

基于CNN和注意力机制的深度学习建模和个性化推荐系统

张贤斌*, 王 晰, 李天星, 杨茗惠, 李星慧

长春理工大学数学与统计学院, 吉林 长春

收稿日期: 2025年4月28日; 录用日期: 2025年5月21日; 发布日期: 2025年5月30日

摘 要

随着信息时代的到来, 互联网成为学习和共享的重要平台, 但庞大的电影数据量和信息复杂性使用户查询变得困难。为帮助用户快速找到感兴趣的电影, 我们提出了一个结合深度学习和注意力机制的个性化电影推荐系统。通过分析用户基本信息和评分数据, 该系统利用特征嵌入、卷积、自注意力机制、LSTM层以及残差模块, 训练出个性化推荐模型, 以提升推荐效果和准确性。在特征处理方面, 用户特征通过嵌入层将年龄、性别(one-hot编码)及职业映射为稠密向量; 电影特征借助multi-hot编码与卷积层捕捉类型组合的特征共现模式, 再结合电影ID嵌入向量保留独特性。特征交互模块通过卷积提取用户与电影的偏好关联, 自注意力机制动态调整特征权重以适应兴趣变化, LSTM层捕捉观影行为的长期依赖与短期变化, 残差模块则有效融合兴趣信息并避免梯度消失, 确保模型的稳定性与高效训练。

关键词

卷积神经网络, 注意力机制, LSTM, 残差网络, 冷启动

Deep Learning Modeling and Personalized Recommendation System Based on CNN and Attention Mechanism

Xianbin Zhang*, Xi Wang, Tianxing Li, Minghui Yang, Xinghui Li

School of Mathematics and Statistics, Changchun University of Science and Technology, Changchun Jilin

Received: Apr. 28th, 2025; accepted: May 21st, 2025; published: May 30th, 2025

Abstract

With the advent of the information age, the Internet has become an important platform for learning

*通讯作者。

文章引用: 张贤斌, 王晰, 李天星, 杨茗惠, 李星慧. 基于 CNN 和注意力机制的深度学习建模和个性化推荐系统[J]. 应用数学进展, 2025, 14(5): 568-579. DOI: 10.12677/aam.2025.145283

and sharing. However, the vast amount of movie data and the complexity of information make it difficult for users to search. To help users quickly find movies they are interested in, we propose a personalized movie recommendation system that combines deep learning and attention mechanisms. By analyzing users' basic information and rating data, the system uses feature embedding, convolution, self-attention mechanisms, LSTM layers, and residual modules to train a personalized recommendation model to improve the recommendation effect and accuracy. In terms of feature processing, user features are mapped to dense vectors through an embedding layer, including age, gender (one-hot encoded), and occupation; movie features capture the co-occurrence patterns of genre combinations through multi-hot encoding and convolutional layers and combine movie ID embedding vectors to retain uniqueness. The feature interaction module extracts the preference correlation between users and movies through convolution, the self-attention mechanism dynamically adjusts feature weights to adapt to changes in interest, the LSTM layer captures the long-term dependencies and short-term changes in viewing behavior, and the residual module effectively fuses interest information and avoids gradient vanishing, ensuring model stability and efficient training.

Keywords

Convolutional Neural Network, Attention Mechanism, Long Short-Term Memory (LSTM), Residual Network, Cold Start

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 绪论

随着信息技术的快速发展,尤其是互联网和大数据的普及,用户面临的信息量急剧增加,如何从海量信息中提取有价值的内容并满足用户的个性化需求,成为了当前研究的核心课题。在这一背景下,深度学习技术,尤其是卷积神经网络(CNN)和注意力机制,逐渐成为推荐系统中的核心技术。这些技术能够有效捕捉用户与物品之间复杂的关系,从而提升推荐准确性和系统的鲁棒性。基于CNN和注意力机制的深度学习模型,不仅能够处理高维数据,自动提取特征,减少人工干预,还能深度挖掘用户兴趣,提供更加个性化的推荐服务,提升用户满意度与平台的商业价值。因此,这一研究不仅推动了学术进展,还为电商、社交媒体、在线教育、音乐和视频推荐等多个领域的应用提供了新的思路和方法,具有广泛的应用前景和巨大的市场潜力。

2. 基于 CNN 及注意力机制的电影推荐算法模型

本章节的研究重点是基于 CNN 及注意力机制的电影推荐算法。我们将首先介绍用户和电影的相关特征,并分析如何有效地将这些特征转化为适合模型处理的形式。接着,我们将讨论如何利用卷积神经网络(CNN)提取特征中的局部信息,结合自注意力机制挖掘全局依赖关系,从而提升推荐系统的性能[1]。

2.1. 特征处理模块

在本研究中,我们提出了一种基于卷积神经网络(CNN)和注意力机制的电影推荐模型,旨在通过多模态特征融合和深度优化,显著提升电影推荐的准确性和个性化程度。模型的整体框架由用户特征嵌入、电影特征嵌入、卷积模块、自注意力机制、LSTM 层、残差模块以及评分预测等关键模块构成。下面介绍两大特征处理的部分,如图 1 所示。

其一，对于用户特征嵌入部分，我们对年龄、性别和职业等用户特征进行了精细化处理。年龄通过离散化分桶的方式被划分为多个区间，例如将 10~19 岁归为“青少年”类别，随后通过嵌入层映射为稠密向量。性别采用 one-hot 编码，分别用[1, 0]和[0, 1]表示男性和女性，同样通过嵌入层转换为稠密向量。职业作为多分类特征，直接通过嵌入层映射为稠密向量。这一系列处理将离散特征和连续特征统一映射为相同维度的稠密向量，有效解决了特征异构性问题，为后续的特征交互提供了坚实基础。

其二，对于电影特征嵌入部分，我们对电影的类型和 ID 进行了处理。电影类型采用 multi-hot 编码，例如“科幻|动作”会被编码为一个向量，然后通过嵌入层和卷积层处理，以捕捉类型组合模式，如“科幻 + 冒险”的联合特征。卷积核在滑动窗口内捕获特征共现模式，每个卷积核的输出对应一个潜在语义空间，例如“青少年偏好类型组合”，通过梯度加权类激活映射(Grad-CAM)可以可视化关键特征区域。同时，每部电影的唯一 ID 通过嵌入层映射为稠密向量，保留了电影个体的独特性。

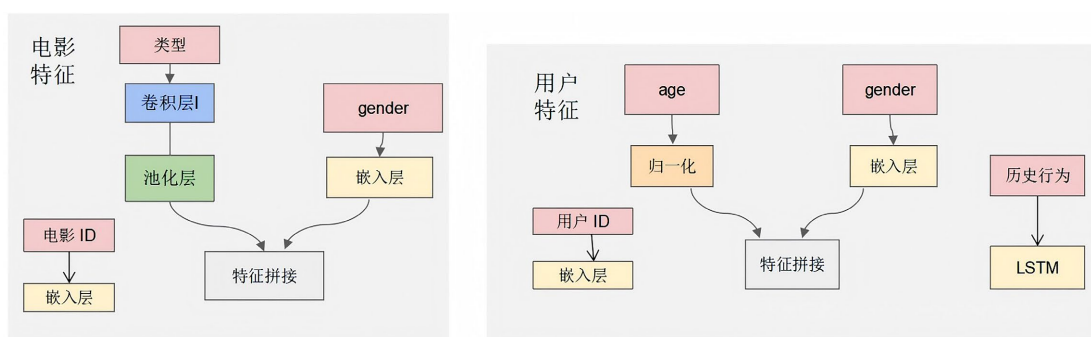


Figure 1. Feature processing structure diagram

图 1. 特征处理结构图

综上，我们将模型的详细数学描述和总体设计展示如下：

$$r_{ui} = \sigma \left(W_0 \cdot \left[\text{ResBlock} \left(\text{LSTM} \left(\text{Attention} \left(\text{Conv} (E_u \oplus E_i) \right) \right) \right) \right] \right) \quad (1)$$

① 用户特征嵌入：

$$E_i = \text{Embedding}(\text{年龄, 性别, 历史评分均值}) \quad (2)$$

年龄离散化分桶(如 10~19 岁→青少年)、性别(one-hot 编码)、职业(嵌入层)将离散特征(如职业编码)和连续特征(如年龄)映射为统一维度的稠密向量，解决特征异构性问题。

② 电影特征嵌入：

$$E_i = \text{Conv}(\text{类型, 年份}) \oplus \text{Embedding}(\text{电影ID}) \oplus \text{Norm}(\text{平均评分}) \quad (3)$$

卷积层处理多类型标签(如“科幻|动作”)，捕捉类型组合模式，ID 嵌入保留个体独特性。

符号关联：

$\oplus$ ：用户与电影特征的拼接，构建交互空间。

σ ：Sigmoid 函数将输出映射为 0~5 的评分值。

在本研究提出的电影推荐模型中，我们通过一系列关键设计实现了对用户和电影特征的深度挖掘与融合，从而显著提升了推荐系统的性能。首先，我们采用了多模态输入的设计，将用户特征(如年龄、性别、职业等)和电影特征(如类型、ID 等)分别通过嵌入层映射为稠密向量，解决了特征异构性问题，为后续的特征交互提供了统一的基础。

为了捕捉用户与电影之间的复杂关联，我们引入了特征交叉模块。通过卷积模块，模型能够提取局

部交互模式，例如“青少年用户”与“科幻电影”之间的潜在关联。这种局部模式的捕捉使得推荐系统能够更精准地理解用户对不同类型电影的偏好。在处理用户观影行为的时序特征方面，我们采用了时序建模。LSTM 层能够有效处理用户观影序列的时序依赖关系，例如连续观看同一导演的电影。通过这种方式，模型不仅能够捕捉用户的长期偏好，还能及时发现用户的兴趣漂移，从而为用户提供更贴合其动态兴趣的推荐[2]。

2.2. 模型的架构与实现

在深入解析该模型的架构设计时，我们绘制了框架流程图，以直观呈现模型的运行机制与各模块之间的协同关系，如图 2 所示。

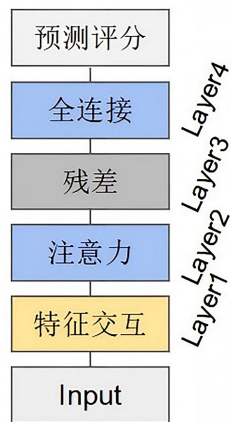


Figure 2. Movie recommendation system framework flowchart
图 2. 电影推荐系统框架流程图

第一步对局部特征进行提取。基于 CNN 的细粒度特征捕捉在局部特征提取环节，以卷积神经网络 (CNN) 作为核心技术，专门针对电影类型等具有空间局部关联的特征展开处理。电影类型特征并非孤立存在，而是存在如“动作 + 冒险”“剧情 + 爱情”等组合模式，CNN 通过卷积核在特征空间上的滑动操作，能够精准捕捉这类局部组合模式。例如，针对电影类型特征矩阵，卷积核可提取不同类型间的潜在关联特征，挖掘出隐藏的类型搭配偏好。经过卷积操作后，再通过池化层降维，在保留关键局部信息的同时压缩特征维度，最终形成细粒度的局部特征表达，为后续模型分析提供坚实的基础特征输入。

第二步进行全局关系挖掘，注意力机制驱动的特征关联分析。在全局关系挖掘层面，模型引入注意力机制 (Layer1) [3]，通过动态计算不同特征间的关联权重，打破特征独立处理的局限。例如，在用户历史行为与电影特征的交互过程中，注意力机制[4]可识别对用户决策影响更大的特征——若用户近期频繁观看科幻类电影，模型会自动提升科幻电影相关特征(如科幻类型标签、科幻电影导演等)的权重。在此基础上，叠加自注意力模块增强全局特征挖掘能力，该模块能基于用户实时行为数据，持续调整特征权重分配，使模型更灵活地适应用户兴趣变化，深度挖掘跨特征的全局依赖关系，从而捕捉用户潜在的综合偏好。

第三步，LSTM 对用户兴趣动态的捕捉。针对用户观影行为的时序特性，模型采用 LSTM 层进行时序建模[5]。用户观影序列蕴含着丰富的时序依赖关系，如连续观看同一导演的多部作品、短期内集中观看某一题材电影等。LSTM 凭借其独特的门控机制，能够有效记忆长序列中的关键信息，不仅可以捕捉用户长期稳定的观影偏好(如长期喜爱悬疑题材)，还能及时发现兴趣漂移(如突然对纪录片产生关注)。例如，当用户连续观看某位导演的电影时，LSTM 会强化对该导演相关电影特征的学习，将时序依赖转化为特征表达，为动态兴趣推荐提供核心依据，使推荐结果更贴合用户实时兴趣。

第四步,在此基础上进行深度特征融合与优化。残差连接与全连接层的协同作用在深度特征融合与优化阶段,模型通过残差连接(Layer2)和全连接层(Layer3~Layer4)实现特征的高效整合与输出。残差连接[6]作为深层网络训练的“稳定器”,通过跳跃连接融合用户长期兴趣(如基于基础画像的历史评分模式)与短期兴趣(LSTM 提取的近期行为特征),避免因网络层数增加导致的梯度消失或爆炸问题,确保底层特征信息不丢失。经过残差连接优化后,全连接层作为最终融合环节,对处理后的特征进行加权整合,通过多层非线性变换,捕捉特征间复杂的高阶交互关系——例如,将用户年龄、历史行为、电影类型等特征深度融合,最终输出精准的预测评分,使模型能够更准确地捕捉用户偏好,提升电影推荐的精度与有效性。

2.3. 模型概况

在电影推荐系统中,为了增强推荐结果的强相关性,我们从特征工程、评估指标、消融实验、注意力可视化以及冷启动处理机制等多个方面进行了针对性的设计与优化,具体模块如表 1 所示。

在特征工程方面,电影推荐系统通过采用 1D 卷积对 multi-hot 类型编码进行处理,以显式建模不同类型组合的协同效应,精准捕捉用户对电影类型的偏好[7]。同时,引入时间上下文嵌入,将电影年份按照年代进行分段,以捕捉用户对经典电影与新片的偏好差异。在评估指标上,使用 RMSE 和 MAE 衡量评分预测的准确性,HR@K 和 NDCG@K 用于 Top-K 推荐任务的评估。此外,通过消融实验验证各模块的重要性,进一步分析注意力机制、LSTM 等模块对推荐系统性能的贡献。为了应对深度网络中的梯度消失问题,系统引入了残差连接,这有助于避免信息在深层网络中逐渐消失,确保梯度有效传播,从而提高模型的训练效率和稳定性。在电影推荐的可解释性方面,采用电影属性注意力可视化机制[8],展示推荐原因,增强系统透明度。对于冷启动问题,采用针对新用户和新电影的不同策略,结合人口统计特征和类型、年份嵌入,缓解冷启动问题并提高推荐质量。

Table 1. Module function analysis table
表 1. 模块功能分析表

步骤	数据处理	数学操作	电影推荐关联性
输入	用户 ID、年龄、性别、历史行为 电影 ID、类型、年份	X_{user}, X_{movie}	原始特征对齐 MovieLens 数据字段
嵌入	用户年龄分桶嵌入 电影类型卷积嵌入	$E_i = \phi(X_{user}), E_u = \phi(X_{movie})$	类型卷积适配 multi-hot 编码特性
交叉	拼接用户与电影特征	$E_{iu} = E_i \oplus E_u$	构建联合交互空间
注意力	计算用户 - 电影匹配度	$Z = A(E_{iu})$	动态突出关键特征交互
时序模型	处理用户观影序列	$H_t = \Gamma(Z)$	捕捉兴趣演化轨迹
残差融合	长期 + 短期兴趣	$H_{final} = R(H_t)$	防止深层网络丢失基础特征

通过以上多方面的综合优化与设计,我们的电影推荐系统能够显著提升推荐结果的相关性和个性化程度,增强用户体验,同时在面对冷启动问题时也能提供有效的解决方案,确保推荐系统的稳定性和可靠性。

3. 推荐模型的实验与分析

3.1. 数据处理与特征处理

3.1.1. 数据集

在本章中,为了更好地检验该电影推荐系统的性能以及各指标水准,我们选用了较为知名的 MovieLens 网站电影评分数据作为本次实验的数据集,其中包括 100 多万条电影评分数据以及不同用户

的基本信息数据。该数据集中，一共有三个.dat 后缀文件(movies、ratings 和 users)，在 movies 的表中包含电影名称、ID 及所属类别，ratings 文件中包含用户 ID 以及分别对应的评分数据，users 文件则是所有已统计用户的基本信息。以下是该数据集中各个文件的随机五条数据，具体情况如表 2 所示。

Table 2. Movie information table

表 2. 电影信息表

电影 ID	电影名称	电影类别
3	Grumpier Old Men (1995)	Comedy Romance
8	Tom and Huck (1995)	Adventure Children's
9	Sudden Death (1995)	Action
10	GoldenEye (1995)	Action Adventure Thriller
15	Cutthroat Island (1995)	Action Adventure Romance

Table 3. Scoring information table

表 3. 评分信息表

用户 ID	电影 ID	评分	时间戳
1	608	4	978301398
1	1246	4	978302091
2	1357	5	978298709
2	3068	4	978299000
5	2987	4	978243170

如表 3 所示，在评分信息表中有一列为时间戳，它是用于在数据中记录事件发生的时间点，而在该系统中，我们在电影数据集中引入时间戳有助于评估模型的实时预测能力，验证其稳定性，应用于时间序列预测，同时支持数据分析与可视化，从而全面提升推荐系统的性能和用户体验。

Table 4. User information table

表 4. 用户信息表

用户 ID	性别	年龄	职业	邮编
12	M	25	12	32793
13	M	45	1	93304
14	M	35	0	60126
15	M	25	7	22903
16	F	35	0	20670

用户信息如表 4 所示，其中“F”和“M”分别代表女性和男性，并且为了后续数据分析和模型检验，我们将年龄以及职业分别用不同的数据来替代，对此要进行必要说明，具体情况见表 5。

Table 5. Comparison table for age information

表 5. 年龄信息对照表

替代值	年龄区间
1	18 岁以下
18	18~24 岁

续表

25	25~34 岁
35	35~44 岁
45	45~49 岁
50	50~55 岁
56	56 岁以上

由于该模型暂无进行对特定职业的敏感特征抓取的功能并鉴于对后续模型检验并无影响，在这里对不同编号所对应的职业不再详细描述。

3.1.2. 数据预处理

经过上述对数据集的详细介绍以及数据的选用，我们从 MovieLens 数据集中筛选了有效数据，将 movies、ratings 和 users 三张表进行合并，分别通过将电影编号和用户编号作为合并的索引将后续实验验证所需数据整合在一张大表中，无效信息则选择删除[9]，合并并且经过处理后的数据集如表 6 所示。

Table 6. Merged dataset

表 6. 合并后的数据集

用户 ID	电影 ID	评分	时间戳	电影标题	电影类别	性别	年龄	职业	邮编
2015	986	1	975035760	Fly Away Home (1996)	Adventure Children's	M	18	4	1003
4077	539	4	965444857	Sleepless in Seattle (1993)	Comedy Romance	F	35	2	22030
2237	3514	3	974596081	Joe Gould's Secret (2000)	Drama	F	35	6	14905
3391	2093	3	1046190153	Return to Oz (1985)	Adventure Children's Fantasy Sci-Fi	M	18	4	48135
822	2987	3	981481471	Who Framed Roger Rabbit (1988)	Adventure Animation Film-Noir	F	25	1	7666

在上表中，由于一部电影可能存在多个类别标签，为了能够将其划分清楚，我们通过在中间插入“|”来进行分割。

3.1.3. 特征提取

数据预处理的过程中，我们筛选了部分有效的数据，其中大部分均为提取到的特征，例如用户 ID、电影 ID、用户信息等，保留了这些电影及用户的基本属性信息，具体所提取到的特征信息总览如下表 7 所示。

Table 7. Feature selection

表 7. 特征选取

特征类别	特征选取	特征编码
电影特征	电影 ID	MovieID
	电影类别	Genres
用户特征	用户 ID	UserID
	性别	Gender
	年龄	Age
	职业	Occupation

以上提取到的特征在模型代码中均有体现，具体的特征处理过程和步骤已在前文详述，在此不再过多赘述，直接进行模型验证实验。

3.2. 实验设计

3.2.1. 参数选取

现在进行实验参数设置[10]，在实验分析时我们会采用超参数初步选取的方式进行，从而得到最优值，具体参数设置如表 8 所示。

Table 8. Experimental parameter setting
表 8. 实验参数设置

参数名称	参数值
嵌入层(output_dim)	64
卷积层数	64
LSTM 层	64
残差块(dropout_rate)	0.2
学习率(learning_rate)	0.0001
批量大小(batch_size)	64
训练参数(epochs)	100
训练参数(patience)	10

3.2.2. 实验过程描述

后续模型训练过程中，使用预设参数进行训练，并通过自动调整超参数、监控训练集和验证集误差变化来优化模型。当验证集损失在 10 个 epoch 内未改善时，训练将提前终止。训练完成后，使用测试集评估模型性能，计算 MSE、RMSE 和 MAE。通过参数敏感性分析，调整嵌入层的 output_dim、卷积层的 filters 和 kernel_size、LSTM 的 units 等参数，观察其对模型性能的影响。此外，通过分析残差分布、注意力机制输出、LSTM 层激活值和梯度，进一步优化模型结构和参数设置。最终，通过检验测试集上的 MSE、RMSE 和 MAE 值，评估不同参数设置对模型性能的影响，为模型优化提供参考。

3.3. 实验结果及分析

3.3.1. 参数选取实验

为了优化模型性能，本节对模型中多个关键超参数进行了系统性实验分析。通过调整这些超参数，我们旨在找到最佳配置，以提高模型在评分预测任务上的表现。以下是各超参数的实验设置、结果分析以及对实验结果的解释。下面实验中我们不断地更新参数值大小来进行多次实验检测，直到找到超参数最合适的值，具体参数值变更如表 9 所示。

通过一系列详细的实验，我们探讨了不同超参数设置对模型性能的影响，包括嵌入层的大小、卷积层数、LSTM 层大小、残差块的 dropout_rate、学习率、批量大小以及训练参数 patience。每个实验都旨在平衡模型的复杂度、特征提取能力与过拟合风险，从而找到最佳的预测模型。例如，调整嵌入层的 output_dim 可以影响特征提取的能力，同时控制过拟合；改变卷积层数则影响模型对局部特征的捕捉能力；而 LSTM 层的设置决定了模型对时间序列特征的处理能力。此外，残差块的 dropout_rate、学习率、批量大小及训练参数 patience 都直接影响模型的训练效率、收敛速度和稳定性。以下是经过多次实验后得到的最佳检验数据，如图 3 所示。

Table 9. Hyperparameter setting value
表 9. 超参数设置值

参数名称	参数值
嵌入层(output_dim)	64/128/192
卷积层数	1/2/3
LSTM 层	32/64/128
残差块(dropout_rate)	0.1/0.2/0.3
学习率(learning_rate)	0.0001/0.001/0.01
批量大小(batch_size)	32/64/128
训练参数(patience)	5/10/20

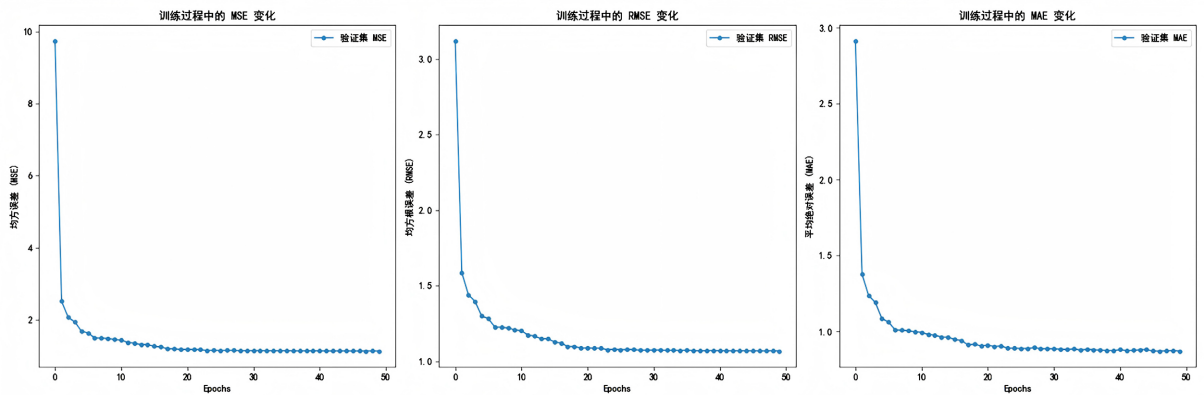


Figure 3. Experimentally verified numerical change plot
图 3. 实验验证数值变化图

从图表中可以看出，随着训练的进行，模型的 MSE、RMSE 和 MAE 均迅速下降，并在后期趋于平稳，表明模型的预测误差逐步减小，训练效果逐渐提高，最终收敛到一个稳定的状态，表现出良好的训练效果。结合终端的实际计算数据，我们可以判断该预测模型的良好性能。

3.3.2. 实验结果分析

在本轮超参数实验中，我们逐步选取了表现较为优异的一组数据并得到以下结论：

- 1) 嵌入层(output_dim): 192 的嵌入维度能够更好地捕捉用户特征，但模型复杂度较高，容易过拟合。64 的嵌入维度模型简单，但特征提取能力较弱，容易欠拟合。128 的嵌入维度在两者之间，可能提供更好的平衡。
- 2) 卷积层数: 128 的卷积层数能够更好地捕捉局部特征，但模型复杂度较高，容易过拟合。32 的卷积层数模型简单，但特征提取能力较弱，容易欠拟合。64 的卷积层数在两者之间，可能提供更好的平衡。
- 3) LSTM 层: 128 的 LSTM 层能够更好地捕捉序列特征，但模型复杂度较高，容易过拟合。32 的 LSTM 层模型简单，但特征提取能力较弱，容易欠拟合。64 的 LSTM 层在两者之间，可能提供更好的平衡。
- 4) 残差块(dropout_rate): 0.3 的残差块能够更好地捕捉残差表达能力，但模型复杂度较高，容易过拟合。0.1 的残差块模型简单，但特征提取能力较弱，容易欠拟合。0.2 的残差块在两者之间，可能提供更好的平衡。
- 5) 学习率: 0.01 的学习率能够更好地收敛，但模型复杂度较高，容易过拟合。0.0001 的学习率模型

简单,但收敛速度较慢,容易欠拟合。0.001 的学习率在两者之间,可能提供更好的平衡。

6) 批量大小: 128 的批量大小能够更好地捕捉数据的多样性,但训练时间较长,容易过拟合。32 的批量大小模型简单,但训练效率较高,容易欠拟合。64 的批量大小在两者之间,可能提供更好的平衡。

7) 训练参数(patience): 20 的 patience 值能够更好地平衡模型的收敛速度和过拟合风险,但训练时间较长,容易过拟合。5 的 patience 值模型简单,但训练效率较高,容易欠拟合。10 的 patience 值在两者之间,可能提供更好的平衡。

3.3.3. 对比实验分析

接下来,我们对不同的推荐模型进行比较,重点考察了引入残差连接对模型性能的影响。我们选择了几种经典的推荐算法,如 K 最近邻(KNN)和奇异值分解(SVD),以及一种基于深度学习的推荐模型。对于深度学习模型,我们分别测试了带有和不带有残差连接的版本,以观察残差连接是否能够有效提升模型的性能。

首先基于梯度消失可能会出现的问题,我们穿插增加了残差网络,为此我们也对增加残差连接后的模型与原模型进行比较,未增加残差时的模型详情如图 4 所示。

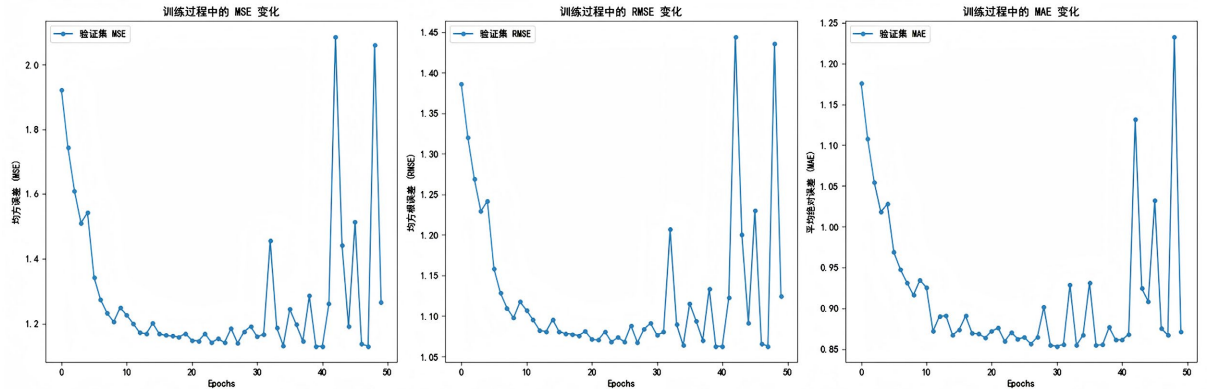


Figure 4. Evaluation diagram of the model without residuals
图 4. 无残差模型评估图

从本次结果与之前结果对比可以明显看出,残差连接在提高模型性能方面发挥了重要作用。深度神经网络通常容易遇到梯度消失或梯度爆炸的问题,尤其是在层数较多时。残差连接通过将输入信号直接传递到后续层,允许梯度在反向传播时更容易地流动,从而减轻了梯度消失的影响。这种结构的优势在深度学习模型中得到了体现,特别是在训练较深的网络时,残差连接显著改善了模型的收敛速度和最终表现。

为了验证我们的推荐模型的优异性能,我们将模型与其他的经典模型性能进行了对比分析,这些建立的模型同样使用了相同的训练集和验证集,评估指标为 Top-K 准确率(HR@K)和标准化折损累积增益(NDCG@K) [11]。这些指标用于衡量推荐系统的准确性和相关性,能够有效反映推荐系统的实际效果,具体评估指标分析图如图 5 所示。

在 HR@K 指标上,深度学习模型(特别是带有残差连接的版本)表现出明显的优势。相比之下,传统的 KNN 和 SVD 模型在这一指标上的表现较为逊色。具体而言, KNN 和 SVD 的 HR@K 值分别为 0.72 和 0.75,而深度学习模型(Plain)的 HR@K 为 0.82,带残差的深度学习模型(Residual)则达到了 0.85。这表明,深度学习模型能够更好地捕捉用户与物品之间复杂的非线性关系,尤其是在大规模数据集上,残差连接有助于避免梯度消失问题,进一步提升了模型的推荐精度。

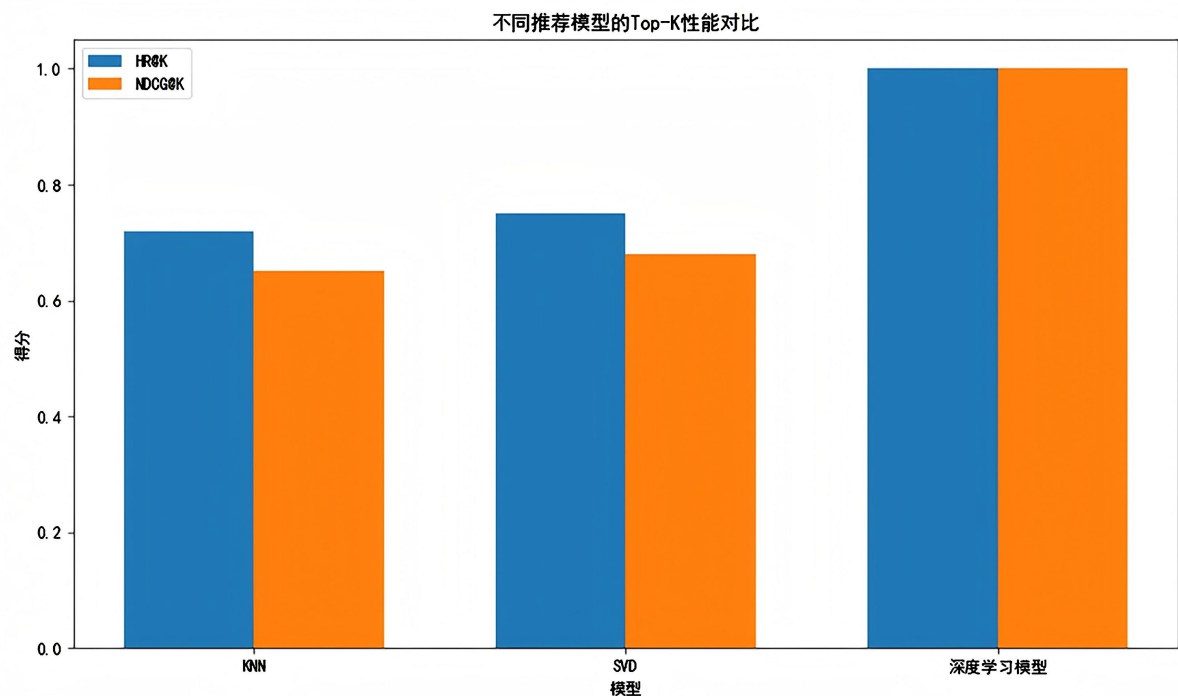


Figure 5. Comparison diagram of the performance of different models
图 5. 不同模型的性能对比图

在 NDCG@K 指标上，我们观察到类似的趋势。KNN 和 SVD 的 NDCG@K 分别为 0.65 和 0.68，而深度学习模型(Plain)达到了 0.75，带残差的深度学习模型(Residual)进一步提升至 0.78。NDCG@K 指标考虑了推荐列表中项目的顺序，因此它更加侧重于排名质量。残差连接的引入使得深度学习模型不仅能够提高准确率，还能够在排序上获得更好的效果，进一步提升了推荐的相关性。

4. 研究总结

综上，本研究提出的基于 CNN 和注意力机制的个性化电影推荐模型在提升推荐效果方面取得了较好的效果，但随着技术的发展，仍然存在进一步优化和拓展的空间。未来的研究将继续探索深度学习技术在推荐系统中的应用，尤其是在多模态数据融合、冷启动问题解决和跨域推荐等方面，以期为用户提供更加精准、个性化和实时的推荐体验。

基金项目

长春理工大学大学生创新创业训练计划项目(项目编号：S202410186248)。

参考文献

[1] 吴长旺. 融合注意力机制的图神经网络推荐算法研究[D]: [硕士学位论文]. 南京: 南京邮电大学, 2022.

[2] 浩庆波. 基于图神经网络的推荐模型研究[D]: [博士学位论文]. 天津: 天津理工大学, 2024.

[3] Zhang, P., Li, W., Ouyang, P., Xu, J., Liu, F., Ren, X., Wang, W. and Huang, J. (2017) Social LSTM: Human Trajectory Prediction in Crowded Spaces. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 146-155.

[4] 陈海涵, 吴国栋, 李景霞, 等. 基于注意力机制的深度学习推荐研究进展[J]. 计算机工程与科学, 2021, 43(2): 370-380.

[5] 于蒙, 何文涛, 周绪川, 等. 推荐系统综述[J]. 计算机应用, 2022, 42(6): 1898-1913.

-
- [6] Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., *et al.* (2017) Residual Attention Network for Image Classification. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 6450-6458. <https://doi.org/10.1109/cvpr.2017.683>
 - [7] 黄海新, 杜亭亭. 基于图神经网络的推荐算法综述[J]. 通信与信息技术, 2023(3): 65-69.
 - [8] 朱昆, 刘姜, 倪枫, 等. 融合上下文和视觉信息的多模态电影推荐模型[J]. 软件工程, 2024, 27(6): 68-73.
 - [9] 王小香. 基于用户兴趣建模的个性化新闻推荐方法及应用[D]: [硕士学位论文]. 济南: 山东师范大学, 2024.
 - [10] 曾彦程. 基于注意力机制的推荐算法研究[D]: [硕士学位论文]. 成都: 电子科技大学, 2021.
 - [11] 谢刚. 基于注意力机制与用户兴趣的序列推荐模型研究与实现[D]: [硕士学位论文]. 桂林: 桂林理工大学, 2022.