

doi: 10.12052/gdutxb.170020

一种融合隐式信任关系的推荐算法

胡惠成, 陈平华

(广东工业大学 计算机学院, 广东 广州 510006)

摘要: 为解决传统协同过滤推荐算法存在的数据稀疏和冷启动问题, 解决显式信任数据难以获取以及数据稀疏问题, 提高推荐系统的准确率, 提出一种融合用户间隐式信任关系的矩阵分解推荐算法, 通过融合皮尔逊相关系数和信任因子, 计算用户间的隐式信任关系, 然后将隐式信任数据融入矩阵分解模型进行评分预测. 实验结果表明新算法能有效提高推荐结果的准确率.

关键词: 隐式信任; 矩阵分解; 推荐算法

中图分类号: TP311

文献标志码: A

文章编号: 1007-7162(2017)03-0043-06

Research on Implicit Trust Relationship Aware Recommendation Algorithm

Hu Hui-cheng, Chen Ping-hua

(School of Computers, Guangdong University of Technology, Guangzhou 510006, China)

Abstract: Collaborative filtering algorithm suffers from data sparsity and cold start, and explicit trust is more difficult to obtain and sparse. In order to solve these problems and improve the accuracy of recommendation systems, a matrix factorization recommendation algorithm is proposed integrating implicit trust information of users. This algorithm calculates implicit trust relationship between users by using Pearson correlation coefficient and factor of trust, and then integrates implicit trust information into a matrix factorization model to predict the ratings. Experimental results show that the algorithm has better recommendation accuracy.

Key words: implicit trust; matrix factorization; recommendation algorithm

随着互联网的迅速发展, 网上聚集了海量的数据信息, 然而, 用户却很难迅速从中找到其需要的信息, 出现了“信息过载”现象. 推荐系统是一种解决信息过载的有效方法. 协同过滤算法因其高效和便捷性被广泛地应用在推荐系统中, 其核心是找到与用户具有相似兴趣的用户群, 然后利用这些用户对物品的评分推测目标用户对物品的感兴趣程度. 协同过滤算法具有较好的推荐效果, 但其推荐准确率受限于数据稀疏性问题和冷启动问题. 数据稀疏性问题指的是在用户物品评分矩阵中用户和物品的数量十分庞大而用户对物品的评分数较少, 通过这样的稀疏矩阵计算得到的相似度不准确, 最终会影响推荐的质量; 而冷启动问题指的是对于新加入的用户或物品, 由于评分记录不存在从而不能为该用户进

行个性化推荐和无法将该物品推荐给对其感兴趣的用户.

矩阵分解技术(Matrix Factorization, 简称MF)是一种解决数据稀疏和冷启动问题的有效方法. 传统的矩阵分解技术包括: 奇异值分解(singular value decomposition, SVD)算法^[1]、概率矩阵分解(probabilistic matrix factorization, PMF)算法^[2]和非负矩阵分解(non-negative matrix factorization, NMF)算法^[3]等. 这些算法都是将一个高维矩阵通过降维的方法分解为两个或多个低维矩阵, 通过低维空间研究高维空间的性质, 这在一定程度上缓解了数据稀疏和冷启动问题.

随着社交网络的发展, 研究人员将信任关系引入推荐系统, 以此缓解数据稀疏和冷启动问题.

收稿日期: 2017-01-25

基金项目: 广东省科技计划项目(2016B030308001, 2016B030306002); 广州市科技计划项目(201604010099)

作者简介: 胡惠成(1988-), 男, 硕士研究生, 主要研究方向为推荐系统.

通信作者: 陈平华(1967-), 男, 教授, 硕士生导师, 主要研究方向为推荐系统、云计算技术. E-mail: phchen@gdut.edu.cn

Yang等^[4]提出的TrustMF模型对引入的用户信任网络信息进行细分,从信任和被信任的角度进行混合推荐;Jamali等^[5]提出的SocialMF方法考虑了信任在社交网络中的传播特性;刘英南等^[6]提出的推荐模型考虑了信任和不信任在社会网络中的传递和动态演化;肖晓丽等^[7]提出了一种融合用户兴趣和信任信息的聚类推荐算法;杜永萍等^[8]通过构建用户声望信任和局部信任的信任网络进行混合推荐.社交网络信任关系的引入,一定程度上缓解了数据稀疏和冷启动问题,并提高了推荐准确率.然而,由于目前大多数基于信任的推荐算法引入的都是用户的显式信任关系,显式信任信息需要用户明确指定,考虑到安全和隐私问题,通常较难获取,而且在存在大量用户的推荐系统中,只有少部分用户会明确给出其信任信息,所以显式信任也存在着数据稀疏的问题.为此,本文提出一种融合隐式信任关系的矩阵分解推荐算法,将从历史数据中挖掘出的隐式信任数据引入推荐系统,以此提高推荐准确率.

1 相关工作

1.1 信任的定义

Guo^[9]给出了信任在推荐系统中的定义,信任被定义为一个人对他人提供有价值评分的能力的信赖程度.根据文献[10],信任具有如下特征:(1)非对称性,比如用户 u 信任用户 v ,但并不代表用户 v 同样也信任用户 u ;(2)传递性,传递性是在基于信任的推荐系统中的一个非常重要的特性,它表示如果用户 u 信任用户 v ,同时用户 v 也信任用户 p ,那么在一定程度上可以认为用户 u 也信任用户 p ;(3)动态性,信任不是一成不变的,它会随着时间而不断变化;(4)上下文相关性,在不同的上下文环境下,用户的信任值是不同的,如一个用户在电影方面值得信任,但不代表他在其他方面也值得信任.

根据信任数据来源不同,可将信任分为显式信任和隐式信任.显示信任指的是用户明确给出对其他用户的信任值,范围一般在 $[0, 1]$ 之间.在公开的数据集中信任值一般都用0和1表示,其中0表示不信任,1表示信任;隐式信任通常是利用用户的历史评分数据来挖掘出用户间的信任关系.在关于隐式信任的研究中,O'Donovan^[11]提出了两种信任:Profile-level信任和Item-level信任.

1.2 矩阵分解算法

矩阵分解算法的基本原理是将一个用户-物品

评分矩阵 $R^{N \times M}$ 分解成两个低阶矩阵,即用户特征矩阵 $P^{d \times N}$ 和物品特征矩阵 $Q^{d \times M}$,其中 d 为特征数,则 $R \approx P^T Q$,如图1所示.

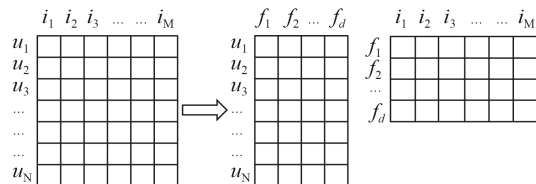


图1 用户-物品评分矩阵分解

Fig.1 The decomposition of user-item rating matrix

基于图1所给的用户-物品评分矩阵分解,可以得到用户 u 对物品 j 的预测评分 $\hat{r}_{u,j}$,如式(1)所示.

$$\hat{r}_{u,j} = \mathbf{q}_j^T \mathbf{p}_u = \sum_{k=1}^d q_{j,k} p_{u,k}. \quad (1)$$

其中向量 $\mathbf{p}_u$ 为用户特征矩阵 P 中每个用户所对应的向量,向量 $\mathbf{q}_j$ 为物品特征矩阵 Q 中每个物品所对应的向量, $p_{u,k}$ 表示用户 u 在第 k 个特征的兴趣度, $q_{j,k}$ 表示物品 j 拥有第 k 个特征的程度.另外由于在用户评分矩阵中,每个用户存在着不同的评分标准,即有些用户的评分比较严格,导致某些物品的评分较低,而有些用户的评分比较宽松,导致某些物品的评分较高.为了解决这个问题,Koren Y^[12]提出基准偏移量 $b_{u,j}$,更新后的评分预测模型为

$$\hat{r}_{u,j} = b_{u,j} + \mathbf{q}_j^T \mathbf{p}_u. \quad (2)$$

其中 $b_{u,j} = b_u + b_j + \mu$; b_u 、 b_j 分别代表用户 u 和物品 j 的评分偏置; μ 表示全局平均评分.

用户-物品评分矩阵中不仅包含用户对物品的评分,还包含一些额外的用户隐式反馈信息,这些隐式反馈信息告诉我们用户对哪些物品进行了评分,它反映了用户的偏好,从中可挖掘出用户对某个特征的喜好程度.通过使用这些隐式反馈信息,可以更好地探索用户的行为,能有效缓解数据稀疏和冷启动问题,提高推荐的准确度.为此,SVD++算法^[12]在SVD算法的基础上,引入了用户隐式反馈信息,通过增加另一个物品特征集合,即对每一个物品 i ,都有一个与之相关联的 d 维特征向量 $\mathbf{y}_i \in \mathbf{R}^d$,其评分预测模型为

$$\hat{r}_{u,j} = b_{u,j} + \mathbf{q}_j^T (\mathbf{p}_u + |\mathbf{I}_u|^{-\frac{1}{2}} \sum_{i \in \mathbf{I}_u} \mathbf{y}_i). \quad (3)$$

其中 $\mathbf{I}_u$ 为用户 u 评分过的物品集, $\mathbf{p}_u + |\mathbf{I}_u|^{-\frac{1}{2}} \sum_{i \in \mathbf{I}_u} \mathbf{y}_i$ 表示用户 u 的特征偏好程度.

为了进一步缓解数据稀疏和冷启动问题,提高

推荐准确率,信任关系被引入了推荐系统. TrustSVD^[13]在SVD++算法的基础上引入用户间的显式信任关系,其基本原理是认为用户更倾向于喜欢他们信任的用户的推荐,在评分上表现为用户的评分更受其信任用户的影响,TrustSVD将这种影响引入用户的评分预测中. 算法的基本流程如图2所示.

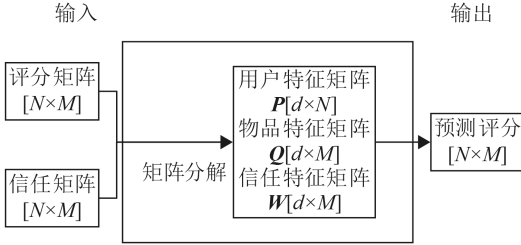


图2 TrustSVD算法流程图

Fig.2 Flowchat of TrustSVD

通过对用户评分矩阵 R 和信任矩阵 T 进行矩阵分解,即 $R \approx P^T Q$, $T \approx P^T W$, $\hat{r}_{u,v} = w_v^T p_u$, 其评分预测模型为

$$\hat{r}_{u,j} = b_{u,j} + q_j^T (p_u + |I_u|^{-\frac{1}{2}} \sum_{i \in I_u} y_i + |T_u|^{-\frac{1}{2}} \sum_{v \in T_u} w_v). \quad (4)$$

其中 w_v 为信任特征矩阵 W 中每个被信任者所对应的向量, $q_j^T w_v$ 可解释为用户 u 受其信任用户 v 对其在物品 j 上的评分的影响, T_u 表示用户 u 的显式信任集合.

2 融合隐式信任的矩阵分解推荐算法

由于在TrustSVD算法中引入的信任关系是显式信任关系,而显式信任存在着较难获取和数据稀疏的问题,为此,本文提出一种融合隐式信任关系的矩阵分解推荐算法,在SVD++算法的基础上引入从用户的历史数据中挖掘出的隐式信任信息来提高推荐准确率. 该算法的基本流程如图3所示,整个流程可分为两个部分,其中第一部分是隐式信任度的计算,主要是根据用户-物品评分矩阵计算用户间的皮尔

逊相关系数和每个用户在整个推荐系统中的信任因子^[14],然后通过调和信任度得到用户间的隐式信任度,最终形成隐式信任矩阵;第二部分是对评分矩阵和信任矩阵进行矩阵分解,最终得到用户特征矩阵、物品特征矩阵和信任特征矩阵.

2.1 隐式信任的度量

2.1.1 皮尔逊相关系数

在传统的推荐算法中,皮尔逊相关系数是计算用户相似度的一种经典算法,其相关公式为

$$S_{u,v} = \frac{\sum_i (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_i (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_i (r_{v,i} - \bar{r}_v)^2}}. \quad (5)$$

其中, $r_{u,j}$ 和 $r_{v,j}$ 分别表示用户 u 和用户 v 对项目 i 的评分, $\bar{r}_u$ 和 $\bar{r}_v$ 表示用户 u 和用户 v 的总体评分的平均值.

Papagelis M等^[15]通过皮尔逊相关系数来度量用户间的信任值. 对于给定阈值,当皮尔逊相关系数大于它时,则认为该值为用户间的信任值,否则为0.

$$T'_{u,v} = \begin{cases} S_{u,v}, & \text{if } S_{u,v} > \theta_s, |I_{u,v}| > \theta_t; \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

其中 θ_s 表示用户相似度的阈值, $I_{u,v}$ 表示用户 u 和用户 v 的共同评分项目数, θ_t 表示共同评分项目数阈值. Sotos等^[16]通过实验证明只有当 $\theta_s = 0.707$ 时,即 $S_{u,v} > 0.707$, 该信任值才具有传递性,因此取 $\theta_s = 0.707$. 另外取 $\theta_t = 2$, 表示当用户共同评分数大于2时,该相关系数才有意义.

2.1.2 信任因子

信任因子^[14]可以用来表示用户在整个推荐系统中被其他用户信任的程度,其取值范围在 $[0, 1]$ 之间,信任因子越大,则表示该用户越值得信任.

在实际生活中,新人往往倾向于向别人提出咨询. 对于在系统中积极评价的用户,其信任值要高于消极评价的用户的信任值. 假设 f_u 为用户 u 对项目的评价个数,当 f_u 的值越高时,则说明用户 u 越受其他用户信任;反之,当 f_u 的值越低时,则说明用户 u 受其他用户的信任程度也越低. 设 q_u 为用户 u 对其他用户的最近邻次数,次数越多,表明越受其他用户信任;反之,则说明该用户越不受其他用户信任.

设 T_u 为用户 u 的信任因子,其计算公式可表示为

$$T_u = \frac{2(1 - 1/\ln f_u)(1 - 1/\ln(q_u + 3))}{2 - 1/\ln f_u - 1/\ln(q_u + 3)}. \quad (7)$$

2.1.3 调和信任度

本文的调和信任度同时考虑皮尔逊相关系数和信任因子两方面对权重的影响,采用调和平均数公式,其计算公式为

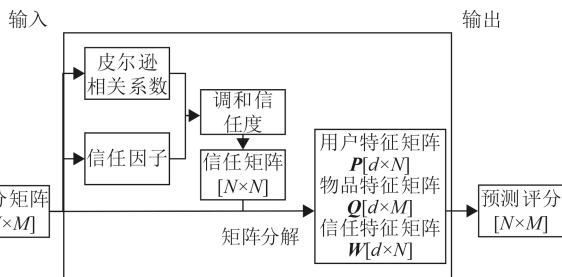


图3 融合隐式信任关系的矩阵分解算法流程图

Fig.3 Flowchat of matrix factorization algorithm integrated implicit trust relationship

$$T_{u,v} = \frac{2T'_{u,v} \times T_v}{T'_{u,v} + T_v}. \quad (8)$$

其中, $T_{u,v}$ 表示用户 u 对用户 v 的调和信任度, $T'_{u,v}$ 表示用户 u 和用户 v 的皮尔逊相关系数, T_v 表示用户 v 的信任因子. 隐式信任度计算算法的具体描述如下所示.

算法1 隐式信任度计算算法

输入: 用户物品评分矩阵 $\mathbf{R}$

输出: 隐式信任度

```

1 根据公式(1)计算用户  $u$  和用户  $v$  的相似  $sim(u, v)$ 
2 if  $sim(u, v) > \theta_s$ 
3  $trust1(u, v) = sim(u, v)$ 
4 else
5  $trust1(u, v) = 0$ 
6 计算用户  $v$  的评分项目数  $count$ 
7  $q = neighbourTimes[v]$  // 用户  $v$  作为其它用户的最近邻次数
8  $trust2(v) = 2 * (1 - 1 / \log(count)) * (1 - 1 / \log(q + 3)) / (2 - 1 / \log(count) - 1 / \log(q + 3))$ 
9  $trust(u, v) = 2 * trust1(u, v) * trust2(v) / (trust1(u, v) + trust2(v))$ 

```

2.2 评分预测

本文提出的算法的评分预测模型为

$$\hat{r}_{u,j} = b_{u,j} + \mathbf{q}_j^T (\mathbf{p}_u + |\mathbf{I}_u|^{-\frac{1}{2}} \sum_{i \in I_u} \mathbf{y}_i + |\mathbf{T}_u^I|^{-\frac{1}{2}} \sum_{v \in T_u^I} \mathbf{w}_v^I). \quad (9)$$

其中 T_u^I 表示用户 u 的隐式信任集合; $\mathbf{q}_j^T$ 、 $\mathbf{w}_v^I$ 可解释为用户 u 受其隐式信任用户 v 对其在物品 j 的评分的影响.

为了避免拟合问题, 该模型的损失函数将采用 TrustSVD 算法的正则化策略, 即给予评分数较多的用户和流行物品更小的惩罚, 给予冷启动用户和物品更大的规范, 得到的目标函数为

$$\begin{aligned}
L = & \frac{1}{2} \sum_u \sum_{j \in I_u} (\hat{r}_{u,j} - r_{u,j})^2 + \frac{\lambda}{2} \sum_u \sum_{v \in T_u^I} (\hat{t}_{u,v} - t_{u,v}^I)^2 + \\
& \frac{\lambda}{2} \sum_u |\mathbf{I}_u|^{-\frac{1}{2}} b_u^2 + \frac{\lambda}{2} \sum_j |\mathbf{J}_j|^{-\frac{1}{2}} b_j^2 + \\
& \sum_u \left(\frac{\lambda}{2} |\mathbf{I}_u|^{-\frac{1}{2}} + \frac{\lambda}{2} |\mathbf{T}_u^I|^{-\frac{1}{2}} \right) \|\mathbf{p}_u\|_F^2 + \\
& \frac{\lambda}{2} \sum_j |\mathbf{J}_j|^{-\frac{1}{2}} \|\mathbf{q}_j\|_F^2 + \frac{\lambda}{2} \sum_i |\mathbf{U}_i|^{-\frac{1}{2}} \|\mathbf{y}_i\|_F^2 + \\
& \frac{\lambda}{2} |\mathbf{T}_v^I|^{-\frac{1}{2}} \|\mathbf{w}_v^I\|_F^2. \quad (10)
\end{aligned}$$

其中, U_i 、 U_j 分别为对物品 i 和物品 j 评过分的用户集合; T_v^I 为隐式信任用户 v 的用户集合.

为了得到损失函数在最小值情况下的 $\mathbf{P}$ 、 $\mathbf{Q}$ 和 $\mathbf{W}$, 采用随机梯度下降法进行求解, 算法的具体描述如下所示.

算法2 融合隐式算法的矩阵分解推荐算法

输入: 用户-物品评分矩阵 $\mathbf{R}$, 隐式信任矩阵 $\mathbf{T}$, 学习率 α , 最大迭代次数 K

输出: 潜在特征矩阵 $\mathbf{P}$ 、 $\mathbf{Q}$ 、 $\mathbf{W}$

```

1 初始化  $\mathbf{P}$ ,  $\mathbf{Q}$ ,  $\mathbf{W}$ ,  $b_u$ ,  $b_j$ 
2 for  $iter = 1, 2, \dots, K$  do
3 for each  $\langle u, j \rangle \in \mathbf{R}$ , each  $\langle u, v \rangle \in \mathbf{T}$ 
4 根据公式(9)计算用户  $u$  对物品  $j$  的预测评分  $\hat{r}_{u,j}$ 
5 计算预测评分和实际评分的差值  $e_{u,j} = \hat{r}_{u,j} - r_{u,j}$ 
6 计算预测信任度和实际信任度的差值  $e_{u,v} = \hat{t}_{u,v} - t_{u,v}^I$ 
7  $b_u \leftarrow b_u - \alpha(e_{u,j} + \lambda |\mathbf{I}_u|^{-\frac{1}{2}} b_u)$ 
8  $b_j \leftarrow b_j - \alpha(e_{u,j} + \lambda |\mathbf{J}_j|^{-\frac{1}{2}} b_j)$ 
9  $\mathbf{p}_u \leftarrow \mathbf{p}_u - \alpha(e_{u,j} \mathbf{q}_j + \lambda_t^I e_{u,v} \mathbf{w}_v^I + (\lambda |\mathbf{I}_u|^{-\frac{1}{2}} + \lambda_t^I |\mathbf{T}_u^I|^{-\frac{1}{2}}) \mathbf{p}_u)$ 
10  $\mathbf{q}_j \leftarrow \mathbf{q}_j - \alpha(e_{u,j} (\mathbf{p}_u + |\mathbf{I}_u|^{-\frac{1}{2}} \sum_{i \in I_u} \mathbf{y}_i + |\mathbf{T}_u^I|^{-\frac{1}{2}} \sum_{v \in T_u^I} \mathbf{w}_v^I + \lambda |\mathbf{J}_j|^{-\frac{1}{2}} \mathbf{q}_j))$ 
11  $\forall i \in I_u: \mathbf{y}_i \leftarrow \mathbf{y}_i - \alpha(e_{u,j} |\mathbf{I}_u|^{-\frac{1}{2}} \mathbf{q}_j + \lambda |\mathbf{U}_i|^{-\frac{1}{2}} \mathbf{y}_i)$ 
12  $\forall v \in T_u^I: \mathbf{w}_v^I \leftarrow \mathbf{w}_v^I - \alpha(e_{u,j} |\mathbf{T}_u^I|^{-\frac{1}{2}} \mathbf{q}_j + \lambda_t^I e_{u,v} \mathbf{p}_u + \lambda |\mathbf{T}_v^I|^{-\frac{1}{2}} \mathbf{w}_v^I)$ 
13 根据上面所计算出的值更新潜在特征矩阵  $\mathbf{P}$ 、 $\mathbf{Q}$ 、 $\mathbf{W}$ 
14 end for
15 end for

```

3 实验结果及分析

3.1 数据集

本实验所采用的数据集是 FilmTrust^[17], 表1给出了该数据集的具体信息.

表1 数据集 FilmTrust 统计信息

Tab.1 statistics of data set FilmTrust

评分数据				信任数据			
用户数	物品数	评分数	评分密度/%	信任者	被信任者	信任数	信任密度/%
1 508	2 071	35 497	1.14	609	732	1 853	0.42

3.2 评价标准

本文采用平均绝对偏差 MAE 和均方根误差 RMSE 作为度量标准, 当 MAE 或者 RMSE 越小时, 则表明推荐准确率越高. 设预测的用户评分集合为 $\{P_1, P_2, \dots, P_N\}$, 实际的用户评分集合为 $\{R_1, R_2, \dots, R_N\}$, 则计算公式为

$$MAE = \frac{\sum_{i=1}^N |P_i - R_i|}{N}, \quad (11)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (P_i - R_i)^2}{N}}. \quad (12)$$

3.3 实验结果及分析

为了验证本文推荐算法 ITrustSVD 的效果, 本实验将与 SocialMF 方法、TrustMF 方法、SVD++ 方法和 TrustSVD 方法进行比较. 本文的推荐算法 ITrustSVD 的参数设置如下: $\lambda=0.9$, $\lambda_t=0.8$, 学习率设为 0.01, 迭代次数 100 次, 特征数 d 分别取值 5 和 10. 另

外对于其他方法的参数设置如下:对于SocialMF方法, $\lambda_u=\lambda_v=0.001$, $\lambda_i=1$; 对于TrustMF方法, $\lambda=0.001$, $\lambda_i=1$; 对于SVD++方法, $\lambda=0.1$; 对于TrustSVD方法, $\lambda=1.2$, $\lambda_i=0.9$. 本实验采用5折交叉验证, 每次试验随机选取80%的数据作为训练集, 并将剩下的20%的数据作为测试集, 实验结果如表2所示.

表2 实验结果对比
Tab.2 Comparison of experiment results

	度量	SocialMF	TrustMF	SVD++	TrustSVD	ITrustSVD
$d=5$	MAE	0.638	0.631	0.613	0.611	0.608
	RMSE	0.837	0.810	0.806	0.792	0.790
$d=10$	MAE	0.642	0.631	0.611	0.609	0.606
	RMSE	0.844	0.819	0.802	0.789	0.788

由实验结果可知, 在不同的特征数下, 算法ITrustSVD相比其它算法在MAE和RMSE都有一定的提升, 特别是与传统的加入隐式反馈信息的SVD++算法和基于显式信任的TrustSVD算法相比, 算法ITrustSVD都稍优于上述两种算法. 分析原因可知, 本文的隐式信任关系是由皮尔逊相关系数和用户在整个推荐系统中的信任度共同推断得出的, 它与用户的偏好和用户在推荐系统中让人信赖的程度有关. 在显式信任较难获取和存在数据稀疏的情况下, 引入用户的隐式信任关系对提升整体推荐效果有积极的作用, 能在一定程度上提高推荐的准确率.

3.4 近邻个数对推荐结果的影响

本文的隐式信任值由皮尔逊相关系数和信任因子两部分构成, 而信任因子的影响因素包含用户对项目的评价个数和用户作为其他用户的最近邻次数, 不同的近邻个数会导致最近邻次数的不同, 最终影响信任因子的结果. 在实验中分别选择近邻个数在5、10、15、20、25、30的情况下, 对MAE进行计算分析, 实验结果如图4所示. 当近邻个数为30时, MAE的值最低, 因此在实验中将近邻个数的值设为30.

算得到的调和信任度来度量用户之间的隐式信任关系, 在一定程度上缓解显式信任较难获取和数据稀疏的缺点. 实验结果表明该算法能提高推荐结果准确率, 能更好地为用户做出推荐. 然而, 随着数据规模的扩大、迭代次数的增加, 算法所需的计算时间也不断增加, 因此下一步的工作重点是对算法进行分布式实现.

参考文献:

[1] FUNK S. Netflix update: try this at home[EB/OL]. (2011)[2017-01-20]. <http://sifter.org/~simon/journal/20061211.html>.

[2] SALAKHUTDINOV R, MNIH A. Probabilistic matrix factorization [J]. Advances in Neural Information Processing Systems, 2007: 1257-1264.

[3] LEE D D, SEUNG H S. Algorithms for non-negative matrix factorization [J]. Proceedings of Advances in Neural and Information Processing Systems, 2000, 32(6): 556-562.

[4] YANG B, LEI Y, LIU D, *et al*. Social collaborative filtering by trust[C]//Proceedings of the 23rd International Joint Conference on Artificial Intelligence. [S.l.]: AAAI Press, 2013: 2747-2753.

[5] JAMALI M, ESTER M. A matrix factorization technique with trust propagation for recommendation in social networks[C]//Proceedings of the Fourth ACM Conference on Recommender Systems, Recsys 2010, Barcelona, Spain: [s.n.], 2010: 135-142.

[6] 刘英南, 谢瑾奎, 张家利, 等. 社交网络中基于信任的推荐算法[J]. 小型微型计算机系统, 2015, 36(6): 1165-1170.

LIU Y N, XIE J K, ZHANG J L, *et al*. Recommendation algorithm based on trust in social network [J]. Journal of Chinese Computer Systems, 2015, 36(6): 1165-1170.

[7] 肖晓丽, 钱娅丽, 李旦江, 等. 基于用户兴趣和社交信任的聚类推荐算法[J]. 计算机应用, 2016, 36(5): 1273-1278.

XIAO X L, QIAN Y L, LI D J, *et al*. Clustering recommendation algorithm based on user interest and social trust [J]. Journal of Computer Applications, 2016, 36(5): 1273-1278.

[8] 杜永萍, 黄亮, 何明. 融合信任计算的协同过滤推荐方法

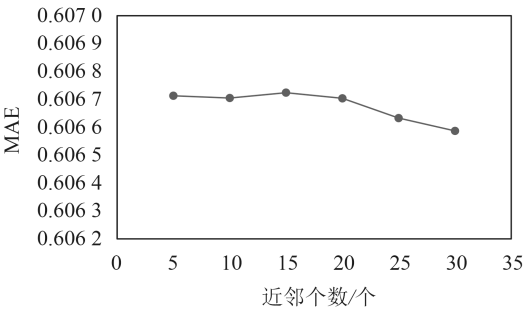


图4 近邻个数对MAE的影响

Fig.4 Influence of neighbor number for MAE

4 结束语

本文提出一种融合用户隐式信任关系的矩阵分解推荐算法, 通过由皮尔逊相关系数和信任因子计

[J]. 模式识别与人工智能, 2014, 27(5): 417-425.

DU Y P, HUANG L, HE M. Collaborative filtration recommendation algorithm based on trust computation [J]. Pattern Recognition and Artificial Intelligence, 2014, 27(5): 417-425.

[9] GUO G B. Integrating trust and similarity to ameliorate the data sparsity and cold start for recommender systems[C]// Proceedings of the 7th ACM Conference on Recommender Systems. Hong Kong: ACM, 2013: 451-454.

[10] GUO G B, ZHANG J, THALMANN D, *et al.* From ratings to trust: an empirical study of implicit trust in recommender systems[C]//Proceedings of the 29th Annual ACM Symposium on Applied Computing. Gyeongju, Korea: ACM, 2014: 248-253.

[11] O'DONOVAN J, SMYTH B. Trust in recommender systems[C]//Proceedings of the 2005 International Conference on Intelligent User Interfaces. San Diego, California, USA: ACM, 2005.

[12] KOREN Y. Factorization meets the neighborhood: a multifaceted collaborative filtering model[C]//Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Las Vegas, USA: ACM, 2008: 426-434.

[13] GUO G, ZHANG J, YORKE-SMITH N. TrustSVD: collaborative filtering with both the explicit and implicit influence of user trust and of item ratings[C]//Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. [S.l.]: AAAI Press, 2015: 123-129.

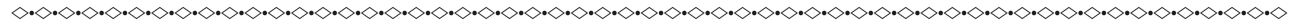
[14] 郭艳红, 邓贵仕, 雒春雨. 基于信任因子的协同过滤推荐算法[J]. 计算机工程, 2008, 34(20): 1-3.

GUO Y H, DENG G S, LUO C Y. Collaborative filtering recommendation algorithm based on factor of Trust [J]. Computer Engineering, 2008, 34(20): 1-3.

[15] PAPAGELIS M, PLEXOUSAKIS D, KUTSURAS T. Alleviating the sparsity problem of collaborative filtering using trust inferences[C]// Trust Management, Third International Conference, iTrust 2005. Paris, France: Springer, 2005: 224-239.

[16] SOTOS A, VANHOOF S, VAN DEN NOORTGATE W, *et al.* The transitivity misconception of pearson's correlation coefficient [J]. Statistics Education Research Journal, 2009, 8(2): 33-55.

[17] GUO G B, ZHANG J, YORKE-SMITH N. A novel Bayesian similarity measure for recommender systems[C]// Proceedings of the 23rd International Joint Conference on Artificial Intelligence. Beijing, China: AAAI, 2013: 2619-2625.



(上接第20页)

[16] 曾碧, 林展鹏, 邓杰航. 自主移动机器人走廊识别算法研究与改进[J]. 广东工业大学学报, 2016, 33(5): 9-21.

ZENG B, LIN Z P, DENG J H. Algorithm research on recognition and improvement for corridor of autonomous mobile robot [J]. Journal of Guangdong University of Technology, 2016, 33(5): 9-21.

[17] 陆靖桥, 傅秀芬, 蒙在桥. 复杂网络的鲁棒性与中心性指标的研究[J]. 广东工业大学学报, 2016, 33(4): 302-308.

LU J Q, FU X F, MENG Z Q. Research on robustness and centrality metrics of complex networks [J]. Journal of

Guangdong University of Technology, 2016, 33(4): 302-308.

[18] PAGEL, BRINS, MOTWASNIR, *et al.* The PageRank citation ranking: bringing order to the web[R]. [S.l.]: Stanford InfoLab, 1999.

[19] 金宏桥, 董一鸿. 大数据下图三角计算的研究进展[J]. 电信科学, 2016(6): 153-162.

JIN H Q, DONG Y H. Research progress of triangle counting in big data [J]. Telecommunications Science, 2016(6): 153-162.



(上接第35页)

[20] WORBOYS M F. Communicating geographic information in context[C]//DUCKHAM M, GOODCHILD M, WORBOYS M F, *et al.* Foundations of Geographic Information Science. London and New York: Taylor & Francis, 2003: 33-45.

[21] 董振东, 董强. 知网和汉语研究[J]. 当代语言学, 2001, 3(1): 33-44.

DONG Z D, DONG Q. Hownet and Chinese research [J]. Contemporary linguistics, 2001, 3(1): 33-44.

[22] 董振东, 董强, 郝长伶. 知网的理论发现[J]. 中文信息学

报, 2007, 21(4): 3-9.

DONG Z D, DONG Q, HAO C L. Theoretical findings of HowNet [J]. Journal of Chinese information processing, 2007, 21(4): 3-9.

[23] 李景, 苏晓鹭, 钱平. 构建领域本体的方法[J]. 计算机与农业, 2003(7): 7-10.

LI J, SU X L, QIAN P. The methodology of developing domain ontology [J]. Computer and Agriculture, 2003(7): 7-10.

[24] 陈珊珊. 基于语义的大学生就业推荐系统研究[D]. 武汉: 武汉科技大学信息科学与工程学院, 2014: 1-46.