

# 融合注意力机制的改进神经协同过滤模型

田晓婧, 谢颖华

东华大学信息科学与技术学院, 上海

收稿日期: 2021年10月24日; 录用日期: 2021年11月22日; 发布日期: 2021年11月29日

## 摘要

注意力机制如今被广泛的应用到各个领域, 注意力机制的引入可以使模型的不同部分占据不同的权重从而为最终预测做出不同的贡献。本文基于何向南博士提出的神经协同过滤模型做出改进, 在此模型基础上融合长短期记忆网络与广义矩阵分解以捕获用户的长期偏好和短期偏好, 同时引入注意力机制, 关注用户属性, 例如性别, 年龄, 职业等对推荐效果的影响, 为用户属性分配不同的权重, 通过用户历史观看序列和用户个人信息进行学习来提升推荐模型的性能。本文基于提出的模型, 在MovieLens-1M数据集上进行实验验证, 并与其他推荐模型进行对比。实验证明本文提出的融合注意力的长短期记忆网络矩阵分解模型(ALSMF)拥有更好的推荐效果。

## 关键词

注意力机制, 协同过滤, 矩阵分解, 长短期记忆网络

# Improved Neural Collaborative Filtering Model Based on Attention Mechanism

Xiaojing Tian, Yinghua Xie

College of Information Science and Technology, Donghua University, Shanghai

Received: Oct. 24<sup>th</sup>, 2021; accepted: Nov. 22<sup>nd</sup>, 2021; published: Nov. 29<sup>th</sup>, 2021

## Abstract

The attention mechanism is widely used in various fields. The introduction of the attention mechanism can make different parts of the model occupy different weights to make different contributions to the final prediction. This article is based on the neural collaborative filtering model proposed by Doctor He. On the basis of this model, the long and short-term memory network and generalized matrix decomposition are combined to capture the long-term and short-term preferences of users. At the same time, the attention mechanism is introduced to focus on user's attributes, such as gender, age, occupation, etc. on the recommendation effect, different

weights are assigned to user attributes, and the performance of the recommendation model is improved by learning from the user's historical viewing sequence and user personal information. This article is based on the proposed model. We use the MovieLens-1M dataset to verify the new model and compare this model, we propose with other recommendation models. Experiments prove that the attention-integrated long and short-term memory network matrix factorization model (ALSMF) proposed in this paper has better recommendation effects.

## Keywords

Attention Mechanism, Collaborative Filtering, Matrix Factorization, Long Short-Term Memory Networks

Copyright © 2021 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

在如今的个性化时代, 推荐算法无疑是当今算法领域的一大热点。它被应用在电商, 新闻等各个领域, 在这些领域中也发挥着至关重要的作用。近年来, 深度学习技术迅速发展, 很多推荐算法模型都把深度学习技术与协同过滤模型相结合。

新加坡国立大学的何向南博士首次提出了神经协同过滤模型(Neural Collaborative Filtering, NCF) [1]。文中何博士在此框架下融合了广义矩阵分解(GMF)和多层感知机模型(MLP), 前者通过利用线性内核来建模用户与项目潜在特征向量之间的交互关系, 后者则是使用非线性内核学习交互函数, 进一步优化推荐模型, 此模型被命名为神经矩阵分解模型(Neural Matrix Factorization, NMF)。针对 NMF 无法同时捕获用户长期偏好和短期偏好的问题, 论文[2]提出在神经协同过滤模型的框架下, 将长短期记忆网络和广义矩阵分解进行融合, 同时捕捉用户的短期偏好和长期偏好。利用长短期记忆网络对时序数据的强拟合能力, 学习用户的短期偏好信息, 捕捉序列的长依赖关系, 通过广义矩阵分解学习用户的长期偏好信息, 这对推荐性能有了进一步的提高。

近年来, 注意力机制被广泛应用到深度学习模型中, 针对目标项目的不同进行更有针对性的推荐。2017年, 浙江大学提出了注意力因子分解机(AFM) [3], 在因子分解机的基础上, 对每个特征交叉结果加入注意力得分, 并利用注意力网络计算注意力得分。但是这个注意力网络的训练过程比较复杂。2019年阿里巴巴提出了深度兴趣网络(DIN) [4], 在传统深度学习推荐模型基础上引入注意力机制, 利用用户行为历史和目标项目的相关性计算注意力得分。

综合深度学习和注意力机制在推荐算法模型上的优良表现, 本文在 LSMF 模型上进行进一步的改进, 由于 LSMF 没有充分利用除了用户行为序列以外的特征, 对于用户的个人信息特征没有做提取, 浪费了很多已知的关键信息, 导致推荐精度不够。因此在 LSMF 模型的基础上, 本文提出了融合注意力机制的 LSMF 模型(ALSMF)。通过引入注意力机制, 为各个项目和特征分配注意力权重来更好的表达推荐场景中用户的个性化喜好, 提高推荐质量和推荐的可解释性。

## 2. 相关知识

### 2.1. 神经协同过滤模型

神经协同过滤模型利用隐式反馈数据, 即不关注具体评分, 只关注用户与项目是否有交互行为, 利

用深度神经网络结构来学习任意交互函数, 为每个用户生成一个推荐列表, 其框架结构如图 1 所示。NCF 将传统矩阵分解模型中的内积操作替换为“多层神经网络 + 输出层”的结构, 这样可以使用户和项目向量之间进行更充分的交叉, 从而获得更多的特征组合信息, 而这种非线性特征的加入也将使模型的表达能力获得提升[5]。

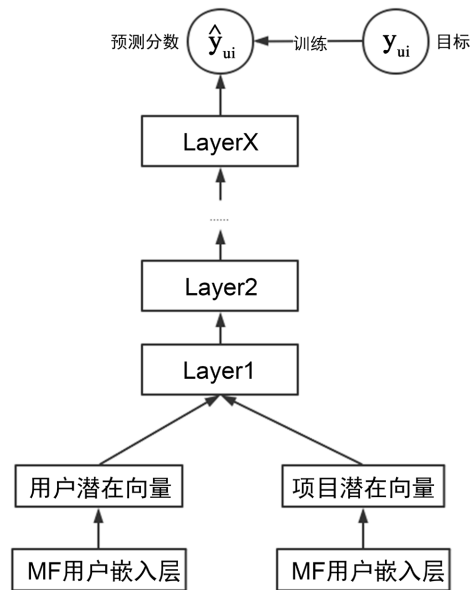


Figure 1. NCF framework

图 1. NCF 框架

## 2.2. 注意力机制

注意力机制(Attention Mechanism) [6] [7]在神经网络中有着非常优秀的表现, 在众多领域的研究中都非常重要。追溯其根源, 注意力机制借鉴了人类独有的快速筛选能力, 简单说便是在繁杂的信息中极为迅速地找出价值高的信息, 这是通过不断学习获得的成果, 可以放大研究部分的细节, 使得观察更加精细。例如人在超市中寻找需要的货品型号, 会迅速地观察全部区域, 找到区域中的重要部分, 之后对这一部分继续使用注意力寻找, 来得到更多的与所需物品的细致信息, 在此同时还会自动忽略其它的干扰信息。

在翻译和自动文摘中, 只有输入序列中特定的某些单词可能对预测下一个单词有帮助。注意力机制通过允许模型动态地聚焦于输入的特定部分来提高任务的效率而将这种相关性概念结合起来。特别在比较长的序列之中, 注意力机制发挥的作用更加重大, 可以在长序列中找到其中的重点, 增加其预测效率与准确率。它们对神经网络的研究也有重要贡献, 在此之前一直将神经网络看作为一个黑盒, 并不知道其中的过程, 而注意力机制可以在一定程度解释神经网络, 提高对机器学习的透明性和公平性的认知[8]。

使用注意力机制的模型是对人脑注意力的模仿, 参考人脑将注意力集中在重要部分, 在此部分投入更多的注意力, 降低在其他部分的注意力投入, 利用有限的资源, 更加高效地寻找到深度学习中起作用的重要因素, 与此同时减少其它因素的不利影响[9]。

注意力机制大多作为深度学习模型中的一个组成部分, 负责捕获数据的相互依赖性和重要性, 进行重新组合后得到更具有代表性和泛化性的数据。

注意力机制权重计算的基本公式如下[10]:

$$e_{i,j} = a(H_i, H_j) \quad (1)$$

$$a_{i,j} = \frac{\exp(e_{i,j})}{\sum_{k=1}^H \exp(e_{i,k})} \quad (2)$$

$$C_i = \sum_{j=1}^H a_{i,j} H_j \quad (3)$$

其中,  $H_i$  表示嵌入向量中的第  $i$  个向量,  $a$  表示注意力分布函数,  $C_i$  表示经过注意力机制处理后的第  $i$  个新向量。

### 3. 融合注意力机制的改进神经协同过滤模型

#### 3.1. 实验数据集

本文选择 MovieLens-1M 作为实验数据集, 其中包含 users.dat, movies.dat 和 ratings.dat 三个数据文件。其中 ratings.dat 拥有来自 6040 位在 2000 年加入 MovieLens 的用户对 3706 部电影的 100 多万条从 1 星到 5 星评级信息, 其中每个用户评级过的电影数据均大于 20。文件 users.dat 中含有 6040 个用户的基本特征属性, 包括用户 id, 性别, 年龄, 职业类别, 所在地区邮编的信息。

#### 3.2. 数据预处理

对数据集进行预处理, 提取出每一次交互中的用户的 id, 年龄, 性别, 职业, 最近十次的观看历史的电影 id, 下一次预测的电影 id, 以及预测评分。模型采用隐式数据, 即所有与用户有交互的项目预测评分都设置为 1。本文选择利用留一法(leave-one-out)进行实验验证。把每个用户最新的评分信息作为测试集, 把其他的交互信息作为训练集。同时为每一次正交互对应 4 个负样本, 即 4 个没有与用户有过交互行为的电影, 这些项目评分设置为 0。为了避免对所有项目排序造成的耗时过长, 所以实验是从所有的项目样本中随机选择 99 个没有评分的电影, 然后将测试集中的电影与这些电影一起输入到模型进行排序。交互信息取每个用户近十次的电影 id, 不足 10 次则添 0。

#### 3.3. 模型框架

融合注意力机制的长短期记忆网络矩阵分解模型(简称 ALSMF)是在神经协同过滤模型的基础上, 将注意力机制, 广义矩阵分解和长短期记忆网络相结合, 除了同时捕获用户的长期偏好和短期偏好以外, 还可以通过用户历史观看序列和用户个人信息进行学习来提升推荐模型的性能。如图 2 所示, 左侧模型利用注意力机制针对不同的用户属性对于推荐结果的影响, 赋予它们不同的权重来凸显用户偏好关系。右侧模型通过长短期记忆网络捕捉用户历史记录中的短期偏好。引入注意力机制的核心思想是在原有的广义矩阵分解模型的基础上加上用户个性化信息向量以此对用户特征向量进行重构。ALSMF 模型框架如图 2 所示。

##### 1) 嵌入层

本文利用按照时间戳排序后的用户和项目 id 作为模型的输入, 而模型根据实际可以有不同的输入。首先模型将用户 id 和项目 id 进行 one-hot 编码, 然后将这种稀疏向量映射成稠密向量作为用户和项目的特征向量。将用户历史交互记录进行 one-hot 编码, 作为用户交互特征向量输入到 LSTM 层。将用户的性别年龄, 职业输入作为用户特征属性输入, 同样进行 one-hot 编码后, 输入到 Attention 层。

##### 2) LSTM 模块

模型通过 LSTM 来捕获用户和项目的时序依赖性, 对时序信息循环计算以挖掘上下文关系, 通过构造用户状态和电影状态进行自回归计算, 对推荐系统中的用户行为序列进行分析。

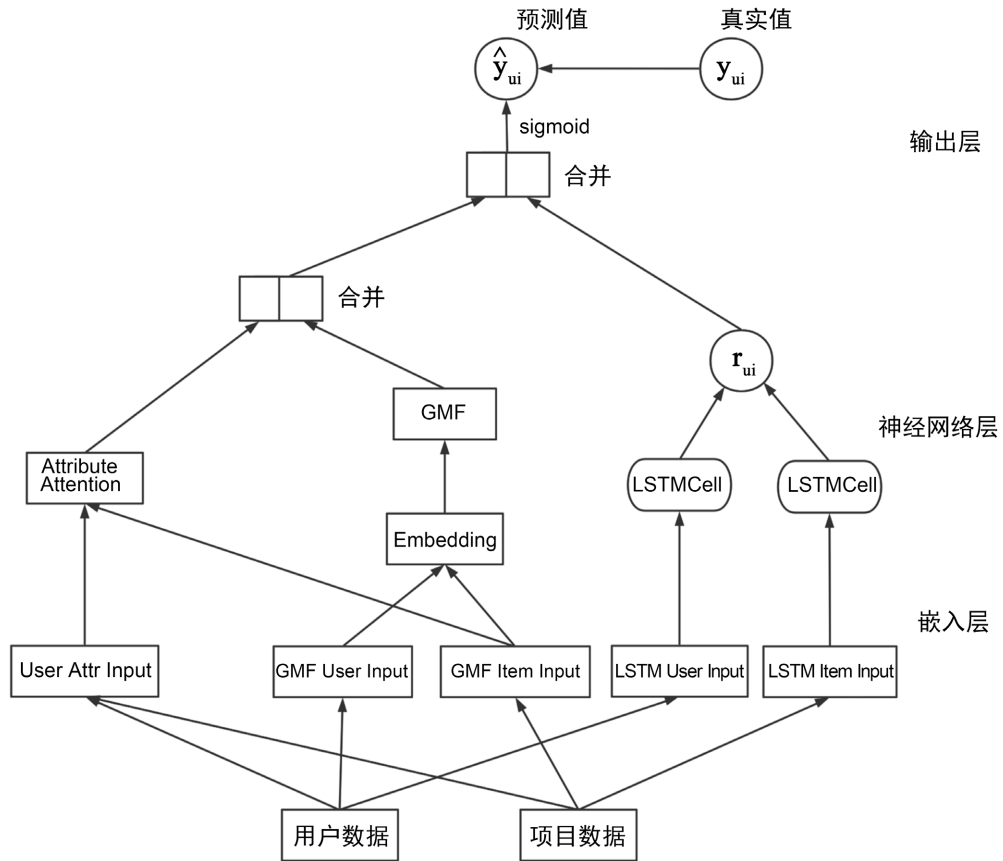


Figure 2. ALSMF model  
图 2. ALSMF 模型

模型通过上一时刻的用户和项目状态预测此时的用户和电影状态[11], 然后预测此时用户和项目进行交互地可能性。

$$u_{u,t+1} = g(u_u, \{r_{ui|t}\}) \tag{4}$$

$$m_{i,t+1} = h(m_i, \{r_{ui|t}\}) \tag{5}$$

基于用户评分行为预测评分问题表示为:

$$\hat{r}_{ui|t} = f(u_u, m_i) \tag{6}$$

式中,  $r_{ui|t}$  和  $\hat{r}_{ui|t}$  分别表示用户  $u$  在时刻  $t$  对电影  $i$  的实际评分和预测评分。

### 3) GMF

通过嵌入层得到了用  $P^T V_u^U$  表示的用户  $p_u$  的特征向量,  $Q^T V_i^I$  表示的项目的特征向量  $q_i$ , 然后将神经协同过滤模型中的第一层的映射函数[12]定义为:

$$\phi_1(p_u, q_i) = p_u \odot q_i \tag{7}$$

上式中  $\odot$  表示向量的点积, 然后将向量映射到输出层。

$$\hat{y}_{ui} = a_{out}(h^T(p_u \odot q_i)) \tag{8}$$

其中  $a_{out}$  表示输出层的激活函数,  $h$  表示边缘权重。

#### 4) 注意力模块

将用户的 id, 性别, 年龄, 职业等信息进行 one-hot 编码后, 输入到 Attention 网络中, 自动捕捉用户关键属性, 并计算每个属性的个性化权重。利用用户的属性信息来丰富用户的潜在向量表示。注意力网络定义如下:

$$a(p_u, q_{attr_n}) = \text{softmax}\left(\text{Relu}\left(w(p_u \odot q_{attr_n}) + b\right)\right) \quad (9)$$

其中,  $w$  和  $b$  分别为权重矩阵和偏置向量,  $\text{softmax}$  为激活函数输出注意力权重,  $a(p_u, q_{attr_n})$  表示用户  $u$  对每个属性的注意力系数。

#### 5) 输出层

注意力模块的输出向量与 GMF 的输出向量共同得到用户的长期偏好信息, 然后将此向量与 LSTM 层输出的向量进行拼接。利用 sigmoid 作为激活函数, 使用对数损失函数(log loss)学习  $h$  进行优化。

$$\hat{y}_{ui} = \sigma\left(h^T [\phi_{\text{ATT-GMF}}, \phi_{\text{LSTM}}]\right) \quad (10)$$

## 4. 实验结果及分析

### 4.1. 评估标准

本实验选择留一法(leave-one-out)进行实验验证。即取用户交互的最新信息作为测试集, 其他历史交互信息作为训练集。为了避免对所有项目排序造成的耗时过长, 所以实验是从所有的项目样本中随机选择 99 个没有评分的电影, 然后将测试集中的电影与这些电影一起输入到模型进行排序。

本文将命中率(Hits Ratio, HR)和归一化折损累积增益(Normalized Discounted Cumulative Gain, NDCG) [13]作为性能评估标准。

$$HR@K = \frac{\text{NumberOfHits}@K}{|GT|} \quad (11)$$

$$NDCG@K = Z \sum_{i=1}^K \frac{2^{x_i} - 1}{\log(1+i)} \quad (12)$$

在式(11)中,  $GT$  所表示的是所有用户测试集合中的项目总数, 分子表示推荐列表中属于其测试集合的项目总数。式(12)中,  $Z$  是归一化因子, 表示理想推荐列表的折损累计增益。  $x_i$  是第  $i$  个项目的预测相关性。两式中  $K$  表示排名个数, 本文取  $K = 10$ 。

### 4.2. 实验结果及分析

本文对四种不同的模型进行对比, 分别为神经网络矩阵分解模型(NMF), 深度兴趣网络(DIN)长短期记忆网络矩阵分解模型(LSMF)和融合注意力的长短期记忆网络矩阵分解模型(ALSMF)。将模型训练 50 轮后, 取不同模型的性能平均值, 得出结果如表 1 所示。图 3, 图 4 显示了 NMF, DIN, LSMF, ALSMF 的性能对比。

**Table 1.** Comparison of HR and NDCG for different models

**表 1.** 不同模型的 HR 和 NDCG 对比

|       | HR@10   | NDCG@10 |
|-------|---------|---------|
| NMF   | 0.70347 | 0.42392 |
| DIN   | 0.74586 | 0.49664 |
| LSMF  | 0.77996 | 0.52267 |
| ALSMF | 0.80976 | 0.57021 |



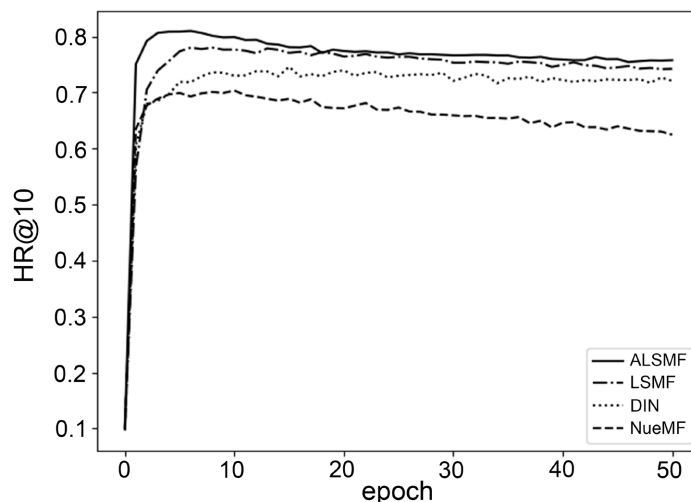


Figure 3. Hit ratio performance

图 3. 命中率表现

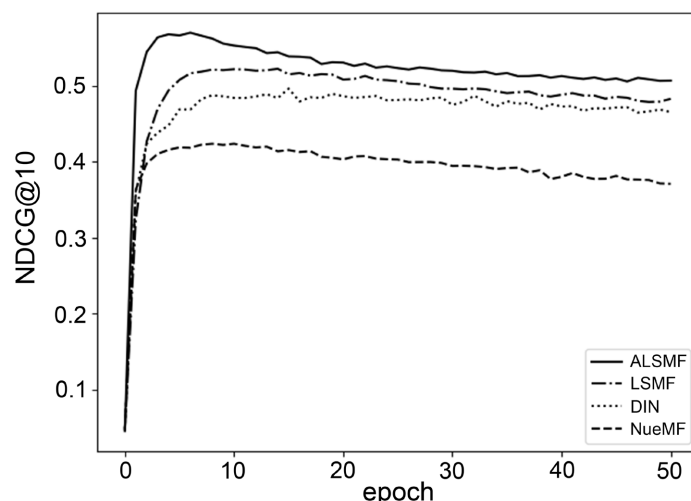


Figure 4. Normalized loss cumulative gain performance

图 4. 归一化折损累积增益表现

通过以上实验数据, 本文可以得出结论, ALSMF 模型在命中率和归一化折损累积增益两个评价指标上相对于另外三个模型都有明显的提高。NMF, DIN 和 LSMF 大约在在训练十轮后达到最优性能模型, 而 ALSMF 在训练第 5 轮时即达到最优性能模型。因此 ALSMF 的推荐性能相较于另外三个模型更加优秀。

## 5. 总结

本文首先对于推荐算法目前的研究情况做了分析和介绍, 并重点针对神经协同过滤模型和注意力机制做了比较详细的介绍。针对论文提出的长短期记忆矩阵分解模型没有充分利用用户的属性信息的问题, 本文提出了在此模型基础上引入注意力机制, 命名为融合注意力的长短期矩阵分解模型(ALSMF)。利用注意力机制自动捕获用户属性对推荐模型的影响, 赋予不同属性不同的权重。最终在 MovieLens-1M 数据集上进行实验验证, 对比 NMF, DIN 和 LSMF 模型, 证明本文提出的 ALSMF 在推荐性能上有较大的提升。

在接下来的工作中, 可以加入更加丰富的属性特征, 例如项目属性, 电影的类型, 主演人员等等, 从而挖掘更多的信息进一步提高模型性能。

## 参考文献

- [1] He, X.N., Liao, L.Z., Zhang, H.W., *et al.* (2017) Neural Collaborative Filtering. *Proceedings of the 26th International Conference on World Wide Web*, Pearce, 3-7 April 2017, 173-182. <https://doi.org/10.1145/3038912.3052569>
- [2] 田晓婧, 谢颖华. 融合长短期记忆网络与广义矩阵分解的神经协同过滤模型[J]. 计算机系统应用, 2021. (已录用)
- [3] Xiao, J., Ye, H., He, X.N., *et al.* (2017) Attentional Factorization Machines: Learning the Weight of Feature Interactions via Attention Networks. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, Melbourne, 19-25 August 2017, 3119-3125. <https://doi.org/10.24963/ijcai.2017/435>
- [4] Zhou, G.R., Zhu, X.Q., Song, C.R., *et al.* (2018) Deep Interest Network for Click-Through Rate Prediction. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, London, 19-23 August 2018, 1059-1068. <https://doi.org/10.1145/3219819.3219823>
- [5] 刘晶. 基于深度学习的协同过滤算法的研究[D]: [硕士学位论文]. 北京: 北京邮电大学, 2019.
- [6] 郑波. 基于属性注意力神经矩阵分解的推荐算法和应用[D]: [硕士学位论文]. 广州: 华南理工大学, 2020.
- [7] 尚庚. 基于注意力机制的深度协同过滤推荐算法研究[D]: [硕士学位论文]. 阜新: 辽宁工程技术大学, 2019.
- [8] 张全贵, 李志强, 张新新, 曹志强. 融合元数据及 attention 机制的深度联合学习推荐[J]. 计算机应用研究, 2019, 36(11): 3290-3293.
- [9] 王喆. 深度学习推荐系统[M]. 北京: 电子工业出版社, 2020.
- [10] 徐旻韬. 基于自注意力机制的长短期神经网络推荐模型研究[D]: [硕士学位论文]. 上海: 东华大学, 2021.
- [11] Wu, C.Y., Ahmed, A., Beutel, A., *et al.* (2017) Recurrent Recommender Networks. *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, Cambridge, 6-10 February 2017, 495-503. <https://doi.org/10.1145/3018661.3018689>
- [12] 李川. 基于改进神经网络协同过滤的研究[D]: [硕士学位论文]. 西安: 西安电子科技大学, 2019.
- [13] 武海龙. 基于卷积神经网络的协同过滤算法研究[D]: [硕士学位论文]. 武汉: 华中科技大学, 2019.