

基于一维卷积神经网络的序列推荐算法

黄康鹏, 冯 锋

(宁夏大学 信息工程学院, 宁夏 银川 750021)

摘要:近年来,针对信息过载问题,推荐算法已成为关键解决方法之一,能够有效地提供用户个性化内容。在序列推荐研究中,卷积神经网络因其能有效提取序列信息中的局部特征而受到广泛关注。然而,卷积神经网络在捕捉时序信息方面存在局限性。为了解决这一问题,提出了一种基于一维卷积的序列推荐算法。该算法首先通过卷积操作提取序列的局部特征,然后通过池化操作提取序列的长期特征,并将两者进行加权相加获得用户特征信息,使得其能够综合提取局部特征和长期特征。随后,将用户特征信息与线性变化的序列信息进行点乘,以引入时序信息。此外,将用户特征信息通过前馈网络,以实现非线性转换和跨维度交互增强。最后,对用户特征向量与项目特征向量进行计算获得相关性,并以此进行推荐。实验结果表明,在 MovieLens 电影数据集和 KuaiRand 短视频数据集上的测试中,该算法在推荐命中率和归一化折损累计增益等指标上相比四种基线算法均有显著提升,表明该算法能够更有效地进行推荐。

关键词:推荐算法;序列推荐;卷积神经网络;前馈网络;用户特征

中图分类号:TP391.3

文献标识码:A

文章编号:1673-629X(2025)03-0172-07

doi:10.20165/j.cnki.ISSN1673-629X.2024.0335

Sequence Recommendation Algorithm Based on One-dimensional Convolutional Neural Network

HUANG Kang-peng, FENG Feng

(School of Information Engineering, Ningxia University, Yinchuan 750021, China)

Abstract: In recent years, recommendation algorithms have become a key solution to the problem of information overload, providing users with personalized content effectively. In the field of sequential recommendation, convolutional neural networks have gained widespread attention for their ability to extract local features from sequential information. However, convolutional neural networks have limitations in capturing temporal information. To address this issue, we propose a temporal recommendation algorithm based on one-dimensional convolution. The proposed algorithm first extracts local features from the sequence through convolution operations, then extracts long-term features through pooling operations, and combines the two using weighted addition to effectively capture both local and long-term features. Next, the extracted information is multiplied pointwise with linearly transformed sequential information to introduce temporal information. Additionally, a feedforward network is used to achieve nonlinear transformation and enhance cross-dimensional interactions. Finally, the algorithm calculates the correlation between user feature vectors and item feature vectors to make recommendations. Experimental results show that in tests on the MovieLens movie dataset and the KuaiRand short video dataset, the proposed algorithm significantly improves metrics such as hit rate and normalized discounted cumulative gain compared to four baseline algorithms. It is demonstrated that the proposed algorithm is more effective in making recommendations.

Key words: recommendation algorithm; sequential recommendation; convolutional neural network; feedforward network; user feature

0 引言

随着社交网络、新闻媒体及电子商务等互联网服务领域信息量的爆炸性增长,用户正面临着日益严重的“信息过载”问题。为了应对这一挑战,推荐算法^[1-3]已成为帮助用户高效筛选并精准对接其个性化

需求的关键技术手段。然而,由于隐私保护等原因,许多用户信息在实际应用中无法获取。因此,在缺乏用户显式信息的情况下,需要根据用户的历史行为来建模其动态偏好,并预测用户可能感兴趣的下一个项目。这一研究方向被称为序列推荐^[4-5](Sequential Recom-

收稿日期:2024-07-02

修回日期:2024-11-05

基金项目:宁夏重点研发计划重点项目(2022BEG02016);宁夏自然科学基金项目(2023AAC03031)

作者简介:黄康鹏(2000-),男,硕士研究生,研究方向为推荐算法;冯 锋(1971-),男,教授,博士,博导,研究方向为信息系统集成及应用、物联网技术。

mendation, SR), 已成为推荐算法研究领域的重点和热点。

近年来,随着深度学习技术的迅猛发展,越来越多的相关技术被应用于序列推荐系统的研究中。其中,基于卷积神经网络的方法^[6-8]已经成为具有竞争性能的主流解决方案。然而,卷积神经网络不能编码序列中元素的时间或顺序信息,只是将序列信息进行聚合。受到 Yang Jianwei 等人^[9]在图像识别的研究和焦点自注意力^[10]的启发,该文提出了一种基于一维卷积网络的信息提取模块 (Causal Convolutional Feature Extractor, CCFE)。首先,该模块通过线性变化将输入序列投影到查询信息和新特征域,在新特征域上通过一维卷积操作提取用户兴趣信息,再将兴趣信息与查询信息进行结合,以加入时序信息。

主要工作包括:

(1)提出了一种基于一维卷积的信息提取模块 CCFE。该模块首先通过卷积操作聚合特征,然后通过逐个元素相乘的方式进行加权求和,从而提取序列信息中的用户兴趣,同时具有较低的时间复杂度。

(2)基于 CCFE 模块,提出了序列推荐模型 CCFERec。该模型首先通过堆叠的 CCFE 模块和前馈网络进行兴趣提取,其中 CCFE 模块负责用户兴趣的复杂提取,前馈网络负责非线性处理。然后将提取的兴趣特征与项目信息相结合,以获得项目评分预测。该模型能够有效地提取序列信息中的用户特征,并通过序列建模进行项目预测。

(3)在多个公开推荐数据集上对模型进行了对比分析实验,实验结果表明,CCFERec 模型是有效的。进一步的消融研究解释了该模块的优越性。

1 相关研究

1.1 序列推荐

序列推荐系统旨在根据用户过往与项目交互形成的序列信息,预测其下一步可能感兴趣的目标项目。早期的非深度学习方法主要依赖于马尔可夫链理论^[11]和矩阵分解技术^[12-13],通过对序列内项目间转换关系的建模来捕捉用户行为的序列模式。随着深度学习在图像识别和自然语言处理等领域的广泛应用与显著成效,这些方法逐渐被应用到推荐系统的研究与实践中,并迅速占据主导地位。代表性的工作包括利用卷积神经网络 (Convolutional Neural Network, CNN)^[6-8]、递归神经网络 (Recurrent Neural Network, RNN)^[14-16]以及图神经网络 (Graph Neural Network, GNN)^[17-18]等架构来实现高效的序列推荐。

1.2 卷积神经网络

卷积神经网络是一类专门用于处理具有网格拓扑

结构数据的深度学习模型。卷积神经网络已在深度学习领域取得了显著成果,并在推荐算法领域得到了广泛应用。

其中,一维卷积是一种特殊的卷积操作,主要用于时间序列数据或序列数据的处理。通过适当的填充,确保当前时间步的输出只依赖于当前及之前的时间步,而不依赖于未来的时间步。该操作可以保持序列数据的顺序关系,符合实际应用场景中的时间顺序要求。Yang Wenbiao 等人^[19]提出了基于一维卷积的模型用于时间序列分类,取得了优异的性能。尹梓诺等人^[20]将一维卷积用于流量信息特征分析,在异常流量识别方面取得了显著进展。然而,卷积网络无法编码时序信息。针对该问题,该文提出 CCFE 模块,将卷积后的信息与序列信息进行点乘,以实现耦合时序信息的效果。

1.3 自注意力机制

自注意力机制是一种广泛应用于深度学习模型中的技术,特别是在自然语言处理和图像处理等领域。Vaswani 等人^[21]首次提出了自注意力机制,并将其应用于 Transformer 模型中。自注意力机制允许模型动态且位置敏感地聚合信息,使其既能聚焦于序列中的重点信息,即局部兴趣,也能综合考虑整个序列的全局依赖关系,即全局兴趣。

在序列推荐研究领域,Wang-Cheng Kang 等人^[22]提出的 SASRec 模型依靠堆叠的自注意力机制模块,在项目序列中提取用户兴趣,表现出色。孟志鹏等人^[23]提出的 SNTSR 模型,则将时序信息融合自注意力网络中,同样取得了优异的效果。受此启发,该文提出了 CCFE 模块。不同于自注意力机制先求权重后聚合特征的流程,该模块在进行线性变换后,首先通过卷积操作聚合特征,然后通过逐个元素相乘的方式进行加权求和。图 1 分别展示了自注意力机制与 CCFE 模块的流程。

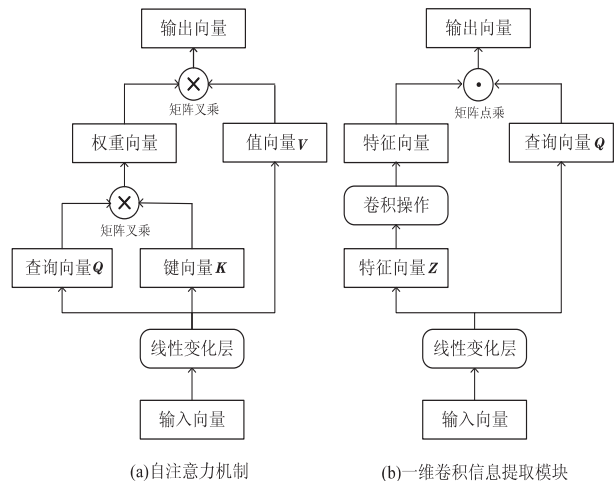


图 1 具体流程对比

2 序列推荐模型 CCFERec

2.1 问题定义

在该研究中,设用户集合为 $U = \{u_1, u_2, \dots, u_m\}$, 项目集合为 $I = \{i_1, i_2, \dots, i_n\}$, 其中 m 和 n 表示用户集合数量和项目集合数量。对于目标用户 $u \in U$, 在某一时刻,其历史交互项目序列可以表示为 $S = \{s_1, s_2, \dots, s_l\}$, 其中 l 表示项目序列的长度。

该文主要研究在给定用户 u 的行为序列 S , 在时间步 $l-1$ 时,根据前 $l-1$ 个项目预测目标用户 u 与目标项目 s_l 的相关性 r , 以进行推荐。

2.2 CCFERec 模型

CCFERec 模型主要包括三部分:信息嵌入层、基于卷积神经网络的信息提取层、项目预测层。具体模型结构如图 2 所示。

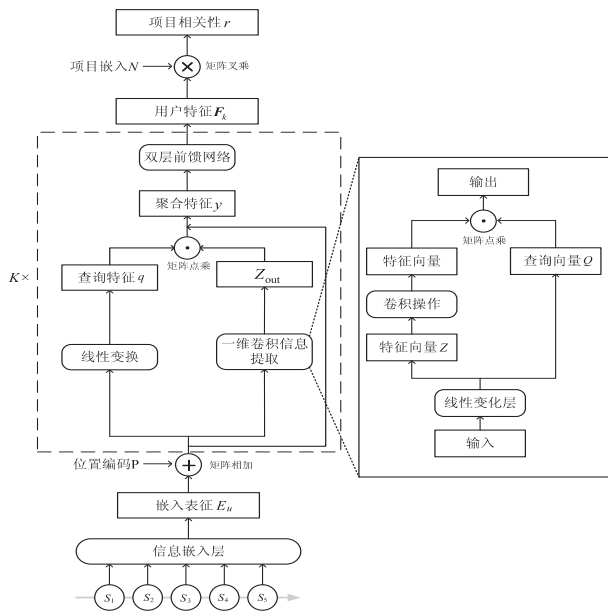


图 2 CCFERec 模型结构

信息嵌入层主要将离散的、高维的项目序列 S 数据信息转化为连续的、低维的向量表示。基于卷积神经网络的信息提取层通过堆叠的基于一维卷积网络的信息提取模块 (CCFE) 和基于前馈网络的非线性处理模块 (Feed-Forward Network, FFN) 来实现用户特征提取。项目预测层将项目嵌入与用户特征相乘,得到最终的项目相关性。

2.2.1 信息嵌入层

首先,对于项目集合建立一个嵌入矩阵 $E \in R^{n \times v}$, 其中 n 为项目集合数量, v 为嵌入维度。对于目标用户 u , 将历史交互序列 $S \in R^{n \times l}$ 通过独热编码转化为 $S_1 \in R^{n \times v}$ 。再通过嵌入矩阵 E , 将 S_1 转化为嵌入表征 $E^u \in R^{n \times v} = \{e_1^u, e_2^u, \dots, e_l^u\}^T$ 。

最后,将嵌入表征 E^u 与位置编码 P 相加,并将其输入基于卷积神经网络的信息提取层。

2.2.2 基于卷积神经网络的信息提取层

焦点信息处理层由两个子模块组成:基于一维卷积神经网络的信息提取模块 (CCFE) 和基于前馈网络的非线性处理模块 (FFN)。基于一维卷积神经网络的信息提取模块,是为了深入挖掘序列信息。基于前馈网络的非线性处理模块,为了赋予模型非线性并考虑不同潜在维度之间的相互作用。最后,通过将基于卷积神经网络的信息提取层进行堆叠,以实现学习更复杂的特征。

在基于卷积的信息聚合模块中,将输入特征 F 通过线性变化,形成查询特征 q 和新的特征域 Z_0 。在新的特征域 Z_0 上,利用卷积网络和门控机制来聚合项目上下文信息,获得调制信息 Z_{out} 。最后,将查询特征 q 和调制信息 Z_{out} 进行点乘,获得最终信息 y 。具体流程如下:

第一步,线性变化:给定输入特征矩阵 F , 再通过线性映射 f_z 转换至新的特征,从而得到 $Z_0 = f_z(F) \in R^{n \times v}$, 为后续处理奠定基础。同时,通过线性映射生成查询特征 $q \in R^{n \times v}$ 。

第二步,上下文信息提取:对于新特征 Z_0 , 首先依靠一维卷积操作提炼出序列数据中的上下文表示。具体而言,特征映射 Z_0 经过一种右端补零策略预处理,旨在确保信息抽取专注于每个项目之前的序列内容。随后,采用具有核大小 k 的一维卷积神经网络,以捕捉局部上下文特征,并结合 GeLU 作为激活函数,以引入必要的非线性。计算过程如式 1 所示。

$$Z_1 = \text{GeLU}(\text{ODConv}(Z_0)) \quad (1)$$

其中, ODConv 表示一维卷积操作。此外,为了全面整合序列的全局上下文,实施全局平均池化 (Global Average Pooling, GAP) 操作,以产生第 2 层的特征 Z_2 。计算过程如式 2 所示。

$$Z_2 = \text{GAP}(Z_1) \quad (2)$$

这一系列操作汇聚了两个特征集合 $\{Z_1, Z_2\}$, 协同覆盖了从细粒度的瞬时上下文到宏观的序列全局信息,实现了对序列内复杂依赖结构的高效编码与解析。

在此基础上,该文使用门控机制来聚合调制特征,通过对特征矩阵 E^u 使用线性层来获得相应的门控权重 $G^l \in R^{n \times v}$ 。计算过程如式 3 所示。

$$G^l = f_g^l(E^u) \quad (3)$$

其中, f_g^l 是一个线性变换层, l 表示层数。然后,通过元素相乘执行加权和,以获得与输入 E^u 大小相同的单个特征映射 $Z_{out}^l \in R^{n \times v}$ 。计算过程如式 4 所示。

$$Z_{out} = \sum_{l=1}^2 G^l \odot Z^l \quad (4)$$

其中, G^l 是第 l 层的门控权重, Z^l 表示第 l 层的卷积特征, \odot 符号表示矩阵元素对应相乘操作。

第三步,聚合特征:使用查询特征 q 与特征映射 Z_{out} 进行矩阵点乘操作,同时,为了避免后续堆叠策略带来的影响,此处引入残差网络^[24]设计来提升模型泛化能力,从而形成最终信息 y_t 。计算过程如式 5 所示。

$$y_t = (q \odot Z_{out}) + F \quad (5)$$

其中, \odot 符号表示矩阵元素对应相乘操作, F 表示输入特征矩阵。

获得最终信息 y_t 后,将其输入基于前馈网络的非线性处理层中。而在基于前馈网络的非线性处理层中,该文采纳前馈神经网络(FFN)实现对信息的非线性转化及跨维度的交互增强。具体而言,通过部署双层前馈网络,以点对点的方式对最终整合信息 y_t 进行操作,旨在捕捉非线性特征并促进特征间的协同作用,最终获得用户特征 F_t 。计算过程如式 6 所示。

$$F_t = \text{ReLU}(y_t W^1 + b^1) W^2 + b^2 \quad (6)$$

其中, W^1, W^2 是可学习的矩阵, b^1, b^2 是偏置项, ReLU 是激活函数,旨在引入必要的非线性关系。

在初步实现序列信息的信息提取之后,为深化模型对复杂项目序列的信息提取能力,该文创新性地采用了堆叠策略,将信息提取层堆叠 k 层,得到最终的用户特征 F_t^k 。对于第 b 层的信息提取层,其输入为 $b-1$ 层的特征输出,计算过程如式 7、式 8 所示。

$$y^b = \text{CCFE}(F_{b-1}) \quad (7)$$

$$F_t^b = \text{FFN}(y^b) \quad (8)$$

其中, $b \in \{1, 2, \dots, k\}$ 表示层数, k 表示堆叠总块数。同时,对于第一层,输入 F_0 为嵌入表征 E^u 。

2.2.3 项目预测层

在经历 k 个层的信息处理层,对用户过往消费项目的深层次特征进行充分提取之后,针对序列中的前 i 个项目,该文利用用户特征表示 F_t^k , 结合预测项目嵌入 N_i , 通过特定的交互机制来量化项目与用户兴趣的相关性。计算过程如式 9 所示。

$$r_{i,t} = F_t^k N_i \quad (9)$$

其中, $r_{i,t}$ 表示在给定性序列 (s_1, s_2, \dots, s_t) 时,对于项目 i 的相关性预测评分。 N_i 表示项目 i 的嵌入表达。高交互分数 $r_{i,t}$ 意味着高相关性,因此通过对分数进行排序来产生项目推荐序列。

2.2.4 损失函数

该文使用二进制交叉熵作为模型的损失函数。损失函数的具体形式如公式 10 所示。

$$L = - \sum_{S^* \in S} \sum_{t \in [1, 2, \dots, n]} [\log(\sigma(r_{o,t})) + \sum_{j \in S^*} \log(1 - (r_{j,t}))] \quad (10)$$

其中, S 表示训练集中的用户序列集合, n 是序列的最大长度, o 是在时间步 t 的目标项目, $r_{o,t}$ 是模型预测在

时间步 t 预测目标项目 o 的相关性得分, σ 是用于将相关性得分转换为概率的 Sigmoid 函数。 $r_{j,t}$ 是模型在时间步 t 预测负样本项目 j 的相关性得分。负样品项目 j 是随机采样,即不在用户实际交互的项目集合中。该文对每一个目标项目 o , 随机采样一个负样本项目 j 。

3 实验与分析

3.1 实验设置

3.1.1 数据集

该文在两个公开数据集 MovieLens 和 KuaiRand^[25]上进行实验。MovieLens 是一个广泛用于评估协同过滤算法的基准数据集,其数据包含了用户评分、用户标签等信息。该文使用 MovieLens-1M 版本。KuaiRand 数据集是一个无偏见的顺序推荐数据集,收集自视频共享移动应用程序快手的推荐日志,推荐日志包含用户 id、视频 id、观看时间、视频长度等信息。该文使用 KuaiRand-27K 版本。对于 KuaiRand-27K, 仅使用 2022 年 4 月 8 日的相关信息。

对于所有数据集,该文将评论或评分的存在视为隐式反馈(即用户与项目进行交互),并使用时间戳来确定动作的顺序并丢弃了少于 5 个相关动作的用户和项目。同时将每个用户 u 的历史交互序列 S_u 拆分为三部分:训练集、验证集、测试集。测试集包含序列的最后一个交互,验证集包含序列的倒数第二个交互,训练集包含其余剩下交互。两个数据集的统计数据展示在表 1 中。

表 1 数据集统计信息

数据集	用户数	项目数	交互数	平均序列长度
MovieLens-1M	6 040	3 416	987 531	163.5
KuaiRand-27K	23 528	1 171 045	3 845 681	163.4

为了避免在所有项目评分对上进行繁重的计算,该文遵循了文献[4]中的策略。对于每个用户 u , 随机抽取 100 个负样本项目,并将这些项目与真实的项目一起进行排序。基于这 101 个项目的排名,计算评价指标。

3.1.2 评价指标

该文采用命中率(Hit Rate, HR)和归一化折损累计增益(Normalized Discounted Cumulative Gain, NDCG)作为实验结果的评价指标。HR@ N 的计算方式如公式 11 所示。

$$\text{HR}@N = \frac{1}{N} \sum_{i=1}^N \text{hit}(i) \quad (11)$$

其中, N 为项目数量, $\text{hit}(i)$ 表示用户真实访问的项目是否存在于推荐列表中,存在时, $\text{hit}(i) = 1$, 不存在, $\text{hit}(i) = 0$ 。

NDCG@N 的计算方式如公式 12 所示。

$$NDCG@N = \frac{1}{N} \sum_{i=1}^N \frac{1}{\log_2(p_i + 1)} \quad (12)$$

其中, N 为项目数量, p_i 为用户真实访问的项目在推荐列表中的位置。受推荐列表长度的限制, 实验中取前 10 个项目作为推荐列表, 则两个指标分别表示为 HR@10 和 NDCG@10。

3.1.3 对比方法

为验证文中模型的效能, 选择 4 种代表性模型进行对比:

RBPR^[12]: 该模型是从隐式反馈中学习个性化排名的经典方法。偏置矩阵分解被用作基础推荐算法。

GRU4Rec^[16]: 该模型运用 GRU 网络, 对用户顺序交互进行深度建模。

SR-GNN^[17]: 该模型基于图神经网络, 提取捕获项目之间的转换信息。

Caser^[6]: 该模型基于卷积神经网络, 提取用户行为中的序列模式。

SASRec^[22]: 该模型基于自注意力机制, 捕获用户序列偏好以进行推荐。

3.1.4 参数设置

实验中, 对于 CCFERec 模型的堆叠架构, 该文使用了两个基于卷积神经网络的焦点信息处理层 ($k = 2$)。位置编码 P 是可训练的参数矩阵。嵌入层和预测层中的项目嵌入矩阵是共享的。焦点信息处理层中的卷积核长度在 [3, 5, 7, 9] 中选择。

为了公平对比, 所有网络模型优化器为 Adam 优化器, 学习率设置为 0.001, 批处理大小为 128。数据集的数据差异, drop 率在 MovieLens-1M 数据集上设为 0.3, 在 KuaiRand-27K 数据集上设为 0.2。正则化权重在 MovieLens-1M 为 0.001, 在 KuaiRand-27K 设为 0.01。嵌入维度在 [5, 10, 15, 25, 30, 40, 50] 中选择。为确保模型达到最优效能并防止过拟合现象, 采用了早停策略 (Early Stopping)。在 MovieLens-1M 数据集上, 进行了为期 50 个轮次的初步训练, 一旦检测到验证集的损失值在连续 5 个轮次训练内无显著下降, 则立即终止训练流程。而对于 KuaiRand-27K 数据集, 初步训练设定为 25 个轮次。

3.2 实验结果及分析

3.2.1 不同模型结果对比分析

表 2 展示了所有方法在两个数据集上的推荐系统性能评估。CCFERec 方法性能表现都优于其他方法。在数据集 MovieLens-1M 上, 与 SASRec 相比, CCFERec 在 HR@10 指标上实现了 0.85 个百分点的提升, 在 NDCG@10 指标上实现了 1.1 个百分点的提升。这一结果表明, CCFERec 在相对密集的数据环境中更

能有效地捕获并反映用户的兴趣。在较为稀疏的 KuaiRand-27K 数据集上, CCFERec 在 HR@10 指标上实现了 0.85 个百分点的提升, 在 NDCG@10 指标上实现了 0.72 个百分点的提升。结果表明 CCFERec 通过全局池化操作和非线性处理层的设计, 进一步优化了对用户长期兴趣模式的挖掘能力。

表 2 性能表现统计信息

模型	MovieLens-1M		KuaiRand-27K	
	HR@10	NDCG@10	HR@10	NDCG@10
BPR	0.656 9	0.388 0	0.378 5	0.205 3
GRU4Rec	0.558 1	0.376 1	0.361 4	0.198 6
SR-GNN	0.681 8	0.467 2	0.371 9	0.185 9
Caser	0.753 2	0.512 6	0.549 2	0.386 2
SASRec	0.785 3	0.546 8	0.570 2	0.418 5
CCFERec	0.793 8	0.557 8	0.578 7	0.425 7

3.2.2 超参数影响分析

为评估模型中模型参数对推荐结果的影响, 分别对卷积层长度和项目嵌入维度设计对比实验并分析, 测试时保持其它参数不变。图 3 和图 4 展示了卷积核长度对模型性能的影响。根据实验结果, 在 MovieLens-1M 数据集上, 随着卷积核长度的增加, 模型性能逐渐提升, 并在长度为 7 维时达到最佳。在较

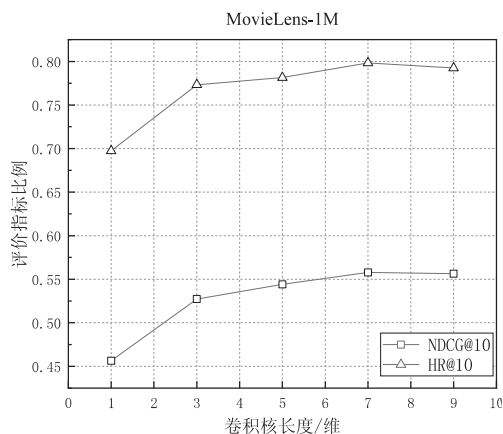


图 3 MovieLens-1M 卷积核长度对模型性能的影响

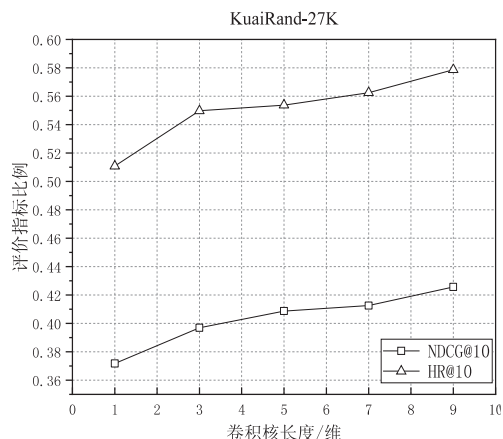


图 4 KuaiRand-27K 卷积核长度对模型性能的影响

为稀疏的 KuaiRand-27K 数据集上,卷积核长度的增加对模型性能的提升效果较为明显,在长度为9维时,模型性能达到最佳。

图5和图6展示了嵌入维度对模型性能的影响。根据实验结果,在 MovieLens-1M 数据集上,随着嵌入维度的增加,模型性能逐渐提升,并在嵌入维度为40时达到最佳。特别是当嵌入维度从5增加到15时,性能提升最为显著。之后,嵌入维度的进一步增加对性能提升的速度有所减缓。在较为稀疏的 KuaiRand-27K 数据集中,嵌入维度的增加对模型性能的提升效果较小,在嵌入维度为10时,模型性能达到最佳。再增加嵌入维度不仅没有显著提升性能,反而可能导致模型性能的下降。

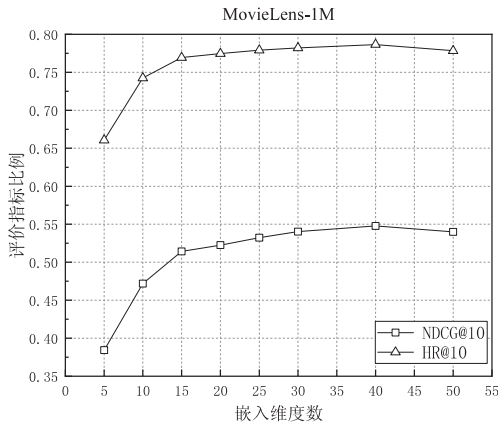


图5 MovieLens-1M 嵌入维度对模型性能的影响

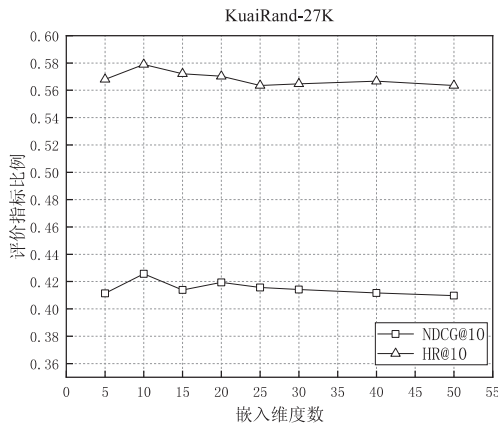


图6 KuaiRand-27K 嵌入维度对模型性能的影响

3.2.3 消融实验

在本小节中,该文通过一系列严谨的实验探讨了模型中若干核心组件的功效,具体涉及基于卷积的信息聚合层(CCFE)和基于前馈网络的非线性处理层(FFN)。在两个数据集上进行了消融实验,相关实验结果详列于表3之中。

首先,该文考察了基于卷积的信息聚合层组件的贡献。移除该组件后,在数据集 MovieLens-1M 上,模型性能表现 HR@10 和 NDCG@10 指标分别降低了 9.38 个百分点和 8.53 百分点。在数据集 KuaiRand-

27K 上,模型性能表现 HR@10 和 NDCG@10 指标分别降低了 8.61 个百分点和 9.10 百分点。实验结果显示,移除信息聚合层会大大损害模型提取用户兴趣的能力。

表3 消融实验性能表现统计信息

模型	MovieLens-1M		KuaiRand-27K	
	HR@10	NDCG@10	HR@10	NDCG@10
CCFERec	0.793 8	0.557 8	0.578 7	0.425 7
without CCFE	0.701 0	0.472 5	0.492 6	0.334 7
without FFN	0.778 1	0.541 2	0.498 1	0.353 2

其次,该文考察了基于前馈网络的非线性处理层的贡献。该组件移除后 HR@10 和 NDCG@10 指标分别降低了 1.57 个百分点和 1.66 百分点。在数据集 KuaiRand-27K 上,HR@10 和 NDCG@10 指标分别降低了 8.06 个百分点和 7.25 百分点。实验结果表明,非线性处理层在通过非线性转化及跨维度的交互增强,能大幅提高模型的泛化能力。

4 结束语

该文提出了一种基于一维卷积的序列推荐模型,旨在解决标准卷积神经网络在序列推荐任务中无法编码时序信息的问题。通过引入一维卷积网络的信息提取模块(CCFE),模型能够有效捕捉用户历史行为中的时序特征,并通过前馈网络(FFN)进行非线性处理,进一步提升了推荐的准确性和效果。在 MovieLens 和 KuaiRand 两个真实推荐场景中的实验结果表明,该模型在推荐命中率和归一化折损累计增益等指标上均表现出显著的提升。

参考文献:

- [1] KO H, LEE S, PARK Y, et al. A survey of recommendation systems: recommendation models, techniques, and application fields[J]. Electronics, 2022, 11(1):141.
- [2] 胡琪, 朱定局, 吴惠舜, 等. 智能推荐系统研究综述[J]. 计算机系统应用, 2022, 31(4):47-58.
- [3] 赵晔辉, 柳林, 王海龙, 等. 知识图谱推荐系统研究综述[J]. 计算机科学与探索, 2023, 17(4):771-791.
- [4] 任豪, 刘柏嵩, 孙金杨, 等. 基于时间和关系感知的图协同过滤跨域序列推荐[J]. 计算机研究与发展, 2023, 60(1):112-124.
- [5] LI J, WANG M, LI J, et al. Text is all you need: Learning language representations for sequential recommendation [C]//Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining. Long Beach: ACM, 2023:1258-1267.
- [6] TANG J, WANG K. Personalized top-n sequential recommendation via convolutional sequence embedding[C]//Pro-

- ceedings of the eleventh ACM international conference on web search and data mining. Marina Del Rey, ACM, 2018; 565–573.
- [7] JIANG J, ZHANG P, LUO Y, et al. AdaMCT: adaptive mixture of CNN – transformer for sequential recommendation [C]//Proceedings of the 32nd ACM international conference on information and knowledge management. Birmingham: ACM, 2023: 976–986.
- [8] 钱忠胜, 赵 畅, 俞情媛, 等. 结合注意力 CNN 与 GNN 的信息融合推荐方法 [J]. 软件学报, 2023, 34(5): 2317–2336.
- [9] YANG J, LI C, DAI X, et al. Focal modulation networks [J]. Advances in Neural Information Processing Systems, 2022, 35: 4203–4217.
- [10] YANG J, LI C, ZHANG P, et al. Focal attention for long-range interactions in vision transformers [J]. Advances in Neural Information Processing Systems, 2021, 34: 30008–30022.
- [11] 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: International World Wide Web Conference Committee, 2010; 811–820.
- [12] RENDLE S, FREUDENTHALER C, GANTNER Z, et al. BPR: Bayesian personalized ranking from implicit feedback [J]. arXiv: 1205. 2618, 2012.
- [13] 刘泽京, 邬 楠, 黄抚群, 等. 基于知识图谱与协同过滤混合策略的在线编程评测系统题目推荐模型 [J]. 计算机科学, 2023, 50(2): 106–114.
- [14] 李伟玥, 朱志国, 董 昊, 等. 循环神经网络和注意力增强的门控图神经网络会话推荐模型 [J]. 模式识别与人工智能, 2024, 37(3): 191–206.
- [15] XU C, FENG J, ZHAO P, et al. Long – and short – term self – attention network for sequential recommendation [J]. Neurocomputing, 2021, 423: 580–589.
- [16] HIDASI B, KARATZOGLOU A, BALTRUNAS L, et al. Session – based recommendations with recurrent neural networks [J]. arXiv: 1511. 06939, 2015.
- [17] WU S, TANG Y, ZHU Y, et al. Session – based recommendation with graph neural networks [C]//Proceedings of the AAAI conference on artificial intelligence. Vancouver: Association for the Advancement of Artificial Intelligence, 2019: 346–353.
- [18] 章 琪, 于双元, 尹鸿峰, 等. 基于图注意力的神经协同过滤社会推荐算法 [J]. 计算机科学, 2023, 50(2): 115–122.
- [19] YANG W, XIA K, WANG Z, et al. Self – attention causal dilated convolutional neural network for multivariate time series classification and its application [J]. Engineering Applications of Artificial Intelligence, 2023, 122: 106151.
- [20] 尹梓诺, 马海龙, 胡 涛. 基于联合注意力机制和一维卷积神经网络 – 双向长短期记忆网络模型的流量异常检测方法 [J]. 电子与信息学报, 2023, 45(10): 3719–3728.
- [21] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [J]. Advances in Neural Information Processing Systems, 2017, 30: 5998–6008.
- [22] KANG W C, MCAULEY J. Self – attentive sequential recommendation [C]//2018 IEEE international conference on data mining (ICDM). Singapore: IEEE, 2018: 197–206.
- [23] 孟志鹏, 成卫青. 基于自注意力网络的时间感知序列化推荐 [J]. 计算机系统应用, 2023, 32(1): 197–205.
- [24] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas: IEEE, 2016: 770–778.
- [25] GAO C, LI S, ZHANG Y, et al. Kuairand: an unbiased sequential recommendation dataset with randomly exposed videos [C]//Proceedings of the 31st ACM international conference on information & knowledge management. Atlanta: ACM, 2022: 3953–3957.