

# 基于融合 GMM 聚类与 FOA-GRNN 模型的推荐算法

李毅鹏, 阮叶丽, 张杰

(中南财经政法大学信息与安全工程学院, 湖北 武汉 430073)

**摘要:** 针对传统基于物品的推荐算法由于数据稀疏性导致的低推荐精度问题, 提出了一种融合 GMM 聚类和 FOA-GRNN 模型的推荐算法。该算法首先使用高斯混合模型 (GMM) 方法对物品特征进行聚类; 然后根据聚类结果分别构造评分矩阵, 并使用 Slope One 算法填充评分矩阵; 最后计算用户对物品的相似度预测评分作为输入, 通过 FOA-GRNN 模型输出最终的评分。基于 movielens-2k 数据集的实验结果表明, 与其他 3 种算法相比, 该算法能够更好地处理高稀疏性数据, 推荐精度更优, 并能够在一定程度上解决冷启动问题。

**关键词:** 推荐算法; GMM 聚类; 果绳优化; 广义回归神经网络; Slope One 算法

**中图分类号:** TP311

**文献标识码:** A

**doi:** 10.11959/j.issn.2096-109x.2018097

## Recommendation algorithm based on GMM clustering and FOA-GRNN model

LI Yipeng, RUAN Yeli, ZHANG Jie

School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan 430073, China

**Abstract:** Aiming at the problem of low recommendation accuracy caused by sparse data in traditional item-based recommendation algorithm, a recommendation algorithm based on GMM clustering and FOA-GRNN model is proposed. The algorithm firstly uses Gaussian mixture model (GMM) to cluster the item features, then constructs the score matrix according to the clustering results, and fills the score matrix with slope one algorithm. Finally, the user's score based on similarity prediction is taken as input, and the final score is output through FOA-GRNN model. Experimental results based on movielens-2k dataset show that the proposed algorithm can deal with highly sparse data better and has better recommendation accuracy than the other three algorithms.

**Key words:** recommendation algorithm, GMM clustering, FOA, GRNN, Slope One algorithm

## 1 引言

随着互联网技术的发展, 人们在线获取的数据越来越丰富, 然而大量无关冗余数据严重干扰了人们对有用信息的选择, 这种现象被称为“信

息过载”。因此, 推荐系统成为帮助用户获取有用信息的必要工具, 其原理是基于用户的兴趣偏好、历史评分、需求、行为等信息产生一个排序序列, 将排序靠前的商品或信息推荐给用户。近年来, 越来越多的个性化推荐方法被相继提出, 协同过

收稿日期: 2018-11-10; 修回日期: 2018-11-29

通信作者: 阮叶丽, yeli\_ruan@126.com

基金项目: 校级中央高校基本科研基金资助项目; 教育教学改革基金资助项目 (No.2018-9)

**Foundation Items:** The Fundamental Research Funds for the Central Universities, The Education and Teaching Reform Fund (No.2018-9)

滤推荐算法由于其推荐过程自动高效,并且仅利用用户对物品的评分信息而被广泛使用,成为互联网时代的新宠。协同过滤推荐算法有2种不同的推荐途径,分为基于用户的协同过滤和基于物品的协同过滤。

虽然协同过滤算法由于其简单、有效和准确性而被广泛应用,在信息时代节约了用户获取信息的时间代价,但是用户评分数据的稀疏性和冷启动问题成为制约其推荐精度的主要原因。对此,国外学者率先进行了相关研究,Deshpande等<sup>[1]</sup>提出SVD奇异值分解法可以对用户-项目评分矩阵进行降维,通过删除冗余信息实现降低数据稀疏度的目的。Rendle<sup>[2]</sup>则从贝叶斯原理出发,使用贝叶斯概率修正用户对项目的预测评分,降低了算法的预测偏差。Ocepek<sup>[3]</sup>就冷启动的问题,在传统协同过滤算法基础上加入用户认知信任信息,提升了推荐精度。国内学者在协同过滤推荐算法的研究上虽起步较晚,但已取得了可观的研究成果。张锋等<sup>[4]</sup>首先将机器学习的思想融入协同过滤算法中,利用BP网络降低数据稀疏度,取得了不错的实证效果。詹增荣<sup>[5]</sup>则通过RBF函数插值法对稀疏数据进行填补,接着采用SVM进行评分预测,使算法克服了数据质量的不足,增加了稳定性。张玉连等<sup>[6]</sup>另辟蹊径,提供了一种简单高效的优化方案。用最小二乘法对Slope One算法进行加权,从而实现对稀疏数据的填充。郭兰杰等<sup>[7]</sup>考虑到“朋友关系”可能影响不同用户对于物品的选择,故引入用户社交网络信息对物品相似度计算和最终预测评分阶段的缺失值进行填充,在降低评分误差的同时提高了分类准确率,但却增加了相关数据获取的难度以及算法整体的复杂性。针对传统协同过滤算法未充分重视用户反馈信息的问题,张宇等<sup>[8]</sup>提出了基于正负反馈的支持向量机协同过滤,筛选了用户较为喜欢的物品进行排序,加快了算法的推荐速度。庞海龙等<sup>[9]</sup>将真实评分与预测评分的插值作为线性回归模型的输入训练模型,从而产生最终推荐,为缓解数据稀疏性提供了另一种解决方案。与此同时,丁永刚等<sup>[10]</sup>提出码本聚类的思想获取用户评分风格信息,降低了推

荐成本。龚敏等<sup>[11]</sup>利用K-means方法对用户进行聚类,提高了推荐算法的可扩展性。陶维成等<sup>[12]</sup>则引入灰色关联度的聚类方法对物品特征进行聚类,取得了类似的实证效果。杨大鑫等<sup>[13]</sup>在上述研究的基础上,利用最小方差对K-means算法进行优化后对用户聚类,对协同过滤算法做了进一步改进。

综上所述,国内外学者对协同过滤算法的改进大多是从稀疏矩阵的填充以及用户信息的扩展两方面提出,对物品特征之间的联系兼顾较少,或考虑到物品类别,但未对用户项目评分做进一步优化,并且对用户信息的扩展增加了数据获取的难度。因此,本文从物品特征出发,提出了融合高斯混合模型(GMM)聚类和果蝇优化-广义回归神经网络(FOA-GRNN)模型的推荐算法,结合物品特征信息以及广义回归神经网络(GRNN)模型的非线性拟合能力产生推荐。

## 2 融合GMM聚类和FOA-GRNN模型的推荐算法

基于高斯混合模型聚类算法、果蝇优化算法、广义回归神经网络,本文构建了融合GMM聚类和FOA-GRNN模型的协同过滤算法。整套算法以基于物品的协同过滤算法为核心加以优化,算法构建的基本流程细化如下。

### 2.1 基于物品特征的GMM聚类

设 $A = \{a_1, a_2, \dots, a_m\}$ 表示含有 $m$ 个物品的向量, $F = \{f_1, f_2, \dots, f_n\}$ 表示每个物品的特征向量,则物品特征矩阵可以表示为

$$B = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

显然,矩阵 $B$ 中每一列可以视为物品特征 $f_j$  ( $j=1,2,\dots,n$ )的 $m$ 条记录,GMM聚类假定每一个 $f_j$ 的数据服从高斯分布,计算 $B$ 中每列的均值,可得均值向量 $\mu = \{\mu_1, \mu_2, \dots, \mu_n\}$ ,计算方差后可得协方差阵为 $\Sigma$ 。将其作为EM算法的初始化参数,通过EM算法迭代后可得最终的阈值,

选取样本点属于各个高斯分布的概率取得最大值时物品  $a_i (i=1,2,\dots,m)$  服从的分布作为最终的聚类结果, 则最终将  $m$  个物品根据各自特征聚成  $K$  类。

相较于  $K$ -means 等传统的聚类方法, GMM 聚类不用指定类别  $K$  的个数, 将样本特征假设为服从高斯分布, 在大样本条件下更符合中心极限定理, 最终的聚类结果是每个样本分属于不同类别的概率, 相较于  $K$ -means 聚类更为合理, 同时 GMM 算法的时间复杂度较  $K$ -means 有所减小, 降低了算法运行开销。

## 2.2 计算物品相似度以及用户-物品预测评分

### 2.2.1 结合 GMM 聚类结果计算物品间相似度

利用用户物品评分信息计算物品之间相似度, 其中评分信息可以整合到表 1 所示的矩阵中, 用户组成的用户向量为  $U=\{u_1, u_2, \dots, u_l\}$ , 物品组成的物品向量为  $A=\{a_1, a_2, \dots, a_m\}$ 。由于推荐算法数据集具有高稀疏度的特点, 评分数据一般含有大量的缺失值。

| 表 1 用户-物品评分矩阵 |       |       |       |         |       |
|---------------|-------|-------|-------|---------|-------|
| $u-a$         | $a_1$ | $a_2$ | $a_3$ | $\dots$ | $a_M$ |
| $u_1$         | 3     | 1     |       |         |       |
| $u_2$         |       |       | 4     |         | 5     |
| $u_3$         | 2     |       |       |         |       |
| $\vdots$      |       |       | 5     |         |       |
| $u_l$         | 1     | 3     |       |         | 1     |

与传统协同过滤算法只通过用户评分矩阵计算物品之间相似度不同, 本文提出的算法结合了对物品特征进行 GMM 聚类的结果, 将物品的特征信息深入融合到物品相似度计算中。根据 2.1 节的假设, 所有物品被聚为  $K$  类, 对于每一类中的物品, 分别计算其与同属于一类中的其他物品的相似度, 相似度采用常用的皮尔逊相似度度量, 即

$$\text{sim}(p, q) = \frac{\sum_{u \in U(p, q)} (S_{u,p} - \bar{s}_p)(S_{u,q} - \bar{s}_q)}{\sqrt{\sum_{u \in U(p, q)} (S_{u,p} - \bar{s}_p)^2} \sqrt{\sum_{u \in U(p, q)} (S_{u,q} - \bar{s}_q)^2}} \quad (1)$$

其中,  $U(p, q)$  表示对物品  $p$  和  $q$  都评分过的用户

集合,  $S_{u,p}$  表示用户  $u$  对物品  $p$  的评分,  $\bar{s}_p$  表示物品  $p$  所有评分的均值。将用户对物品的评分数据代入式(1)计算完成后, 根据计算结果的大小排序, 选取前  $N$  个物品作为  $N$  最相似邻居, 即选取前  $N$  个与指定物品相似度最高的其他物品组成一个集合。

### 2.2.2 基于物品相似度预测评分

在相似度计算的基础上, 通过用户  $u$  对物品  $p$  的  $N$  最相似邻居的评分值进行相似度加权后得到用户  $u$  对物品  $p$  的评分预测值, 计算公式如下。

$$P_{u,p} = \bar{s}_p + \frac{\sum_{q \in Z} \text{sim}(p, q)(S_{u,q} - \bar{s}_q)}{\sum_{q \in Z} \text{sim}(p, q)} \quad (2)$$

其中,  $Z$  表示物品  $p$  的  $N$  最相似邻居集合。

### 2.2.3 Slope One 算法填充缺失值

由于用户评分矩阵的高稀疏性, 对式(1)、式(2)的计算有所不便。在此采用 Lemire 等<sup>[14]</sup>于 2005 年提出的一种有效的缺失值填充方法, 即 Slope One 算法<sup>[14]</sup>, 本文在式(1)、式(2)的计算过程中采用 Slope One 算法对可能的缺失值进行选择填充, Slope One 算法重复利用用户-物品评分矩阵中已评分数据, 计算相关的未评分数据以改善数据的稀疏性。

## 2.3 FOA-GRNN 模型预测最终评分

广义回归神经网络是径向基函数网络的一种, 具备非线性拟合能力突出、学习速度较快的特点, 与传统 BP 神经网络的拟合路径不同, GRNN 的训练过程中无须调整各种权值和阈值, 传递函数的改变只需调整一个平滑参数  $\sigma$ , 大大简化了优化的难度, 这也是 GRNN 的优势之一。果蝇优化 (FOA) 算法是受果蝇觅食行为启发而设计出的进行全局迭代寻优的群体算法, 其具有估计参数少、迭代速度快、寻优能力强等特点。FOA 优化 GRNN 的核心进程是将 GRNN 拟合后的均方根误差 RMSE 以 FOA 味道浓度适应度函数的形式代入 FOA 中进行迭代。当达到最小阈值或者满足迭代次数标准后结束算法, 此时算法输出的食物位置即最终得出的最优平滑因子, 然后代入 GRNN 模型中进行预测, 得到最优的 GRNN 预测模型。

用户-物品评分矩阵中的缺失值由式(2)的计算结果代替,将所有相似度预测评分作为 GRNN 模型的输入,以聚类结果为依据,按物品所属类分别训练 FOA-GRNN 神经网络模型,以 FOA-GRNN 神经网络模型的输出结果作为最终的预测评分。

设第  $k$  个物品类别中含有  $i$  个物品,其用户评分信息如表 2 所示,假定相关缺失信息已经回填完毕,现将  $u_1, u_2, \dots$  对  $a_1, a_2, \dots, a_i$  的评分信息作为 GRNN 神经网络的输入,应用 FOA 算法寻找最优的 GRNN 模型平滑因子  $\sigma$ ,因为 GRNN 模型具有多项输出的能力,可将输出向量的维数设置为 2,得出的预测结果即可作为  $u_1, u_2, \dots$  对  $a_i$  的最终预测评分。最后根据最终评分的大小排序,选择前  $H$  个物品形成对用户  $u_j$  的推荐列表。

表 2 第  $k$  类用户对物品评分信息

| $u-a$    | $a_1$ | $a_2$ | $\dots$ | $a_i$ |
|----------|-------|-------|---------|-------|
| $u_1$    | 1     | 4     | 3       | ?     |
| $u_2$    | 2     | 2     | 2       | ?     |
| $\vdots$ | 1     | 2     | 2       | ?     |
| $u_j$    | 2     | 4     | 3       | ?     |

## 2.4 评价指标

本文选用推荐算法常用的 2 种评价指标来评价算法推荐结果的优劣<sup>[15-16]</sup>,这 2 种评价指标分别为平均绝对误差 (MAE) 和均方根误差 (RMSE),即

$$MAE = \frac{1}{N_s} \sum_{p,i} |S_{p,i} - \hat{S}_{p,i}| \quad (3)$$

$$RMSE = \sqrt{\frac{1}{N_s} \sum_{p,i} (S_{p,i} - \hat{S}_{p,i})^2} \quad (4)$$

其中,  $N_s$  代表测试数据集中的评分个数,  $S_{p,i}$ 、 $\hat{S}_{p,i}$  分别表示用户  $p$  对物品  $i$  的真实评分以及预测评分。MAE 与 RMSE 均为衡量预测误差大小的指标,都是在值较小时表示推荐算法取得较小的预测误差和较高的预测准确度。

## 3 实验分析

### 3.1 实验数据集

本文采用的实验数据集是推荐系统常用的

movielens-2k 数据集,包含在 HetRec2011 Dataset 中。该数据集是 movielens 10M 数据集的扩展,包含更多的电影特征信息。除了基本的“release time”“genre”外,还有“Top Critics Rating”“Top Critics NumReviews”“Top Critics Score”“Audience Rating”等连续型数据,相比 movielens 10M 数据集更加利于项目聚类算法的开展。该数据集的核心是 2 113 位用户对 10 197 部电影的评分,共 855 598 个记录,而每部电影只有 84 个评分记录,保证了高稀疏度的数据要求,因此,本文选用该数据集作为融合 GMM 聚类和 FOA-GRNN 模型推荐算法的实验数据集。

### 3.2 实验结果分析

对 movielens-2k 数据集的电影信息进行整理后发现,其可供聚类使用的有效特征数为 16 个,剔除其中含有的无效记录后可供研究的电影数目为 10 105 个,即整个实验的物品数目为 10 105 个,用户数目为 2 111 个,涵盖 855 598 个评分记录,其中不难发现实验数据集的评分稀疏度已达到 96%。本文所提算法的第一步根据电影特征对 10 105 部影片进行聚类,由于电影特征数较多,在此先用 PCA 主成分分析法对电影特征进行降维,并识别出 3 个主成分,再对主成分进行 GMM 聚类。GMM 聚类能够自动生成聚类的簇数,并不需要自己设定,最终的聚类结果如图 1 所示。

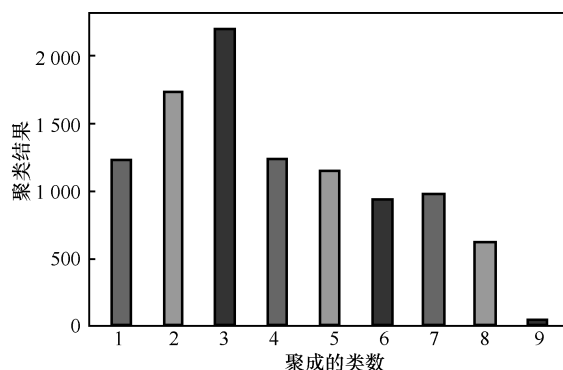


图 1 GMM 聚类结果

从图 1 中可以看出, GMM 聚类根据电影特征信息将 10 105 部电影聚成 9 类,各个电影所属的类别根据 EM 算法计算出的最终概率而定。聚类结果中各类含有的影片数目大体较为均匀,其中含有电影数目最多的类是第 3 类,包含 2 202 部

影片，而所含电影数目最少的类别为第 9 类，仅有 39 部电影。

根据上述聚类结果，分别对每个类别中的物品计算相似度，再根据每个物品的相似度大小进行排序，构建每个物品的  $N$  最相似物品邻居集合，计算过程的缺失值使用 Slope One 算法选择性填充，以相似度为基础计算出的用户-物品预测评分作为 FOA-GRNN 神经网络模型的输入，得到最终的预测评分信息。其中，FOA 算法对 GRNN 模型平滑因子进行迭代寻优的过程如图 2 所示。

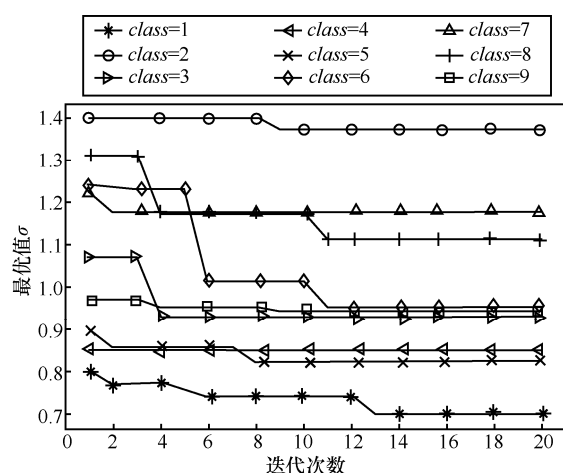


图 2 FOA 优化 GRNN 迭代路径

图 2 表示 FOA 算法优化 GRNN 模型的平滑因子  $\sigma$  的路径，其中迭代次数由横轴体现，纵轴显示的是最优的平滑参数  $\sigma$  值。基于 GMM 聚类的结果，分别在每个类别中训练 FOA-GRNN 神经网络模型，共训练了 9 个 FOA-GRNN 模型。根据 FOA 优化 GRNN 模型平滑参数的迭代截止条件，本文选择的最大迭代次数为 20，各个模型的 FOA 寻优路径相似，从图中可以看到，上述 9 个模型最多经历 13 次迭代后达到了最优，GRNN 模型的平滑参数  $\sigma$  值不再改变，不难看出，FOA 算法具有较强的寻优能力和计算效率。

根据 MAE 和 RMSE 这 2 种推荐算法优劣的评价指标，本文计算了在每一类中训练出的 FOA-GRNN 模型的 MAE 值和 RMSE 的值，根据文献[7]，本文将物品的相似度邻居数设置为 30 进行对比，每一类的 MAE 和 RMSE 值均为 5 折交叉验证的平均结果，如图 3 所示。

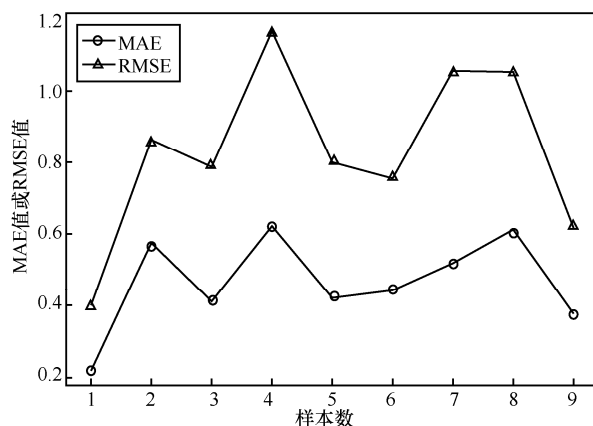


图 3 各个类中的 RMSE 和 MAE 值

从图 3 可以看出，各个类别的 MAE 值和 RMSE 值较为稳定，其中的少许波动一定程度上是由于各类样本数大小不同引起的，其中，第 4 类中的用户-电影评分矩阵的稀疏程度较大，故得到较大的 MAE 和 RMSE，相对正常。总体来看，RMSE 值略大于 MAE，RMSE 和 MAE 指标均未出现极端值，这也说明本文提出的推荐算法具有较高的泛化能力。

为了进一步比较验证本文提出的算法与已有的推荐算法性能的优劣，这里分别将新算法与如下 3 种算法的性能进行对比分析，所用的评价指标依然是 MAE 和 RMSE。

第 1 种算法为传统的基于物品的协同过滤算法（UCF），只根据用户-物品评分矩阵计算物品之间的相似度，进而根据物品的  $N$  最相似集合计算用户对物品的预测评分。

第 2 种算法为采用  $K$ -means 算法对物品聚类，然后采用 FOA 优化 GRNN 模型得出最终预测评分的算法，即在本文算法的基础上，将 GMM 聚类算法替换为  $K$ -means 算法后得到的算法，以比较  $K$ -means 算法和 GMM 算法在处理本文问题的表现优劣，记为  $K$ -FOA-GRNN。

第 3 种算法为本文算法中的 FOA 优化 GRNN 模型替换为传统的 BP 神经网络模型，以预测最终的评分值，以此比较 FOA-GRNN 模型同 BP 模型在本文问题上的预测效果，记为 GMM-BP。

采用相同的数据集实验后，4 种算法的对比结果如图 4 和图 5 所示。其中，横轴代表相似度邻居数  $N$ ，纵轴则表示误差 MAE 和 RMSE 值。

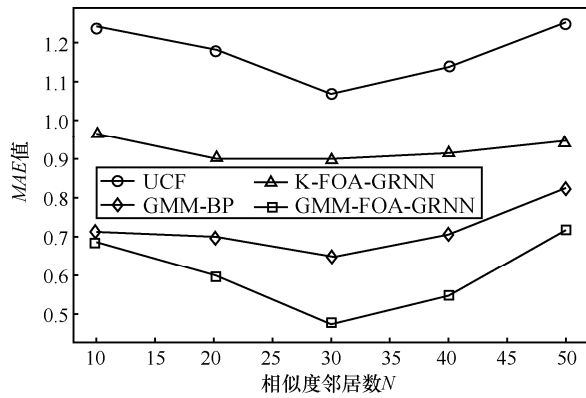


图4 4种算法MAE值对比

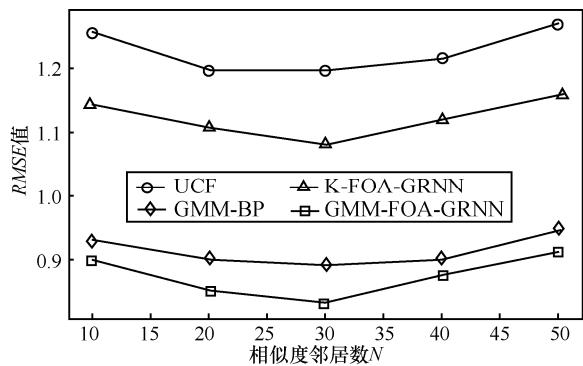


图5 4种算法RMSE值对比

由图4和图5可以看出,随着物品相似度邻居 $N$ 的增加,每种算法的MAE值和RMSE值都有一个先降后升的趋势,在邻居数为30时,MAE和RMSE都达到一个最小值。与其他3种已有算法相比较,本文提出的算法无论在MAE还是RMSE上,均小于其他3种已有算法,MAE和RMSE值均有较大幅度的降低,在各个邻居数下皆是如此。因此,本文提出的新算法较已有的协同过滤算法,有更好的推荐性能,适用范围更加广泛。

## 4 结束语

本文在已有研究的基础上,提出了一种融合GMM聚类与FOA-GRNN模型的协同过滤算法。该算法以传统面向物品的协同过滤算法为核心,充分利用物品特征的信息反馈,先通过GMM聚类方法将物品进行聚类,再根据聚类结果计算各个类别的物品之间的相似度,并用Slope One算法对计算过程的确切值进行填充,缓解了数据的稀疏性问题。最后将基于物品相似度邻居集合计算出的用户对物品的预测评分作为GRNN模型

的输入,利用FOA算法对GRNN模型的平滑因子进行迭代寻优,从而获得最终的用户物品预测评分,得益于GRNN模型多变量输出功能,可以较为有效地解决评分稀疏性问题。在movielens-2k数据集上的实证分析显示,改进的算法推荐误差较小,精度较高,并且较其他3种协同过滤算法有更好的推荐能力和泛化能力。

## 参考文献:

- [1] DESHPANDE M, KARYPIS G. Item-based top- $N$ , recommendation algorithms[M]. ACM, 2004.
- [2] RENDLE S, FREUDENTHALER C, GANTNER Z, et al. BPR: Bayesian personalized ranking from implicit feedback[M]. AUAI Press, 2009:452-461.
- [3] OCEPEK U, RUGELJ J, BOSNIĆ Z. Improving matrix factorization recommendations for examples in cold start[J]. Expert Systems with Applications, 2015, 42(19):6784-6794.
- [4] 张锋, 常会友. 使用BP神经网络缓解协同过滤推荐算法的稀疏性问题[J]. 计算机研究与发展, 2006, 43(4):667-672.  
ZHANG F, CHANG H Y. Employing bp neural networks to alleviate the sparsity issue in collaborative filtering recommendation algorithms[J]. Journal of Computer Research and Development, 2006, 43(4):667-672.
- [5] 詹增荣, 曾青松. 基于协方差矩阵表示的图像集匹配[J]. 湖南师范大学自然科学学报, 2015, 38(4):74-79.  
ZHAN Z R, ZENG Qing-song image set matching based on covariance matrix[J]. Journal of Natural Science of Hunan Normal University, 2015, 38(4): 74-79.
- [6] 张玉连, 邹思思, 梁顺攀. 融合机器学习的加权Slope One算法[J]. 小型微型计算机系统, 2016, 37(8):1712-1716.  
ZHANG Y L, HUAN S S, LIANG S P. Integrating machine learning into weighted Slope One algorithm[J]. Journal of Chinese Computer Systems, 2016, 37(8): 1712-1716.
- [7] 郭兰杰, 梁吉业, 赵兴旺. 融合社交网络信息的协同过滤推荐算法[J]. 模式识别与人工智能, 2016, 29(3):281-288.  
GUO L J, LIANG J Y, ZHAO X W. collaborative filtering recommendation algorithm incorporating social network information[J]. PR&AI, 2016, 29(3): 281-288.
- [8] 张宇, 王文剑, 赵胜男. 基于正负反馈的SVM协同过滤Top- $N$ 推荐算法[J]. 小型微型计算机系统, 2017, 38(5):961-966.  
ZHANG Y, WANG W J, ZHAO S N. SVM collaborative filtering Top- $N$  recommendation algorithm based on positive and negative feedback[J]. Journal of Chinese Computer Systems, 2017, 38(5): 961-966.
- [9] 庞海龙, 赵辉, 李万龙, 等. 融合协同过滤的线性回归推荐算法[J]. 计算机应用研究, 2019(5).

- PANG H L, ZHAO H, LI W L, et al. Linear regression recommendation algorithm with collaborative filtering[J]. Application Research of Computers, 2019(5).
- [10] 丁永刚, 李石君, 余伟, 等. 基于码本聚类和因子分解机的多指标推荐算法[J]. 计算机科学, 2017, 44(10):182-186.
- DING Y G, LI S J, YU W, et al. Multi-criteria recommendation algorithm based on codebook-clustering and factorization machines[J]. Computer Science. 2017, 44(10): 182-186.
- [11] 龚敏, 邓珍荣, 黄文明. 基于用户聚类与 Slope One 填充的协同推荐算法[J]. 计算机工程与应用. 2018, 20(8):102-110.
- GONG M, DENG Z R, HUANG W M. Collaborative recommendation algorithm based on user clustering and Slope One filling[J]. Computer Engineering and Applications, 2018, 20(8): 102-110.
- [12] 陶维成, 党耀国. 基于灰色关联聚类的协同过滤推荐算法[J]. 运筹与管理, 2018(1):84-88.
- TAO W C, DANG Y G. Collaborative filtering recommendation algorithm based on grey incidence clustering[J]. Operations Research and Management Science, 2018(1):84-88.
- [13] 杨大鑫, 王荣波, 黄孝喜, 等. 基于最小方差的 K-means 用户聚类推荐算法[J]. 计算机技术与发展, 2018(1):104-107.
- YANG D X, WANG R B, HUANG X X, et al. K-means user clustering recommendation algorithm based on minimum variance[J]. Computer Technology and Development, 2018(1): 104-107.
- [14] LEMIRE D, MACLACHLAN A. Slope One Predictors for online: rating-based collaborative filtering[C]//Proc of SIAM Data Mining. 2005: 471-476
- [15] BREESE J S, HECKERMAN D, KADIE C. Empirical analysis of predictive algorithm for collaborative filtering[R]. MSR-TR-98-12, Redmond, Microsoft, 1998.
- [16] SEHAFAER J, FRANKOWSKI D, HERLOCKER J. Collaborative filtering recommender systems[M]. Berlin: Springer. 2007.

#### [作者简介]



李毅鹏 (1980-), 男, 湖北武汉人, 中南财经政法大学副教授, 主要研究方向为管理信息系统、DSS、Mobile 电子商务。

阮叶丽 (1994-), 女, 贵州仁怀人, 中南财经政法大学硕士生, 主要研究方向为推荐系统。

张杰 (1995-), 男, 安徽六安人, 中南财经政法大学硕士生, 主要研究方向为金融统计。