

史加荣, 马文华. 基于矩阵补全的 BPAM 推荐算法[J]. 智能计算机与应用, 2025, 15(5): 75–81. DOI: 10. 20169/j. issn. 2095–2163. 250510

基于矩阵补全的 BPAM 推荐算法

史加荣, 马文华

(西安建筑科技大学 理学院, 西安 710055)

摘要: 受深度学习显著成功的启发, 研究人员尝试将深度神经网络(DNN)引入到推荐系统中, 以学习用户对项目的偏好。然而, 稀疏的评分数据容易导致过拟合, 且 DNN 通常导致较高的计算复杂度和存储成本。为了解决这些问题, 提出一种融合矩阵补全(MC)与带注意力机制的反向传播神经网络(BPAM)的推荐算法: MC-BPAM。首先, 将原始用户-项目评分矩阵的未知元素通过低秩矩阵补全模型进行填充; 然后, 根据补全后的矩阵, 利用反向传播(BP)神经网络来学习目标用户与其邻居的复杂关系。此外, 还加入了一种注意力机制, 以捕捉每个用户对所有最近目标用户的全局影响。最后, 在 ml-la, filmtrust 和 A-Digital-Music 这 3 个真实数据集上进行了实验, MC-BPAM 的均方根误差(RMSE)达到了 0.269 5~0.574 7, 平均绝对误差(MAE)达到了 0.180 4~0.345 9, 相对于 BPAM 模型, MC-BPAM 的 RMSE 和 MAE 分别降低了至少 54.59% 和 47.94%。MC-BPAM 算法显著地提高了推荐系统的性能。

关键词: 推荐系统; 矩阵补全; BP 神经网络; 注意力机制; 深度学习

中图分类号: TP391.3

文献标志码: A

文章编号: 2095-2163(2025)05-0075-07

Recommendation algorithm of BPAM based on matrix completion

SHI Jiarong, MA Wenhua

(School of Science, Xi'an University of Architecture and Technology, Xi'an 710055, China)

Abstract: Inspired by the remarkable success of deep learning, researchers have attempted to introduce Deep Neural Networks (DNN) into recommendation systems to learn user preferences for items. However, sparse scoring data is prone to overfitting, and high computational complexity and storage costs are usually required in DNN. In order to solve these problems, a recommendation algorithm, MC-BPAM, is proposed by fusing Matrix Completion (MC) and Back Propagation neural network with Attention Mechanism (BPAM). Firstly, the unknown elements of the original user-item scoring matrix are filled in by the low-rank matrix completion model. Then, the Back Propagation (BP) neural network is used to learn the complex relationship between the target user and its neighbors according to the completed matrix. In addition, an attention mechanism has been incorporated to capture the global impact of each user on all nearest target users. Finally, experiments are carried out on the three real datasets of ml-la, filmtrust and A-Digital-Music, and the Root Mean Squared Error (RMSE) of MC-BPAM reaches 0.269 5~0.574 7, and the Mean Absolute Error (MAE) reaches 0.180 4~0.345 9, which decrease by at least 54.59% and 47.94% respectively compared with the BPAM model. The MC-BPAM algorithm significantly improves the performance of recommendation systems.

Key words: recommendation system; matrix completion; BP neural networks; attention mechanisms; deep learning

0 引言

随着时代的发展, 人们从海量的数据中找到自己真正感兴趣的内容是极其困难的。为了解决这个问题, 研究人员提出了个性化推荐系统, 其基本思想是用信息过滤的方法发现用户的偏好, 从而为用户

提供合适的项目^[1]。作为推荐系统中最经典的技术之一, 协同过滤(Collaborative Filtering, CF)是根据用户的历史行为数据挖掘用户的偏好^[2-4]。

传统的协同过滤算法通常分为 2 类: 矩阵分解(Matrix Factorization, MF)^[5]和基于邻域的方法^[6-7]。前者先学习用户和项目的潜在空间, 并将

基金项目: 陕西省自然科学基金(2021JM-378)。

作者简介: 马文华(1999—), 女, 硕士研究生, 主要研究方向: 机器学习, 数据挖掘。

通信作者: 史加荣(1979—), 男, 博士, 教授, 硕士生导师, 主要研究方向: 机器学习, 数据挖掘。Email: shijiarong@xauat.edu.cn。

收稿日期: 2023-10-17

哈尔滨工业大学主办 ◆ 学术研究与应用

两者映射到公共空间中,最后利用用户和项目之间的相似性为每个用户的偏好进行预测。然而,真实的评分数据会出现长尾分布,很容易在 MF 方法中出现稀疏性问题^[8-9]。基于邻域的 CF 方法则将相似用户或项目进行平均或加权后,再对目标用户的评分进行预测。Liang 等学者^[10]将自编码器应用于基于隐式反馈的 CF 推荐任务,通过非线性概率模型克服线性因子模型的局限。由于用户不可能为所有的项目进行评分,故由评分数据构成的矩阵通常都是非常稀疏的,从而导致推荐性能不高。

在过去几年里,为了能够更好地缓解评分数据的稀疏性问题^[11],人们提出了许多策略。任永功等学者^[12]提出一种基于社交网络能量扩散的协同过滤推荐算法,在一定程度上缓解了数据的稀疏性。陈佩武等学者^[13]在含隐式反馈信息的矩阵奇异值分解语义模型中引入了隐性信任关系,降低了数据的稀疏性。高子建等学者^[14]提出的谱聚类和隐语义模型有效缓解了稀疏性。Lemire 等学者^[15]提出了 Slope One 算法,通过对稀疏矩阵进行填充,在预测时使用线性回归思想以缓解推荐系统中的稀疏性问题。向小东等学者^[16]基于 Slope One 算法提出了一种改进评分稀疏矩阵的方法,通过评分预测公式对目标用户进行评分预测。

基于注意力机制的反向传播神经网络 (Back Propagation neural network with Attention Mechanism, BPAM) 在基于邻域的 CF 算法中引入了反向传播 (Back Propagation, BP) 神经网络,能够有效地捕获目标用户与其邻居之间的非线性关系,但该算法不能克服数据的稀疏性^[17-18]。为了降低评分数据的稀疏性和捕获目标用户与其邻居之间的复杂关系,本文提出了一种新的融合矩阵补全 (Matrix Completion, MC) 与 BPAM 的推荐算法: MC-BPAM。该算法先对评分矩阵进行补全,再将恢复后的矩阵作为 BPAM 模型的输入进行评分预测。

1 相关工作

1.1 矩阵补全

在矩阵存在缺失元素的情况下,恢复其缺失值的问题被称为矩阵补全^[19-20]。在一般的 MC 问题中,需要根据近似低秩性来恢复缺失元素。对于用户集 U 和项目集 V 所组成的原始评分矩阵 $\mathbf{Z} = (z_{ui})_{|U| \times |V|}$, 根据低秩性对缺失数据进行补全,优化模型为:

$$\min_{\mathbf{R}} \text{rank}(\mathbf{R}), \text{ s. t. } r_{ui} = z_{ui}, (u, i) \in \Omega \quad (1)$$

其中, $\mathbf{R} = (r_{ui})_{|U| \times |V|}$ 表示补全后的评分矩阵; $\Omega \subseteq \{1, 2, \dots, |U|\} \times \{1, 2, \dots, |V|\}$ 表示原始评分矩阵 $\mathbf{Z}$ 已知元素的下标集合; $|U|$ 和 $|V|$ 分别表示用户和项目集合的势。

式(1)的求解是一个 NP 难问题,秩函数 $\text{rank}(\mathbf{R})$ 在 $\|\mathbf{R}\|_2 \leq 1$ 约束下的凸包络函数为 $\Phi(\mathbf{R}) = \|\mathbf{R}\|_*$, 其中 $\|\cdot\|_2$ 和 $\|\cdot\|_*$ 表示矩阵的谱范数和核范数。将最优化模型(1)凸松弛到矩阵核范数最小化模型:

$$\min_{\mathbf{R}} \|\mathbf{R}\|_*, \text{ s. t. } r_{ui} = z_{ui}, (u, i) \in \Omega \quad (2)$$

上述矩阵补全方法已被广泛应用在推荐系统、计算机视觉、图像修复和图像分类等领域中。

1.2 样本集的构建

对每个用户 $u \in U$, 计算用户评分向量之间的余弦相似度来获得 k 个最近邻用户 (K Nearest Users, KNU), 其中 k 是预先指定的最近邻用户数量。用户 u 的 KNU 评分矩阵 $\mathbf{N}^u \in \mathbb{R}^{|V| \times (k+1)}$ 由目标用户 u 和其 k 个最近邻用户对所有项目的评分信息所构成。 $\mathbf{N}^u$ 的前 k 列是 u 的 KNU 对所有项目的评分, 记为 $\mathbf{X}^u$; u 对所有项目的评分为 $\mathbf{N}^u$ 中的最后一列, 记为 $\mathbf{y}^u$ 。对于用户 u 和项目 i 构建输入输出样本对 $(\mathbf{X}_i^u, \mathbf{y}_i^u)$, 其中 $\mathbf{X}_i^u$ 是矩阵 $\mathbf{X}^u$ 的第 i 行, $\mathbf{y}_i^u$ 是向量 $\mathbf{y}^u$ 的第 i 个元素。在训练过程中, 根据预测评分 $\hat{y}_i^u$ 和真实评分 y_i^u 之间的误差来调整参数, 其中 $y_i^u = z_{ui}$ 。

1.3 带注意力机制的反向传播神经网络

带注意力机制的反向传播神经网络 (BPAM) 模型的架构如图 1 所示, 其中 BP 神经网络包括输入层、隐藏层和输出层。将注意力机制引入 BP 神经网络, 不仅可以考虑最近邻用户对目标用户的局部注意力, 还能够捕捉目标用户的最近邻用户对其最近邻用户集的全局注意力^[21]。

对于目标用户 $u \in U$, 其 KNU 集合表示为 $T^u = \{n_1^u, n_2^u, \dots, n_k^u\}$, 其中 n_j^u 为 u 的第 j 个最近邻用户, $j = 1, \dots, k$ 。在对用户 u 进行预测评分时, 将 T^u 中的用户 n_j^u 对 u 的影响视为局部注意力, 对应的矩阵记为 $\mathbf{W}^u \in \mathbb{R}^{k \times o}$, 其中 o 为隐藏层中神经元的数量。将用户 n_j^u 对 T^u 的影响视为全局注意力, 所有用户的全局注意力矩阵由 $\mathbf{A} \in \mathbb{R}^{|U| \times o}$ 表示。 $\mathbf{A}_{T^u} \in \mathbb{R}^{k \times o}$ 表示 $\mathbf{A}$ 的子矩阵, 由其第 $n_1^u, n_2^u, \dots, n_k^u$ 行组成, 即对应于用户 u 的 k 个最近邻用户。全局注意力机制使模型不容易陷入局部最优。

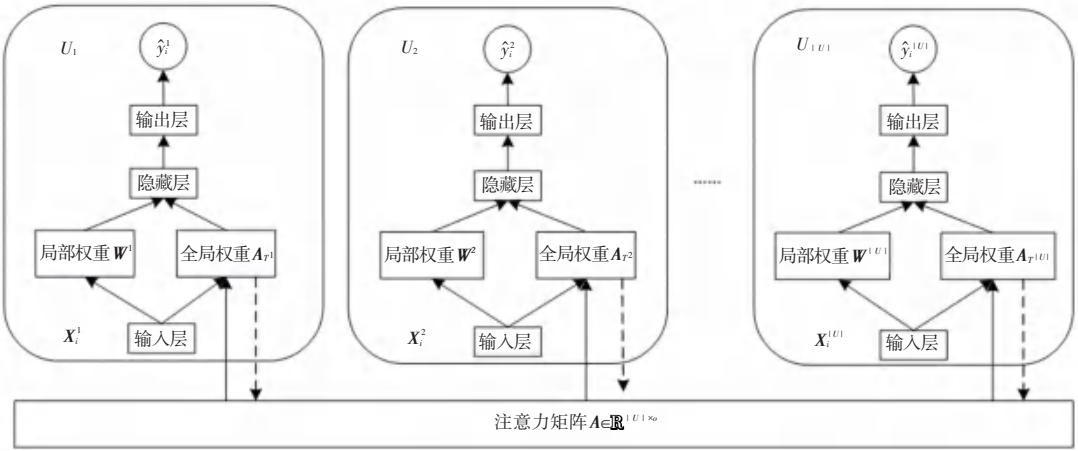


图 1 BPAM 模型的体系结构

Fig. 1 Architecture of the BPAM model

2 MC 与 BPAM 的融合

为了缓解矩阵的稀疏性,通过求解优化模型(2)将原始评分矩阵进行补全,形成新的评分矩阵 R 。用户 u 对应的训练样本集 D^u 由 KNU 评分矩阵

N^u 构成,即 $D^u = \{(X_1^u, y_1^u), \dots, (X_{|V_u|}^u, y_{|V_u|}^u)\}$, 其中 V_u 表示用户 u 的所有评分项目的子集。MC-BPAM 包括原始评分矩阵的补全、输入层、注意力机制层、反向传播层和输出层,如图 2 所示。

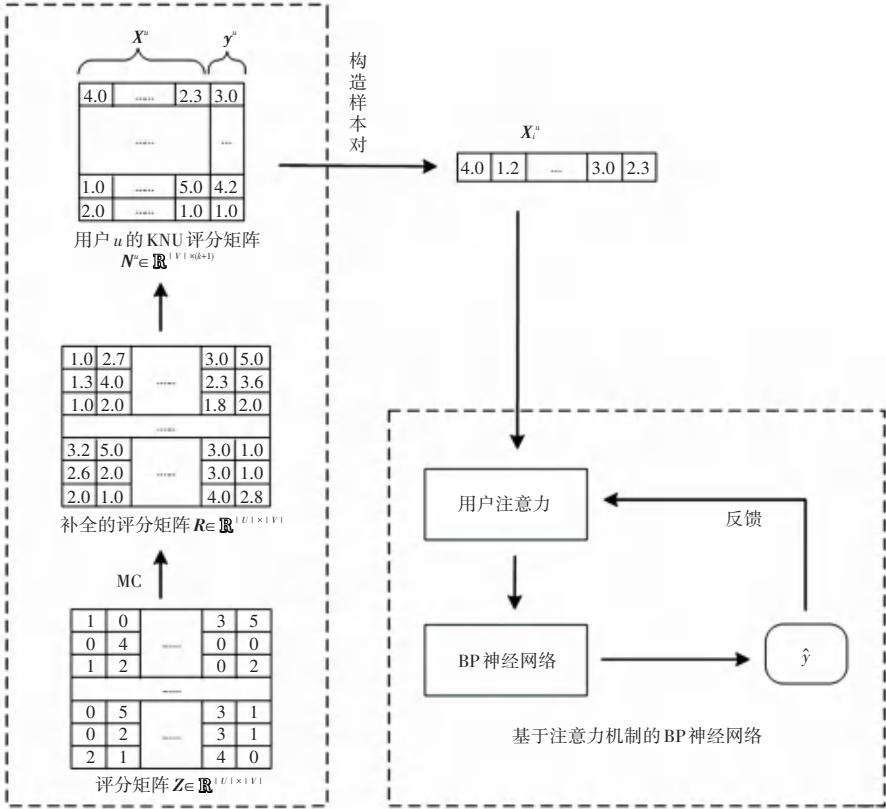


图 2 MC-BPAM 的结构

Fig. 2 Structure of MC-BPAM

MC-BPAM 的训练过程可分为 2 个阶段:正向传播和误差反向传播。用户 u 、项目 i 的基于正向传播的评分预测为:

$$\mathbf{h}_i^u = \sigma((\mathbf{W}^u + \alpha \mathbf{A}_{T^u})^T \mathbf{X}_i^u + \mathbf{b}_i^u) \quad (3)$$

$$\hat{y}_i^u = \sigma((\mathbf{W}_h^u)^T \mathbf{h}_i^u + \mathbf{b}_h^u) \quad (4)$$

其中, $\mathbf{h}_i^u \in \mathbb{R}^{o \times 1}$ 表示隐藏层的值; $\sigma(\cdot)$ 表示 Sigmoid 激活函数; $\mathbf{W}^u \in \mathbb{R}^{k \times o}$ 和 $\mathbf{b}_i^u \in \mathbb{R}^{o \times 1}$ 分别表示输入层与隐藏层之间的连接权重矩阵与偏置向量; $\mathbf{W}_h^u \in \mathbb{R}^{o \times 1}$ 和 $b_h^u \in \mathbb{R}^{1 \times 1}$ 分别表示隐藏层与输出层之间的权重向量与偏置; α 表示用于调整全局注意力的权衡参数。

对于 MC-BPAM 模型, 正则化的目标函数如下:

$$E = \frac{1}{2 \times |U|} \sum_{u=1}^{|U|} E^u + \frac{\lambda_1}{2} \|\mathbf{A}\|_F^2 \quad (5)$$

同时推得:

$$E^u = \frac{1}{|V_u|} \sum_{i \in V_u} (y_i^u - \hat{y}_i^u)^2 + \lambda_2 (\|\mathbf{W}^u\|_F^2 + \|\mathbf{W}_h^u\|_F^2 + \|\mathbf{b}_i^u\|_F^2 + (b_h^u)^2) \quad (6)$$

其中, λ_1 和 λ_2 表示非负的正则化系数, $\|\cdot\|_F$ 表示矩阵的 Frobenius 范数。

对于模型参数 $\mathbf{A}$ 和 $\theta = \{\mathbf{W}^u, \mathbf{W}_h^u, \mathbf{b}_i^u, b_h^u : u \in U\}$, 均按照均值为 0、标准差为 0.01 的高斯分布随机初始化, 基于随机梯度下降方法更新模型参数。对于训练样本 $(\mathbf{X}_i^u, y_i^u)$, 采用误差反向传播策略。取学习率 $\eta \in (0, 1)$, 则参数更新公式为:

$$\begin{aligned} \uparrow \mathbf{A}_{r^u} &: = \mathbf{A}_{r^u} - \eta(\alpha \mathbf{e}_i^u \mathbf{X}_i^u + \lambda \mathbf{A}_{r^u}) \\ \uparrow \mathbf{W}^u &: = \mathbf{W}^u - \eta(\mathbf{e}_i^u \mathbf{X}_i^u + \lambda \mathbf{W}^u) \\ \uparrow \mathbf{b}_i^u &: = \mathbf{b}_i^u - \eta \mathbf{e}_i^u \\ \uparrow \mathbf{W}_h^u &: = \mathbf{W}_h^u - \eta(\mathbf{g}_i^u \mathbf{h}_i^u + \lambda \mathbf{W}_h^u) \\ \uparrow b_h^u &: = b_h^u - \eta g_i^u \end{aligned} \quad (7)$$

在式(7)中,

$$g_i^u = \hat{y}_i^u (1 - \hat{y}_i^u) (y_i^u - \hat{y}_i^u) \quad (8)$$

$$\mathbf{e}_i^u = \mathbf{h}_i^u \odot (\mathbf{I} - \mathbf{h}_i^u) \odot \mathbf{W}_h^u \mathbf{g}_i^u \quad (9)$$

其中, “ $\odot$ ” 表示哈达玛乘积; $\mathbf{I}$ 表示元素都是 1 的 o 维向量。

综上, 通过引入注意力机制, MC-BPAM 结合局

部权重和全局权重来预测缺失评分。所提模型的训练过程交替迭代前向传播和误差反向传播, 直到达到收敛条件。算法 1 总结了 MC-BPAM 的整个过程。

算法 1 MC-BPAM 的算法框架

1. 输入评分矩阵 $\mathbf{Z}$, 最近邻用户数 k , 权衡参数 α ;
2. 通过求解式(2)得到补全后的评分矩阵 $\mathbf{R}$;
3. 计算用户 u 的 k 个最近邻用户 (KNU), 集合 $T^u = \{n_1^u, n_2^u, \dots, n_k^u\}, u \in U$;
4. 根据 T^u 构造用户 u 的评分矩阵 $\mathbf{N}^u, u \in U$;
5. 根据 $\mathbf{N}^u$ 建立训练集 $D^u = \{(\mathbf{X}_1^u, y_1^u), \dots, (\mathbf{X}_{|V_u|}^u, y_{|V_u|}^u)\}, u \in U$;
6. 随机初始化模型参数 $\mathbf{A}$ 和 θ ;
7. 当目标函数(5)不满足终止条件
8. 对于 $u \in U$
9. 对于 $(\mathbf{X}_i^u, y_i^u) \in D^u$
10. 由式(4)预测评分 $\hat{y}_i^u$;
11. 由式(7)更新 $\mathbf{A}_{r^u}$ 和 θ ;
12. 结束
13. 结束
14. 结束
15. 输出模型参数 $\mathbf{A}$ 和 θ 。

3 实验结果与分析

3.1 数据集与评估指标

在 3 个公开的数据集上进行实验: MovieLens 中的 ml-la、filmtrust, Amazon 中的 A-Digital-Music。这 3 个数据集的基本信息见表 1。其中, 数据稀疏度是指在评分矩阵中, 无评分数据的元素占整个矩阵空间的比率。将每个数据集的用户按 3 : 1 随机分成训练集和测试集。

表 1 推荐系统数据集的基本信息

Table 1 Basic information about the recommendation system datasets

数据集	用户数	项目数	评分数	稀疏度/%	评分范围
ml-la	610	9 724	100 836	98.30	[0, 5]
filmtrust	1 508	2 071	35 497	98.86	[0, 5]
A-Digital-Music	1 000	4 796	11 677	99.90	[0, 5]

在推荐系统中, 通常使用均方根误差 (Root Mean Squared Error, $RMSE$) 和平均绝对误差 (Mean Absolute Error, MAE) 来衡量预测模型在测试集上的性能。 $RMSE$ 的计算公式为:

$$RMSE = \sqrt{\frac{\sum_{(i,j) \in \Omega_{\text{test}}} (r_{ij} - \hat{r}_{ij})^2}{|\Omega_{\text{test}}|}} \quad (10)$$

其中, $\hat{r}_{ij}$ 表示预测的评分, r_{ij} 表示实际评分。 Ω_{test} 表示测试集中已评分元素的指标集。 MAE 的计算公式为:

$$MAE = \frac{\sum_{(i,j) \in \Omega_{\text{test}}} |r_{ij} - \hat{r}_{ij}|}{|\Omega_{\text{test}}|} \quad (11)$$

$RMSE$ 和 MAE 值越小, 预测性能越好。

3.2 模型参数设置

对于 MC-BPAM 模型, 在集合 {3, 5, 7, 9, 11} 中调整最佳的最近邻用户的数量; 以 0.2 的间隔测试了从 0 到 2 之间的注意力比率 α 。将所提出的模型与以下 5 种模型进行比较。

(1) 概率矩阵分解 (Probabilistic Matrix Factorization, PMF)。是一种经典的概率算法, 只考虑潜在因素, 并使用矩阵分解来建立用户和项目之间的双线性关系^[22]。

(2) 深度矩阵分解 (Deep Matrix Factorization, DMF)。将矩阵分解中的线性嵌入操作替换为双路径神经网络^[23]。

(3) 深度协同过滤 (Deep Collaborative Filtering, DeepCF)。将表示学习与匹配函数学习结合到协同过滤方法中, 以捕获用户和项目之间的低阶关系^[24]。

(4) 宽度协同过滤 (Broad Collaborative Filtering, BroadCF)。将用户项目评分数据输送到宽度学习系统 (Broad Learning System, BLS)^[25]。

(5) BPAM。引入了全局注意力矩阵, 考虑了目标用户对其最近邻用户集的全局影响^[26]。

本文对上述 5 种模型的参数进行优化调整。对于 PMF 模型, 设置用户和项目的因子数均为 30, 用户和项目的正则化参数分别为 0.001 和 0.000 1。对于 DMF 模型, 设置批量大小为 256, 学习率为 0.000 1。对于 DeepCF 模型, 设置表示学习部分的多层感知器的大小为 [512, 256, 128, 64], 批量大小和学习率与 DMF 的设置相同。在 BroadCF 模型中, 设置最近邻的用户和项目数均为 5, 映射的特征组和增强特征组的数量均为 25, 映射特征维度和增强特征维度均为 10。对于 BPAM 模型, 在集合 {3, 5, 7, 9} 中选择最佳的最近邻用户的数量。所有实验都是在 Python 中实现的。

3.3 实验结果比较

对于所有数据集, 将 MC-BPAM 与现有的最新模型作比较, 每个模型运行 20 次, 求得平均结果。表 2 和表 3 分别比较了不同模型的 $RMSE$ 和 MAE ,

其中最佳结果以粗体显示。

实验结果表明, 与 PMF 相比, MC-BPAM 在 $RMSE$ 和 MAE 方面至少改进了 74.63% 和 59.08%, 性能提升相对显著。主要原因可能是 PMF 仅捕捉线性潜在因素, 即用户和项目之间的线性关系; 而 MC-BPAM 利用了矩阵补全算法, 缓解了数据的稀疏性, 且所使用的 BP 神经网络能捕捉目标用户和其近邻之间的非线性关系。与 DMF 和 DeepCF 两种基于深度神经网络 (Deep Neural Networks, DNN) 的算法相比, MC-BPAM 在 $RMSE$ 和 MAE 方面至少提升了 78.5% 和 83.76%, 改善效果更加显著。尽管基于 DNN 的算法和 MC-BPAM 都能够捕捉用户和项目之间的非线性关系, 但前者容易受过拟合问题的影响。与 BroadCF 和 BPAM 相比, MC-BPAM 的 $RMSE$ 和 MAE 至少提升了 15.31% 和 0.63%。综上, 实验结果证实了所提出的 MC-BPAM 在 3 个数据集上的有效性和优越性。

表 2 不同模型的 $RMSE$ 值比较

Table 2 Comparison of $RMSE$ values for different models

模型	ml-la	filmtrust	A-Digital-Music
PMF	1.152 1	1.139 2	2.156 0
DMF	1.190 8	1.211 6	2.262 4
DeepCF	1.846 1	1.648 2	2.091 3
BroadCF	0.816 5	0.919 2	0.727 8
BPAM	0.968 3	0.856 3	1.120 6
MC-BPAM	0.405 8	0.269 5	0.574 7

表 3 不同模型的 MAE 值比较

Table 3 Comparison of MAE values for different models

模型	ml-la	filmtrust	A-Digital-Music
PMF	0.802 7	0.915 8	1.485 6
DMF	1.049 5	1.169 2	1.942 1
DeepCF	1.658 7	1.528 3	1.867 2
BroadCF	0.516 2	0.605 6	0.352 2
BPAM	0.766 4	0.676 4	0.825 3
MC-BPAM	0.211 9	0.180 4	0.345 9

3.4 不同超参数对 MC-BPAM 的影响

以 $RMSE$ 和 MAE 为评价指标, 在 ml-la、filmtrust 和 A-Digital-Music 数据集上比较不同超参数对 MC-BPAM 的推荐性能的影响。

3.4.1 最近邻用户的数量

本节分析最近邻用户的数量 k 对 MC-BPAM 的推荐性能的影响。设置 $k \in \{3, 5, 7, 9, 11\}$, 图 3 报告了 MC-BPAM 的 $RMSE$ 和 MAE 。观察图 3 可以

看出:较小的 k 值具有较大的误差,而较大的 k 值往往有助于提高预测性能;但当 k 值过大时,预测性能会略有下降。这可能是由于过多的近邻数目错误地引入了更多的噪声,从而导致过拟合的发生。从评

价误差来看,ml-la 和 A-Digital-Music 的最佳 k 值为 7, filmtrust 的最佳 k 值为 9。考虑到运行效率和精度之间的权衡,本文根据上述结果设置参数 k 。

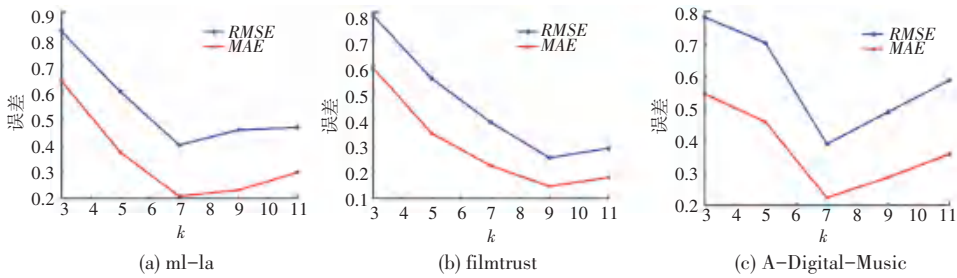


图3 不同 k 值下的 RMSE 和 MAE 值
Fig. 3 RMSE and MAE values under different k values

3.4.2 注意力比率

除了最近邻用户数 k 之外,注意力比率 α 的选择也是影响推荐性能的关键因素。取注意力比率 $\alpha \in \{0, 0.2, 0.4, \dots, 1.8, 2.0\}$, 计算 MC-BPAM 的 RMSE 和 MAE 值。当 $\alpha = 0$ 时,所提出的 MC-BPAM 模型不考虑目标的最近邻用户的全局影响。在 3 个数据集上运行 20 次实验,平均后的结果如图 4 所

示。从图 4 可以看出,ml-la 的最佳 α 值为 0.8, filmtrust 的最佳 α 值为 0.6, A-Digital-Music 的最佳 α 值为 0.4,且性能显著优于 $\alpha = 0$ 时的预测结果,即注意力能够提高模型的性能。较大的 α 往往产生较差的预测性能,这可能是模型过度关注近邻用户所致。在 3 个数据集上,本文根据上述结果设置注意力比率 α 。

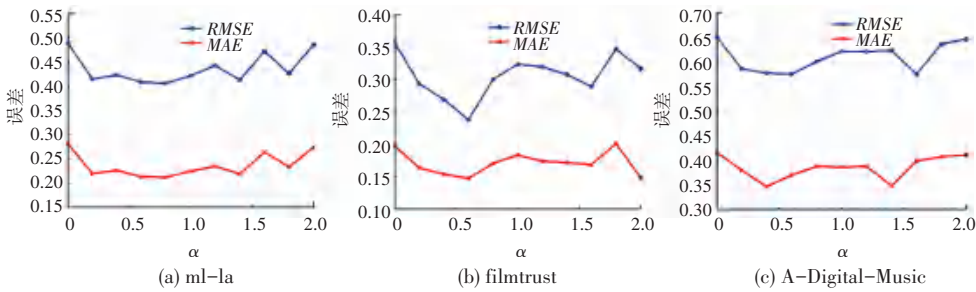


图4 不同注意力比率 α 下的预测误差
Fig. 4 Prediction errors under different attention ratios α

4 结束语

本文提出了一种基于矩阵补全的 BP 神经网络推荐算法 MC-BPAM,该算法融合了注意力机制。在 MC-BPAM 算法中,先利用最小化核范数模型对原始用户-项目矩阵中的未知元素进行补全,得到完整的评分矩阵;再通过 BP 神经网络学习目标用户与其近邻之间的复杂关系。与 DNN 相比,浅层的 BP 神经网络不仅可以降低计算和存储成本,还可以防止由于评分数量少而导致的模型过拟合。此外,所提模型引入了注意力机制,以获取用户影响其最近邻用户集的全局注意力。最后,在 3 个数据集上进行了实验,结果表明所提出的 MC-BPAM 算法比单一的 BPAM 算法具有更好的推荐效果。本文仅根据用户对项目的评分进行推荐,评分数据中的项

目信息、浏览时间和社交网络等辅助内容对推荐性能也有影响。因此,在今后的研究中可以把上述相关信息融入到推荐模型中,以提高用户满意度和推荐性能。

参考文献

[1] ZHANG J, JIN G, MA Y, et al. A diversity-enhanced resource allocation strategy for decomposition - based multiobjective evolutionary algorithm [J]. IEEE Transactions on Cybernetics, 2019, 48(8): 2388-2401.

[2] CHAE D K, KANG J S, KIM S W, et al. Rating augmentation with generative adversarial networks towards accurate collaborative filtering [C]// Proceedings of the World Wide Web Conference on Association for Computing Machinery. New York: ACM, 2019:2616-2622.

[3] DING Rui, GUO Guibing, YANG Xiaochun, et al. BiGAN: Collaborative filtering with bidirectional generative adversarial

- networks [C]// Proceedings of the 2020 SIAM International Conference on Data Mining. Philadelphia, USA: SIAM, 2020: 82–90.
- [4] 罗洁, 王力. 基于用户和项目的协同过滤算法的比较研究[J]. 智能计算机与应用, 2023, 13(1): 195–197.
- [5] HE Yuan, WANG Cheng, JIANG Changjun. Correlated matrix factorization for recommendation with implicit feedback[J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 31(3): 451–464.
- [6] VALCARCE D, LANDIN A, PARAPAR J, et al. Collaborative filtering embeddings for memory-based recommender systems[J]. Engineering Applications of Artificial Intelligence, 2019, 85: 347–356.
- [7] 胡文洁, 杨凯祥, 谭宗元. 基于 k-means 特征的适应性近似最近邻搜索算法[J]. 智能计算机与应用, 2023, 13(9): 80–84.
- [8] HU L, CAO L, CAO J, et al. Learning informative priors from heterogeneous domains to improve recommendation in cold-start user domains[J]. ACM Transactions on Information Systems, 2016, 35(2): 1–37.
- [9] 黄承宁, 孙自梅, 张斌, 等. 基于麻雀算法优化的神经网络推荐算法研究[J]. 智能计算机与应用, 2023, 13(9): 95–102.
- [10] LIANG Dawen, KRISHNAN R G, HOFFMAN M D, et al. Variational autoencoders for collaborative filtering [C]// Proceedings of the 2018 World Wide Web Conference on World Wide Web. Geneva: IWWW, 2018: 689–698.
- [11] NATARAJAN S, VAIRAVASUNDARAM S, NATARAJAN S, et al. Resolving data sparsity and cold start problem in collaborative filtering recommender system using linked open data [J]. Expert Systems with Applications, 2020, 149: 113248.
- [12] 任永功, 王瑞霞, 张志鹏, 等. 基于社交网络能量扩散的协同过滤推荐算法[J]. 模式识别与人工智能, 2021, 34(6): 561–571.
- [13] 陈佩武, 束方兴. 基于 SVD++ 隐语义模型的信任网络推荐算法[J]. 大数据, 2021, 7(4): 105–116.
- [14] 高子建, 张晗睿, 窦万春, 等. 基于谱聚类和隐语义模型的智能协同推荐方法[J]. 计算机集成制造系统, 2021, 27(9): 2517–2524.
- [15] LEMIRE D, MACLACHLAN A. Slope one predictors for online rating-based collaborative filtering [C]// Proceedings of the 2005 SIAM International Conference on Data Mining. Philadelphia, USA: SIAM, 2005: 471–475.
- [16] 向小东, 邱梓威. 基于 slope-one 算法改进评分矩阵填充的协同过滤算法研究[J]. 计算机应用研究, 2019, 36(4): 1064–1067.
- [17] XU K, BA J, KIROS R, et al. Show, attend and tell: Neural image caption generation with visual attention [C]// Proceedings of the 32nd International Conference on Machine Learning. New York: JMLR, 2015: 2048–2057.
- [18] WANG Shaonan, ZHANG Jiajun, ZONG Chengqing. Learning sentence representation with guidance of human attention [C]// Proceedings of the 26th International Joint Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2017: 4137–4143.
- [19] CHATTERJEE S. A deterministic theory of low rank matrix completion[J]. IEEE Transactions on Information Theory, 2020, 66(12): 8046–8055.
- [20] 史加荣, 郑秀云, 魏宗田, 等. 低秩矩阵恢复算法综述[J]. 计算机应用研究, 2013, 30(6): 1601–1605.
- [21] CHEN Jingyuan, ZHANG Hanwang, HE Xiangnan, et al. Attentive collaborative filtering: Multimedia recommendation with item-and component-level attention[C]// Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2017: 335–344.
- [22] MNIEH A, SALAKHUTDINOV R R. Probabilistic matrix factorization[J]. Advances in Neural Information Processing Systems, 2007, 20: 1257–1264.
- [23] XUE Hongjian, DAI Xinyu, ZHANG Jianbing, et al. Deep matrix factorization models for recommender systems [C]// Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne: AAAI Press, 2017, 17: 3203–3209.
- [24] DENG Zhihong, HUANG Ling, WANG Changdong, et al. DeepCF: A unified framework of representation learning and matching function learning in recommender system [C] // Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2019, 33(1): 61–68.
- [25] HUANG Ling, GUAN Canrong, HUANG Zhenwei, et al. Broad recommender system: An efficient nonlinear collaborative filtering approach[J]. arXiv preprint arXiv, 2204.11602, 2022.
- [26] XI Wudong, HUANG Ling, WANG Changdong, et al. BPAM: Recommendation based on BP neural network with attention mechanism [C]// Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao: AAAI Press, 2019: 3905–3911.