

# 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<sup>①</sup>



田晓婧, 谢颖华

(东华大学 信息科学与技术学院, 上海 201620)  
通信作者: 田晓婧, E-mail: 15001702826@163.com

**摘要:** 在实现推荐的过程中, 用户对项目的浏览和关注的时间顺序是推荐算法中重要的数据信息, 同一用户在不同时间对项目的喜好不同对推荐结果也有着一定的影响. 本文在神经协同过滤模型的框架下, 提出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, 从而优化推荐算法, 提高推荐性能. 通过 MovieLens-1M 数据集进行试验后, 结果表明, 本文提出的新模型在收敛速度和推荐性能方面都有提升.

**关键词:** 神经网络; 推荐算法; 神经协同过滤; 矩阵分解; 长短期记忆网络

引用格式: 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. 计算机系统应用, 2022, 31(1): 190-194. <http://www.c-s-a.org.cn/1003-3254/8281.html>

## Neural Collaborative Filtering Model Integrating LSTM and Generalized Matrix Factorization

TIAN Xiao-Jing, XIE Ying-Hua

(School of Information Science and Technology, Donghua University, Shanghai 201620, China)

**Abstract:** In the process of implementing recommendations, the browsing order of users is important information for the recommendation algorithm. The same user's different preferences for items at different times also affect the recommendation results. Under the framework of the neural collaborative filtering model, this study proposes to integrate long short-term memory networks with generalized matrix factorization and capture both the user's long-term and short-term preferences. The new model utilizes the strong fitting ability of long short-term memory networks to time series data to learn the user's short-term preference and capture the long dependence relationship of the sequence. The user's long-term preference is learned through generalized matrix factorization. The recommendation algorithm is thereby optimized, and the recommendation performance is improved. Experiments are carried out on the MovieLens-1M dataset and the results show that the new model has a higher convergence rate and better recommendation performance.

**Key words:** neural network; recommendation algorithm; neural collaborative filtering; matrix factorization; long short-term memory (LSTM)

近年来, 随着深度学习的飞速发展, 深度学习技术被大规模地运用到推荐系统领域<sup>[1]</sup>. 深度学习技术能够有效地捕获非线性和特殊性的用户和项目之间的关系,

并通过引用神经网络, 将其作为更高层中的数据表示, 来表达更复杂抽象的数据形式以克服数据稀疏性的问题<sup>[2]</sup>. 因此, 许多研究人员都通过将深度神经网络与传

① 收稿时间: 2021-03-18; 修改时间: 2021-04-16, 2021-05-07; 采用时间: 2021-05-11; csa 在线出版时间: 2021-12-17

统协同过滤算法结合,改善传统推荐算法的推荐性能。

2017年,He等首次提出了神经协同过滤模型(neural collaborative filtering, NCF)<sup>[3]</sup>,同时在此框架下融合了广义矩阵分解和多层感知机模型,广义矩阵分解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。多层感知机则是使用非线性内核学习交互函数,进一步优化推荐模型。

近几年,许多研究人员也根据NCF模型做出了改进以进一步提升推荐效果。例如刘晶<sup>[4]</sup>在2019年提出基于Inception-ResNet网络的外积交互协同过滤算法,将NCF中的神经网络层定义为Inception网络,复杂的Inception对特征的非线性映射能力更强,同时添加ResNet结构解决训练中梯度弥散的问题,使模型具有更好的效果。王骏等<sup>[5]</sup>在2020年提出将多层感知机与贝叶斯个性化排序算法融合,提取所需要的排序信息,使模型性能得到提高。纪强<sup>[6]</sup>在2020年提出基于循环神经网络的序列推荐模型,通过循环神经网络来学习用户交互的序列历史中的短期用户偏好。

然而这些都不能很好地同时捕捉用户的短期偏好和长期偏好,针对这个问题,本文提出融合长短期记忆网络和广义矩阵分解的神经协同过滤模型,这对实现更好的推荐性能有很重要的意义。

## 1 相关知识

### 1.1 神经协同过滤

神经协同过滤模型基于隐式反馈数据,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。

### 1.2 矩阵分解

推荐算法中传统的协同过滤算法存在着一定的局限性,例如头部效应较明显,泛化能力较弱<sup>[7]</sup>等问题。针对这个问题,矩阵分解(matrix factorization, MF)算法<sup>[8]</sup>被提出。MF在协同过滤算法中“共现矩阵”的基础上,加入了隐向量的概念,加强了模型处理稀疏矩阵的能力,解决了协同过滤存在的主要问题<sup>[9]</sup>。

矩阵分解算法为每一个用户和项目生成一个隐向量,然后将用户和项目定位到隐向量的表示空间上,距

离相近的用户和项目表明兴趣特点接近,在推荐过程中,距离相近的项目表明相似度更高,更应该被推荐给目标用户。

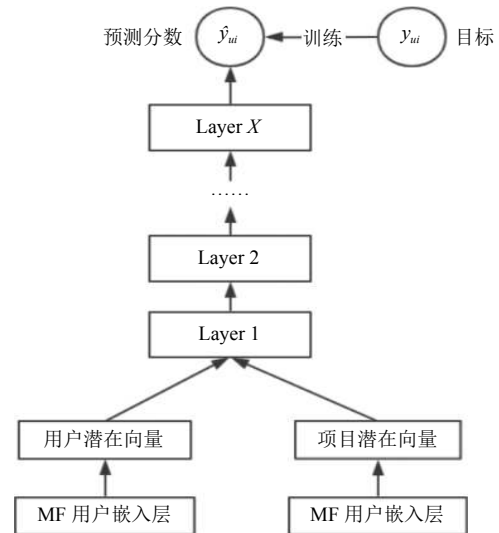


图1 NCF框架

计算用户  $u$  对项目  $i$  的预估评分如式(1)所示。

$$\hat{r}_{ui} = q_i^T p_u \quad (1)$$

其中,  $p_u$  是用户  $u$  在用户矩阵  $U$  中的对应行向量,  $q_i$  是项目  $i$  在项目矩阵  $V$  中的对应列向量。

式(2)是求解矩阵分解的目标函数,该目标函数的目的在于最小化实际评分  $r_{ui}$  与用户向量和项目向量之积  $q_i^T p_u$  的差,这样才能最大限度地保存共现矩阵地原始信息。

$$\min_{q^*, p^*} \sum_{(u,i) \in K} (r_{ui} - q_i^T p_u)^2 \quad (2)$$

### 1.3 长短期记忆网络

长短期记忆网络(long short-term memory, LSTM)<sup>[10]</sup>是针对循环神经网络(recurrent neural networks, RNN)产生的梯度消失和梯度爆炸问题而进行改进的,它的结构<sup>[11]</sup>如图2所示。

LSTM 状态更新满足以下公式:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (3)$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (4)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tanh(W_c[h_{t-1}, x_t] + b_c) \quad (5)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (6)$$

$$h_t = o_t \cdot \tanh(c_t) \quad (7)$$

其中,  $f_t$ 、 $i_t$ 、 $o_t$  分别表示遗忘门、输入门和输出门,  $c_t$  表示隐藏层神经元状态,  $x_t$  表示当前输入,  $\sigma$  是 Sigmoid 函数.  $[h_{t-1}, x_t]$  表示  $t-1$  时刻隐藏层的输出和  $t$  时刻的输入的连接向量,  $W$  和  $b$  分别表示权重矩阵和偏置.

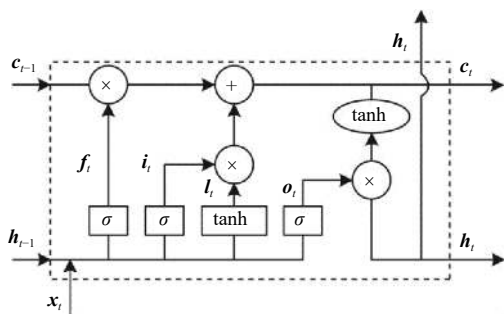


图2 LSTM 结构

## 2 长短期记忆网络矩阵分解模型

### 2.1 总体框架

每个用户对项目的浏览顺序是推荐系统中的一项重要数据, 随着时间的不同, 用户的喜好也会有着一定的变化. 然而, 传统的推荐算法默认用户的喜好是一成不变的, 仅仅对用户的偏好进行全局预测, 通过静态的方式建模用户与项目之间的交互, 这样忽略了用户兴趣动态变化的过程.

而在实际的推荐场景中, 由于用户所处的上下文不同应该进行不同的推荐. 一个好的推荐系统不能仅仅捕捉用户的长期偏好, 而应该把时效特性考虑其中. 对时间序列数据有很强拟合能力的循环神经网络就刚好可以做到这一点, 捕捉用户的短期偏好, 进一步改善推荐效果.

但是, 通常在进行推荐时, 推荐的预测目标与实际所需要的信息并不是连续的, 甚至会出现超长间隔, 而普通的循环神经网络难以学习这种长依赖关系<sup>[12]</sup>. 所以在类似电影推荐这种行为建模场景中, 普通的 RNN 难以建模这种关系. 于是, LSTM 被提出, 它是一种特殊的 RNN, 专门针对数据的长短期依赖关系, 对于实际推荐系统中序列间的这种长依赖关系进行很好地建模.

因此本文对 NCF 中的神经网络结构进行了改进, 通过 LSTM 学习用户的短期偏好信息, 捕捉序列的长依赖关系, 同时通过广义矩阵分解学习用户的长期偏好信息, 然后在 NCF 框架下构建两个模型, 并把他们

进行融合得到预测数据. 本文把这个改进模型称为长短期记忆网络矩阵分解 (LSMF) 模型, 结构如图 3 所示.

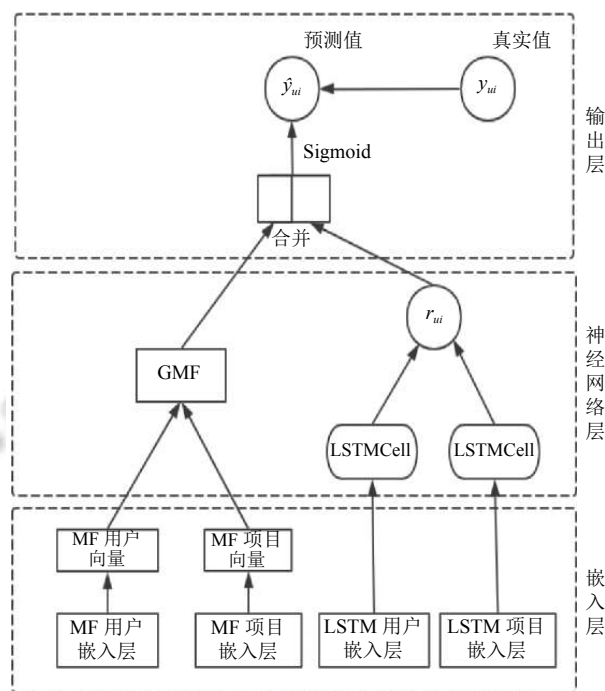


图3 LSMF 模型

### 2.2 嵌入层

嵌入层的作用在于将数据映射成为特征向量. 本文中, 它将输入层的用户 ID 和项目 ID 进行了 one-hot 编码后, 分别通过 MF 嵌入层和 LSTM 嵌入层映射成输入到 MF 层和 LSTM 层的用户和项目特征向量. 其中, 用户 ID 和电影 ID 需要通过时间排序后再输出到嵌入层.

### 2.3 神经网络层

#### 2.3.1 广义矩阵分解

通过嵌入层得到了用  $P^T V_u^U$  表示的用户的特征向量  $p_u$ ,  $Q^T V_i^I$  表示的项目特征向量  $q_i$ , 然后将神经协同过滤模型中的第一层的映射函数<sup>[13]</sup> 定义为:

$$\phi_1(p_u, q_i) = p_u \odot q_i \quad (8)$$

其中,  $\odot$  表示向量的点积, 然后将向量映射到输出层.

$$\hat{y}_{ui} = a_{out}(h^T(p_u \odot q_i)) \quad (9)$$

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

#### 2.3.2 长短期记忆网络

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。

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<sup>[14]</sup>, 然后预测此时用户和项目进行交互地可能性。

$$u_{u,t+1} = g(u_{ut}, \{r_{ui|t}\}) \quad (10)$$

$$m_{i,t+1} = h(m_{it}, \{r_{ui|t}\}) \quad (11)$$

基于用户评分行为预测评分问题表示为:

$$\hat{r}_{ui|t} = f(u_{ut}, m_{it}) \quad (12)$$

其中,  $r_{ui|t}$  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。

## 2.4 输出层

将 GMF 层和 LSTM 层输出的向量进行拼接。

$$\hat{y}_{ui} = \sigma(h^T[\phi_{GMF}, \phi_{LSTM}]) \quad (13)$$

本文使用 Sigmoid 作为激活函数。由于本文专注于隐式反馈的研究, 所以将 LSMF 视为概率模型, 把推荐作为一个二分类任务。因此本文使用对数损失函数 (log loss) 学习  $h$  进行优化。

## 3 实验分析

### 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 三个表。它拥有来自 6 040 位在 2000 年加入 MovieLens 的用户对 3 706 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。

### 3.2 实验评估标准

为了验证模型的推荐效果, 我们选用留一法 (leave-one-out) 进行验证。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。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,  $NDCG$ )<sup>[15]</sup> 作为性能评估标准。

$$HR@K = \frac{NumberOfHits@K}{|GT|} \quad (14)$$

$$NDCG@K = Z \sum_{i=1}^K \frac{2^{x_i} - 1}{\log(1 + i)} \quad (15)$$

在式 (14) 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 (15) 中,  $Z$  是归一化因子, 表示理想推荐列表的折损累积增益。  $x_i$  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  $K=10$ 。

### 3.3 实验结果验证

本文对 5 种不同的模型进行对比, 这些模型分别为 NCF 框架下的广义矩阵分解模型 (GMF), 多层感知机模型 (MLP), 神经网络矩阵分解模型 (NueMF), 长短期记忆网络模型 (LSTM) 和长短期记忆网络矩阵分解模型 (LSMF)。训练 50 轮, 取不同模型的性能平均值, 得出结果如表 1 所示。

表 1 不同模型的  $HR$  和  $NDCG$  对比

模型	$HR@10$	$NDCG@10$
GMF	0.65711	0.37574
MLP	0.68824	0.40928
NueMF	0.69470	0.41789
LSTM	0.70215	0.41160
LSMF	0.71138	0.42233

图 4 和图 5 显示了广义矩阵分解 (GMF) 模型, 长短期记忆网络 (LSTM) 模型以及两者融合后的长短期记忆网络矩阵分解 (LSMF) 模型的性能对比, 从图中可以看出, 前 10 个迭代的更新是最有效的。从以上实验数据可以明显看出, LSTM 模型的收敛速度较慢, 但收敛后的推荐性能比 GMF 要好, 而融合了 LSTM 和 GMF 的 LSMF 模型, 不仅收敛速度比较快, 而且推荐性能也是最好的。

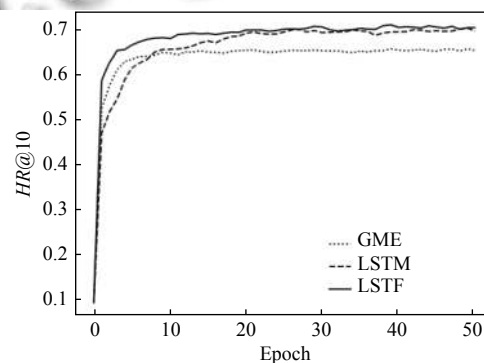


图 4 各个模型在  $HR$  上的表现

从以上实验结果中, 我们可以清楚地得到结论, 这 5 种模型在推荐性能上的优劣对比为 LSMF > LSTM > NueMF > MLP > GMF。因此本文提出 LSMF 模型在推荐性能上有了提高。

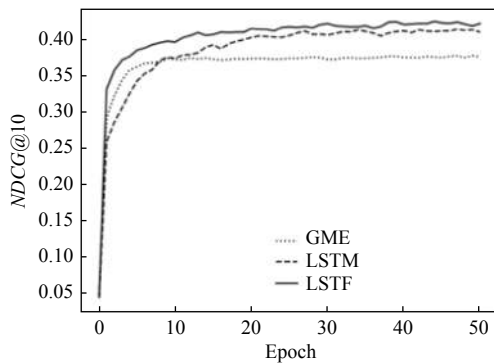


图5 各个模型在NDCG上的表现

## 4 结论

本文首先介绍了神经协同过滤模型, 然后针对模型中没有考虑时间因素, 无法同时捕捉用户的短期偏好和长期偏好的问题, 提出了长短期记忆网络与广义矩阵分解相结合的神经协同过滤模型, 通过长短期记忆网络学习用户的短期偏好信息, 捕捉序列的长依赖关系, 同时通过广义矩阵分解学习用户的长期偏好信息. 最后在 MovieLens-1M 数据集上与其他模型进行对比, 根据实验结果发现在命中率和归一化折现累积增益这两个指标上, 本文提出的改进模型有明显的提升, 收敛速度更快, 而且具备更好的推荐性能.

### 参考文献

- 1 李同欢. 基于深度学习的推荐模型研究 [硕士学位论文]. 重庆: 西南大学, 2019.
- 2 王喆. 深度学习推荐系统. 北京: 电子工业出版社, 2020.
- 3 He XG, Liao LZ, Zhang HW, *et al.* Neural collaborative filtering. Proceedings of the 26th International Conference on World Wide Web. Australia: ACM, 2017. 173–182.

- 4 刘晶. 基于深度学习的协同过滤算法的研究 [硕士学位论文]. 北京: 北京邮电大学, 2019.
- 5 王骏, 虞歌. 基于改进神经协同过滤模型的电影推荐系统. 计算机工程与设计, 2020, 41(7): 2069–2075.
- 6 纪强. 基于循环神经网络的深度推荐模型研究 [硕士学位论文]. 合肥: 安徽大学, 2020.
- 7 郭辰. 基于概率矩阵分解的协同过滤推荐 [硕士学位论文]. 大连: 大连理工大学, 2019.
- 8 Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems. Computer, 2009, 42(8): 30–37. [doi: 10.1109/MC.2009.263]
- 9 袁志远. 基于协同过滤的个性化推荐算法研究 [硕士学位论文]. 南京: 南京邮电大学, 2018.
- 10 Devooght R, Bersini H. Long and short-term recommendations with recurrent neural networks. Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization. Bratislava: ACM, 2017. 13–21.
- 11 高茂庭, 徐彬源. 基于循环神经网络的推荐算法. 计算机工程, 2019, 45(8): 198–202, 209.
- 12 Song Y, Elkahky AM, He XD. Multi-rate deep learning for temporal recommendation. Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval. Pisa: ACM, 2016. 909–912.
- 13 李川. 基于改进神经网络协同过滤的研究 [硕士学位论文]. 西安: 西安电子科技大学, 2019.
- 14 Wu CY, Ahmed A, Beutel A, *et al.* Recurrent recommender networks. Proceedings of the 10th ACM International Conference on Web Search and Data Mining. Cambridge: ACM, 2017. 495–503.
- 15 武海龙. 基于卷积神经网络的协同过滤算法研究 [硕士学位论文]. 武汉: 华中科技大学, 2019.