

•信息工程•

DOI:10.15961/j.jsuese.201900637



本刊网刊

## 基于社会信任正则化的排名推荐算法

张 俐<sup>1,2</sup>

(1.江苏理工学院 计算机工程学院, 江苏 常州 213001; 2.北京邮电大学 可信分布式计算与服务教育部重点实验室, 北京 100876)

**摘要:**随着在线商品交易额逐年增大和社交网络不断深入发展,推荐系统已成为解决信息过载的重要工具之一。当评分矩阵数据稀疏性较大时推荐精度就会显著下降,特别是用户冷启动时该问题更加明显。因此,本文提出一种新的基于隐式反馈信息的社会化排序推荐算法。该算法首先利用矩阵分解方法计算不同项目间的用户偏好。其次,将用户偏好信息融入贝叶斯个性化排名(Bayesian personalized ranking, BPR)算法。然后,挖掘用户之间的相似关系以及信任用户的直接和间接关系,并量化用户之间的信任关系,从而研究不同项目之间用户的偏好差异。最后,将以上信任关系和BPR算法进行融合,进而构建出社会化排序推荐模型。为了验证所提出的社会化排序推荐算法,在DouBan数据集和FilmTrust数据集上,进行算法的有效性验证。通过Precision、MAP和NGCD这3种排序评估指标分别在全数据集和用户冷启动中验证本文算法与SBPR、TBPR、BPR和MostPopular等算法之间排序推荐的优劣性。实验结果表明本文算法明显优于其他对比的排序推荐算法,并可以获得更好的推荐准确率。可见本文算法可以有效改善由于数据稀疏性和用户冷启动所引起的推荐效果差的问题。

**关键词:**推荐系统; 排序推荐算法; 贝叶斯个性化排名算法; 相似关系; 信任关系

中图分类号:TP181

文献标志码:A

文章编号:2096-3246(2020)05-0201-08

### Ranking Recommendation Algorithm Based on Social Trust Regularization

ZHANG Li<sup>1,2</sup>

(1.School of Computer Eng., Jiangsu Univ. of Technol., Changzhou 213001, China; 2.Key Lab. of Trustworthy Distributed Computing and Service (Ministry of Education), Beijing Univ. of Posts and Telecommunications, Beijing 100876, China)

**Abstract:** With the increase of online commodity transaction volume and the further development of social network, recommendation system has become one of important tools to solve information overload. However, when the data sparsity of the score matrix is relatively large, the recommendation accuracy will decrease significantly, especially when the user starts cold. A new social ranking recommendation algorithm based on implicit feedback information was proposed in this paper. The user preferences among different items using a matrix factorization method were first calculated by the proposed algorithm. Secondly, the user preference information was incorporated into the Bayesian Personalized Ranking (BPR) algorithm. Then, the similar relationships between users as well as direct and indirect relationships between trusting users were mined and quantified in order to study the differences of user preferences across projects. Finally, these trust relationships were fused with the BPR algorithm to build a social ranking recommendation model. To validate the proposed social ranking recommendation algorithm, the validity of the algorithm was verified on the DouBan dataset and FilmTrust dataset. The ranking evaluation metrics of Precision, MAP, and NGCD were used to verify the merits of ranking recommendation between the proposed algorithm and SBPR, TBPR, BPR, and MostPopular algorithms. The ranking recommendation tests were performed on these two specific social datasets for the full dataset and user cold start, respectively. The experimental results demonstrated that the proposed algorithm significantly outperforms other ranking recommendation algorithms and achieves better recommendation accuracy. It can be seen that the algorithm can improve the poor recommendation performance caused by data sparsity and user cold start.

**Key words:** recommendation system; ranking recommendation algorithm; BPR algorithm; similar relationship; trust relationship

收稿日期:2019-06-26

基金项目:国家重点研发计划项目(2017YFC1307705);江苏理工学院博士科研启动基金项目(KYY19042)

作者简介:张 俐(1977—),男,副教授,博士。研究方向:机器学习和推荐系统。E-mail: zhangli\_3913@163.com

网络出版时间:2020-09-14 14:14:14

网络出版地址:https://kns.cnki.net/kcms/detail/51.1773.TB.20200914.0916.001.html

不同用户对于商品或者服务存在着许多不同需求, 如何对用户偏好进行建模, 及对用户潜在需求进行预测和推荐, 满足他们个性化需求, 成为社会化推荐亟需解决的问题<sup>[1-2]</sup>。

推荐算法<sup>[3-4]</sup>主要分为两类: 显式反馈推荐算法和隐式反馈推荐算法。显式反馈推荐算法只单独考虑每个项目, 预测单个用户对单个项目的偏好, 并给出精确的评分。常见的协同过滤算法就是最基本的显式反馈算法。其又可分为基于内存的算法和基于模型的算法。矩阵分解算法是一种经典的基于模型的协同过滤算法。由于矩阵分解算法具有很强的可扩展性, 所以研究者提出了许多矩阵分解算法的扩展算法, 具体有SVD++算法<sup>[5]</sup>、SocialReg算法<sup>[6]</sup>、BPR算法<sup>[7]</sup>和SVD算法<sup>[8]</sup>。

当没有显式反馈信息时, 只能从用户操作中获得隐式反馈信息(如浏览商品和购买商品), 进而进行商品或服务的推荐。相比显式反馈技术, 隐式反馈技术由于客户体验更好、应用广泛等优点, 逐渐变成研究热门<sup>[9]</sup>。在隐式反馈推荐算法<sup>[3]</sup>中主要分为对级和列表级推荐。对级推荐中有RankSVD算法<sup>[10]</sup>和CCF模型<sup>[11]</sup>, 但是这两种算法由于需要不断计算项目之间的相对位置, 导致计算复杂度较高。列表级排序推荐中有LCR算法<sup>[12]</sup>、CLiMF算法<sup>[13]</sup>和ListRank-MF算法<sup>[14]</sup>, 但是, 当用户项目评分矩阵 $\mathbf{R}$ 数据稀疏性较大时其推荐精度下降, 特别是在面临用户冷启动时该问题更加明显。

如何降低计算复杂度并解决由于数据稀疏和用户冷启动所引起的推荐质量下降的问题<sup>[15]</sup>。社交网络可以识别用户之间的信任关系, 并量化用户对物品的偏好差异<sup>[4]</sup>。因此许多研究者利用社交化网络和对级个性化排名推荐提升排序推荐的准确性<sup>[15-17]</sup>。Pan等<sup>[18]</sup>提出组贝叶斯个性化排名(group Bayesian personalized ranking, GBPR)算法, 该算法认为群体偏好可以更加丰富群体用户之间的交互关系, 并且这种关系有助于提升排序推荐。Zhao等<sup>[19]</sup>提出社交化贝叶斯个性化排名(social Bayesian personalized ranking, SBPR)算法, 该算法认为用户喜欢将更好的物品推荐给其朋友, 并以此提高排序推荐的准确性。Wang等<sup>[20]</sup>在SBPR算法的基础上提出TBPR(BPR with strong and weak ties)算法, 该算法认为朋友之间有强弱关系, 并在社交网络中对这种强弱关系进行分类, 因为不同的关系可以推荐不同物品, 这样能进一步提升推荐的准确性。Guo等<sup>[9]</sup>提出社会信任贝叶斯个性化排名(social based Bayesian personalized ranking, SocialBPR), 该算法是一种具有多方位信任关系的隐式反馈个性化排名算法。但是, 上述这些隐

式反馈的社会化推荐算法都只是将社会化关系作为约束条件对目标函数进行优化。

本文认为用户项目评分矩阵 $\mathbf{R}$ 中的用户间的相似信任关系、用户信任矩阵 $\mathbf{T}$ 中的信任用户之间的直接和间接关系, 对提高排序推荐有着重要的作用。因此, 提出一种新的社会化排序推荐STrustBPR(social trust recommendation algorithm based on Bayesian personalized ranking)算法。首先, 通过矩阵分解计算用户的偏好; 其次, 利用BPR算法对用户偏好进行个性化排序; 同时, 利用用户信任矩阵 $\mathbf{T}$ 对不同用户的直接信任关系和间接信任关系进行度量, 利用用户项目评分矩阵 $\mathbf{R}$ 对用户之间相似信任关系进行度量, 并挖掘不同用户间的偏好; 最后, 将这些信任关系和BPR算法进行融合, 从而构建出一个用户信任的排序推荐模型。在DouBan、FilmTrust两个具体社会化数据集集中进行多种排序指标的实验验证, 实验结果表明社会化推荐排序算法STrustBPR是有效的。

## 1 基于社会正则化的排序推荐算法

设用户项目评分矩阵为 $\mathbf{R}$ ,  $\mathbf{R} = [r_{i,j}]$ 为一个 $m \times n$ 的评分矩阵, 其中,  $r_{i,j} \in [0,5]$ 。同时, 设信任矩阵为 $\mathbf{T}$ ,  $\mathbf{T} = [t_{i,j}]$ 为一个 $m \times m$ 的矩阵, 其中,  $t_{i,j}$ 为用户 $u_i$ 和用户 $u_j$ 之间的信任关系。当 $t_{i,j} = 1$ 时,  $\mathbf{T}$ 表示直接好友关系; 当 $t_{i,j} = 0$ 时,  $\mathbf{T}$ 表示间接好友关系。

### 1.1 用户偏好计算

一般来讲, 矩阵分解<sup>[5]</sup>将用户项目评分矩阵 $\mathbf{R}$ 分解为用户隐特征矩阵 $\mathbf{U}$ 和项目隐特征矩阵 $\mathbf{V}$ 。现利用矩阵分解技术预测 $\mathbf{U}$ 和 $\mathbf{V}$ 之间的偏好值。设 $\mathbf{U}_u$ 为用户 $u$ 的隐特征向量,  $\mathbf{V}_i$ 为项目 $i$ 的隐特征向量,  $\varphi_i^{(u)}$ 为 $\mathbf{U}_u$ 和 $\mathbf{V}_i$ 之间的偏好值, 具体公式如下:

$$\varphi_i^{(u)} = \langle \mathbf{U}_u^T, \mathbf{V}_i \rangle = \sum_{d=1}^k \mathbf{U}_{u,d}^T \mathbf{V}_{d,i} \quad (1)$$

式中,  $k$ 为隐特征向量的维数 $k \ll \min(m, n)$ 。

### 1.2 贝叶斯个性化排名

Salakhutdinov等<sup>[7]</sup>为了解决项目排序推荐的问题提出贝叶斯个性化排名(BPR)算法, 其是一种对级排序推荐算法。如果 $\varphi_i^{(u)} > \varphi_j^{(u)}$ 则表示用户 $u$ 更喜欢项目 $i$ 而不是项目 $j$ 。因此, 从 $\varphi_{i,j}^{(u)}$ 中可以得出用户 $u$ 对项目 $i$ 、 $j$ 的具体量化偏序关系, 具体公式如下:

$$\varphi_{i,j}^{(u)} := \varphi_i^{(u)} - \varphi_j^{(u)} \quad (2)$$

通过式(2)可以知道如果 $\varphi_{i,j}^{(u)} > 0$ , 那么 $\varphi_i^{(u)} > \varphi_j^{(u)}$ , 则可以描述为 $i >_u j$ , 其中,  $>_u$ 为用户 $u$ 对项目 $i$ 、 $j$ 的偏序关系,  $\{i_1, i_2, \dots, i_n\} \in I$ ,  $I$ 为项目集合。

那么, 用户 $u$ 的项目偏好集合 $\pi(u)$ 定义如下:

$$\pi(u) := \{(u; i_1, i_2, \dots, i_n) | i_1 > i_2, i_3 > i_4, \dots, i_{n-1} > i_n\} \quad (3)$$

根据贝叶斯推论<sup>[12-13]</sup>可以得到后验概率:

$$p(\Theta|\pi) \propto p(\pi|\Theta)p(\Theta) \quad (4)$$

其中,

$$p(\pi|\Theta) = \prod_{u \in m} p(u; i, j|\Theta) = \prod_{(u; i, j) \in m \times n \times n} p(i >_u j|\Theta)^{\delta((u; i, j) \in \pi(u))} (1 - p(i >_u j|\Theta))^{\delta((u; i, j) \notin \pi(u))} \quad (5)$$

式中:  $\delta(\cdot)$  为指示器, 如果  $(u; i, j) \in \pi(u)$  成立, 则  $\delta((u; i, j) \in \pi(u))$  表示为1, 否则为0;  $\Theta$  为BPR算法中的参数, 设  $\Theta = \{U_u \in R^{k \times m}, V_i \in R^{k \times n}, V_j \in R^{k \times n}\}$ 。

为了最大化  $p(\Theta|\pi)$  的后验概率, 利用函数  $\sigma(x) = 1/(1 + \exp(-x))$  计算  $p(\varphi_{ij}^{(u)})$ , 同时, 使用对数函数  $\ln(x)$  减少计算过程的复杂度。

因此, 可以得到最优目标损失函数  $J_{\text{BPR}}(\Theta)$ , 具体公式如下:

$$J_{\text{BPR}}(\Theta) = \sum_{u \in m} \sum_{i \in n} \sum_{j \in n} \ln \sigma(\varphi_i^{(u)} - \varphi_j^{(u)}) + \frac{\lambda_u}{2} \|U_u\|^2 + \frac{\lambda_v}{2} \|V_i\|^2 + \frac{\lambda_v}{2} \|V_j\|^2 \quad (6)$$

式中,  $\|U_u\|^2$ 、 $\|V_i\|^2$  和  $\|V_j\|^2$  用于防止过拟合, 参数  $\lambda_u$  和  $\lambda_v$  分别用于控制正则化  $U_u$ 、 $V$  和  $V_j$  的影响。

### 1.3 用户信任关系量化

#### 1.3.1 直接和间接信任关系量化

在用户信任矩阵  $T$  中, 用户总是信任自己的朋友, 如果朋友推荐给用户某个商品, 那么该商品就会受到该用户关注。根据社会化理论<sup>[21-22]</sup>, 上述可以理解为用户  $u_u$  受直接用户  $u_j$  的影响。如果用户  $u_u$  有若干个朋友, 那么, 可以进一步描述为用户  $u_u$  受直接社会关系集合  $N_u^+$  影响, 其中,  $|N_u^+|$  为用户  $u_u$  的直接信任用户数。同理, 如果用户  $u_u$  的朋友  $u_j$  也有朋友  $u_g$ , 而用户  $u_u$  和用户  $u_g$  不是朋友关系, 那么, 可以认为用户  $u_u$  也会受到间接关系集合  $N_u^-$  影响, 其中,  $|N_u^-|$  为用户的  $u_u$  间接信任用户数。

设社会化用户定义为  $u_u \in X$  且  $u_u \in Q$ , 其中,  $X$  为用户项目评分  $R$  中用户集合,  $Q$  为用户信任矩阵  $T$  中用户集合。那么, 用户  $u_u$  和用户  $u_j$  之间的直接信任关系可靠度 ( $W_{u,j}^+$ )、用户  $u_u$  和用户  $u_g$  之间的间接信任关系可靠度 ( $W_{u,g}^-$ ) 定义式如下:

$$W_{u,j}^+ = \sqrt{\frac{|N_u^+|}{\sum_{j \in N_u^+} |N_u^+|}} \quad (7)$$

$$W_{u,g}^- = \sqrt{\frac{|N_u^-|}{\sum_{g \in N_u^-} |N_u^-|}} \quad (8)$$

#### 1.3.2 用户之间相似信任关系的量化

在用户项目评分矩阵  $R$  中, Jaccard相似度方法从全局的角度描述用户之间的相似信任关系。设有用户  $u_i$  和用户  $u_j$ , 如果二者之间的共同项目较多则用户  $u_i$  和用户  $u_j$  之间的相似程度越高, 反之, 则用户  $u_i$  和用户  $u_j$  之间相似程度越低, 具体公式如下:

$$S_{\text{Jaccard}}(u_i, u_j) = \frac{|N(u_i) \cap N(u_j)|}{|N(u_i) \cup N(u_j)|} \quad (9)$$

式中,  $N(u_i)$  为用户  $u_i$  的项目集合,  $N(u_j)$  为用户  $u_j$  的项目集合,  $|\cdot|$  表示项目集合数。

皮尔逊相似性 (Person correlation coefficient, PCC) 从局部的角度描述用户之间的相似信任关系, 定义式如下:

$$S_{\text{pcc}}(u_i, u_j) = \frac{\sum_{g \in N(u_i, u_j)} (\bar{r}_{u_i} - r_{g, u_i}) \cdot (\bar{r}_{u_j} - r_{g, u_j})}{\sqrt{\sum_{g \in N(u_i)} (\bar{r}_{u_i} - r_{g, u_i})^2} \cdot \sqrt{\sum_{g \in N(u_j)} (\bar{r}_{u_j} - r_{g, u_j})^2}} \quad (10)$$

式中,  $r_{g, u_i}$  为用户  $u_i$  对项目  $g$  的评分,  $\bar{r}_{u_i}$  为  $N(u_i)$  中的均值,  $r_{g, u_j}$  为用户  $u_j$  对项目  $g$  的评分,  $\bar{r}_{u_j}$  为  $N(u_j)$  中的均值。

如果只考虑一种用户之间相似信任关系的度量方法, 则无法兼顾从全局、局部角度度量用户信任关系的优势。因此, 本文提出利用  $S_{\text{pcc}}$  和  $S_{\text{Jaccard}}$  的特点, 度量用户  $u_i$  和用户  $u_j$  之间相似信任关系, 具体定义式如下:

$$S_{ij}(u_i, u_j) = S_{\text{Jaccard}}(u_i, u_j) \cdot S_{\text{pcc}}(u_i, u_j) \quad (11)$$

式中,  $S_{\text{pcc}}$  和  $S_{\text{Jaccard}}$  互为自适应加权系数。通过这种方法可以改进用户  $u_i$  和用户  $u_j$  间度量的准确性。

#### 1.3.3 多元信任关系的融合

吴宾等<sup>[23]</sup>认为在大规模数据集中, 正则化方法有利于快速搜索到合适的参数和较高的模型预测准确率。通过式(7)、(8)、(11)描述可知, 用户  $u_u$  和用户  $u_j$  受到直接和间接信任关系及用户之间相似关系的影响。因此用户信任的正则化模型定义如下:

$$C_{u,j}^+ = \sum_{u=1}^m \sum_{j \in N_u^+} (W_{u,j}^+ + S_{u,j}(u_u, u_j)) \|u_u - u_j\|^2 \quad (12)$$

$$C_{u,g}^- = \sum_{u=1}^m \sum_{g \in N_u^-} (W_{u,g}^- + S_{u,g}(u_u, u_g)) \|u_u - u_g\|^2 \quad (13)$$

式中,  $C_{u,j}^+$  为直接信任的用户社会化正则模型,  $C_{u,g}^-$  为间接信任的用户社会化正则模型。由式(12)、(13)可知:  $W_{u,j}^+ + S_{u,j}(u_u, u_j)$  描述用户  $u_i$  和用户  $u_j$  之间距离大小关系, 如果用户  $u_i$  和用户  $u_j$  之间距离较大, 那么,  $W_{u,j}^+ + S_{u,j}(u_u, u_j)$  值就偏小; 反之, 如果用户  $u_u$  和用户  $u_j$  之间距离较小, 那么,  $W_{u,j}^+ + S_{u,j}(u_u, u_j)$  值就偏大。同理,

$W_{u,g}^- + S_{u,g}(u_u, u_g)$ 也有类似的表述。

#### 1.4 构建社会化排序模型和伪代码实现

根据式(6)、(12)、(13)可以得到社会化排序推荐目标函数 $J_{\text{STrustBPR}}(\Theta)$ ,具体如下:

$$\begin{aligned} J_{\text{STrustBPR}}(\Theta) = & J_{\text{BPR}}(\Theta) + C_{u,j}^+ + C_{u,g}^- = \\ & \sum_{u \in m} \sum_{i \in n} \sum_{j \in n} \ln \sigma(\varphi_i^{(u)} - \varphi_j^{(u)}) + \\ & \frac{\lambda_u}{2} \|U_u\|^2 + \frac{\lambda_v}{2} \|V_i\|^2 + \frac{\lambda_v}{2} \|V_j\|^2 + \\ & \frac{\lambda_\alpha}{2} \sum_{u=1}^m \sum_{j \in N_u^+} (W_{u,j}^+ + S_{u,j}(u_u, u_j)) \|u_u - u_j\|^2 + \\ & \frac{\lambda_\alpha}{2} \sum_{u=1}^m \sum_{g \in N_u^-} (W_{u,g}^- + S_{u,g}(u_u, u_g)) \|u_u - u_g\|^2 \end{aligned} \quad (14)$$

现采用随机梯度下降方式求解 $J_{\text{STrustBPR}}(\Theta)$ 局部最小解,具体如下:

$$\begin{aligned} \frac{\partial J_{\text{STrustBPR}}}{\partial U_u} = & (1 - \sigma(\varphi_{i,j}^{(u)}))(V_i - V_j) + \lambda_u U_u + \\ & \lambda_\alpha \sum_{j \in N_u^+} (W_{u,j}^+ + S_{u,j}(u_u, u_j))(U_u - U_j) + \\ & \lambda_\alpha \sum_{g \in N_u^-} (W_{u,g}^- + S_{u,g}(u_u, u_g))(U_u - U_g) \end{aligned} \quad (15)$$

$$\frac{\partial J_{\text{STrustBPR}}}{\partial V_i} = U_u^T (1 - \sigma(\varphi_{i,j}^{(u)})) + \lambda_v V_i \quad (16)$$

$$\frac{\partial J_{\text{STrustBPR}}}{\partial V_j} = U_u^T (1 - \sigma(\varphi_{i,j}^{(u)})) + \lambda_v V_j \quad (17)$$

STrustBPR算法的伪代码如下:

Step 1 输入用户项目评分矩阵 $\mathbf{R}$ 、社会化关系矩阵 $\mathbf{T}$ 、隐特征向量的维数 $k$ 、最大迭代次数 $O$ 、学习率 $H$ 和参数 $\lambda_\alpha$ 。

Step 2 初始化 $U$ 、 $V$ 、 $C_{u,j}^+$ 和 $C_{u,j}^-$ 。

Step 3 While  $i \leq O$

从 $\pi(u)$ 中抽取 $(u; i, j)$ ;

根据式(15)更新 $U_u$ ,即

$$U_u \leftarrow U_u - H \times \frac{\partial J_{\text{STrustBPR}}}{\partial U_u};$$

根据式(16)更新 $V_i$ ,即

$$V_i \leftarrow V_i - H \times \frac{\partial J_{\text{STrustBPR}}}{\partial V_i};$$

根据式(17)更新 $V_j$ ,即

$$V_j \leftarrow V_j - H \times \frac{\partial J_{\text{STrustBPR}}}{\partial V_j};$$

$i \leftarrow i + 1$ 。

Step 4 输出 $U$ 和 $V$ 。

## 2 实验与分析

### 2.1 实验描述和算法参数设置

实验环境是Windows7-64位操作系统,程序语

言是python2.7。选择两个公开的数据集DouBan和FilmTrust进行STrustBPR算法的验证和性能对比。DouBan数据集是从中文社交网站豆瓣电影中抓取的一个电影评分数据集,其评分范围是[1,5],步长是1。FilmTrust数据集是从FilmTrust网站抓取的一个电影评分数据集,其评分范围是[0.5,4.0],步长是0.5。对DouBan和FilmTrust数据集,采用80%的数据作为训练数据集,其余20%的数据作为测试数据集。

DouBan数据集和FilmTrust数据集均包括用户评分数据和信任评分数据,具体见表1。从表1可以得出,DouBan数据集中的用户数、项目数和用户信任数均高于FilmTrust数据集的,这是因为DouBan数据集比FilmTrust数据集规模更大。

表 1 两个数据集的统计

Tab. 1 Statistics of two data sets

数据集	用户数	项目数	评分数
DouBan	129 490	58 541	16 830 839
FilmTrust	1 508	2 071	35 497
数据集	信任用户数	信任数	稀疏度/%
DouBan	129 490	1 692 952	99.778
FilmTrust	1 642	1 853	98.860

选择3种经典的排名评价指标<sup>[1,3]</sup>进行算法的对比分析。3种指标分别是精确率、平均准确率和归一化折损累计增益。

1)用 $P$ 表示精确率(precision),具体指推荐物品排序列表在 $L$ 之前有多少是准确的,其中, $L$ 为推荐物品列表的长度即文献[24]中的 $\text{top}_N$ 。具体公式如下:

$$P_u = \sum_{u \in U_{\text{test}}} \frac{|L_u \cap L_{\text{test},u}|}{L} \quad (18)$$

式中, $U_{\text{test}}$ 为测试集的用户集合, $L_u$ 为用户 $u$ 产生的推荐列表长度, $L_{\text{test},u}$ 为在测试集上用户 $u$ 的预测推荐列表长度。

2)用 $M$ 表示平均准确率(mean average precision, MAP),具体指对所有用户平均准确率的平均值。公式具体如下:

$$M_u = \frac{\sum_{u \in U} A_u}{|U_{\text{test}}|} \quad (19)$$

式中: $M_u$ 对排序位置极为敏感, $M_u$ 值越大,说明该算法的排序效果越好; $A$ 为平均精确率(average precision, AP),具体指用户 $u$ 平均准确率,即

$$A_u = \frac{\sum_{i=1}^L P_u}{L}。$$

3)用 $D_{\text{NDCG}}$ 表示归一化折损累计增益(normalized discounted cumulative gain, NDCG),具体公式如下:

$$D_{\text{NDCG},u} = \sum_{i \in U_{\text{test}}} \frac{D_{\text{DCG},u}}{D_{\text{NDCG},u}^{\max}} \quad (20)$$

式中: $D_{\text{NDCG},u}$ 为 $D_{\text{DCG},u}$ 值与最大的 $D_{\text{NDCG},u}^{\max}$ 之间的比率, $D_{\text{DCG},u}$ 值越大,说明算法的整体排序性能越好; $D_{\text{DCG}}$ 为折损累计增益(discounted cumulative gain, DCG), $D_{\text{DCG}}$ 计算公式如下:

$$D_{\text{DCG},u} = \sum_{i=1}^L \frac{2^e - 1}{\text{lb}(i+1)} \circ$$

式中, $e$ 为用户 $u$ 与项目 $i$ 之间的相关性。当 $e=1$ 时,表示用户 $u$ 与项目 $i$ 之间是相关的;当 $e=0$ 时,表示用户 $u$ 与项目 $i$ 之间是不相关。

选择以下5种常见算法作为对比算法:SBPR认为社会化推荐优于非社会化推荐;TBPR认为在社会化关系中信任之间的强弱关系可以对推荐结果产生重要的影响,并以此影响推荐列表的顺序;BPR通过重新定义用户项目矩阵,得到用户项目偏序关系,从而得出预测推荐列表的顺序;MostPopular<sup>[19]</sup>是根据项

目流行度对项目进行排序,项目流行度的高低决定项目的排列顺序;Random<sup>[19]</sup>指在测试集合中随机为用户选择项目进行推荐。

对比算法的设置参数如表2所示,其中,MostPopular和Random不需要设置参数。

表2 对比算法中参数设置

算法	DouBan	FilmTrust
SBPR	$\lambda_U = \lambda_V = 0.01, \lambda_S = 0.01$	$\lambda_U = \lambda_V = 0.01, \lambda_S = 0.01$
TBPR	$\lambda_U = \lambda_V = 0.01, \lambda_S = 0.01, \lambda_T = 0.01$	$\lambda_U = \lambda_V = 0.01, \lambda_S = 0.01, \lambda_T = 0.01$
BPR	$\lambda_U = \lambda_V = 0.01$	$\lambda_U = \lambda_V = 0.01$
STrustBPR	$\lambda_U = \lambda_V = 0.01, \lambda_\alpha = 0.1$	$\lambda_U = \lambda_V = 0.01, \lambda_\alpha = 0.1$

## 2.2 全数据集下L推荐测试

对比测试本文提出的STrustBPR算法与SBPR、TBPR、BPR和MostPopular算法的推荐效果。表3、4分别给出6种算法在DouBan和FilmTrust数据集及两个推荐列表长度L下的3种指标P、M和 $D_{\text{NDCG}}$ 结果。其中,L分别设置为5和10,隐特征维数k值设为10。对比算法的参数设置详见表2。

表3 不同算法在FilmTrust数据集上的指标评价结果

Tab. 3 Index evaluation results of different algorithms on FilmTrust dataset

算法	P		M		$D_{\text{NDCG}}$	
	L=5	L=10	L=5	L=10	L=5	L=10
STrustBPR	0.321 900 161	0.174 798 712	0.334 040 973	0.261 752 823	0.418 302 148	0.368 038 034
SBPR	0.303 338 633	0.167 488 076	0.312 129 703	0.240 892 817	0.389 309 868	0.341 596 319
TBPR	0.268 265 388	0.150 039 968	0.259 566 569	0.192 139 404	0.326 549 126	0.283 594 316
BPR	0.282 315 113	0.156 189 711	0.283 473 562	0.214 025 345	0.353 956 088	0.307 603 685
MostPopular	0.311 111 111	0.168 425 260	0.323 652 411	0.247 542 835	0.402 785 279	0.348 762 451
Random	0.002 535 658	0.001 901 743	0.001 278 394	0.000 974 364	0.002 699 941	0.002 617 803

表4 不同算法在DouBan数据集上的指标评价结果

Tab. 4 Index evaluation results of different algorithms on DouBan dataset

算法	P		M		$D_{\text{NDCG}}$	
	L=5	L=10	L=5	L=10	L=5	L=10
STrustBPR	0.191 004 498	0.156 821 589	0.142 125 187	0.101 985 750	0.205 986 514	0.179 124 441
SBPR	0.077 186 453	0.064 756 234	0.048 595 191	0.033 023 075	0.084 253 568	0.074 777 704
TBPR	0.121 500 938	0.103 377 111	0.079 138 837	0.056 130 377	0.126 158 781	0.113 466 107
BPR	0.067 484 207	0.063 805 277	0.036 662 125	0.026 867 219	0.068 041 333	0.066 583 466
MostPopular	0.159 211 016	0.123 595 087	0.115 137 597	0.077 278 911	0.172 396 698	0.144 325 342
Random	0.001 497 566	0.001 535 006	0.000 475 477	0.000 349 224	0.001 224 201	0.001 358 981

从表3和4中可以看出:

1)在DouBan和FilmTrust数据集中,STrustBPR、SBPR、TBPR、BPR和MostPopular算法的3种指标评价结果均优于Random。说明有效的排序算法完全优于随机排序算法。

2)3种排序评价指标中,所有算法在FilmTrust数

据集的推荐性能优于DouBan数据集的算法推荐性能。因为DouBan数据集比FilmTrust数据集更加稀疏,因此,不同数据集的稀疏性会影响到项目推荐的准确度。

3)在DouBan和FilmTrust数据集中,本文提出的STrustBPR算法的3种指标评价结果优于其他算法。

这说明本文提出的STrustBPR算法并不依赖某个具体的数据集。以 $L=5$ 为例: FilmTrust数据集中, 本文提出的STrustBPR算法与SBPR、TBPR、BPR、MostPopular和Random相比,  $P$ 指标分别提升了1.856%、5.363%、3.958%、1.079%和31.936%,  $M$ 分别提升了2.191%、7.447%、5.057%、1.039%和33.276%,  $D_{\text{NDCG}}$ 指标分别提升了4.302%、7.214%、5.16%、2.435%和22.586%; 在DouBan数据集中, 本文提出的STrustBPR算法与SBPR、TBPR、BPR、MostPopular和Random相比,  $P$ 指标分别提升了11.381%、6.95%、12.352%、3.179%和18.95%,  $M$ 指标分别提升了9.353%、6.299%、10.547%、2.699%和14.165%,  $D_{\text{NDCG}}$ 指标分别提升了12.174%、7.983%、13.795%、3.359%和20.477%。

表 5 不同算法在FilmTrust数据集上的用户冷启动的指标评价结果

Tab. 5 Table 5 Index evaluation results of different algorithms for user cold-start on FilmTrust dataset

算法	$P$		$M$		$D_{\text{NDCG}}$	
	$L=5$	$L=10$	$L=5$	$L=10$	$L=5$	$L=10$
STrustBPR	0.095 852 535	0.054 377 880	0.177 163 338	0.183 366 707	0.227 480 541	0.241 234 042
SBPR	0.075 675 676	0.047 747 748	0.141 773 023	0.151 475 583	0.184 457 959	0.204 961 283
TBPR	0.057 894 737	0.039 473 684	0.124 298 246	0.133 642 439	0.155 344 567	0.177 035 658
BPR	0.070 689 655	0.043 534 483	0.136 709 770	0.143 533 200	0.175 877 167	0.193 463 232
MostPopular	0.078 378 378	0.046 396 396	0.166 516 517	0.173 083 798	0.203 133 656	0.217 409 662
Random	0.000 556 566	0.000 454 545	0.000 757 576	0.000 757 576	0.001 619 124	0.001 619 124

表 6 不同算法在DouBan数据集上的用户冷启动的指标评价结果

Tab. 6 Index evaluation results of different algorithms for user cold-start on the Douban dataset

算法	$P$		$M$		$D_{\text{NDCG}}$	
	$L=5$	$L=10$	$L=5$	$L=10$	$L=5$	$L=10$
STrustBPR	0.009 022 556	0.006 015 038	0.011 027 569	0.012 698 413	0.017 345 754	0.021 280 849
SBPR	0.001 492 537	0.003 731 343	0.002 487 562	0.005 123 194	0.003 731 343	0.010 867 626
TBPR	0.002 877 698	0.003 305 785	0.003 597 122	0.007 231 405	0.005 506 192	0.011 976 962
BPR	0.001 481 481	0.002 222 222	0.000 370 370	0.001 031 746	0.001 118 664	0.003 307 423
MostPopular	0.008 450 704	0.004 225 352	0.005 221 387	0.005 639 098	0.010 101 688	0.011 489 475
Random	0.000 863 680	0.000 781 250	0.000 868 056	0.000 868 056	0.002 351 797	0.002 351 797

从表5和6中可以看出:

1) 当用户冷启动时, 用户评分信息在急剧下降, 只凭用户评分信息或者社会关系提升项目推荐性能是有限的。本文提出的STrustBPR算法在一定程度上, 可以通过用户直接关系和间接关系改善推荐列表的推荐结果。说明STrustBPR算法可以在一定程度上解决用户冷启动问题。

2) 本文提出的STrustBPR算法的3种指标评价结果完全优于SBPR、TBPR、BPR、MostPopular和Random算法的评价结果, 本文算法的 $P$ 、 $M$ 和 $D_{\text{NDCG}}$ 指标的推荐性能均有大幅度提升。以 $L=5$ 为例: 在FilmTrust数据集中, 本文提出的STrustBPR算法与SBPR、TBPR、

4) 对于本文提出的STrustBPR算法, 当 $L$ 不断增大时, 3种评价指标的度量性能会有不同程度的减弱, 但与其他5种算法相比本文算法的推荐性能仍然具有明显的优势。这也说明用户之间的直接关系和间接关系对于项目推荐有重要的影响。

### 2.3 在用户冷启动下 $L$ 推荐测试

根据文献[6,8], 评分数少于5的用户是冷启动用户。对比测试本文提出的STrustBPR算法与SBPR、TBPR、BPR和MostPopular算法在用户冷启动下的推荐效果。表5、6给出了用户冷启动时6种算法在DouBan和FilmTrust数据集及两个推荐列表长度 $L$ 下的3种指标 $P$ 、 $M$ 和 $D_{\text{NDCG}}$ 结果。其中,  $L$ 分别设为5和10, 隐特征维度值 $k$ 设为10。对比算法参数设置详见表2。

BPR、MostPopular和Random相比,  $P$ 指标分别提升了2.017%、3.796%、2.516%、1.747%和9.529%,  $M$ 指标分别提升了3.539%、5.286%、4.045%、1.064%和17.64%,  $D_{\text{NDCG}}$ 指标分别提升了4.302%、7.214%、5.16%、2.435%和22.586%; 在DouBan数据集中, 本文提出的STrustBPR算法与SBPR、TBPR、BPR、MostPopular和Random相比,  $P$ 指标分别提升了0.753%、0.614%、0.754%、0.057%和0.816%,  $M$ 指标分别提升了0.854%、0.743%、1.066%、0.581%和1.016%,  $D_{\text{NDCG}}$ 指标分别提升了1.362%、1.184%、1.623%、0.725%和1.5%。

3) 本文提出的STrustBPR算法在FilmTrust数据集的推荐性能表现明显优于在DouBan数据集中的表现。

### 3 结 论

本文提出了社会化信任排序推荐算法(STrustBPR)。首先,该算法通过矩阵分解技术预测用户对不同项目的偏好值,同时,通过BPR算法对不同项目进行个性化排序;最后,将用户之间相似关系以及信任用户之间的直接和间接关系融入进BPR算法中。在DouBan数据集和FilmTrust数据集上通过3个经典排序推荐指标 $P$ 、 $M$ 和 $D_{NDCG}$ 进行实验验证,结果表明本文提出的STrustBPR算法通过引入多元化的社会化信任关系能够取得更好的推荐结果,特别是在解决用户冷启动问题上推荐效果明显。

在后续的研究中,作者将进一步研究社交化网络中不同用户的社会化影响力和信任关系对用户偏序关系的影响,并提高STrustBPR算法的推荐性能。

#### 参考文献:

- [1] Liu Haiyang,Wang Zhihai,Huang Dan,et al.Listwise collaborative ranking based on the assumption of locally low-rank rating matrix[J].*Journal of Software*,2015,26(11):2981–2993.[刘海洋,王志海,黄丹,等.基于评分矩阵局部低秩假设的成列协同排名算法[J].*软件学报*,2015,26(11):2981–2993.]
- [2] Liu Hongzhi,Wu Zhonghai,Zhang Xing.CPLR:Collaborative pairwise learning to rank for personalized recommendation[J].*Knowledge-Based Systems*,2018,148:31–40.
- [3] Huang Zhenhua,Zhang Jiawen,Tian Chunqi,et al.Survey on learning-to-rank based recommendation algorithms[J].*Journal of Software*,2016,27(3):691–713.[黄震华,张佳雯,田春岐,等.基于排序学习的推荐算法研究综述[J].*软件学报*,2016,27(3):691–713.]
- [4] Gonzalez Camacho L A,Alves-Souza S N.Social network data to alleviate cold-start in recommender system:A systematic review[J].*Information Processing & Management*,2018,54(4):529–544.
- [5] Koren Y.Factorization meets the neighborhood:A multifaceted collaborative filtering model[C]//Proceedings of the Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'08).*New York:ACM*,2008:426–434.
- [6] Ma Hao,Zhou Dengyong,Liu Chao,et al.Recommender systems with social regularization[C]//Proceedings of the Fourth ACM International Conference on Web Search and Data Mining (WSDM'11).*New York:ACM*,2011:287–296.
- [7] Salakhutdinov R,Mnih A.Bayesian probabilistic matrix fac-

torization using Markov chain Monte Carlo[C]//Proceedings of the 25th International Conference on Machine Learning (ICML'08).*New York:ACM*,2008:880–887.

- [8] Koren Y.Collaborative filtering with temporal dynamics[J].*Communications of the ACM*,2010,53(4):89–97.
- [9] Guo Lei,Ma Jun,Jiang Haoran,et al.Social trust aware item recommendation for implicit feedback[J].*Journal of Computer Science and Technology*,2015,30(5):1039–1053.
- [10] Cao Yunbo,Xu Jun,Liu Tiejian,et al.Adapting ranking SVM to document retrieval[C]//Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'06).*New York:ACM*,2006:186–193.
- [11] Yang Shuanghong,Long Bo,Smola A J,et al.Collaborative competitive filtering:Learning recommender using context of user choice[C]//Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information (SIGIR'11).*New York:ACM*,2011:295–304.
- [12] Lee J,Bengio S,Kim S,et al.Local collaborative ranking[C]//Proceedings of the 23rd International Conference on World Wide Web (WWW'14).*New York:ACM*,2014:85–96.
- [13] Shi Yue,Karatzoglou A,Baltrunas L,et al.CLiMF:Learning to maximize reciprocal rank with collaborative less-is-more filtering[C]//Proceedings of the Sixth ACM Conference on Recommender Systems (RecSys'12).*New York:ACM*,2012:139–146.
- [14] Shi Yue,Larson M,Hanjalic A.List-wise learning to rank with matrix factorization for collaborative filtering[C]//Proceedings of the Fourth ACM Conference on Recommender Systems (RecSys'10).*New York:ACM*,2010:269–272.
- [15] Ardissono L,Mauro N.A compositional model of multi-faceted trust for personalized item recommendation[J].*Expert Systems with Applications*,2020,140:112880.
- [16] Zhao Feng,Shen Yu,Gui Xiangyu,et al.SDBPR:Social distance-aware Bayesian personalized ranking for recommendation[J].*Future Generation Computer Systems*,2019,95:372–381.
- [17] Liu Qinghua,Reiner A H,Frigessi A,et al.Diverse personalized recommendations with uncertainty from implicit preference data with the Bayesian Mallows model[J].*Knowledge-Based Systems*,2019,186:104960.
- [18] Pan Weike,Chen Li.GBPR:Group preference based Bayesian personalized ranking for one-class collaborative

- filtering[C]//Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI'13). [New York:ACM](#),2013:2691–2697.
- [19] Zhao Tong,McAuley J,King I.Leveraging social connections to improve personalized ranking for collaborative filtering[C]//Proceedings of the 23rd ACM International Conference on Information and Knowledge Management (CIKM'14).[New York:ACM](#),2014:261–270.
- [20] Wang Xin,Lu Wei,Ester M,et al.Social recommendation with strong and weak ties[C]//Proceedings of the 25th ACM International on Conference on Information and Knowledge Management.[New York:ACM](#),2016:5–14.
- [21] Wang Tianchun,Jin Xiaoming,Ding Xuetao,et al.User interests imbalance exploration in social recommendation:A fitness adaptation[C]//Proceedings of the 23rd ACM International Conference on Information and Knowledge Management (CIKM'14).[New York:ACM](#),2014:281–290.
- [22] Ma Hao.On measuring social friend interest similarities in recommender systems[C]//Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR'14).[New York:ACM](#), 2014:465–474.
- [23] Wu Bin,Lou Zhengzheng,Ye Yangdong.Co-regularized matrix factorization recommendation algorithm[J].[Journal of Software](#),2018,29(9):2681–2696.[吴宾,娄铮铮,叶阳东.联合正则化的矩阵分解推荐算法[J].[软件学报](#),2018,29(9):2681–2696.]
- [24] Guo Guibing,Zhang Jie,Zhu Feida,et al.Factored similarity models with social trust for top-N item recommendation[J].[Knowledge-Based Systems](#),2017,122:17–25.

(编辑 赵 婧)

引用格式:Zhang Li.Ranking recommendation algorithm based on social trust regularization[J].[Advanced Engineering Sciences](#), 2020,52(5):201–208.[张俐.基于社会信任正则化的排名推荐算法[J].[工程科学与技术](#),2020,52(5):201–208.]