邓妍, 王诚. 融合用户聚类和信任关系权重的推荐算法[J]. 智能计算机与应用,2026,16(6):90-96. DOI:10.20169/j.issn.2095-2163.24091001

融合用户聚类和信任关系权重的推荐算法

邓 妍, 王 诚
(南京邮电大学 通信与信息工程学院, 南京 210003)

摘 要: 协同过滤算法是推荐系统广泛应用的算法之一,通过用户或物品的数据记录进行相似度计算产生推荐。但在大规模系统中,随着用户和物品的数量不断增加,协同过滤算法的计算复杂度不断上升,同时用户与物品的评分矩阵往往比较稀疏,这会影响相似度计算的准确度,进而影响系统的推荐质量。对此,提出了一种融合用户聚类和信任关系权重的推荐算法,为降低计算复杂度,构建用户行为特征向量,由此对用户进行层次聚类;在各用户聚类簇中,分别采用加权 Slope One 算法对初始用户和物品评分矩阵预测填充,从而降低数据稀疏性;用户信任关系建模,借鉴 PageRank 算法思想,通过图结构表示用户之间的信任关系;在加权 Slope One 计算物品评分差异时,引入用户的 PR 值作为权重进行加权计算,提高预测填充的精度。选取 Ciao 数据集进行算法验证,实验结果表明,本文提出的算法能有效提高推荐质量。

关键词: 协同过滤; 数据稀疏性; Slope One 算法; 层次聚类; PageRank 算法; 信任权重
中图分类号: TP399 **文献标志码**: A **文章编号**: 2095-2163(2026)06-0090-07

A recommendation algorithm integrating user clustering and trust weight

DENG Yan, WANG Cheng
(School of Communications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: Collaborative filtering is one of the widely used algorithms in recommendation systems, generating recommendations based on user or item data records by calculating similarity. However, in large-scale systems, as the number of users and items increases, the computational complexity of collaborative filtering rises, and the user-item rating matrix is often sparse, which affects the accuracy of similarity calculations and thus the quality of recommendations. To address this, a recommendation algorithm that integrates user clustering and trust weight is proposed. To reduce computational complexity, user behavior feature vectors are constructed, and hierarchical clustering is performed on the users. In each user cluster, the weighted Slope One algorithm is used to predict and fill the initial user-item rating matrix, thereby reducing data sparsity. Trust relationships between users are modeled using the idea of the PageRank algorithm, representing trust relationships through a graph structure. When calculating item rating differences using weighted Slope One, the user´s PageRank value is introduced as a weight to improve the accuracy of the predictions. The proposed algorithm is validated using the Ciao dataset, and experimental results show that the algorithm effectively improves recommendation quality.

Key words: collaborative filtering; data sparsity; Slope One algorithm; hierarchical clustering; PageRank algorithm; trust weight

0 引 言

随着互联网和数字技术的迅猛发展,用户面临的信息量呈指数级增长,而从海量的数据中筛选出用户最感兴趣、最有价值的内容成为一大挑战。由此,推荐系统[1]应运而生。推荐系统的核心技术是推荐算法,通过分析用户的行为、偏好和特征,推荐算法能够在海量信息中精确地识别出用户可能感兴趣的内容,从而提升用户体验。其中,协同过滤算法[2]是推荐系统中最经典和广泛应用的算法之一。根据处理方式的不同,主要分为基于模型和基于内存的协同过滤算法。其中,基于内存的协同过滤推荐算法的核心思想是利用用户的行为数据,通过识别用户之间的相似性或物品之间的相似性,来预测

基金项目: 国家自然科学基金(61801240)。
作者简介: 王 诚(1970—),男,副教授,硕士生导师。主要研究方向:互联网大数据挖掘。
通信作者: 邓 妍(2000—),女,硕士研究生。主要研究方向:互联网大数据挖掘。Email:dengy2829@163.com。
收稿日期: 2024-09-10

用户可能感兴趣的内容或物品。因此基于内存的协同过滤推荐算法又可以分为基于用户和基于物品的协同过滤算法两种类型。协同过滤算法的优势在于不需要依赖物品的内容信息，纯粹基于用户的行为数据进行推荐，这就决定了该算法具有广泛适用性。然而，却也存在一些局限性，如冷启动、拓展性较差、流行度偏向和数据稀疏性问题，这些问题会影响推荐系统的有效性，降低推荐质量，进而使用户体验变差。

针对上述局限性，研究者们提出了多种改进措施，以提高协同过滤算法的性能和适用性。Volkovs 等学者[3]提出了一种基于神经网络的方法 DropoutNet，通过模拟数据缺失来增强模型的鲁棒性，从而有效应对冷启动问题。He 等学者[4]提出了一种通用的基于神经网络的协同过滤框架 NFC，在该框架下推广矩阵分解，在保持推荐精度的同时，提高了系统的可拓展性。杜东舫等学者[5]研究了信任用户间接影响用户偏好和直接影响用户评分两种不同机制，提出了基于用户间信任关系融合建模的概率矩阵分解模型 TPMF，实现对上述 2 种机制的有效融合，该模型在各项指标上都有较好的表现。陈彩蓉等学者[6]提出了一种融合用户影响力和灰关联分析的改进加权 Slope One 算法，综合考虑了用户相似性和用户影响力，有效缓解冷启动问题的同时，具有较高的推荐准确率。Liu[7]引入时间衰减函数对用户评分进行预处理，并采用 K-means 聚类算法对用户和项目分别进行聚类，较好地解决了数据稀疏、项目新增等问题。陈梅梅等学者[8]考虑到不同活跃度的用户话语权差异对评分预测的影响，提出了兼顾用户话语权的加权 Slope One 算法，有效降低数据稀疏性并提高了推荐精度。余永红等学者[9]对特定领域的用户社交网络进行推导，计算用户社会地位，提出了融合用户社会地位的矩阵分解推荐算法，提高了评分预测的准确度。韩胜宝等学者[10]利用用户人口学信息、评分信息以及项目信息计算得到融合相似度，再使用层次聚类确定用户的初始近邻集，一定程度上缓解了用户冷启动的问题，提高了算法的推荐质量。Sun 等学者[11]提出了信任相似度，根据用户历史记录建立用户兴趣特征向量，采用稀疏子空间聚类算法对用户进行聚类，算法的效率和精度有了一定的提高。Li 等学者[12]通过 LSTM 计算用户对评论的情感倾向，综合用户评分的情感值，再采用矩阵分解推荐算法，有效提高了推荐的质量。

基于推荐系统可扩展性和数据稀疏局限的考虑，本文提出融合用户聚类和信任关系权重的推荐算法。首先分析用户的评分行为记录，提取项目的所有类型标签，统计评分次数作为用户行为特征向量，以此构造用户向量矩阵，对用户采用层次聚类算法进行聚类。然后在各聚类簇中，利用引入 PageRank 算法的 Slope One 填充算法，对初始用户-物品评分矩阵进行预测填充，在此基础上采用基于用户的协同过滤算法对目标用户进行推荐。

1　相关工作

1.1　传统的协同过滤算法

协同过滤算法[13]是一种常用于推荐系统的经典算法，依赖于用户与物品的历史交互信息来做出推荐。传统的协同过滤推荐算法主要分为基于用户和基于物品两种，本文采用基于用户的协同过滤推荐算法，其实现步骤为构建用户-物品评分矩阵、用户相似度计算和评分预测产生推荐。

1.1.1　构建用户-物品评分矩阵

收集用户对物品的评分数据，将其对应转化为用户-物品评分矩阵 $\boldsymbol{R}_{m\times n}$ 见表 1。其中，m 表示用户的数量，n 表示物品的数量，则矩阵中的元素 $r_{i,j}$ 即为用户 u_i 对物品 p_j 的评分值。

表 1　用户-物品评分矩阵 *R*

Table 1　User-item rating matrix *R*

用户	p_1	p_2	p_3	p_4
u_1	3	4		5
u_2	4		5	4
u_3	5	2	3	

1.1.2　用户相似度计算方法

根据上述用户-物品评分矩阵 $\boldsymbol{R}_{m\times n}$，将每一行看作该用户的评分数据向量，计算用户之间的相似度就是计算用户的评分数据向量之间的距离。用户相似度计算方法主要可以分为余弦相似度、修正余弦相似度和皮尔逊相似度[14-15]三种。

(1)余弦相似度。假设 2 个用户 u 和 v 的评分数据向量分别为 $\boldsymbol{u}$ 和 $\boldsymbol{v}$，用户之间的余弦相似度可以用 2 个向量夹角的余弦值表示：

$$\mathrm{sim}(\boldsymbol{u},\boldsymbol{v})=\cos(\boldsymbol{u},\boldsymbol{v})=\frac{\boldsymbol{u}\cdot\boldsymbol{v}}{\|\boldsymbol{u}\|\cdot\|\boldsymbol{v}\|}\tag{1}$$

(2)修正余弦相似度。由于余弦相似度计算过程中，忽略了不同用户可能采用多样化的评分标准，会不可避免地导致误差。为修正这一误差，研究提出了修正余弦相似度，在计算过程中对用户进行中

心化处理,即减去各用户的评分均分,以消除评分标准不同带来的影响,计算公式如下:

$$\mathrm{sim}(u,v)=\frac{\sum_{i\in I_{u,v}}(r_{u,i}-\overline{r_u})(r_{v,i}-\overline{r_v})}{\sqrt{\sum_{i\in I_u}(r_{u,i}-\overline{r_u})^2}\times\sqrt{\sum_{i\in I_v}(r_{v,i}-\overline{r_v})^2}} \tag{2}$$

其中,I_u 和 I_v 分别表示用户 u 和 v 的评分物品集合;$I_{u,v}$ 表示用户 u 和 v 都评分过的物品集合;$r_{u,i}$ 和 $r_{v,i}$ 分别表示用户 u 和 v 对物品 i 的评分。

(3)皮尔逊相似度。与修正余弦相似度的区别在于中心化的处理方式不同,即用户 u 和 v 的评分均分计算的取值范围不同,只计算两者都评分过的物品均分,本文采用该相似度计算方法计算用户之间的相似度。数学公式为:

$$\mathrm{sim}(u,v)=\frac{\sum_{i\in I_{u,v}}(r_{u,i}-\overline{r_u})(r_{v,i}-\overline{r_v})}{\sqrt{\sum_{i\in I_{u,v}}(r_{u,i}-\overline{r_u})^2}\times\sqrt{\sum_{i\in I_{u,v}}(r_{v,i}-\overline{r_v})^2}} \tag{3}$$

1.1.3 预测推荐

由上述相似度计算公式得到用户之间的相似度,选取目标用户相似度最大的前 k 个用户作为近邻集 U_n,采用加权平均预测,则目标用户 u 对物品 i 的预测评分可以表示为:

$$P_{u,i}=\overline{r_u}+\frac{\sum_{v\in U_n}\mathrm{sim}(u,v)(r_{v,i}-\overline{r_v})}{\sum_{v\in U_n}\mathrm{sim}(u,v)} \tag{4}$$

在完成目标用户对未评分物品预测评分后,取出前 N 个评分最高的物品,推荐给用户。

1.2 Slope One 算法

Slope One 算法是一种简单高效且易于实现的算法,在 2005 年由 Lemire 等学者[16]提出,其实现原理基于用户对物品的评分之间的线性关系,利用物品之间的评分差值来进行预测,其主要步骤为计算评分差异和评分预测。

(1)计算评分差异。对于每一对物品 i 和 j,用户对两物品的评分分别为 $r_{u,i}$ 和 $r_{u,j}$,对物品 i 和 j 都评分过的用户集合表示为 $U_{i,j}$,$|U_{i,j}|$ 表示集合 $U_{i,j}$ 中元素的个数,计算所有对两物品评过分的用户评分差异的均值,就是物品 i 和 j 的评分差异。推得的公式为:

$$\mathrm{diff}(j,i)=\frac{\sum_{u\in U_{i,j}}(r_{u,j}-r_{u,i})}{|U_{i,j}|} \tag{5}$$

(2)评分预测。对于给定的目标用户 u,需要预测其对物品 j 的评分,则可以基于用户 u 对其他物品 i 的评分以及上述计算得到的评分差异进行预测。

综合分析后,推得的数学定义公式为:

$$P_{u,j}=\frac{\sum_{i\in I_u}(\mathrm{diff}(j,i)+r_{u,i})}{|I_u|} \tag{6}$$

由于 Slope One 算法忽略了评分差异样本数量对预测结果的影响,每个评分差异都被同等对待,使得预测精度较低。因此 Lemire 等学者[16]又提出加权的 Slope One 算法,引入评分差异样本的数量为权重,如果物品 i 和 j 同时被大量用户共同评分,则其评分权重就会被赋予更高的权重,具体预测公式如下:

$$P^{\mathrm{wso}}_{u,j}=\frac{\sum_{i\in I_u}(\mathrm{diff}(j,i)+r_{u,i})\,|U_{i,j}|}{\sum_{i\in I_u}|U_{i,j}|} \tag{7}$$

1.3 PageRank 算法

PageRank 算法是基于图论的数学算法,属于图数据的无监督学习算法。最初由 Brin 等学者[17]在 1998 年提出,作为计算互联网上网页重要性的方法,被用于 Google 搜索引擎的网页排名。PageRank 算法的核心思想是通过分析网页之间的链接关系,来评估每个网页的重要性,一个网页的重要性不仅取决于其入链的数量,也取决于这些网页本身的重要性。具体体现为,一个网页如果被多个重要的网页链接,则该网页重要性也高。PageRank 基本公式如下:

$$\mathrm{PR}(i)=\frac{1-d}{n}+d\sum_{j\in M(i)}\frac{\mathrm{PR}(j)}{L(j)} \tag{8}$$

其中,$\mathrm{PR}(i)$ 表示网页 i 的 PageRank 值;n 表示网页总数;d 为阻尼系数,表示用户点击当前网页链接进入下一网页的概率,通常为 0.85;$M(i)$ 表示链接到网页 i 的网页集合;$L(j)$ 表示从网页 j 链出的网页数量。

PageRank 算法将网页视为节点,网页的超链接视作边,构建有向图,通过简单的链接分析实现了对海量网页的重要性排序,是应用于搜索引擎排序的重要算法。由于 PageRank 算法强大的衡量节点重要性的能力,在其他领域也得到了广泛的应用,如通信网络优化[18]、社交网络[19]、生物信息学[20]、推荐系统[21]等。

2　融合用户聚类和信任关系权重的推荐算法

2.1　融合用户信任权重的加权 Slope One 算法

为了降低用户-物品评分矩阵的数据稀疏性，选取加权 Slope One 算法对初始用户-物品评分矩阵进行预测填充。由于加权 Slope One 算法在预测过程中，没有考虑到用户信任关系对预测结果的影响，本文提出融合用户信任权重的加权 Slope One 算法，在物品评分差异的计算过程中，引入用户信任关系权重。即对于被大量用户信任的强信任用户，其评分数据的可信度更高，反之则更低，因此相应地增加强信任用户评分数据的权重。

在对初始用户-物品评分矩阵进行预测填充前，需要先得到各用户的信任关系权重值。获取用户信任数据构建用户信任关系矩阵 $\boldsymbol{T}$ 见表 2。假设用户 u_i 信任用户 u_j、$t_{i,j}$ 的值为 1，反之不信任、则 $t_{i,j}$ 的值为 0。

表 2　用户信任关系矩阵 $\boldsymbol{T}$

Table 2　User trust relationship matrix $\boldsymbol{T}$

用户	u_1	u_2	u_3	u_4	u_5
u_1	0	1	1	0	0
u_2	0	0	1	0	1
u_3	1	0	0	0	0
u_4	0	0	1	0	1
u_5	0	0	0	1	0

基于 PageRank 算法的实现原理，将每个用户视为一个节点，用户的信任指向视为边，构建用户信任关系有向图，如图 1 所示。

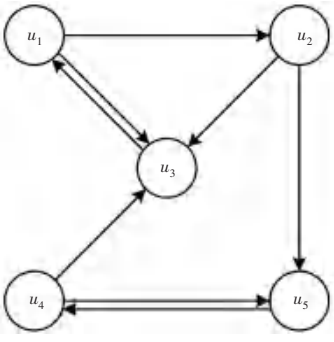


图 1　用户信任关系有向图

Fig. 1　Directed graph of user trust relationship

由式(8)可以计算得到每个用户的 PR 值。显而易见，用户的信任关系权重与用户的 PR 值呈正相关，即用户的 PR 值越大，用户的权重就越高，因此使用用户的 PR 值来代表用户的信任关系权重。考虑到在实际应用中，用户的数量通常较为庞大，用户的 PR 值就会较小，其权重影响也相应较小，因此将 PR 值放大 500 倍后再引入，则加权 Slope One 算法中物品评分差异的计算公式更新为：

$$\mathrm{diff}(j,i)=\frac{\sum_{u\in U_{i,j}}(r_{u,j}-r_{u,i})\times \mathrm{PR}(u)\times 500}{|U_{i,j}|} \tag{9}$$

2.2　基于用户行为特征向量的层次聚类

大数据时代，用户数量不断增加，推荐系统的预测计算量与时间复杂度急剧增大，这不仅会使得推荐结果出现偏差，而且会降低推荐系统的实时性和可扩展性。因此，本文提出基于用户行为特征向量的层次聚类方法，在利用加权 Slope One 算法预测填充之前，先将用户进行聚类，将相似度高的用户聚类到一起，再在每个用户聚类簇中分别对初始用户-物品评分矩阵预测填充，从而大大降低计算成本消耗同时，提高预测填充评分的有效性。

层次聚类[22]是一种基于分层结构的聚类方法，通过构建一颗层次树来描述数据对象之间的聚类关系。层次聚类主要分为 2 种：自底向上的凝聚层次聚类和自顶向下的分裂层次聚类。本文采用凝聚层次聚类，从每个用户开始，单独作为一个聚类，然后逐步合并距离最近的 2 个聚类，直至合并到设定的合适的聚类簇数时停止聚类。

基于用户行为特征向量的凝聚层次聚类，需要先构建用户的行为特征向量，再进行层次聚类，具体过程如下：

(1)用户行为特征向量构建。用户的兴趣偏好往往是复杂多样的，难以进行准确描述。因此本文尝试从用户历史评分行为数据出发，依据用户对不同物品评分次数的统计，作为用户行为特征向量来标记用户。由于物品的类型标签是不可重复的，假设物品类型集合为 $C(c_1,c_2,c_3,\cdots,c_n)$，共 n 个类型，则可以构建成一个 n 维的向量空间 $V(C)$，从而得到用户的行为特征向量。例如用户 u_i 的行为特征向量可以表示为 $v_i(10,7,9,0,4,2)$，即物品类型总数为 6，统计用户 u_i 对所有 6 个类型物品的评分次数，未评分的记为 0。

(2)用户层次聚类。由上述得到的用户行为特征向量，用式(1)计算用户向量之间的余弦相似度，作为用户之间的距离矩阵。簇与簇合并的链接方法选择 Ward 法，通过最小化合并后簇内的总方差来确定簇的合并顺序。具体来说，在合并簇时，选择合并后簇内方差增量最小的那对簇进行合并。假设有

2 个簇 C_i 和 C_j，簇内分别有 n_i 和 n_j 个用户，簇的质心分别为 μ_i 和 μ_j，合并后的新簇 C_{ij} 的质心为 μ_{ij}，其计算公式如下：

$$\mu_{ij} = \frac{n_i \mu_i + n_j \mu_j}{n_i + n_j} \tag{10}$$

Ward 法计算的是合并后簇内方差的增量，假设 x 为用户数据点，$\| x - \mu_k \|$ 表示数据点到质心的欧式距离，则簇内方差(ESS)定义如下：

$$\mathrm{ESS}(C_k) = \sum_{x \in C_k} \| x - \mu_k \|^2 \tag{11}$$

Ward 法的目标是最小化簇内方差的增量，则增量 ΔESS 的计算公式为：

$$\Delta\mathrm{ESS} = \mathrm{ESS}(C_{ij}) - \mathrm{ESS}(C_i) - \mathrm{ESS}(C_j) = \frac{n_i n_j}{n_i + n_j} \| \mu_i - \mu_j \|^2 \tag{12}$$

对实验数据集，提取用户数据，进行预处理后，不断改变设定的聚类簇数，分别观察其推荐结果 MAE 的变化，从而进一步确定最优的聚类簇数。通过实验，观察到当聚类簇数为 30 时，推荐结果的 MAE 较小，推荐准确度较高。

2.3 本文算法流程

针对传统协同过滤算法存在的数据稀疏和可扩展性差的问题，本文提出了融合用户聚类和信任关系权重的推荐算法，其具体流程如图 2 所示。

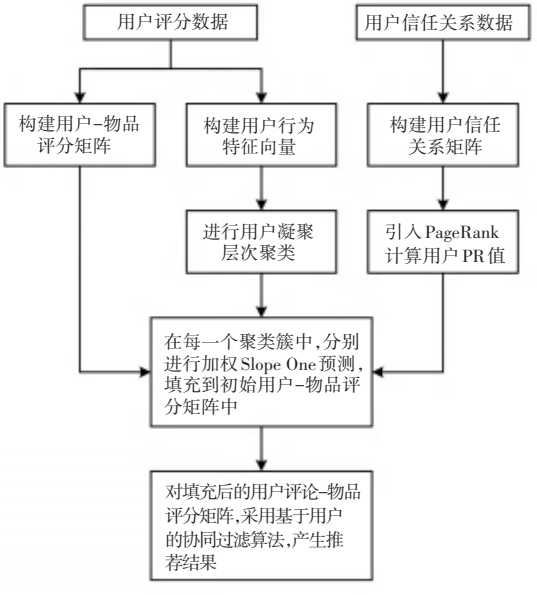


图 2 算法流程图

Fig. 2 Flowchart of the algorithm

步骤 1 收集用户对物品的评分数据、物品类型信息和用户信任关系，构建初始用户-物品评分矩阵 $\boldsymbol{R}$，用户行为特征向量 $\boldsymbol{v}_i$，用户信任关系矩阵 $\boldsymbol{F}$。

步骤 2 结合用户信任关系矩阵 $\boldsymbol{F}$，根据式(8)计算各用户节点的 PR 值。

步骤 3 收集各用户的行为特征向量，根据式(1)计算其余弦相似度，进行用户凝聚层次聚类，设定最终聚类簇数，将相似度较高的用户聚类到一个簇中。

步骤 4 在用户聚类得到的各用户簇中，分别使用融合用户信任权重的加权 Slope One 算法对初始用户-物品评分矩阵 $\boldsymbol{R}$ 进行预测填充，引入各用户的 PR 值作为用户信任权重，根据式(9)计算物品评分差异。

步骤 5 对填充后的用户-物品评分矩阵 $\boldsymbol{R}'$，根据式(3)计算得到用户的皮尔逊相似度，通过对该相似度排序得到用户的近邻用户集，根据式(4)进行评分预测，得到最终推荐物品集。

3 实验比较与分析

3.1 实验数据集

实验数据采用公开数据集 Ciao[23-24]，从流行的社交网络电子商务平台收集。数据集包含了 2 248 个用户对 16 861 个物品的 35 841 条真实评分记录，评分区间为[1,5]，步长为 1，则评分数据稀疏度为 0.095%。除此之外，数据集中还包含用户之间的信任关系数据，共有 57 544 条信任关系，信任关系数据稀疏度为 1.14%。本文实验采用五折交叉验证法，将数据集随机地划分为 5 个互不相交的子集，各子集依次作为测试集，剩余 4 个子集作为训练集，通过 5 次实验，将预测评分与实际评分进行比较，计算得到 2 个评分的 MAE 和 RMSE 值，取 5 次均值作为最终结果。

3.2 评价标准

实验采用平均绝对误差(MAE)和均方根误差(RMSE)两个指标来衡量算法预测性能的优劣，MAE 和 RMSE 是目前广泛使用、认可度较高的 2 个评估推荐系统质量的评估标准。两者均为误差值，误差值越小，预测评分与实际评分就越接近，预测性能越好。具体计算公式为：

$$\mathrm{MAE} = \frac{\sum_{u,i \in N} | r_{u,i} - P_{u,i} |}{| N |} \tag{13}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{u,i \in N} (r_{u,i} - P_{u,i})^2}{| N |}} \tag{14}$$

其中，N 表示测试集。

3.3　实验结果分析

为验证本文算法的有效性，分别对加权 Slope One 推荐算法（WSO）、融合用户特征向量聚类的加权 Slope One 推荐算法（CWSO）、融合用户信任权重的加权 Slope One 推荐算法（PWSO）、以及本文的融合用户聚类和信任关系权重的推荐算法（CPWSO）进行实验。研究中选取公开数据集 Ciao，设置近邻用户集的数量 K 从 10 到 100，步进为 10，用户相似度计算方法采用皮尔逊相似度。实验结果如图 3、图 4 所示。

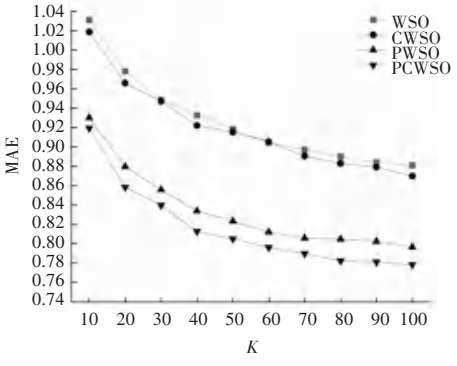


图 3　各算法的 MAE 值对比

Fig. 3　Comparison of MAE values for different algorithms

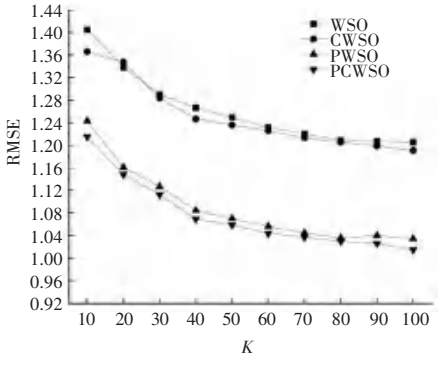


图 4　各算法的 RMSE 值对比

Fig. 4　Comparison of RMSE values for different algorithms

由图 3、图 4 可知，随着近邻集 K 的增大，各算法的 MAE 和 RMSE 误差值都逐渐减小，并在 K > 60 后，逐渐趋于平稳。其中，CWSO 算法相比 WSO 算法的提升较小，而引入 PageRank 算法计算用户信任关系权重的 PWSO 和 PCWSO 算法，无论在 K 取何值的情况下，相比 WSO 算法都有较大的提升。本文提出的 PCWSO 算法，在 K = 100 时，MAE 值能达到 0.77 左右，RMSE 值能达到 1.01 左右，在 MAE 和 RMSE 上的表现均优于其他算法。说明此算法的推荐质量较高，同时也验证了该算法的有效性。

4　结束语

本文主要从评分数据稀疏性以及系统可扩展性欠缺方面切入，考虑到这些问题对推荐系统的推荐质量的影响。在采用加权 Slope One 算法对初始用户-物品评分矩阵预测填充的基础上，进一步在计算物品评分差异过程中，引入用户信任关系权重，从而提高加权 Slope One 算法预测填充的精确度。同时，构建用户行为特征向量，在预测填充前对用户进行层次聚类，这在一定程度上减小了计算量，也提高了推荐系统的可扩展性。最后，对填充后的矩阵采用基于用户的协同过滤算法，计算用户之间的皮尔逊相似度，得到了最终推荐结果。经过实验分析对比，本文在 MAE 和 RMSE 上的表现均优于其他算法，因此，本文算法有效提高了推荐的准确度。

参考文献

[1] AGGARWAL C C. Recommender systems[M]. Cham: Springer, 2016.

[2] ANDIKA H G, HADINATA M T, HUANG W, et al. Systematic literature review: Comparison on collaborative filtering algorithms for recommendation systems [C]//Proceedings of the IEEE International Conference on Communication, Networks and Satellite (COMNETSAT). Piscataway, NJ: IEEE, 2022: 56-61.

[3] VOLKOVS M, YU G, POUTANEN T. DropoutNet: Addressing cold start in recommender systems[C]//Proceedings of Advances in Neural Information Processing Systems. Long Beach: NIPS Foundation, 2017: 496-505.

[4] HE X, LIAO L, ZHANG H, et al. Neural collaborative filtering [J]. arXiv preprint arXiv,1708.05031, 2017.

[5] 杜东舫，徐童，鲁亚男，等. 基于信任机制下概率矩阵分解的用户评分预测[J]. 软件学报，2018，29(12)：3747-3763.

[6] 陈彩蓉，刘虹，张岐山. 融合用户影响力和灰关联分析的 Slope One 推荐算法[J]. 小型微型计算机系统，2023，44(11)：2401-2407.

[7] LIU X J. An improved clustering-based collaborative filtering recommendation algorithm [J]. Cluster Computing, 2017, 20(2): 1281-1288.

[8] 陈梅梅，董晨光，王淇，等. 兼顾用户话语权的改进加权 Slope-One 协同过滤推荐[J]. 小型微型计算机系统，2022，43(9)：1814-1819.

[9] 余永红，高阳，王皓，等. 融合用户社会地位和矩阵分解的推荐算法[J]. 计算机研究与发展，2018，55(1)：113-124.

[10] 韩胜宝，伊华伟，李晓会，等. 基于融合相似度和层次聚类的冷启动推荐算法[J]. 小型微型计算机系统，2022，43(5)：985-991.

[11] SUN P, YIN S, MAN W, et al. Research of personalized recommendation algorithm based on trust and user's interest[C]// Proceedings of the International Conference on Robots & Intelligent System (ICRIS). Piscataway, NJ: IEEE, 2018: 153-156.

[12] LI X J, DENG G S, WANG X Z, et al. A hybrid

recommendation algorithm based on user comment sentiment and matrix decomposition [J]. Information Systems, 2023, 117: 102244.
[13]李嵩, 李书琴, 刘斌. 改进的协同过滤算法及其并行化实现[J]. 计算机工程与设计, 2018, 39(12): 3853-3859.
[14]孙辉, 马跃, 杨海波, 等. 一种相似度改进的用户聚类协同过滤推荐算法[J]. 小型微型计算机系统, 2014, 35(9): 1967-1970.
[15]顾明星, 黄伟建, 黄远, 等. 结合用户聚类与改进用户相似性的协同过滤推荐[J]. 计算机工程与应用, 2020, 56(22): 185-190.
[16]LEMIRE D, MACLACHLAN A. Slope one predictors for online rating-based collaborative filtering[C]//Proceedings of the 2005 SIAM International Conference on Data Mining. Newport Beach: Society for Industrial and Applied Mathematics, 2005: 471-475.
[17]BRIN S, PAGE L. The anatomy of a large-scale hypertextual web search engine[J]. Computer Networks and ISDN Systems, 1998, 30(1-7): 107-117.
[18]EEDI H, PERI S, RANABOTHU N, et al. An efficient practical non-blocking PageRank algorithm for large scale graphs[C]// Proceedings of 2021 29th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP). Piscataway, NJ: IEEE, 2021: 35-43.
[19]HONG Lei, QIAN Yiji, GONG Chaofan, et al. Improved key node recognition method of social network based on PageRank algorithm[J]. Computers, Materials & Continua, 2023, 74(1): 1887-1903.
[20]KOVACS I A, LUCK K, SPIROHN K, et al. Network-based prediction of protein interactions[J]. Nature Communications, 2019, 10(1): 1240.
[21]TAO M, YANG X, GU G, et al. Paperrecommend based on LDA and PageRank[C]//Proceedings of Artificial Intelligence and Security. Cham: Springer, 2020: 571-584.
[22]RAN Xingcheng, XI Yue, LU Yonggang, et al. Comprehensive survey on hierarchical clustering algorithms and the recent developments[J]. Artificial Intelligence Review, 2023, 56(8): 8219-8264.
[23]TANG Jiliang, GAO Huiji, LIU Huan, et al. eTrust: Understanding trust evolution in an online world[C]//Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2012: 253-261.
[24]TANG Jiliang, GAO Huiji, LIU Huan. mTrust: Discerning multi-faceted trust in a connected world[C]//Proceedings of the Fifth ACM International Conference on Web Search and Data Mining. New York: ACM, 2012: 93-102.