doi:10.3969/j.issn.2095-0411.2018.04.012

融合用户评论的矩阵分解推荐算法

胡胜利,谭 青

(安徽理工大学 计算机科学与工程学院,安徽 淮南 232001)

摘要:针对传统协同过滤算法中存在的数据稀疏性和单一利用用户的评分行为进行推荐的问题,提出了一种融合用户评论的矩阵分解推荐算法(USRMF)。该算法首先利用主题模型产生用户评论文本的主题分布,并结合评分提取出准确的用户兴趣和物品特征,然后结合用户兴趣和物品特征,通过余弦相似度计算分别得到用户和物品的最近邻,最后将最近邻的正则化项引入到矩阵分解模型中。实验中将 USRMF 算法与传统的协同过滤算法、正则化矩阵分解算法进行比较,结果表明 USRMF 算法在稀疏的数据集上能够提高推荐的准确度。

关键词:矩阵分解;用户评论;主题模型;正则化项;推荐算法

中图分类号:TK 8 **文献标志码**:A **文章编号**:2095-0411(2018)04-0069-07

Matrix Factorization Recommendation Algorithm Combining Users' Reviews

HU Shengli, TAN Qing

(School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan 232001, China)

Abstract: In order to solve the problems of data sparseness and only using the user's rating behavior to recommend in the traditional collaborative filtering algorithm, a kind of matrix factorization recommendation algorithm combining users' reviews was proposed. First, the algorithms utilized topic model to generate review topics distribution, and extracted accurate user interests and items characteristics by integrating the score. Then, combined with users' interests and item characteristics, the nearest neighbors of users and items were calculated according to the cosine similarity. Last, the nearest regularization terms were introduced into the matrix decomposition model. The USRMF was compared with the traditional collaborative filtering algorithm and the regularization matrix decomposition algorithm. The experimental result shows that the USRMF can improve the accuracy of recommendation in

收稿日期:2018-02-21。

基金项目:安徽理工大学硕士研究生创新基金项目(2017CX2112)。

作者简介:胡胜利(1978—),男,回族,安徽淮南人,硕士,副教授。E-mail:slhu@aust.edu.cn

引用本文:胡胜利,谭青. 融合用户评论的矩阵分解推荐算法[J]. 常州大学学报(自然科学版),2018,30(4):69-75.

sparse datasets.

Key words: matrix factorization; textual review; topic model; regularization term; recommendation algorithm

推荐系统的核心是通过推荐算法,利用用户对商品的反馈信息挖掘出用户的喜好。推荐算法主要有协同过滤推荐、基于内容的推荐、基于知识的推荐和混合推荐[1]。目前应用最广泛的推荐算法就是协同过滤算法,其主要分为两类:基于记忆的和基于模型的。基于记忆的算法通过相似度计算寻找相似的用户或物品,也就是最近邻,再根据最近邻对目标进行评分预测[2-3];基于模型的算法首先通过建立模型来表示用户评分的规律,然后通过"学习过"的模型进行预测推荐,如矩阵分解、关联规则挖掘、基于概率分析的模型算法[4-5]等。

矩阵分解算法可以有效缓解数据稀疏性问题,但是传统的矩阵分解推荐算法只是单独利用用户的评分行为进行预测评分[6],这种方法没有充分利用用户的其它反馈信息,比如用户的评论信息等,导致预测评分不够准确。针对这个问题,研究人员提出了许多解决办法,文献[7]提出了在矩阵分解模型中利用用户的一系列反馈信息来改进模型。文献[8]提出了通过上下文信息来扩展矩阵分解模型的方法。文献[9]提出了通过加入社会化正则化的方法来扩展模型,这些方法对矩阵分解模型进行了相应的扩展改进,取得了很好的效果。此外,用户的评论信息中包含了大量有用的信息,在推荐系统中的作用也越来越受重视。文献[10]使用情感分析的方法来分析用户评论,获取有用信息,从而提高推荐效果。文献[11]使用主题模型从用户评论信息中挖掘出用户兴趣特征,提升了推荐质量。

本文通过对矩阵分解模型以及用户评论信息的分析,提出了一种融合用户评论的矩阵分解推荐算法(简称 USRMF)。该算法考虑了利用用户评论信息来挖掘出用户兴趣和物品特征,并通过将它们作为正则化项加入到矩阵分解模型中。在稀疏的数据集上,可以达到更好的推荐效果。

1 相关研究工作

1.1 相似性的计算方法

推荐系统中常用的计算相似度的方法有 Pearson 相关系数、余弦相似度和 Jaccard 系数[12]。相似度的计算准确度直接影响用户近邻与物品近邻模型的选择准确度,在本文中主要选取余弦相似度来计算,结合了用户兴趣的用户最近邻和结合了物品特征的物品最近邻。

余弦相似度将用户 u 和用户 v 的评分数据提取为 n 维空间向量 $\boldsymbol{U}$ 和 $\boldsymbol{V}$,通过计算向量之间的夹角余弦来计算相似性,如式(1)

$$s(u,v)=\frac{\boldsymbol{U}\cdot\boldsymbol{V}}{|\boldsymbol{U}|*|\boldsymbol{V}|}=\frac{\sum_{i=1}^{n}r_{u,i}*r_{v,i}}{\sqrt{\sum_{i=1}^{n}r_{u,i}^{2}*\sum_{i=1}^{n}r_{v,i}^{2}}} \tag{1}$$

式中:$s(u,v)$表示用户 u 和用户 v 的余弦相似度;$\boldsymbol{U},\boldsymbol{V}$ 分别表示用户 u 和 v 对物品的评分向量。$s(u,v)$的值越接近于 1,说明用户 u 与用户 v 越相似。

1.2 矩阵分解模型

推荐系统使用矩阵分解的方法从评分模式中抽取出一组潜在的隐藏因子,并通过这些因子向量描述用户和物品,与传统的协同过滤算法相比它可以达到更好的推荐效果。矩阵分解技术在 Netflix

Prize 比赛中得到了广泛应用，它是一种潜在因子模型(Latent Factor Model，LFM)算法，核心思想就是把用户－物品评分矩阵分解成若干个矩阵的组合，用几个低维的矩阵来逼近原来的矩阵，最终目标是通过训练使原来的矩阵与预测矩阵之间的误差平方和最小[13]。在推荐系统领域，矩阵分解模型的基本原理就是将用户－物品评分矩阵 $\boldsymbol{R}_{U\times I}$ 分解成 2 个矩阵 $\boldsymbol{P}_{U\times K}$ 和 $\boldsymbol{Q}_{K\times I}$ 乘积的形式，如式(2)所示

$$\boldsymbol{R}_{U\times I}=\boldsymbol{P}_{U\times K}\times\boldsymbol{Q}_{K\times I} \tag{2}$$

式中：$\boldsymbol{P}_{U\times K}=[p_1,p_2,\cdots,p_u]$ 是用户因子矩阵，表示用户 u 对因子 K 的喜好程度；$\boldsymbol{Q}_{K\times I}=[q_1,q_2,\cdots,q_i]$ 是物品因子矩阵，表示第 i 个物品的因子 K 的程度。然后利用评分矩阵 $\boldsymbol{R}_{U\times I}$ 中的已知评分训练矩阵 $\boldsymbol{P}_{U\times K}$ 和 $\boldsymbol{Q}_{K\times I}$，使得 $\boldsymbol{P}$ 和 $\boldsymbol{Q}$ 相乘的结果最好地拟合已知的评分，那么未知的预测评分也就可以用矩阵 $\boldsymbol{P}$ 的某一行乘上矩阵 $\boldsymbol{Q}$ 的某一列得到，如式(3)所示

$$\hat{r}_{ui}=p_u\cdot q_i \tag{3}$$

式中：$\hat{r}_{ui}$ 表示预测用户 u 对物品 i 的评分，它等于矩阵 $\boldsymbol{P}$ 的第 u 行乘上矩阵 $\boldsymbol{Q}$ 的第 i 列。

1.3　LDA 主题模型

潜在狄利克雷分配(Latent Dirichlet allocation，LDA)是一种文档主题生成模型，也称三层贝叶斯概率模型，包含词、主题和文档 3 层结构[14]。文档到主题服从狄利克雷分布，主题到词服从多项式分布。LDA 可以给出文档集中的每篇文档的主题概率分布，再根据文档的主题分布进行主题聚类。同时，LDA 也是一种典型的词袋模型，即一篇文档是由一组词构成，词与词之间没有先后顺序的关系。另外一篇文档可能包含多个主题，文档中的每一个词都由其中的一个主题生成。在文本分析领域，使用 LDA 可以通过生成主题提取文本的特征，准确地描述文本内容。

2　融合用户评论的矩阵分解推荐算法

传统的矩阵分解模型，使用随机梯度下降法进行训练把真实值与预测值的总的误差平方和降到最小的过程，并没有考虑到用户评分矩阵的一些局部特性，比如用户兴趣和物品特征等。本文提出的 USRMF 推荐算法将用户的评论信息融合到矩阵分解模型当中，算法的思想是首先利用 LDA 主题模型产生用户评论文本的主题分布，并结合用户评分提取出用户兴趣和物品特征，然后通过余弦相似度计算分别得到用户和物品的最近邻，然后将其融合到矩阵分解模型中，最后再利用随机梯度下降法对目标损失函数进行学习优化，进行预测评分。USRMF 算法的流程图如图 1 所示。

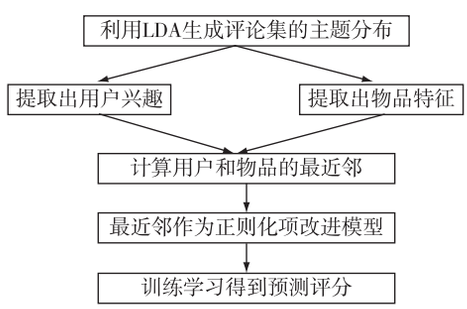

图 1　USRMF 算法流程图

2.1　评论文本中用户兴趣与物品特征的提取

为了把用户评论信息融合进矩阵分解模型中，本文使用 LDA 主题模型对用户的评论文本进行处理。对于每一位用户 u，将其对每一个物品 i 的相关评论文本看做一篇文档 $d_{u,i}$，然后使用 LDA 得到每篇文档 $d_{u,i}$ 在 k 个主题上的概率主题分布 $\Phi_{u,i}$(k 维)。$\Phi_{u,i}$ 代表了每位用户 u 对其评论过的物品 i 的评论文本中主题的分布情况，同时也代表了用户的兴趣特征。用户对其购买的物品不仅会评论，也会对每个物品打出 1～5 分的评分 $r_{u,i}$，所以需结合用户评分提取用户兴趣特征，具体为

$$\boldsymbol{Y}_u = \frac{\sum_{i \in I_u} \Phi_{u,i} r_{u,i}}{\sum_{i \in I_u} r_{u,i}} \tag{4}$$

式中:$\boldsymbol{Y}_u$ 代表用户兴趣,指示了用户 u 对 k 个主题的喜好分布,是一个 k 维的向量,$\boldsymbol{Y}_u = (Y_{u1}, Y_{u2}, Y_{u3}, \cdots, Y_{uk})$;$r_{u,i}$ 为每个用户 u 对物品 i 的评分;I_u 表示用户 u 评论过的物品集合。根据式(4)可以看出,用户 u 对物品 i 的评分越高,说明用户 u 对该物品越喜欢,用户兴趣的特征分布所占比重就越大,越能反映用户的兴趣特征。

对于每一个物品 i,将用户对其所有评论文本看做该物品的评论文档 d_i,然后使用 LDA 得到每个物品 i 在 k 个主题上的概率分布 $\boldsymbol{\Psi}_i$(k 维),$\boldsymbol{\Psi}_i$ 代表了物品 i 的所有评论文本中概率主题分布情况,代表了物品的特征分布。所以结合用户评分提取物品特征的计算为

$$\boldsymbol{Z}_i = \frac{\sum_{i \in I_u} r_{u,i}}{\sum \psi_{i \in I_u} r_{u,i}} \tag{5}$$

式中:$\boldsymbol{Z}_i$ 代表物品特征,指示了物品 i 在 k 个主题上的特征分布,是一个 k 维的向量,$\boldsymbol{Z}_i = (z_{i1}, z_{i2}, z_{i3}, \cdots, z_{ik})$。根据式(5)可以看出,用户 u 对物品 i 的评分越高,说明该物品的特征越受用户喜欢,那么物品对应的特征分布所占比重就越大,越能突出物品的特征。

2.2 计算用户和物品的最近邻

采用余弦相似度式(1)计算用户和物品的最近邻。在 2.1 节中通过 LDA 模型从用户评论文本中提取出了用户兴趣特征向量 $\boldsymbol{Y}_u$ 和物品特征向量 $\boldsymbol{Z}_i$,则结合用户兴趣和物品特征分别定义求解用户相似性和物品相似性,具体如式(6)、式(7)所示:

$$s(u,v) = \frac{\sum_{m=1}^{k} \boldsymbol{Y}_{um} \times \boldsymbol{Y}_{vm}}{\sqrt{\sum_{m=1}^{k} \boldsymbol{Y}_{um}^2 \times \sum_{m=1}^{k} \boldsymbol{Y}_{vm}^2}} \tag{6}$$

式中:$s(u,v)$表示基于用户兴趣求得的用户最近邻;$\boldsymbol{Y}_{um}$ 和 $\boldsymbol{Y}_{vm}$ 分别代表用户 u 和用户 v 在第 m 个主题上的兴趣分布情况。

$$s(i,j) = \frac{\sum_{n=1}^{k} \boldsymbol{Z}_{in} \times \boldsymbol{Z}_{ij}}{\sqrt{\sum_{n=1}^{k} \boldsymbol{Z}_{in}^2 \times \sum_{n=1}^{k} \boldsymbol{Z}_{ij}^2}} \tag{7}$$

式中:$s(i,j)$表示基于物品特征求得的物品最近邻;$\boldsymbol{Z}_{in}$ 和 $\boldsymbol{Z}_{ij}$ 分别代表物品 i 和物品 j 在第 n 个主题上的特征分布情况。

2.3 基于用户评论的矩阵分解

矩阵分解模型具有非常好的可扩展性,能够融合多种特征。但是在模型训练中,过多的变量、同时只有非常少的训练数据时,会导致模型出现过度拟合的问题,结果测试效果会非常差。通过加入正则化项可以有效解决该问题,保留所有的特征变量,提高推荐的准确度。正则化的引入利用了先验知识,在数据稀少的时候,可以防止过拟合。

从用户评论中提取用户兴趣和物品特征,对于具有相似兴趣特征的用户和相似特征的物品,用户的评分也更加相似。所以可以在矩阵分解模型中通过加入用户最近邻和物品最近邻的正则化项,来提高

预测评分的准确度。根据 2.2 节求得的用户和物品最近邻计算得到的正则化项为

$$\sum_{v\in S^k(u)} F_{uv} \| p_u - p_v \|^2 \tag{8}$$

式中：F_{uv} 是结合用户兴趣计算出的用户 u 和用户 v 的相似度（即式(6)的结果）；$s^k(u)$ 表示基于用户兴趣特征得到的用户 u 的最近 k 个邻居的集合，这个正则化项用来惩罚 2 个相似用户的潜在因子向量之间的距离。式(9)是利用物品最近邻求得的正则化项

$$\sum_{j\in T^k(i)} G_{ij} \| q_i - q_j \|^2 \tag{9}$$

式中：G_{ij} 是结合物品特征计算出的物品 i 和物品 j 的相似度（即式(7)的结果）；$T^k(i)$ 表示基于物品特征得到的物品 i 的最近 k 个邻居的集合，这个正则化项用来惩罚 2 个相似物品的潜在因子向量之间的距离。通过引入这 2 个正则化项能够有效避免过度拟合的问题，而且可以有效提高模型预测评分的精准度。

根据式(3)可知矩阵分解模型的预测公式，但是由于用户对物品的打分不仅取决于用户和物品之间的某种关系，还取决于用户和商品独有的性质，所以采用式(10)来计算模型的预测评分。

$$\hat{r}_{ui} = \mu + b_u + b_i + p_u \cdot q_i \tag{10}$$

式中：μ 表示训练集中总的平均分；b_u 代表用户偏置项；b_i 代表物品偏置项。

由 1.2 节可知，计算出真实值与预测值的总的误差平方和后，只要通过训练将该误差平方和降到最小，那么 $\boldsymbol{P}$，$\boldsymbol{Q}$ 就可以最好拟合矩阵 $\boldsymbol{R}$。为了防止过度拟合，需将目标函数中的所有变量都进行惩罚。本文提出的 USRMF 算法在加入基础的正则化项后，将利用用户和物品最近邻求得的正则化项也加入到目标函数中，所以对式(10)参数的求解可以通过对损失函数式(11)进行优化得到。

$$S_{\min} = \frac{1}{2}\sum_{u,i}(r_{ui} - \hat{r}_{ui})^2 + \frac{\lambda_1}{2}\sum_{v\in S^k(u)} F_{uv} \| p_u - p_v \|^2 + \frac{\lambda_2}{2}\sum_{v\in S^k(u)} G_{ij} \| q_i - q_j \|^2 + \frac{\lambda_3}{2}(\| p_u \|^2 + \| q_i \|^2 + \| b_u \|^2 + \| b_i \|^2) \tag{11}$$

采用随机梯度下降法对上述损失函数进行训练学习，根据随机梯度下降法的求解过程，该损失函数的参数更新式见式(12)，将学习得到的正确参数代入到预测公式中就可得到用户的预测评分。

$$\begin{cases} p_u \leftarrow p_u + \varphi\left(e_{ui}q_i - \lambda_1 \sum\limits_{v\in s^k(u)} F_{uv}(p_u - p_v) - \lambda_3 p_u\right) \\ p_v \leftarrow p_v + \varphi \cdot \lambda_1 \cdot F_{uv}(p_u - p_v) \\ q_i \leftarrow q_i + \varphi\left(e_{ui}p_u - \lambda_2 \sum\limits_{j\in T^k(i)} G_{ij}(q_i - q_j) - \lambda_3 q_i\right) \\ q_j \leftarrow q_j + \varphi \cdot \lambda_2 \cdot G_{ij}(q_i - q_j) \\ b_u \leftarrow b_u + \varphi(e_{ui} - \lambda_3 b_u) \\ b_i \leftarrow b_i + \varphi(e_{ui} - \lambda_3 b_i) \end{cases} \tag{12}$$

3　实验结果及分析

3.1　实验数据集

选取 Amazon.com 所提供的 6 个真实数据集进行实验，分别是 Baby，Office products，Health，Electronics，Toys and games 以及 Sports and outdoors，其中数据集中的每条评论都对应一个用户评分。表 1 显示了这些数据集的统计信息，最后一列为数据稀疏度，计算公式见式(13)。

稀疏度＝评论数/(用户数×商品数)(13)

可以看出这些数据集都非常稀疏,使用矩阵分解的方式可以有效缓解数据稀疏性,但是仅依靠用户评分的矩阵分解无法保证较高的推荐质量。实验中随机选择每个数据集,并将其按照 80%的训练集和 20%的测试集进行划分。

表 1 实验数据集信息

数据集	用户数	商品数	评论数	稀疏度
Baby	13 834	1 592	17 053	0.000 70
Office products	135 714	16332	154 862	0.000 06
Health	321 165	39 876	431 786	0.000 04
Electronics	125 386	45 670	172 869	0.000 03
Toys and games	304 726	51 227	403 188	0.000 02
Sports and outdoors	340 857	68 793	548 466	0.000 02

3.2 评价指标

实验采用平均绝对误差(Mean Absolute Error)作为度量标准[15]来验证 USRMF 算法的推荐效率。平均绝对误差的值越小,说明预测评分与真实值越接近,则推荐质量就越高。平均绝对误差为

$$M=\frac{\sum_{u,j\in T}|r_{ui}-\hat{r}_{ui}|}{|T|} \tag{14}$$

式中:M 表示平均绝对误差;T 表示测试集合,$|T|$表示测试集的大小;$r_{u,i}$表示用户 u 对物品 i 的实际评分;$\hat{r}_{ui}$表示用户 u 对物品 i 的预测评分。

3.3 实验结果

3.3.1 参数设置

实验中主要的参数有代表用户兴趣的主题个数 k_1,代表物品特征的主题个数 k_2,迭代步长 φ 和正则化参数 λ。为了便于计算,实验中将 k_1 与 k_2 的值都设置为 k,主题个数 k 分别取 5,10,20,35。通过多次实验,φ 取 0.003,λ_1 取 0.001,λ_2 取 0.001,λ_3 取 0.002,迭代的次数为 20。

表 2 USRMF 算法实验结果

数据集	$k=5$	$k=10$	$k=20$	$k=35$
Baby	1.015	0.968	0.981	0.992
Office products	1.059	1.013	1.024	1.043
Health	1.038	0.965	1.013	1.039
Electronics	0.937	0.872	0.872	0.918
Toys and games	0.836	0.801	0.819	0.806
Sports and outdoors	0.714	0.685	0.697	0.691
平均绝对误差	0.933	0.884	0.901	0.915

3.3.2 不同主题个数对算法预测性能的影响

主题数目的取值影响着 LDA 模型的性能,表 2 显示了 USRMF 算法在各个数据集上不同主题个数下的 MAE 值。从表中可以看出,当主题个数 $k=10$ 时,数据集上的平均 MAE 值为最小,USRMF 算法的性能最佳。但随着 k 值即主题个数的增加,平均 MAE 值逐渐增加。所以在下面的实验中,USRMF 算法的主题数量 k 选择为 10。

3.3.3 与其他算法比较

将 USRMF 算法(主题个数 k 为 10)与传统的协同过滤算法(包括 User-based 和 Item-based)、正则化矩阵分解算法(RSVD)进行对比分析。User-based CF 是利用用户的相似性进行预测评分,Item-based CF 利用物品的相似性进行预测评分,RSVD 在 SVD 模型的基础上加入了正则化项,本文的 USRMF 算法是在 RSVD 模型的基础上改进而来。在表 1 的 6 个数据集上的对比结果如图 2 所示。

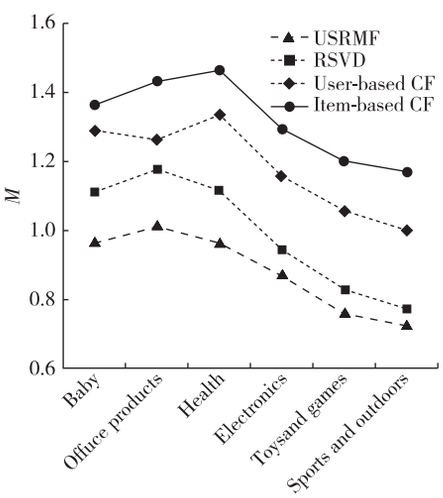

图 2 算法对比实验结果

从图 2 中的实验结果可看出，Item-based CF，User-based CF 和 RSVD 的算法在各个数据集上得到的平均绝对误差值 M 都高于 USRMF 算法得到的误差值，从而说明在稀疏的数据集上，USRMF 算法推荐准确度优于其他 3 种经典算法。

4　结　论

基于用户的评论信息，提出融合用户评论的矩阵分解推荐算法 USRMF。USRMF 利用 LDA 主题模型从用户评论中提取准确的用户兴趣和物品特征，通过相似度计算将用户和物品的最近邻正则化项引入到矩阵分解模型中，改进了模型。实验结果表明，USRMF 算法有效缓解了数据稀疏性问题并且有效提升了预测评分的准确度。但是该算法也存在一些不足，比如只利用用户的历史评论信息与评分行为进行预测，所以存在“冷启动”问题。未来将针对此问题，并结合本文提出的算法进行进一步研究。

参考文献：

[1]XIANG L. Recommendation system practice[M].Beijing: Posts & Telecom Press,2012:41-43.

[2]PARK Y,PARK S,JUNG W,et al.Reversed CF: a fast collaborative filtering algorithm using a k-nearest neighbor graph[J].Expert Systems with Applications,2015,42(8):4022-4028.

[3]荣辉桂，火生旭，胡春华，等.基于用户相似度的协同过滤推荐算法[J].通信学报，2014，35(2)：16-24.

[4]王升升，赵海燕，陈庆奎，等. 基于社交标签和社交信任的概率矩阵分解推荐算法[J]. 小型微型计算机系统，2016(5)：921-926.

[5]熊丽荣，刘坚，汤颖.基于联合概率矩阵分解的移动社会化推荐[J]. 计算机科学，2016(9)：255-260.

[6]张明，郭娣. 一种优化标签的矩阵分解推荐算法[J]. 计算机工程与应用，2015(23)：119-124.

[7]MANZATO M G. Supporting implicit feedback on recommender systems with metadata awareness[C]//SAC 2013: Proceedings of the 28th Annual ACM Symposium on Applied Computing.New York: ACM,2013:908-913.

[8]KRASNOSHCHOK O,LAMO Y.Extended content-boosted matrix factorization algorithm for recommender systems [J].Procedia Computer Science,2014,35: 417-426.

[9] AL-QAHERI H, BANERJEE S. Design and implementation of a policy recommender system towards social innovation:an experience with hybrid machine learning[M]. [S.l.]: Springer International Publishing,2015: 237-250.

[10]JAKOB N,WEBERS H,MLLERM C,et al.Beyond the stars: exploiting free-text user reviews to improve the accuracy of movie recommendations[C]//Proc of the 1st International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion.Hong Kong:[s.n.],2009: 57-64.

[11]王建，黄佳进. LDA-RR：一种基于评分和评论的推荐方法[J]. 计算机科学，2017(2)：267-269，305.

[12]BILGE A,KALELI C. A multi-criteria item-based collaborative filtering framework[C]//International Joint Conference on Computer Science and Software Engineering (JCSSE). [S.l.]: IEEE,2014:18-22.

[13]刘慧婷，陈艳，肖慧慧. 基于用户偏好的矩阵分解推荐算法[J]. 计算机应用，2015(S2)：118-121.

[14]BLEID M,NGA Y,JORDAN M I.Latent dirichlet allocation[J].Journal of Machine Learning Research,2003,3(4/5): 993-1022.

[15]CELMA M,HERRERA P.A new approach to evaluating novel recommendations[C]//RecSys 2008: Proceedings of the 2008 ACM Conference on Recommender Systems.New York: ACM,2008:179-186.

（责任编辑：李艳）