

DOI:10.16652/j.issn.1004-373x.2025.01.013

引用格式:郭东坡,何彬,张明焱,等.基于注意力循环神经网络的联合深度推荐模型[J].现代电子技术,2025,48(1):80-84.

基于注意力循环神经网络的联合深度推荐模型

郭东坡¹, 何彬², 张明焱³, 段超³

(1. 江汉大学, 湖北 武汉 430056; 2. 华中师范大学, 湖北 武汉 430079;
3. 浙江师范大学 浙江省智能教育技术与应用重点实验室, 浙江 金华 321004)

摘要: 为了向用户推荐符合兴趣偏好的项目,设计一种基于注意力循环神经网络的联合深度推荐模型。将双层注意力机制设置于网络中,该模型由五个部分构成,在输入层中生成联合深度推荐模型的输入矩阵,通过序列编码层对项目评论文本语义展开正向和反向编码,获得隐藏状态输出,并将其输入双层注意力机制中,提取项目特征,利用全连接层提取用户偏好特征。在预测层中建立项目与用户的交互模型,获得项目评分,为用户推荐高评分的项目。为了提高模型精度,加权融合MSE损失函数、CE损失函数和RK损失函数建立组合损失函数,对深度联合训练模型展开训练,提高模型的推荐性能。仿真结果表明,所提方法具有良好的推荐效果,能够适应不断变化的市场需求和用户行为。

关键词: 双层注意力机制; 循环神经网络; 用户偏好; 组合损失函数; 交互模型; 联合深度推荐模型

中图分类号: TN711-34; TP183

文献标识码: A

文章编号: 1004-373X(2025)01-0080-05

Joint deep recommendation model based on attention recurrent neural network

GUO Dongpo¹, HE Bin², ZHANG Mingyan³, DUAN Chao³

(1. Jiangnan University, Wuhan 430056, China; 2. Central China Normal University, Wuhan 430079, China;

3. Zhejiang Provincial Key Laboratory of Intelligent Education Technology and Application, Zhejiang Normal University, Jinhua 321004, China)

Abstract: A joint deep recommendation model based on attention recurrent neural network is designed to recommend projects that meet user interests and preferences. The attention mechanism with double layers is set in the network. The designed model consists of five parts. The input matrix of the joint deep recommendation model is generated in the input layer. By the sequence coding layer, the semantics of the project comment text is encoded forward and backward to obtain the hidden state output. The hidden state output is input into the attention mechanism with double layers to extract the project features. The fully-connected layer is used to extract user preference features. An interaction model between projects and users is established in the prediction layer, so as to obtain project ratings and recommend high-rated projects for users. In order to improve the accuracy of the model, the combined loss function is established based on the weighted integration of MSE loss function, CE loss function and RK loss function. The deep joint training model is trained to improve the recommendation performance of the model. The simulation results show that the proposed method has good recommendation effect, so it can adapt to the changing market demand and user behavior.

Keywords: attention mechanism with double layers; recurrent neural network; user preference; combined loss function; interaction model; joint deep recommendation model

0 引言

由于信息过载问题逐渐严重,专家和学者加大了对推荐系统的研究力度,推荐系统的主要作用是分析用户的偏好特征,根据分析结果为用户推荐相关项目。推荐

系统通过采集与分析用户的偏好与交互数据对项目评分,根据评分结果生成推荐列表,实现项目推荐,目前推荐系统已经应用在很多网络服务平台中,实现了个性化服务^[1],在此背景下研究深度联合推荐模型具有重要意义。

文献[2]方法通过单层神经网络融合节点特征,并通

收稿日期:2024-02-04

修回日期:2024-02-26

基金项目: 2021年国家自然科学基金:面向智能化作业辅导的图文混合题可理解机器解答研究(62177024);2023年国家自然科学基金:面向知识整合的碎片化化学学习及学习路径引导研究(62207027);2022年度湖北省教育科学规划课题:基于智慧学习环境的学习者模型构建及个性化学习路径推荐研究(2022GB217);湖北省教育厅科学技术研究计划指导性项目:面向数学课程的自适应学习路径推荐研究(B2022560)

过影响因子调整实体聚合权重,在用户与实体特征上展开评分预测和推荐列表生成,但聚合权重调整结果与节点特征不符,导致推荐结果的 MAE 较高。文献[3]方法通过设置时间权重因子来衡量数据的重要性,并根据隐私差分建立概率矩阵分解模型实现推荐。然而,该方法无法动态调整时间权重因子,导致项目评分的精度降低,且生成的推荐列表的归一化折现累积增益较低。文献[4]方法通过贝叶斯个性化排序算法实现个性化推荐,通过对输入数据进行预处理,并调整模型超参数来确定最佳配置。在数据预处理过程中,容易受到冗余数据的影响,导致推荐结果的 RMSE 较高。文献[5]方法利用元路径采集用户偏好,计算推荐分数,并通过个性化权重学习确定最终推荐结果。但是,该方法的目标函数求解精度较低,导致计算得到的全局推荐分数误差大,降低了用户对推荐项目的点击率。

因此,本文提出基于注意力循环网络的联合深度推荐模型。

1 联合深度推荐模型

通过 Google News 中超过一十亿个单次训练的词向量作为联合深度推荐模型的输入,用词向量矩阵 $B_{u1}, B_{u2}, \dots, B_{uj}$ 表示用户评论, j 表示单个用户的评论数最大值, B_{uy} 表示用户 u 的第 y 条评论中的词构成的矩阵, $1 \leq y \leq j$ 。用单词向量矩阵 $B_{i1}, B_{i2}, \dots, B_{ik}$ 表示项目 i 的全部评论, k 表示项目对应的评论数最大值; B_{iz} 为项目 i 的第 z 条评论中的词构成的矩阵, $1 \leq z \leq k$, 将上述矩阵输入联合深度推荐模型中。

采用 LSTM 网络^[6-7]在序列编码层中对项目评论文本语义展开进行编码处理:

$$\begin{cases} \vec{h}_{LM}^i = \vec{B}_{iLSTM}(\mathbf{E}_{LM}, \vec{h}_{LM-1}^i) \\ \tilde{h}_{LM}^i = \tilde{B}_{iLSTM}(\mathbf{E}_{LM}, \tilde{h}_{LM-1}^i) \end{cases} \quad (1)$$

式中: \vec{h}_{LM}^i 为该层生成的正向隐藏状态; \vec{B}_{iLSTM} 为语义正向编码操作; \mathbf{E}_{LM} 为词向量; \tilde{h}_{LM}^i 为反向隐藏状态; \tilde{B}_{iLSTM} 为语义反向编码操作。通过 $\vec{h}_{LM}^i, \tilde{h}_{LM}^i$ 获得项目 i 的第 L 个评论中存在的第 L 个单词对应的隐藏状态输出 $h_{LM}^i = \vec{h}_{LM}^i \parallel \tilde{h}_{LM}^i$ 。

从项目评论的角度展开分析,在特征提取过程中,每个单词产生的贡献度均不相同,通过注意力机制^[8-9]确定贡献度较高的信息词,设置权重矩阵 \mathbf{W}_{wh}^i , 利用上述提取的信息词表示项目评论 d_{iL} 。

$$\begin{cases} d_{iL} = \sum_{M=1}^n s_{LM}^i h_{LM}^i \\ s_{LM}^i = \frac{\exp[(\mathbf{v}_w^i)^T \tanh(\mathbf{W}_{wh}^i h_{LM}^i + b_w^i)]}{\sum_{M=1}^n \exp[(\mathbf{v}_w^i)^T \tanh(\mathbf{W}_{wh}^i h_{LM}^i + b_w^i)]} \end{cases} \quad (2)$$

式中: s_{LM}^i 为隐藏状态 h_{LM}^i 的注意力权重; \mathbf{v}_w^i 为权重向量; \tanh 属于激活函数; b_w^i 为偏置项。

在模型的第二层注意力层中学习评论对应的权重以此提取代表性较为显著的信息,将编写评论的用户 ID 嵌入 u_{iL} , 评论特征向量 d_{iL} 作为输入,用 ReLU 表示该层的激活函数,通过式(3)定义第二层注意力网络^[10-11] s_{iL}^* :

$$s_{iL}^* = \mathbf{J}^T \text{ReLU}(\mathbf{W}_s d_{iL} + \mathbf{W}_u u_{iL}) + b_2 \quad (3)$$

式中: \mathbf{J} 为注意力机制的权重向量; $\mathbf{W}_s, \mathbf{W}_u$ 分别为 d_{iL} 与用户 ID 嵌入 u_{iL} 对应的权重矩阵; b_2 为第二层注意力层的偏置项。

第二层注意力网络的输出即为注意力得分,采用 softmax 函数对式(3)的输出展开归一化处理获得评论权重 s_{iL} , 通过式(4)确定项目 i 的特征向量 p_i :

$$\begin{cases} s_{iL} = \frac{\exp(s_{iL}^*)}{\sum_{L=1}^k \exp(s_{iL}^*)} \\ p_i = \sum_{L=1}^k s_{iL} d_{iL} \end{cases} \quad (4)$$

设 $\mathbf{W}_x, \mathbf{W}_r$ 分别代表更新门和重置门对应的权重矩阵,由上一时刻的隐藏状态 h_{t-1} 和当前时刻的输入以及上下文信息 v_t 决定 CAGRU 单元中更新门 x_t 与重置门 r_t 的值。

$$\begin{cases} x_t = \sigma(\mathbf{W}_x h_{t-1} + \mathbf{W}_p p_t + \mathbf{W}_v v_t + b_x) \\ r_t = \sigma(\mathbf{W}_r h_{t-1} + \mathbf{W}_p p_t + \mathbf{W}_v v_t + b_r) \end{cases} \quad (5)$$

式中: σ 为 sigmoid 函数; p_t 为当前时刻的 POI; \mathbf{W}_p 为 p_t 对应的权重矩阵; \mathbf{W}_v 表示上下文信息的权重矩阵; b_x, b_r 均代表偏置项,分别位于更新门与重置门中。

针对上一时刻流入全连接层的隐藏状态,采用重置门对其展开调整与控制,通过式(6)对候选隐藏状态 \tilde{h}_t 展开计算。

$$\tilde{h}_t = \tanh[\mathbf{W}_h(v_t \cdot h_{t-1}) + \mathbf{W}_p p_t + \mathbf{W}_v v_t + b_h] \quad (6)$$

式中: \mathbf{W}_h 表示权重矩阵; b_h 属于偏置项。

更新门的主要作用是组合上述过程获取的隐藏状态^[12-13], 输出当前时刻的隐藏状态 h_t , 即用户偏好特征。

$$h_t = x_t \cdot h_{t-1} + (1 - x_t) \cdot \tilde{h}_t \quad (7)$$

将上述获取的项目特征和用户偏好特征扩展为基于评级的特征,向隐藏空间中映射上述特征,建立项目 i 与用户 u 的交互模型 h_0 :

$$h_0 = h_i(q_u + h_i) \cdot (a_i + p_i) \quad (8)$$

式中: a_i, q_u 分别为基于评级的项目特征与用户偏好特征。将式(8)输出结果输入预测层中,获得项目评分 \hat{y}_i 。

$$\hat{y}_i = \mu + n_u + n_i + \mathbf{W}_2^T h_0 \quad (9)$$

式中: μ 代表全局平均偏差; n_i, n_u 描述的是项目与用户

偏差; W_2 代表预测层的权重矩阵。

根据上述评价结果对项目排序,将高分项目推荐给用户。为了提高模型精度,加权融合 MSE 损失函数、CE 损失函数和 RK 损失函数建立组合损失函数,对深度联合训练模型展开训练,提高模型的推荐性能。

1) MSE 损失函数

MSE 损失函数 L_{MSE} 是分析预测值与实际值之间的偏差,将 MSE 损失函数最小作为目标,对建立的模型展开训练。

$$L_{MSE} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (10)$$

式中: y_i 表示实际值; n 表示样本数量。

2) CE 损失函数

CE 损失函数 L_{CE} 用于分析训练模型与真实模型之间存在的概率性误差^[14],将 CE 损失函数最小作为目标,对联合深度推荐模型展开训练,使模型面对海量的项目和用户数据时具有良好的分类能力,CE 损失函数 L_{CE} 的表达式如下:

$$L_{CE} = \frac{1}{n} \sum_{i=1}^n [-\lg(l_i)] \quad (11)$$

式中 l_i 描述的是模型正确预测样本类别的概率。

3) RK 损失函数

RK 损失函数 L_{RK} 由两部分构成:第一部分是中心损失函数 L_c ,用于衡量类内高维特征之间的距离;第二部分是边距排序损失函数 L_{mr} ,用于衡量不同类间特征距离。通过最小化 RK 损失函数来训练推荐模型,可以提高模型的分类精度。根据 L_c 与 L_{mr} 构成 RK 损失函数 L_{RK} ,表达式如下:

$$L_{RK} = L_c + L_{mr} \quad (12)$$

将 RK 损失函数 L_{RK} 最小作为目标对推荐模型展开训练,可提高模型的分类精度。加权组合上述损失函数,建立组合损失函数 L ,对基于注意力循环神经网络的联合深度推荐模型展开整体训练,提高模型的整体性能。

$$L = \alpha L_{MSE} + \beta L_{CE} + \chi L_{RK} \quad (13)$$

式中 α, β, χ 分别表示对应函数的权重系数。

2 实验与分析

为了验证基于注意力循环神经网络的联合深度推荐模型的整体有效性,需要对其展开测试,测试在 PyCharm Community Edition 2022.2.2 集成开发平台中完成,测试所用的数据集如表 1 所示。

采用本文方法、文献[2]方法和文献[3]方法展开推荐测试,通过均方根误差(RMSE)和平均绝对误差(MAE)对推荐结果的精度展开评估。

$$\begin{cases} RMSE = \sqrt{\frac{\sum_{i=1}^N |\hat{y}_i - y_i|^2}{N}} \\ MAE = \frac{\sum_{i=1}^N |\hat{y}_i - y_i|}{N} \end{cases} \quad (14)$$

式中: N 表示样本的总数量; y_i, \hat{y}_i 分别描述的是真实值与预测值。

表 1 测试数据集

数据集	用户数	项目数	评论数
Amazon	29 748	15 146	135 264
ML-1m	6 047	3 516	993 458
ML-10m	69 234	10 058	9 945 264

MAE 与 RMSE 测试结果如图 1 所示。

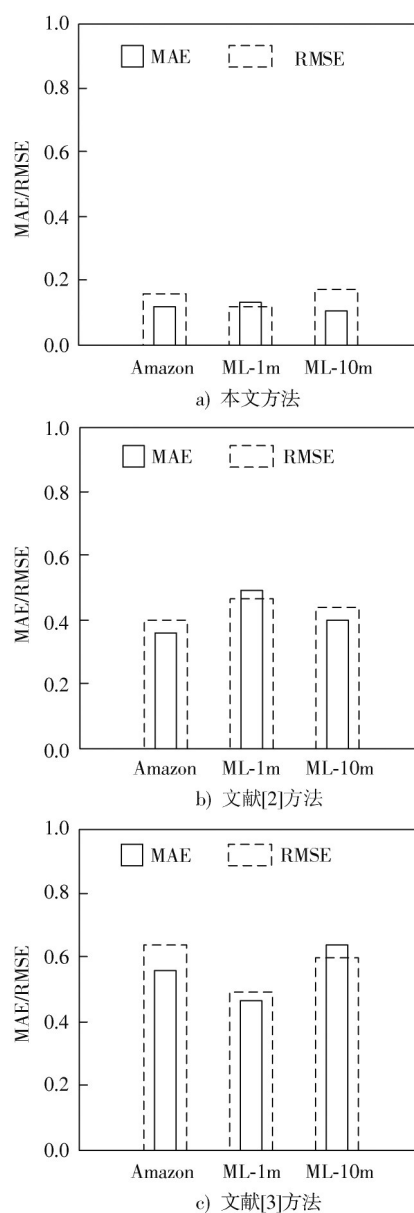


图 1 MAE 与 RMSE 测试结果

根据图 1 中的数据可知,在 Amazon 数据集、ML-1m 数据集与 ML-10m 数据集中,本文方法推荐结果的 MAE 与 RMSE 均低于文献[2]方法和文献[3]方法,控制在 0.2 以内,表明本文方法具有良好的推荐精度,因为本文方法在模型训练过程中加权融合多种损失函数,对联合深度推荐模型展开训练,提高了联合深度推荐模型的精度。

点击率(HR)用来衡量推荐项目与用户之间的交互情况,HR 越高,表明推荐的项目越符合用户偏好。HR 表达式为:

$$HR = \frac{N_u}{M}$$

(15)

式中: M 表示方法推荐的项目数量; N_u 表示项目点击量。三种方法的点击率(HR)如图 2 所示。

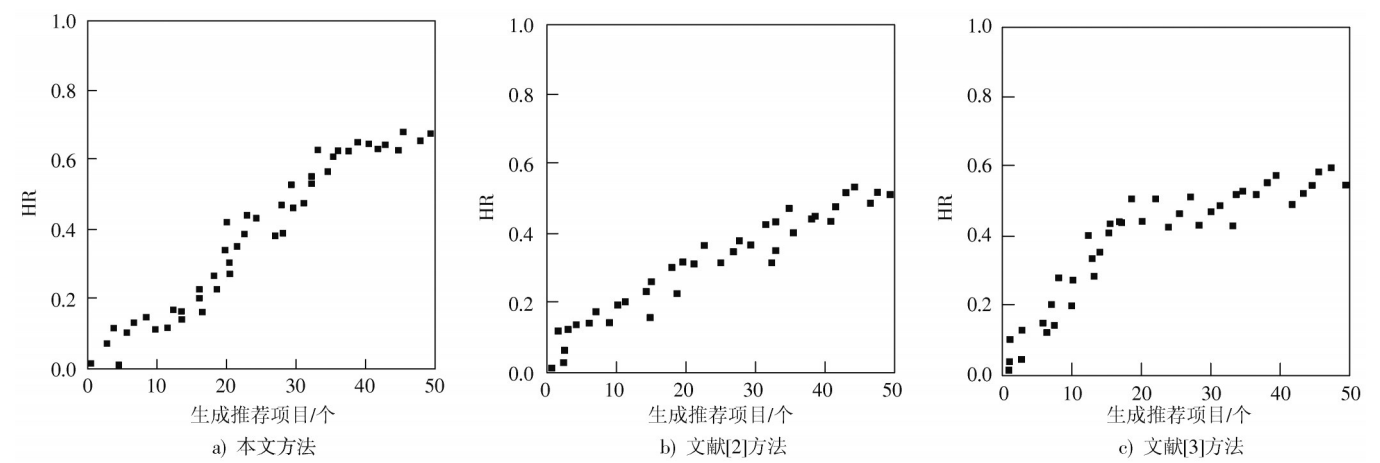


图 2 不同方法的 HR

分析上述测试结果可知,随着生成推荐项目数量的增加,用户点击量随之增高,点击率(HR)整体呈上升趋势,但在相同生成推荐项目数量下,本文方法生成的推荐项目更符合用户偏好,用户更愿意点击本文方法生成的推荐项目,表明本文方法的推荐效果良好。

归一化折现累积增益(NDCG)用于衡量推荐列表中项目是否处于正确的位置,其表达式如下:

$$NDCG = \frac{DCG@N}{Z}$$

(16)

式中:DCG@N代表理想折现累积收益;Z描述的是规范化常数。三种方法生成的推荐列表对应的 NDCG 如表 2 所示。

表 2 不同方法的 NDCG %	
方法	NDCG
本文方法	32.6
文献[2]方法	20.1
文献[3]方法	19.7

根据上述测试结果可知,本文方法的 NDCG 是三种方法中最高的,表明本文方法生成的推荐列表中项目排序正确,侧面验证了本文方法建立的联合深度推荐模型具有较高的预测精度。

3 结 语

针对目前推荐方法存在的问题,本文方法在网络中设置了双层注意力机制,用于学习项目特征,同时通过全连接层学习用户偏好特征。结合上述特征生成推荐列表,该模型将用户行为、内容特征和上下文信息等多维度的数据联合考虑,挖掘数据背后的深层关联,为用户提供更符合其需求和兴趣的推荐结果。经验证所提方法具有良好的推荐性能,在未来的发展中,联合深度推荐模型将继续优化,以适应不断变化的市场需求和用户行为。

注:本文通讯作者为段超。

参 考 文 献

[1] 黄震华,林威,吴正洋,等.基于最优传输和知识回放的新闻推荐模型增量学习方法[J]. 计算机学报, 2023, 46(10): 2161-2177.

[2] 刘欢,李晓戈,胡立坤,等.基于知识图谱驱动的图神经网络推荐模型[J]. 计算机应用, 2021, 41(7): 1865-1870.

[3] 王永,王利,冉珣,等.基于时间权重因子的隐私保护推荐算法[J]. 湖南大学学报(自然科学版), 2022, 49(8): 196-207.

[4] LAK P, BOZANTA A, KAVAKLIOGLU C, et al. A replication study on implicit feedback recommender systems with application to the data visualization recommendation [J]. Expert systems: The international journal of knowledge engineering, 2022,

- 39(4): 12871.
- [5] LI Y, WANG R H, NAN G F, et al. A personalized paper recommendation method considering diverse user preferences [J]. Decision support systems, 2021, 146: 113546.
- [6] 王琛,王颖,郑涛,等.基于ResNet-LSTM网络和注意力机制的综合能源系统多元负荷预测[J].电工技术学报,2022,37(7): 1789-1799.
- [7] 李卓,叶林,戴斌华,等.基于IDSCNN-AM-LSTM组合神经网络超短期风电功率预测方法[J].高电压技术,2022,48(6): 2117-2127.
- [8] 胡从强,曲娜,张帅,等.连续小波变换和具有注意力机制的深度残差收缩网络在低压串联电弧故障检测中的应用[J].电网技术,2023,47(5):1897-1905.
- [9] 蒯红权,吴建华,吴亮.基于注意力机制的深度循环神经网络的语音情感识别[J].电子器件,2022,45(1):139-142.
- [10] 汪璟玢,雷晶,张璟璇,等.基于Triplet注意力的循环卷积神经网络模型[J].模式识别与人工智能,2022,35(2):116-129.
- [11] 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.
- [12] 任剑锋,叶春明.嵌入指针网络的深度循环神经网络模型求解作业车间调度问题[J].计算机应用研究,2021,38(1):120-124.
- [13] 赵伟,王文娟,甘玉芳.基于预训练模型和多视角循环神经网络的电力文本匹配模型[J].重庆邮电大学学报(自然科学版),2023,35(3):545-553.
- [14] 戴永寿,高倩倩,孙伟峰,等.结合改进CNN和双约束损失函数的叠前地震数据低频补偿方法[J].石油地球物理勘探,2022,57(6):1287-1295.
- [15] 魏玉福,陈丽萍.基于注意力机制的深度学习体育运动姿态估计技术[J].电子设计工程,2023,31(2):152-155.

作者简介:郭东坡(1992—),女,河南南阳人,博士研究生,讲师,研究方向为数字化。

何 彬(1981—),男,湖北恩施人,博士研究生,讲师,硕士生导师,研究方向为机器解答和智能导学。

张明焱(1988—),男,湖北随州人,博士研究生,讲师,硕士生导师,研究方向为大数据学习分析。

段 超(1987—),男,湖北黄冈人,博士研究生,讲师,硕士生导师,研究方向为推荐系统和数据挖掘。