局偏好,并与短期偏好相融合得出最终的推荐结果。

4.1.3 评价指标

使用 P@20 和 MMR@20 这两个定量评价指标来评估模型。P@20 表示前 20 个推荐项目中下一个实际点击的项目的比例,MMR@20 是指目标推荐项目 v 在推荐列表中的排名。二者均是值越高,推荐性能越好。

4.1.4 参数设置

为了公平性,实验使用相同的实验环境和参数。实验是在一台配置了 12G 内存、NVIDIA GTX2060 6G 图形处理器和 Intel (R) Core (TM) i7-10875H CPU 2.30 GHz 处理器的计算机上进行,使用 Windows 10, 64 位操作系统。开发语言使用 Python3.7,并采用 PyTorch 1.7.0 作为学习框架。参数设置参考 SR-GNN^[8]的设置方法。对于这两个数据集,都设置了隐藏向量的维数 $d=100$ 。在模型训练方面,将隐藏大小设置为 100,学习率设置为 0.001,每 5 步后的学习率衰减 0.1, L_2 惩罚系数设置为 10^{-5} ,其他的参数采用默认值。

4.2 实验结果与分析

4.2.1 消融实验

本节进行消融实验,以验证图神经位置感知层和图神经霍克斯层的有效性。其中 GNP 为 HPGNN 去掉图神经霍克斯层,仅利用图神经位置感知层中 GGNN 输出的最后一项项目表示作为用户的短期偏好进行推荐;GNH 为 HPGNN 去掉了位置感知层后的模型。其实验结果如表 1 所示。

根据表 1 结果可知,GNP 的性能相比于提出的

表1 消融实验

Tab. 1 Ablation study results

模型	Diginetica		Yoochoose1/64	
	P@20	MMR@20	P@20	MMR@20
GNP	50.73	17.56	70.32	30.57
GNH	51.48	17.61	70.92	31.23
HPGNN	51.84	17.84	71.32	31.55

HPGNN在两个数据集上都有明显的下降,并且在Diginetica数据集上P@20指标下降了2.14%;此外,GNH性能相比HPGNN在两个数据集上也都有下降。实验结果表明,图神经位置感知层和图神经霍克斯层对推荐性能均有积极的促进作用。一方面,引入位置感知注意力网络,将用户的位置向量嵌入,充分捕获了项目在会话内的相对位置,有效地提升了推荐模型的性能;另一方面,将图神经网络和霍克斯过程相结合,不仅有效提取了用户点击量之间的交互关系,并且引入了时序信息,对提升模型的推荐性能起到了至关重要的作用。

4.2.2 对比实验

本节将提出的会话推荐模型HPGNN与其他基线模型进行对比实验,实验结果如表2所示。

表2 对比实验

Tab. 2 Comparative experiment results

模型	Diginetica		Yoochoose1/64	
	P@20	MMR@20	P@20	MMR@20
FPMC ^[16]	26.53	6.95	45.62	15.01
BPR-MF ^[17]	5.24	1.98	31.31	12.06
GRU4Rec ^[4]	29.45	8.33	60.64	22.89
NARM ^[6]	49.70	16.17	68.32	28.63
STAMP ^[7]	45.65	14.32	68.74	29.67
SR-GNN ^[8]	50.41	17.36	69.86	30.51
GC-SAN ^[20]	49.23	16.01	70.05	30.44
DGSR ^[28]	50.67	17.94	70.23	30.45
IHGNN ^[29]	51.56	17.98	71.04	30.53
HPGNN	51.84	17.84	71.32	31.55

由表2可知,本文提出的HPGNN在Yoochoose1/64数据集上展现出了最好的性能,在Diginetica数据集上也展现了较优的性能,其实验结果分析如下:

对于两个基于马尔可夫链的FPMC和BPR-MF,二者都展现了相对较差的结果。这可能是由于马尔可夫链假设当前的交互仅仅取决于一个或最近的几个交互,只能捕获点式依赖关系,而忽视了协同依赖关系;大多数基于深度学习的方法,例

如GRU4Rec、NARM和STAMP,它们的性能相比于传统方法都有了显著提升,这得益于深度学习可以挖掘用户和项目之间复杂的交互能力。与GRU4Rec相比,NARM和STAMP的性能要更加优秀,虽然它们都利用GRU来捕获项目转换表达出的用户的长期偏好,但是NARM同时还利用注意力机制自适应地对历史行为中的不同部分进行加权,增强了模型对关键行为的关注度;STAMP则将最后一次点击的项目表示定义为用户的短期偏好,进一步优化了模型对用户偏好的描述;SR-GNN、GC-SAN、DGSR和IHGNN均是利用GNN对会话进行建模,它们的性能均优于GRU4Rec和STAMP等基于深度学习的会话推荐方法,这是由于基于GNN的会话推荐方法将会话建模成图结构,能够捕捉更加复杂的项目之间的双向交互。GC-SAN虽然利用自注意力机制较好的捕获了项目的全局依赖信息,但也忽略了用户短期偏好对最终推荐结果的影响,性能相比SR-GNN略有下降。DGSR相较于SR-GNN和GC-SAN,在性能上略有提升,这是因为它使用了动态图神经网络来捕捉用户兴趣随时间的动态变化,并利用时间信息来辅助推荐。IHGNN利用特征维度级的细粒度注意力机制挖掘用户的全局偏好,对于长序列而言,序列内部存在大量的信息冗余和纵向依赖关系,所以需要更加细致的注意力机制来准确地捕捉序列内部的关键特征,因此IHGNN在Diginetica数据集上具有较好的性能。在Yoochoose1/64数据集上,提出的HPGNN具有最好的性能,这是由于Yoochoose1/64的平均序列长度相对较短,相对信息较少,用户的短期偏好对推荐结果的影响比重更大。提出的HPGNN利用图神经霍克斯层捕获了项目点击量之间的交互关系,克服了仅将会话中最后一项节点表示作为用户短期偏好的问题,因此取得了最优的性能。在Diginetica数据集上,HPGNN的MMR@20指标略低于IHGNN,其性能优于除IHGNN以外的其他所有对比模型,这是由于HPGNN除了增强了用户短期偏好的表示,还利用位置感知注意力考虑了项目在会话内的相对位置信息。另外,值得注意的是,由于特征维度级的细粒度注意力机制的引入,导致IHGNN的计算复杂度要明显高于HPGNN。因此,综合考虑模型的效果和复杂度,提出的HPGNN在两个数据集上均取得了最优的推荐性能。

5 结论

本文提出了一种基于霍克斯过程和图神经网络的会话推荐系统,利用图神经位置感知层和图神

经霍克斯层分别学习用户的长期和短期偏好,以解决现有模型大都简单的将会话的最后一个项目表示作为用户短期偏好,以及忽视了会话内项目的相对位置信息等问题。一方面,在图神经位置感知层,引入逐次递减的残差网络,解决网络过拟合问题,并将过去的编码信息与当前网络有效结合。此外,还利用位置感知注意力网络,充分提取项目在会话中的相对位置信息,并与会话信息相融合。另一方面,在图神经霍克斯层,将GGNN与霍克斯函数结合,来捕获项目点击量之间的交互信息,以提升推荐性能。最后,在Diginetica和Yoochoose1/64两个开放数据集上进行消融实验和对比实验,实验结果表明,提出的HPGNN相比于其他会话推荐模型展现出了最好的推荐性能。

参考文献

- [1] WANG Shoujin, CAO Longbing, WANG Yan, et al. A survey on session-based recommender systems[J]. *ACM Computing Surveys*, 2022, 54(7): 1-38.
- [2] WU Shiwen, SUN Fei, ZHANG Wentao, et al. Graph neural networks in recommender systems: A survey[J]. *ACM Computing Surveys*, 2023, 55(5): 1-37.
- [3] PANG Yitong, WU Lingfei, SHEN Qi, et al. Heterogeneous global graph neural networks for personalized session-based recommendation [C]//*Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*. Virtual Event, AZ, USA. New York: ACM, 2022: 775-783.
- [4] HIDASI B, KARATZOGLOU A. Recurrent neural networks with top-k gains for session-based recommendations [C]//*Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. Torino, Italy. New York: ACM, 2018: 843-852.
- [5] CHO K, VAN MERRIENBOER B, BAHDANAU D, et al. On the properties of neural machine translation: Encoder-decoder approaches [C]//*Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*. Doha, Qatar. Stroudsburg, PA, USA: Association for Computational Linguistics, 2014: 103-111.
- [6] LI Jing, REN Pengjie, CHEN Zhumin, et al. Neural attentive session-based recommendation [C]//*Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. Singapore, Singapore. New York: ACM, 2017: 1419-1428.
- [7] LIU Qiao, ZENG Yifu, 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*. London, United Kingdom. New York: ACM, 2018: 1831-1839.
- [8] WU Shu, TANG Yuyuan, ZHU Yanqiao, et al. Session-based recommendation with graph neural networks [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, 33(1): 346-353.
- [9] MEI Hongyuan, EISNER J. The neural Hawkes process: A neurally self-modulating multivariate point process [C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach, California, USA. New York: ACM, 2017: 6757-6767.
- [10] HU Liang, CAO Longbing, WANG Shoujin, et al. Diversifying personalized recommendation with user-session context [C]//*Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*. Macao, China. California: International Joint Conferences on Artificial Intelligence Organization, 2017: 1858-1864.
- [11] HE Xiangnan, ZHANG Hanwang, KAN Minren, et al. Fast matrix factorization for online recommendation with implicit feedback [C]//*Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. Pisa, Italy. New York: ACM, 2016: 549-558.
- [12] ZHANG Ruisheng, LIU Qidong, Chun-Gui, et al. Collaborative filtering for recommender systems [C]//*2014 Second International Conference on Advanced Cloud and Big Data*. Huangshan, China. IEEE, 2015: 301-308.
- [13] HE Xiangnan, LIAO Lizi, ZHANG Hanwang, et al. Neural collaborative filtering [C]//*Proceedings of the 26th International Conference on World Wide Web*. New York: ACM, 2017: 173-182.
- [14] 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*. New York: ACM, 2001: 285-295.
- [15] EIRINAKI M, VAZIRGIANNIS M, KAPOGIANNIS D. Web path recommendations based on page ranking and Markov models [C]//*Proceedings of the 7th Annual ACM International Workshop on Web Information and Data Management*. Bremen, Germany. New York: ACM, 2005: 2-9.
- [16] RENDLE S, FREUDENTHALER C, SCHMIDT-THIEME L. Factorizing personalized Markov chains for next-basket recommendation [C]//*Proceedings of the 19th International Conference on World Wide Web*. Raleigh, North Carolina, USA. New York: ACM, 2010: 811-820.
- [17] RENDLE S, FREUDENTHALER C, GANTNER Z, et al. BPR: Bayesian personalized ranking from implicit feedback [C]//*Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*. Montreal, Quebec, Canada. New York: ACM, 2009: 452-461.
- [18] WANG Shoujin, HU Liang, WANG Yan, et al. Modeling multi-purpose Sessions for next-item recommendations via mixture-channel purpose routing networks [C]//

- Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao, China. New York: ACM, 2019: 3771-3777.
- [19] DONKERS T, LOEPP B, ZIEGLER J. Sequential user-based recurrent neural network recommendations [C]//Proceedings of the Eleventh ACM Conference on Recommender Systems. Como, Italy. New York: ACM, 2017: 152-160.
- [20] XU Chengfeng, ZHAO Pengpeng, LIU Yanchi, et al. Graph contextualized self-attention network for session-based recommendation [C]//Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. Macao, China. California: International Joint Conferences on Artificial Intelligence Organization, 2019: 3940-3946.
- [21] WU Zonghan, PAN Shirui, CHEN Fengwen, et al. A comprehensive survey on graph neural networks [J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(1): 4-24.
- [22] WAIKHOM L, PATGIRI R. Graph neural networks: Methods, applications, and opportunities [EB/OL]. 2021: arXiv: 2108.10733. <https://arxiv.org/abs/2108.10733>.
- [23] WANG Ziyang, WEI Wei, CONG Gao, et al. Global context enhanced graph neural networks for session-based recommendation [C]//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. Virtual Event, China. New York: ACM, 2020: 169-178.
- [24] WEN Yan, KANG Shitao, ZENG Qingtian, et al. Session-based recommendation with GNN and time-aware memory network [J]. Mobile Information Systems, 2022, 2022: 1-12.
- [25] LI Yujia, TARLOW Daniel, BROCKSCHMIDT Marc, et al. Gated graph sequence neural networks [C]. Proceedings of the International Conference on Learning Representations (ICLR), 2016: 1-20.
- [26] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]. 31st Conference on Neural Information Processing Systems, 2017: 1-11.
- [27] YU Feng, ZHU Yanqiao, LIU Qiang, 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. Virtual Event, China. New York: ACM, 2020: 1921-1924.
- [28] ZHANG M, WU S, YU X, et al. Dynamic graph neural networks for sequential recommendation [J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(5): 4741-4753.
- [29] 闫昭, 项欣光, 李泽超. 基于交互序列商品相关性建模的图卷积会话推荐 [J]. 中国科学: 信息科学, 2022, 52(6): 1069-1082.
- YAN Zhao, XIANG Xinguang, LI Zechao. Item correlation modeling in interaction sequence for graph convolutional session recommendation [J]. Scientia Sinica (Informationis), 2022, 52(6): 1069-1082. (in Chinese)
- [30] LI Zechao, TANG Jinhui, MEI Tao. Deep collaborative embedding for social image understanding [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(9): 2070-2083.
- [31] 李泽超, 唐金辉. 社交网络图像智能分析方法 [J]. 科学通报, 2023, 68: 1-17.
- LI Zechao, TANG Jinhui. A survey on social image semantic analysis [J]. Chinese Science Bulletin, 2023, 68: 1-17. (in Chinese)
- [32] QIAO Shutong, ZHOU Wei, WEN Junhao, et al. Multi-perspective enhanced representation for effective session-based recommendation [J]. Knowledge-Based Systems, 2023, 263: 110284.

作者简介



杨真真 女, 1984年生, 山东临沂人, 南京邮电大学副教授、博士。研究方向为深度学习、多媒体信息处理。
E-mail: yangzz@njupt.edu.cn



闫孟儒 男, 1998年生, 江苏徐州人, 南京邮电大学研究生。研究方向为深度学习、推荐系统。
E-mail: 15651725282@163.com



杨永鹏 男, 1984年生, 山东高密人, 南京信息职业技术学院讲师、南京邮电大学博士研究生。研究方向为深度学习、多媒体信息处理。
E-mail: yangyp@njcit.cn



陈亚杰 女, 1999年生, 山西晋中人, 南京邮电大学研究生。研究方向为深度学习、多媒体信息处理。
E-mail: 1012004556@qq.com

(责任编辑: 刘建新)