

基于MMR-DeepFM的图书馆读者图书推荐算法研究

王艺涵¹, 冯晨¹, 王晓晓¹, 解登峰², 张莉红^{2*}

¹中国海洋大学信息科学与工程学部, 山东 青岛

²中国海洋大学图书馆, 山东 青岛

收稿日期: 2025年6月16日; 录用日期: 2025年7月9日; 发布日期: 2025年7月16日

摘要

为更为有效地提高图书馆图书资源使用率并为读者提供个性化推荐服务, 本文提出了一种融合最大边缘相关 (Maximal-Marginal-Relevance, MMR) 算法和深度因子分解机 (Deep Factorization Machine, DeepFM) 深度学习神经网络的MMR-DeepFM图书推荐模型。采用中国海洋大学信息学部读者数据进行了模型训练和测试, 与其它图书推荐算法比较, 本文提出的图书推荐算法可以更为准确和多样化地为读者推荐图书, 最终提高图书馆图书的借阅率。

关键词

MMR, DeepFM, 图书推荐算法

Research on Book Recommendation Algorithm for Library Readers Based on MMR-DeepFM

Yihan Wang¹, Chen Feng¹, Xiaoxiao Wang¹, Dengfeng Xie², Lihong Zhang^{2*}

¹Faculty of Information Science and Engineering, Ocean University of China, Qingdao Shandong

²Library of Ocean University of China, Qingdao Shandong

Received: Jun. 16th, 2025; accepted: Jul. 9th, 2025; published: Jul. 16th, 2025

Abstract

To enhance the utilization rate of library book resources and provide personalized recommendation

*通讯作者。

文章引用: 王艺涵, 冯晨, 王晓晓, 解登峰, 张莉红. 基于 MMR-DeepFM 的图书馆读者图书推荐算法研究[J]. 数据挖掘, 2025, 15(3): 271-278. DOI: 10.12677/hjdm.2025.153023

services for readers more effectively, an MMR-DeepFM book recommendation model that integrates the Maximal-Marginal-Relevance (MMR) algorithm and the Deep Factorization Machine (DeepFM) deep learning neural network is proposed in this paper. The model was trained and tested using reader data from the Information Department of Ocean University of China. Compared with other book recommendation algorithms, the proposed algorithm in this paper can recommend books more accurately and diversely to readers, ultimately increasing the borrowing rate of library books.

Keywords

MMR, DeepFM, Book Recommendation Algorithm

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. 引言

图书馆是高校中的文献信息保障中心，为全校师生提供图书借阅、文献检索及科研支持等服务。在传统借阅方式中，师生产生图书借阅行为源于学习需求及兴趣爱好等因素，图书以一种由读者自主检索“人找书”的被动方式被借阅。在前期调查问卷提出“您通常能通过图书馆找到所需的纸本和电子书刊吗？”问题时，结果表明，面对海量的图书资源，高达 67.56% 读者很难快速地找到符合需求的图书，要花费较多的时间查找(图 1)。前期调查还发现在选择图书时仅有 18.78% 读者关注到了上架新书(图 2)，这限制了读者更好地开展拓展性阅读。这些问题的存在进一步使得许多图书借阅率低，空置率高，图书资源得不到充分利用。因此，亟需将这种“人找书”的被动借阅模式转变为“书找人”的主动服务，这就需要图书馆要为读者提供更为精准的个性化推荐服务，因此需基于读者有限的借阅数据，根据有限的隐式和显式反馈发掘读者阅读喜好，预测读者借阅倾向，为图书馆开展工作提供参考依据和决策支持。

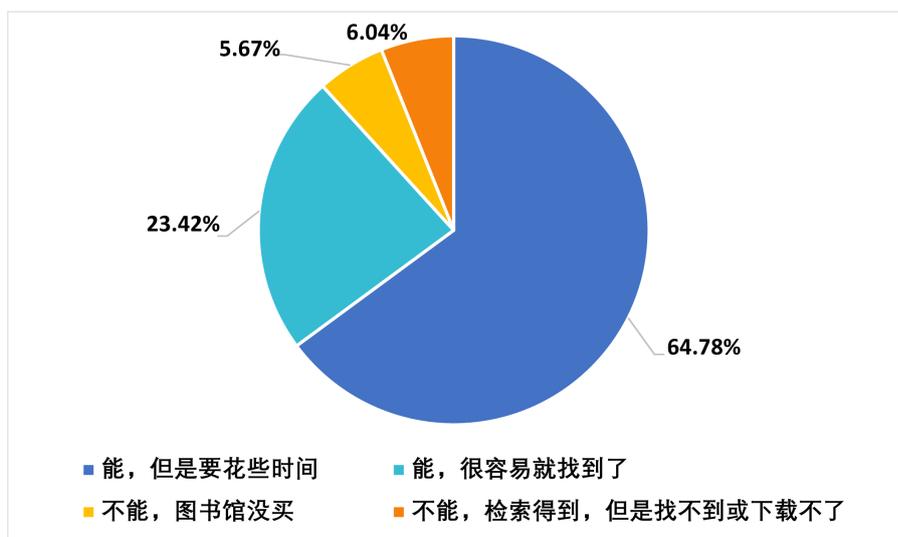


Figure 1. Investigation results of the issue “Can you usually find the paper and electronic books and periodicals you need through the library?”

图 1. “您通常能通过图书馆找到所需的纸本和电子书刊吗？”问题调查结果

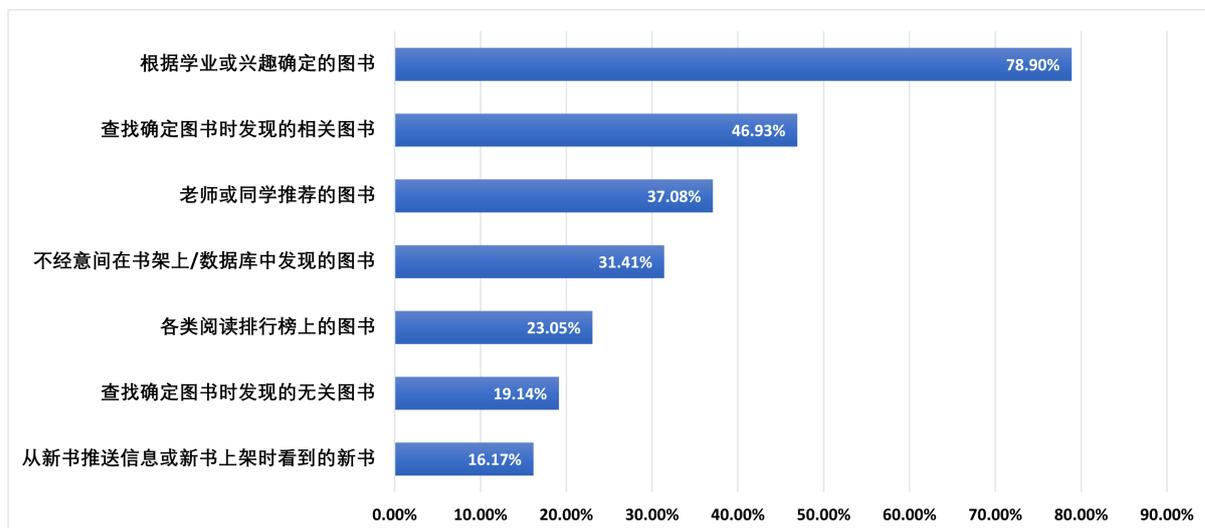


Figure 2. Investigation results of the issue “How do you usually choose to borrow or download paper books and e-books?”
图 2. “您平时怎么选择借阅、下载的纸本、电子图书？”问题调查结果

2. 读者图书推荐算法研究现状

当前，图书推荐算法主要包括基于内容的推荐[1]、关联规则[2]和协同过滤算法[3]等。基于内容推荐算法高度依赖于抽取和表示图书的内容特征，基于关联规则算法通过挖掘用户行为共现模式实现推荐，这两类方法需要对图书内容本身直接分析，限制了为读者提供潜在但未被明确表达的、甚至跨领域的兴趣点图书的推荐可能。而协同过滤算法通过对读者行为模式的相似性，可以提供更个性化的推荐，因此，成为主流的图书推荐算法[3]-[5]。

协同过滤算法通过用户或物品相似性进行推荐，但基于邻域的协同过滤算法依赖于显式评分数据，在用户行为稀疏的场景下容易产生偏差，不能有效地解决数据稀疏和冷启动问题；而基于模型的协同过滤算法虽然降低了计算复杂度，但对隐式反馈的适应性较差。矩阵分解方法通过降维技术将用户-物品交互矩阵映射到潜在空间，Koren 提出的隐语义模型(Latent Factor Model, LFM)通过奇异值分解(Singular Value Decomposition, SVD)将评分矩阵分解为用户潜在向量和物品潜在向量的内积，但其线性特性限制了对复杂关系的建模能力[6]。

随着数据规模的扩大，传统方法的局限性逐渐显现。在处理隐式反馈时，传统矩阵分解对负样本的建模能力不足，而基于采样的负例构造策略易引入噪声；在非线性的特征交互场景中，用户的多模态行为难以被线性潜在空间充分表征。近年深度学习技术为推荐系统带来了突破，因子分解机(Factorization Machine, FM)的提出首次将特征交叉引入推荐模型，但其仅关注两两特征间的乘积关系，对高阶交叉和深层非线性关系的建模能力有限[7]。2017年，Guo 等人提出深度因子分解机(Deep Factorization Machine, DeepFM)，通过并行结合 FM 与深度神经网络(Deep Neural Networks, DNN)，兼顾低阶特征交互和高阶非线性关系，在点击率预测任务中取得显著提升[8]。随后，DeepFM 被用于图书推荐系统设计，得到了较好地效果[9]。

但 DeepFM 算法过度依赖历史行为会导致推荐结果同质化，推荐图书地多样性不足，会限制读者知识视野，降低读者探索兴趣。因此，本文引入最大边缘相关(Maximal-Marginal-Relevance, MMR)算法[10]，通过平衡相关性与多样性改进 DeepFM 算法得到地图书推荐表，提出了 MMR-DeepFM 算法，优化多样性策略，在保证图书推荐效果的同时，实现图书推荐多样性，从而更有效地提高图书馆图书资源利用率。

3. 基于 MMR-DeepFM 模型地图书推荐设计

本文以中国海洋大学信息学部部分活跃读者的阅读记录为数据集，引入基于读者阅读时间的评分来优化生成结果，训练生成 MMR-DeepFM 深度学习模型。具体设计包括读者数据预处理、融合 MMR 算法的 DeepFM 模型构建和对推荐列表优化和算法评估(图 3)。

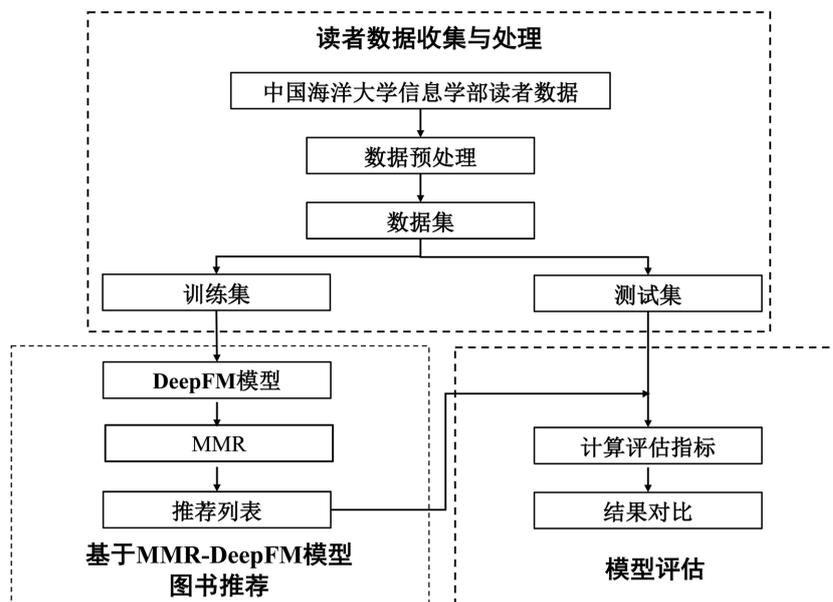


Figure 3. Schematic diagram of the book recommendation algorithm based on the MMR-DeepFM model
图 3. 基于 MMR-DeepFM 模型地图书推荐算法示意图

3.1. 数据预采集及预处理

本文数据集包含中国海洋大学信息科学与工程学部读者 1300 位活跃读者图书借阅数据。每条用户的借阅数据涵盖索书号、条码号、财产号等实体文献标识信息，题名、责任者等书目元数据，借阅时间、续借次数、还书日期等时间序列行为特征，以及借阅地点、馆藏归属等空间维度属性，是一个上下文信息丰富的数据集。

中国海洋大学图书馆主要使用中图法作为索书号分类方法，少部分书籍使用了科图法作为分类方法。因此数据集中数据同时包括了中图法分类号和科图法分类号，为了避免后续相同类型图书索书号冲突情况发生，本文基于科图法和中图法的类目设置表，创建了科图法与中图法映射字典，将使用频率较少的科图法分类号转换成中图法分类号。

在推荐算法中，用户对物品的显式评分通常作为重要的量化指标参与模型训练。因为借阅时长与兴趣强度呈正相关，长期持有某书籍可能反映深度阅读需求，高频次借阅同作者或同主题书籍构成关联规则，因此，本文通过基于借阅时间获取读者对书籍偏好信息，如公式 1 所示，式中 BorrowDate 是读者借书日期，ReturnDate 是读者归还日期，DueDate 是读者应还日期，Ratio 作为评价读者对图书喜爱程度评分，经过尺度扩展后范围为[0, 10]，可以更敏感地区分用户偏好细微差别。通过计算读者实际借阅时长(归还日期减借阅日期与允许借阅时长比率)，量化读者对书籍潜在兴趣强度。

$$\text{Ratio} = \frac{\text{ReturnDate} - \text{BorrowDate}}{\text{DueDate} - \text{BorrowDate}} \quad (1)$$

3.2. DeepFM 推荐算法

DeepFM 是一种融合因子分解机与深度神经网络的混合推荐模型，能够同时捕捉低阶特征交互和高阶非线性关系。DeepFM 模型由输入层、嵌入层、FM 部分、DNN 部分和输出层五部分组成，兼顾模型的表达能力和计算效率，模型结构如图 4 所示。通过融合 FM 与 DNN，兼顾低阶线性关系与高阶非线性交互，显著提升了 DeepFM 模型推荐性能。在实际应用中，DeepFM 可灵活扩展至多特征输入，并可通过调整嵌入维度、隐藏层数等网络结构适应不同任务需求[11]。研究表明 DeepFM 在点击率预测和 Top-N 推荐任务中均优于传统模型，成为工业界与学术界广泛采用的基准模型之一[12]。

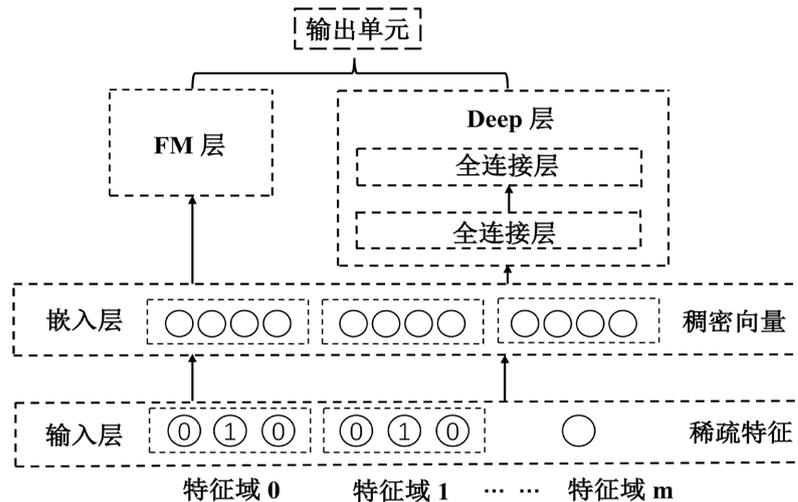


Figure 4. Schematic diagram of the DeepFM model structure [11]

图 4. DeepFM 模型结构示意图[11]

3.3. MMR 算法

MMR 是一种在信息检索和文本处理中平衡结果相关性与多样性的方法，其核心思想是通过动态权衡候选内容与用户查询的匹配度及已选内容的冗余度，优先选择既相关又新颖的信息。算法为每个候选项评分，一方面衡量其与查询内容相似性以保障相关性，另一方面惩罚其与已选内容高相似性以避免重复。该算法广泛应用于搜索引擎结果去重、文本摘要生成和推荐系统等场景[13]。

在模型预测读者对图书种类的推荐列表后，使用 MMR 对得分进行重新排序，确保推荐列表中图书种类间具有一定的差异性，避免内容重复或同质化，生成最终推荐列表[14]。根据数据集已知图书种类总数、嵌入矩阵、相关性分数向量以及推荐列表长度生成候选集，根据相关性分数对图书种类降序排序，生成初始候选索引列表，再排除读者已交互的图书种类索引，得到有效候选集。再初始化一个空的已选图书种类集合，进入迭代流程，只要已选图书种类数量尚未达到推荐列表长度且候选集不为空，便开启遍历。对候选集中每一类图书种类，运用嵌入向量点积相似度，计算 MMR 得分。从中筛选出 MMR 得分最高图书种类，将其纳入已选图书种类集合，并从候选集中移除。循环操作直至满足条件，最终将已选图书种类集合作为最终推荐列表输出。

3.4. 算法评估

对于推荐算法性能评估，主要考虑的指标有准确率(Precision)和召回率(Recall) [15]。准确率用于衡量推荐算法生成的推荐图书列表中，真正符合用户偏好的物品占比情况。定义如公式 2 所示。其中 $R(u)$ 为

推荐算法在训练集上对读者 u 产生的推荐图书种类集合, $T(u)$ 为读者 u 在测试集上有过借阅行为的图书种类集合。

$$\text{precision} = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |R(u)|} \quad (2)$$

召回率反映了测试集读者所有有过借阅行为的图书种类中成功被纳入推荐列表的图书种类所占的比例, 其定义如公式 3 所示。召回率功能和准确率大体相同, 均是直观评估推荐算法优劣的重要指标, 二者的差异在于评估视角不同, 召回率从另一个维度对推荐算法的质量进行量化分析, 体现的是推荐算法成功发掘读者兴趣的能力。

$$\text{recall} = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |T(u)|} \quad (3)$$

在实际应用中, 推荐算法输出的推荐数图书种类量与测试集中读者实际产生行为的图书种类数量, 可能出现严重失衡的极端情况。当算法为每个读者生成的推荐图书种类数量显著多于测试集中读者平均有行为的图书种类数量时, 准确率指标往往会偏低, 此时召回率能够更有效地评估推荐算法的优劣; 反之, 若推荐图书种类数量远少于用户实际产生行为的图书种类数量, 召回率数值通常较低, 准确率则成为衡量算法质量的关键指标。由此可见, 单独依赖准确率或召回率, 都难以全面、客观地评价推荐算法的性能。因此在推荐算法的评估过程中同时使用准确率和召回率这两个指标, 通过综合分析二者的数值表现, 才能更准确地反映算法实际的推荐质量[16]。因此引入指标 F1 定义如公式 4 所示。

$$F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (4)$$

对于 MMR 算法对结果的优化效果评估, 主要考虑的指标有列表内相似性(Intra-List Similarity, ILS)和覆盖率(Coverage) [17]。

ILS 衡量的是推荐列表中图书种类间的平均相似度, 值越低表示推荐列表的多样性越好。ILS 的定义如公式 5 所示, 其中 U 为所有读者集合, $\text{sim}(i, j)$ 为图书种类 i 和 j 使用余弦相似度计算出的相似度。

$$\text{ILS} = \frac{1}{|U|} \sum_{u \in U} \left(\frac{\sum_{i \in R(u)} \sum_{j \in R(u), j \neq i} \text{sim}(i, j)}{|R(u)| \cdot (|R(u)| - 1)} \right) \quad (5)$$

覆盖率反映推荐系统能够覆盖的图书种类占所有图书种类的比例, 值越高表示推荐范围越广。其定义如公式 6 所示, 其中 I 为所有图书种类集合, $\bigcup_{u \in U} R(u)$ 为所有读者推荐图书种类的并集。

$$\text{Coverage} = \frac{|\bigcup_{u \in U} R(u)|}{|I|} \quad (6)$$

4. 实验结果及分析

本文将预处理后的读者数据集按照 80%训练集和 20%测试集比例进行划分。这种划分策略尤其适用于帮助更多样本模型捕捉数据中潜在规律, 有效缓解因训练不足导致的欠拟合问题, 使其拥有能够有效应对小数据集和复杂模型情况。采用训练集训练 DeepFM 算法模型并进行测试。同时, 本文还选取了基于用户的协同过滤算法(User-Based Collaborative Filter, UserCF), 基于物品的协同过滤算法(Item-Based Collaborative Filter, ItemCF), 稀疏线性推荐算法(Sparse Linear Methods, SLIM)和广义矩阵分解算法(Generalized Matrix Factorization, GMF)与 DeepFM 算法及融合了 MMR 的 DeepFM 算法进行比较。UserCF 和

ItemCF 是典型的基于邻域的协同过滤推荐算法, SLIM 是基于机器学习 - 坐标下降法的推荐算法, GMF 是典型的基于神经网络的协同过滤推荐算法。实验结果如表 1 所示。

Table 1. Algorithm comparison experiment results

表 1. 算法对比实验结果

算法	precision	recall	F1	ILS	Coverage
MMR-DeepFM	28.40%	38.24%	32.60%	0.0300	0.4211
DeepFM	23.69%	57.52%	34.47%	0.1094	0.1789
GMF	25.32%	47.51%	29.89%	0.8559	0.2684
SLIM	16.20%	43.90%	20.80%	0.0249	0.3842
ItemCF	18.88%	22.54%	19.38%	0.5597	0.3842
UserCF	17.76%	23.31%	18.84%	0.2916	0.2632

表 1 中 F1 结果表明, 相对其它算法 DeepFM 模型图书推荐效果最好。在引入 MMR 算法后, MMR-DeepFM 算法的 F1 有所下降, 但 ILS 和 Coverage 有所上升, 其根本原因在于 MMR 对候选图书种类排序机制进行了目标函数重构, 通过线性加权组合原始相关性分数与最大相似度惩罚项动态调整图书种类在推荐序列中的优先级。这种重排序过程本质上以牺牲部分短期准确性为代价, 强制降低列表内图书种类同质性, 以促进高多样性特征图书种类进入列表。

在 GMF、SLIM、Item CF 和 UserCF 四种传统算法中, GMF 可以达到较好的效果, 但 GMF 算法 ILS 值最大, Coverage 较小, 表明其图书推荐列表中图书种类相似度较大, 且覆盖种类较少, 此类推荐算法会限制读者搜索范围和视野, 不能为读者提供多样化推荐, 限制了图书馆图书的利用率和流通性。虽然 SLIM 算法的 ISL 最小, 可以较好地保证多样化图书推荐, 但其 F1 比 Deep FM 和 MMR-DeepFM 算法 F1 低 13.67%和 11.8%, 推荐准确性较差。本文提出的 MMR-DeepFM 算法 F1 仅比 DeepFM 算法低 1.87%, 但其 ILS 值仅为 0.03, Coverage 最大。这表明 MMR-DeepFM 算法能够在保证图书推荐效果的同时, 为读者推荐更多种类图书, 因此, 采用本文提出的算法可以有效优化图书多元化推荐策略。

5. 结论

针对高校图书馆资源利用和读者个性化借阅需求, 本文提出了 MMR-DeepFM 图书推荐算法, 并采用中国海洋大学信息科学与工程学部读者数据进行了实验和分析, 实验结果证明了本文算法可以在准确为读者提供图书推荐的同时覆盖更多图书种类, 因此, MMR-DeepFM 图书推荐算法能够为读者提供多样化推荐服务, 有效提升图书借阅率, 提高图书馆图书资源利用率。

基金项目

中央高校基本科研业务专项(“数据驱动下读者个体与群体阅读倾向可解释性分析与预测研究——以信息科学与工程学部读者为例”, 项目编号: 202553002)。

参考文献

- [1] 张麒麟, 姜霖. 基于文献内容的图书推荐机制研究[J]. 图书馆学研究, 2018(1): 78-81, 17.
- [2] 王丽兰. 基于 Apriori 算法的高职院校图书精准推荐系统[J]. 河北软件职业技术学院学报, 2023, 25(3): 6-9.
- [3] 张德青, 程锦. 高校图书馆个性化图书推荐算法研究[J]. 西昌学院学报(自然科学版), 2021, 3(52): 78-81.
- [4] 邢立宁, 孙进强, 谭旭. 基于协同过滤和 TransH 改进的图书智能推荐算法[J]. 深圳信息职业技术学院学报, 2024,

- 22(3): 1-6.
- [5] 马鸿健, 张耘凡, 陈金, 部洪超. 高校图书推荐系统设计与实现[J]. 计算机应用, 2021(12): 49-52.
 - [6] Koren, Y., Bell, R. and Volinsky, C. (2009) Matrix Factorization Techniques for Recommender Systems. *Computer*, **42**, 30-37. <https://doi.org/10.1109/mc.2009.263>
 - [7] Rendle, S. (2010) Factorization Machines. 2010 *IEEE International Conference on Data Mining*, Sydney, 13-17 December 2010, 995-1000. <https://doi.org/10.1109/icdm.2010.127>
 - [8] Guo, H., Tang, R., Ye, Y., *et al.* (2017) DeepFM: A Factorization-Machine Based Neural Network for CTR Prediction. arXiv:1703.04247.
 - [9] 肖伦山. 基于 DeepFM 混合模型的高校图书推荐系统的设计与实现[D]: [硕士学位论文]. 重庆: 西南大学, 2022.
 - [10] Wu, C.H., Wang, Y. and Ma, J. (2021) Maximal Marginal Relevance-Based Recommendation for Product Customisation. *Enterprise Information Systems*, **17**, Article 1992018. <https://doi.org/10.1080/17517575.2021.1992018>
 - [11] 胡克富, 王卫东. 基于编码器增强的 DeepFM 推荐模型[J]. 计算机与数字工程, 2025, 53(1): 158-163.
 - [12] Deng, W., Pan, J., Zhou, T., Kong, D., Flores, A. and Lin, G. (2021) Deeplight: Deep Lightweight Feature Interactions for Accelerating CTR Predictions in Ad Serving. *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, Virtual Event Israel, 8-12 March 2021, 922-930. <https://doi.org/10.1145/3437963.3441727>
 - [13] 李响, 谭静. 融合相关性与多样性的学术论文推荐方法研究[J]. 情报理论与实践, 2017, 40(6): 99-103.
 - [14] Wang, Y., Ma, W., Zhang, M., Liu, Y. and Ma, S. (2023) A Survey on the Fairness of Recommender Systems. *ACM Transactions on Information Systems*, **41**, 1-43. <https://doi.org/10.1145/3547333>
 - [15] Herlocker, J.L., Konstan, J.A., Terveen, L.G. and Riedl, J.T. (2004) Evaluating Collaborative Filtering Recommender Systems. *ACM Transactions on Information Systems*, **22**, 5-53. <https://doi.org/10.1145/963770.963772>
 - [16] 刘君良, 李晓光. 个性化推荐系统技术进展[J]. 计算机科学, 2020, 47(7): 47-55.
 - [17] Vargas, S., Baltrunas, L., Karatzoglou, A. and Castells, P. (2014) Coverage, Redundancy and Size-Awareness in Genre Diversity for Recommender Systems. *Proceedings of the 8th ACM Conference on Recommender systems*, Foster City, 6-10 October 2014, 209-216. <https://doi.org/10.1145/2645710.2645743>