

基于评分填充与信任信息的混合推荐算法

沈学利¹, 李子健^{2*}, 赫辰皓²

(1. 辽宁工程技术大学 软件学院, 辽宁 葫芦岛 125105; 2. 辽宁工程技术大学 电子与信息工程学院, 辽宁 葫芦岛 125105)

(* 通信作者电子邮箱 zijianli1122@163.com)

摘要:针对推荐系统的数据稀疏性导致的推荐效果不佳的问题,提出一种基于评分填充与信任信息的混合推荐的算法 RTWSO(Real-value user item restricted Boltzmann machine Trust WSO)。首先,使用改进的受限玻尔兹曼机模型对评分矩阵进行填充,以缓解评分矩阵的稀疏性问题;其次,从信任关系中提取信任与被信任关系,并通过基于矩阵分解的隐含信任关系相似度来解决信任信息稀疏的问题,而且对原有算法进行了包含信任信息的修正,以提高推荐准确度;最后,通过加权 Slope One(WSO)算法对矩阵填充与信任相似度信息加以整合,并对评分数据进行预测。在 Epinions 与 Ciao 数据集集中验证算法性能,可见所提出混合推荐算法较组成算法在推荐准确度上提升 3% 以上,较现有社会化推荐算法 SocialIT(Social recommendation algorithm based on Implicit similarity in Trust)在推荐准确度上提升 1.2% 以上。实验结果表明,所提出的基于评分填充与信任信息的混合推荐算法在一定程度上提高了推荐准确度。

关键词:受限玻尔兹曼机;加权 Slope One;用户信任相似度;矩阵分解;评分预测

中图分类号:TP391 **文献标志码:**A

Hybrid recommendation algorithm based on rating filling and trust information

SHEN Xueli¹, LI Zijian^{2*}, HE Chenhao²

(1. College of Software, Liaoning Technical University, Huludao Liaoning 125105, China;

2. College of Electronic and Information Engineering, Liaoning Technical University, Huludao Liaoning 125105, China)

Abstract: Aiming at the problem of poor recommendation effect caused by the data sparsity of the recommendation system, a hybrid recommendation algorithm based on rating filling and trust information was proposed namely RTWSO (Real-value user item restricted Boltzmann machine Trust Weighted Slope One). Firstly, the improved restricted Boltzmann machine model was used to fill the rating matrix, so as to alleviate the sparseness problem of the rating matrix. Secondly, the trust and trusted relationships were extracted from the trust relationship, and the matrix decomposition based implicit trust relationship similarity was also used to solve the problem of trust relationship sparsity. The modification including trust information was performed to the original algorithm, improving the recommendation accuracy. Finally, the Weighted Slope One (WSO) algorithm was used to integrate the matrix filling and trust similarity information as well as predict the rating data. The performance of the proposed hybrid recommendation algorithm was verified on Epinions and Ciao datasets. It can be seen that the proposed hybrid recommendation algorithm has the recommendation accuracy improved by more than 3% compared with the composition algorithm, and recommendation accuracy increased by more than 1.2% compared with the existing social recommendation algorithm SocialIT (Social recommendation algorithm based on Implicit similarity in Trust). Experimental results show that the proposed hybrid recommendation method based on rating filling and trust information, improves the recommended accuracy to a certain extent.

Key words: Restricted Boltzmann Machine (RBM); Weighted Slope One (WSO); user trust similarity; matrix decomposition; rating prediction

0 引言

我们生活在一个信息时代,每天都有越来越多的书籍、期刊文章、网页和电影问世。越来越庞大的数据总量造成了信息筛选的困难,这便是信息过载问题。推荐系统是解决信息过载问题的一个重要手段,然而推荐系统长期面临着数据稀疏性的问题。造成此种问题的主要原因是用户仅仅为其交互过的少数项目评分,常用数据集的评分稀疏度均在 5% 以下,而一些真实数据集的评分稀疏度甚至不足 0.1%,评分数据的

大量缺失造成推荐算法提取数据关系时缺乏足够的依据,造成推荐偏差。

为了缓解数据稀疏性问题,研究者们主要从以下两个方面进行改进:一方面,使用混合推荐的方法。利用一种算法对评分矩阵中缺失值进行填充,再使用另一种方法基于填充后的矩阵上生成预测结果。比如杜倩^[1]使用聚类获取用户近邻的评分均值实现评分填充;Liu 等^[2]使用奇异值分解(Singular Value Decomposition, SVD)算法实现矩阵缺失值的填充。另

收稿日期:2020-03-14;修回日期:2020-04-27;录用日期:2020-05-14。 基金项目:国家自然科学基金资助项目(61772249)。

作者简介:沈学利(1969—),男,江苏连云港人,教授,硕士,CCF 会员,主要研究方向:信息安全、推荐系统;李子健(1994—),男,辽宁朝阳人,硕士研究生,主要研究方向:推荐系统;赫辰皓(1992—),男,河北石家庄人,硕士,主要研究方向:深度学习、推荐系统。

一方面,在评分信息之外,通过引入辅助信息(side information),增加算法中的有效信息量,从而达到缓解数据稀疏性的目的。比如,刘方婷等^[3-5]分别将时间因素、社交关系与用户(项目)属性信息引入到推荐算法中。

上述两种方法从两个不同的角度,对于数据稀疏性问题起到了一定的缓解作用。为了进一步提高推荐准确度,本文对两种方法分别加以改进与融合,提出一种基于评分填充与信任信息的混合推荐算法 RTWSO (Real-value user item restricted Boltzmann machine Trust WSO):

首先,使用基于用户与项目的实值受限玻尔兹曼机 (Real-value User Item Restricted Boltzmann Machine, RUI-RBM) 算法对评分矩阵进行填充。受限玻尔兹曼机 (Restricted Boltzmann Machine, RBM) 是一种二层神经网络模型,由 Salakhutdinov 等^[6]于 2007 年首次应用在推荐系统中,取得了良好的推荐效果。本文提出使用一种基于实值的,同时对项目与用户建模的受限玻尔兹曼机 (RUI-RBM) 模型对用户-项目评分矩阵进行填充,以达到更加精确的评分填充效果。

其次,使用填充完整的评分矩阵,通过加权 Slope One (Weighted Slope One, WSO) 算法对评分进行预测。本文算法中的权值将使用信任信息相似度。本文提出一种基于矩阵分解的信任关系相似度计算算法,此算法通过概率矩阵分解^[7]获得用户间的间接信任关系,并且同时考虑信任信息中的信任与被信任情况。

本文算法 RTWSO 通过对两个算法的线性组合,缓解了单一算法所存在的数据稀疏性问题,并对信任信息进行合理化建模,提高了算法的精确度。

1 相关工作

Salakhutdinov 等^[6]提出的应用于推荐系统的受限玻尔兹曼机模型示意图如图 1 所示,模型包含一个可见层与一个隐藏层。对于包含 m 个用户、 n 个项目,评分范围 $1\sim k$ 的推荐情景,此模型对每一个用户建立一个 RBM 模型,每个模型的可见层中包含 n 个包含 k 个节点的 Softmax 单元,用于表示此用户对所有项目的评分,可见层每个节点与隐藏层形成全连接。

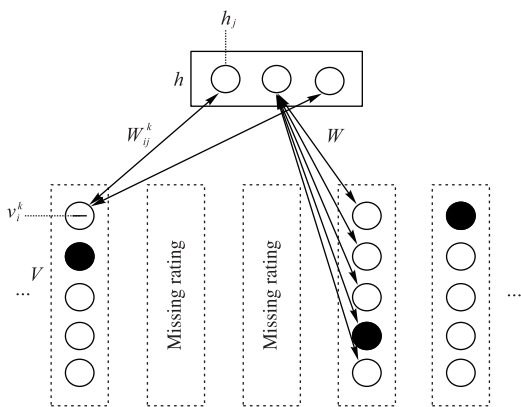


图 1 受限玻尔兹曼机模型

Fig. 1 Constrained Boltzmann machine model

RBM 模型主要存在以下缺陷:

1) 模型的可见单元使用 k 维的 Softmax 单元表示,导致模型所涉及的参数变为原来的 k 倍,大幅降低了模型的运算效率;

2) 由于每个 RBM 模型间共享权值和偏置,这使得如果两用户对同一项目进行评价,那么它们所对应的 RBM 模型将在该项目的可见单元与隐单元得到同样的权重与偏置,缺乏可解释性。

针对上述两个问题,Georgiev 等^[8]提出将评分数值直接应用在可见层,降低了算法的复杂度;霍淑华^[9]提出一种基于项目的 RBM 模型,取得了较优的推荐效果。本文在此基础上,使用一种基于实值的,并同时对项目与用户建模的受限玻尔兹曼机模型 (RUI-RBM),模型如图 2 所示。此模型主要进行了如下两点改进:

1) 在模型的可见层中,评分值直接使用用户评分值表示,未评分数据由 0 表示,相对于 Softmax 单元,削减了参数数量,使训练时间大大缩短;

2) 不同于传统 RBM 模型只针对用户或项目进行建模,本模型同时对用户与项目建模,评分值由两模型权重与偏置共同决定,模型具有更强的可解释性。

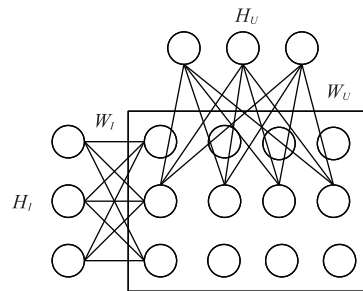


图 2 RUI-RBM 模型

Fig. 2 RUI-RBM model

由图 2 所示,方框内表示包含 m 个用户、 n 个项目所形成的评分矩阵,评分范围 $1\sim k$ 的推荐情景,RUI-RBM 模型可视为基于用户的受限玻尔兹曼机 (Real-value User Restricted Boltzmann Machine, RU-RBM) 模型以及基于项目的受限玻尔兹曼机 (Real-value Item Restricted Boltzmann Machine, RI-RBM) 模型叠加而成,RU-RBM 与 RI-RBM 分别包含各自的连接权重 W_u 与 W_i ,以及各自的隐单元偏置 b_u, b_i ,可见单元偏置 c_u, c_i ,两模型共享评分矩阵。在模型构建过程中,RUI-RBM 模型将会对评分矩阵同时从基于用户以及基于项目的角度进行建模, RU-RBM 与 RI-RBM 共享评分矩阵,但训练过程相互独立,下面以 RI-RBM 为例来对模型训练过程加以说明。

图 3 表示一个 RI-RBM 模型,模型包含一个可见层 V ,层内包含 m 个实值节点,隐藏层包含 n 个二值单元,用于提取数据间特征。可见层与隐层各节点均包含一个偏置,使用 b_j 表示第 j 个隐节点偏置, c_i 表示第 i 个可见节点的偏置,所有偏置值初始值设为 0; w_{ij} 表示第 i 个可见节点与第 j 个隐节点的连接权重,使用满足正态分布 $N(0, 0.01)$ 的随机数进行初始化。由 RBM 的结构可知:当给定可见单元的状态时,隐单元各节点激活条件独立;反之,给定隐单元状态时,可见单元取值也是独立的。可见节点使用实值表示评分,应在原 RBM 能量模型基础上加入一个二次方项,此模型的能量计算公式为:

$$E(v, h|\theta) = \frac{1}{2} \sum_{i=1}^m v_i^2 - \sum_{i=1}^m \sum_{j=1}^n W_{ij} h_j v_i - \sum_{i=1}^m v_i c_i - \sum_{j=1}^n h_j b_j \quad (1)$$

根据能量公式,则当可见层数据已知时,第 j 个隐单元的激活概率为:

$$p(h_j = 1|v, \theta) = \sigma\left(b_j + \sum_i v_i W_{ij}\right) \quad (2)$$

当隐单元的状态确定时,第 i 个可见单元的值为:

$$v_i = c_i + \sum_j h_j W_{ij} \quad (3)$$

参数的更新方式使用对比散度方法,参数更新准则为:

$$\begin{cases} \Delta W_{ij} = \varepsilon \left(\langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{recon}} \right) \\ \Delta c_i = \varepsilon \left(\langle v_i \rangle_{\text{data}} - \langle v_i \rangle_{\text{recon}} \right) \\ \Delta b_i = \varepsilon \left(\langle h_i \rangle_{\text{data}} - \langle h_i \rangle_{\text{recon}} \right) \end{cases} \quad (4)$$

其中: ε 为学习率; $\langle \cdot \rangle_{\text{recon}}$ 为一次重构后模型定义的概率分布。

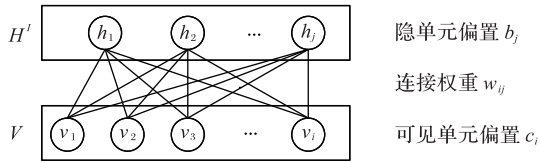


图 3 RI-RBM 模型
Fig. 3 RI-RBM model

RU-RBM 与 RI-RBM 的训练过程类似, RUI-RBM 模型对两模型的训练加以整合, RUI-RBM 的训练过程分为三个阶段:

- 1) 正阶段 (positive phase), 可见单元状态已知, 给定数据样本、训练周期和隐单元数, 得到隐单元激活概率。
- 2) 负阶段 (negative phase), 根据上一阶段获得的隐单元状态, 重构可见单元的值。
- 3) 更新阶段 (update phase), 更新参数的值。

模型参数设置: 使用 CD (Contrastive Divergence) 算法对 RU-RBM 与 RI-RBM 进行训练, CD 算法中学习效率 ε 设置为 0.1, Gibbs 采样步数 T 设置为 1, 权重衰减损失函数中的正则化参数通过实验选取 $\lambda = 0.0002$, momentum 冲量设置为 0.5^[10]。

为了训练一个 RUI-RBM, 用上述方法分别训练 RU-RBM 与 RI-RBM, 负阶段之后, 每一个 RBM 会对可见单元值进行重构, 重构后的可见单元值可视为 RBM 模型对评分的一次预测。在更新阶段, 使用 RU-RBM 与 RI-RBM 分别重构后的评分矩阵的均值来代替原评分矩阵, 并使用新矩阵中数值更新 RU-RBM 与 RI-RBM 的可见单元, 评分更新公式为:

$$r_{ij} = \frac{1}{2} \left(c_i + \sum_{k=1}^n h_k^u W_{ik}^u + c_j + \sum_{l=1}^n h_l^i W_{jl}^i \right) \quad (5)$$

模型的误差计算使用原评分矩阵中所有评级与重构后对应评分来计算, 使 RBM 预测评分接近真实评分。当 RUI-RBM 模型达到收敛或达到训练次数时, 训练过程完成, 此时得到的重构后的评分矩阵 R' 可视为使用 RUI-RBM 模型的预测结果填充完整的评分矩阵。至此, 便完成了对于稀疏评分矩阵的填充。

2 信任关系相似度计算方法

信任关系假设用户与其信任的用户有着相似的好, 在将信任关系引入模型的过程中, 研究者遇到了以下两点问题: 第一, 在信任关系中分为信任与被信任两种关系, 两种关系存在非对称性, 即用户 A 信任用户 B , 并不意味着用户 B 对 A 有着同样的信任程度, 但信任与被信任关系都在一定程度上反

映着用户的好特征; 第二, 在实际应用场景中, 所能提取到的直接信任关系往往十分稀少, 用户可用的信任关系较少 (往往少于 5 个甚至为 0), 较少的可信用户不足以反映用户的好。针对以上两点问题, 学者提出使用双向信任关系^[11-12]以及间接信任关系^[13-14]改善推荐效果。本文信任关系度量将在现有算法基础上, 将两者进一步改进并加以融合, 提出一种融合双向信任关系的间接信任相似度计算方法。

2.1 信任与被信任关系

信任网络中的每个用户, 都扮演着两个不同的角色: 信任者与被信任者, 作为信任者, 评分可能与被信任者一致, 而作为被信任者, 评分可能会影响其信任者。用户担任不同角色时, 表现出的偏好可能不尽相同。比如, 一个摇滚乐手 A , 出于对绘画的喜爱, 关注了画家 B , 同时被乐迷 C 关注, 那么可以预见, 在音乐作品上, A 与 C 的偏好更加相似, 而在绘画作品上, A 与 B 的偏好更加一致。所以, 在对评分进行预测的过程中, 同时考虑用户作为信任者与被信任者两种角色的情形更加合理。故本文中用 $R_{k \times n}$ 和 $Re_{k \times n}$ 分别代表信任矩阵与被信任矩阵。

2.2 隐含信任相似度

针对直接信任信息稀疏性问题, 提出使用基于矩阵分解的隐含信任关系相似度算法。首先由信任关系 T 中提取出信任与被信任矩阵 Rt 和 Re , 使用概率矩阵分解 (Probabilistic Matrix Factorization, PMF) 对矩阵 T 进行分解, 假设信任关系 T 由关于用户的信任向量 Rt 和 Re 的正态分布生成, 则其条件概率分布如下所示:

$$p(T|Rt, Re, \sigma_F^2) = \prod_{u=1}^m \prod_{v \in N_u} N(T_{u,v} | g(Rt_u^T Re_v), \sigma_F^2) \quad (6)$$

其中: $g(x) = \frac{1}{1 + e^{-x}}$ 为 sigmod 函数。同时, 为了避免过拟合,

Rt, Re 分别满足正态分布:

$$p(Rt | \sigma_r^2) = \prod_{u=1}^m N(Rt_u | 0, \sigma_r^2 I) \quad (7)$$

$$p(Re | \sigma_e^2) = \prod_{v=1}^m N(Re_v | 0, \sigma_e^2 I) \quad (8)$$

两者的后验概率分布为:

$$p(Rt, Re | T, \sigma_F^2, \sigma_r^2, \sigma_e^2) = \prod_{u=1}^m \prod_{v \in N_u} [N(T_{u,v} | g(Rt_u^T Re_v), \sigma_F^2) \times \prod_{u=1}^m N(Rt_u | 0, \sigma_r^2 I) \times \prod_{v=1}^m N(Re_v | 0, \sigma_e^2 I)] \quad (9)$$

损失函数如式 (10) 所示:

$$L_F(Rt, Re, T) = \frac{\lambda_F}{2} \sum_{u=1}^m \sum_{v \in N_u} (T_{u,v} - Rt_u^T Re_v)^2 + \frac{\lambda_{Rt}}{2} \sum_{u=1}^m Rt_u^T Rt_u + \frac{\lambda_{Re}}{2} \sum_{v=1}^m Re_v^T Re_v \quad (10)$$

对此损失函数使用梯度下降法求解可得 Rt 与 Re 。

根据所得 Rt 与 Re 向量, 可计算两用户之间的隐含信任相似度, 用户 u 与用户 v 均作为信任角色时的隐含相似度为:

$$S_{u,v}^{Rt} = Rt_u^T Rt_v \quad (11)$$

两用户同时作为被信任角色时的隐含相似度为:

$$S_{u,v}^{Re} = Re_u^T Re_v \quad (12)$$

用户间信任相似度将信任隐含相似度与被信任隐含相似度按一定比例加和, 以权重 $k \in (0, 1)$ 来调整二者所占比重, 隐

含信任相似度计算如式(13)所示:

$$S_{u,v} = kS_{u,v}^{R_u} + (1 - k)S_{u,v}^{R_v} \quad (13)$$

3 使用 WSO 算法生成预测结果

Slope One 算法^[15]是一种基于项目的协同过滤算法,采用简单的一元线性模型 $f(x) = x + b$ 形式进行表示。其中, b 表示用户对两个项目评分矩阵的偏差平均值。基于此种思想,则在一个给定的训练集 x 中,项目 i 对于项目 j 的偏差公式如下:

$$dev_{j,i} = \sum_{u \in O} \frac{u_j - u_i}{card(O)} \quad (14)$$

其中: $s_{j,i}(x)$ 为同时评价过两项目的用户集合; $card(s_{j,i}(x))$ 为 $s_{j,i}(x)$ 中的用户个数。

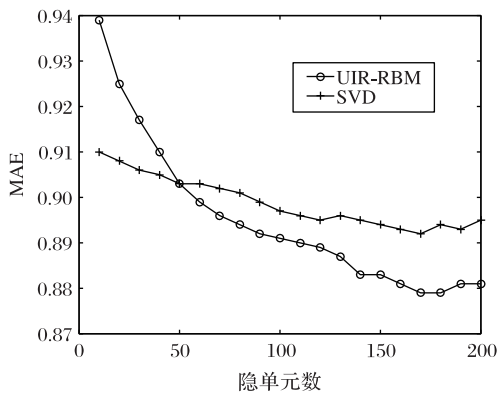
在计算用户 u 对项目 j 的预测评分时,则需选取除 j 外所有被用户 u 评分的集合 R_j 中的项目,取评分的平均值作为用户 u 对项目 j 的预测评分。公式如下所示:

$$P(u,j) = \frac{1}{card(R_j)} \sum_{i \in R_j} (dev_{j,i} + u_i) \quad (15)$$

其中: $|R_j|$ 表示同时评分的项目总数。

由于 Slope One 算法基于线性模型,相较于其他协同过滤算法有着较高的计算效率;并且由于其协同过滤思想,相较于较常使用的取近邻用户评分均值或加权均值的方法,使用偏差项修正不同用户的评分标准不一的问题,使结果更加合理。故本文使用 Slope One 算法进行最终的评分预测。Slope One 算法所存在的数据稀疏性问题,由矩阵填充方法解决,但评分矩阵由稀疏矩阵变为满元矩阵带来运算时间的大幅增加,为了缩短运算时间,通过选择与用户相似度最高的前 N 个用户组成的近邻集 O ,来替代偏差计算公式中 $s_{j,i}(x)$ 。另外,为了引入用户间的信任信息,将隐含信任相似度作为权值加入到算法中。故本文所使用的 WSO 的偏差公式与预测公式如下所示:

$$dev_{j,i} = \sum_{u \in O} \frac{sim_{j,i}(u_j - u_i)}{card(O)} \quad (16)$$



(a) 对模型MAE值的影响

$$P(u,j) = \frac{1}{card(R_j)} \sum_{i \in R_j} (dev_{j,i} + u_i) \quad (17)$$

4 实验与结果分析

4.1 数据集及评价指标

实验数据集选用 Epinions 数据集及 Ciao 数据集,数据集中包含用户对项目的评分信息以及用户间的信任信息。实验采用 5 折交叉验证的方式,将数据集分为 5 份,每次抽取一份作为测试集,其余作为训练集,取 5 次实验平均值作为实验结果。实验分为两个部分,首先,通过 Epinions 数据集上实验结果确定模型各项参数最优值并验证混合算法是否提升了推荐效果;其次,通过 Epinions 与 Ciao 两个数据集上的实验,将本文算法与相关算法进行对比,验证本文算法的性能。

表 1 数据集统计

Tab. 1 Statistics of datasets

参数	用户数	项目数	评分数	评分稀疏度/%	信任数	信任稀疏度/%
Epinions	22 166	296 277	922 267	0.014	335 121	0.07
Ciao	7 375	105 114	284 086	0.037	111 781	0.21

实验使用平均绝对误差 (Mean Absolute Error, MAE) 及均方根误差 (Root Mean Square Error, RMSE) 作为评价指标。

MAE 的计算公式如下:

$$MAE = \frac{1}{N} \sum_j |R_{ij} - \widehat{R}_{ij}| \quad (18)$$

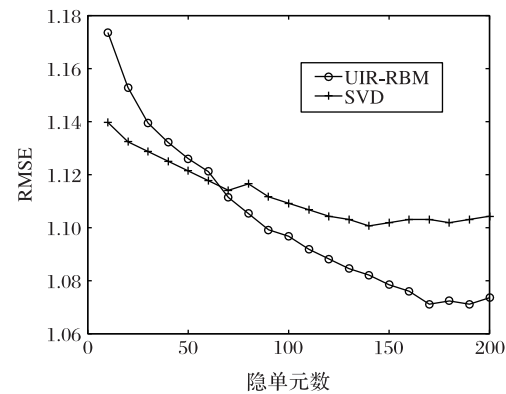
RMSE 的计算公式如下:

$$RMSE = \sqrt{\frac{1}{N} \sum_j |R_{ij} - \widehat{R}_{ij}|^2} \quad (19)$$

4.2 RUI-RBM 模型对矩阵填充效果实验

4.2.1 隐单元数对 RUI-RBM 模型填充效果影响

本实验对 RUI-RBM 中隐单元数对模型填充效果的影响。实验采用 $epochs=100$ 进行实验,探究不同隐单元数对 RUI-RBM 模型推荐效果的影响。实验设置隐单元数范围 10~200,步长设置为 10 进行实验,实验结果如图 4 所示。



(b) 对模型RMSE值的影响

图 4 隐单元数对 RUI-RBM 模型矩阵填充效果的影响

Fig.4 Impact of hidden element number on matrix filling effect of RUI-RBM model

分析图 4 中数据可知,本文所提出的 RUI-RBM 模型,在隐单元数少于 50 的 MAE 值和隐单元数少于 70 的 RMSE 值时,推荐效果弱于 SVD 算法,而随着隐单元数量的增加,推荐效果超过 SVD 算法,并在隐单元数在 120 个左右时,推荐效果趋于稳定,此时相较于 SVD 在 MAE 与 RMSE 上分别下降 3% 左右。本实验验证了本文所提出的 RUI-RBM 算法在数据集

上对于矩阵填充的效果优于 SVD 算法,并在隐单元个数 $n=170$ 时取得较优推荐效果。

4.2.2 隐 RUI-RBM 模型中训练次数对矩阵填充效果实验

本实验建立在隐单元个数 $n=120$ 上,对 RUI-RBM 中训练次数对矩阵填充效果进行实验,实验设置训练次数 $epochs$ 上限为 100,实验结果如图 5 所示。

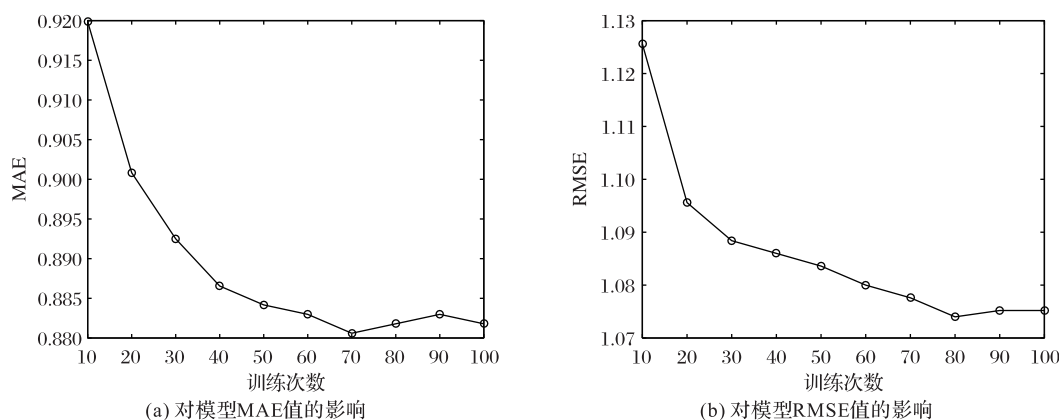


图5 训练次数对矩阵填充的影响

Fig. 5 Impact of training time on matrix filling effect

由图5可见,RUI-RBM模型在训练次数小于60时,模型的MAE与RMSE值呈递减趋势;在训练次数达到70后,模型趋于平稳。故模型选用 $epochs=70$ 为默认参数。

4.3 Trust-WSO算法参数实验

4.3.1 最近邻居数目对推荐效果的影响

本节将对信任WSO(Trust-WSO)算法中,不同最近邻居数目 N 对模型推荐效果的影响进行实验。实验相似度选取为本文所提出的项目混合相似度计算方法,相似度计算方法中 k 值暂取0.5,初始使用步长为10,测试10~100近邻个数时,算法分别在数据集上的MAE值变化情况。实验结果如图6所示。

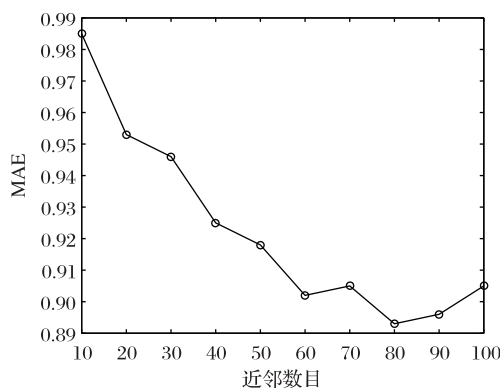


图6 近邻数目对模型推荐效果的影响

Fig. 6 Impact of nearest neighbor number on model recommendation effect

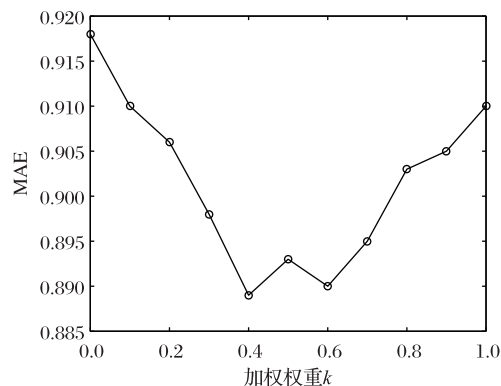
分析图5~6中数据可知,在Epinions数据集中:当 $10 \leq N \leq 30$ 时,由于近邻数量较少,模型产生了比较大的波动,且MAE值表现较差;当 $60 \leq N \leq 80$ 时,推荐效果趋于平稳,并在 $N=80$ 时,得到最佳效果;当 $N > 80$ 时,MAE值呈现平稳上升状态。根据实验结果,选取 $N=80$ 作为默认参数。

4.3.2 社交相似计算中 k 值对推荐效果的影响

由于相似度计算中采取两种相似度加权的形式,两种相似度计算方法以不同的权重相结合,对推荐准确度可能会产生较大的影响,故设计本实验以探究不同的 k 值对实验结果的影响。实验的目的主要有以下两点:首先,验证两种相似度方法的结合是否有效,即使用综合相似度计算方法的推荐准确度是否优于使用单个相似度;其次,在数据集中确定一个最优 k 值作为算法的默认参数。本实验对 k 值在 $[0,1]$ 区间上以

步长为0.1,取得11个数据点来对加权Slope One算法准确度进行实验, k 值表示项目评分相似度在相似度计算方法中所占的比重,特别地,当 $k=0$ 时,表示仅使用信任相似度;当 $k=1$ 时,表示仅使用被信任相似度。

分析图7中数据可知,仅使用项目属性信息时,算法MAE值为0.917;仅使用项目评分相似度时,算法MAE值为0.910,而将两种方法以不同权重进行结合时,即 k 值由坐标轴向中间逐步靠近时,算法MAE值呈现下降的趋势。说明本文将两种相似度方法的结合,有效地提升了推荐的准确度。在 $k=0.4$ 时,算法取得最优MAE值0.889。故本实验确定0.4作为 k 的默认值。

图7 k 值对实验结果的影响Fig. 7 Impact of k value on experimental result

4.4 RTWSO算法对比参数实验

本实验对RTWSO算法的推荐效果进行实验,使用Epinions与Ciao两个数据集,选取组成算法RUI-RBM以及S-WSO算法进行对比,以探究两种算法的混合是否改善了推荐效果;另外,使用文献[15]中提出的基于信任关系隐含相似度的社会化推荐SocialIT(Social recommendation algorithm based on Implicit similarity in Trust)模型;同时对信任者及被信任者建模的TrustMF(Trust Matrix Factorization)模型^[11],使用SVD与Slope One混合推荐算法的SVD-Slope One模型^[16]以及文献[17]中提出的基于用户与项目,并引入属性信息层IC-CRBMF(Item Category-Conditional Restricted Boltzmann Machine Frame)模型进行对比。

实验结果如表2所示。从表2可知:本文所提出的RTWSO算法,相较于其组成算法RUI-RBM与S-WSO,在推荐

准确度上有着较大的提升,说明两算法的混合有效地改善了单一算法的数据稀疏性问题;在与相关算法的比较中,RTWSO分别取得了1.2%~8.4%的提升,说明本文所提出的算法具有较高的推荐效率。

表2 RTWSO算法与其他算法的推荐效果对比

Tab. 2 Recommendation effect comparison of RTWSO algorithm and other algorithms

模型	Epinions		Ciao	
	MAE	RMSE	MAE	RMSE
RUI-RBM	0.874	1.106	0.765	0.989
S-WSO	0.887	1.132	0.781	1.040
SocialIT	0.815	1.048	0.774	1.025
TrustMF	0.828	1.072	0.783	1.038
SVD-Slope One	0.846	1.084	0.779	1.032
IC-CRBMF	0.833	1.067	0.744	0.976
RTWSO	0.796	1.013	0.735	0.963

5 结语

本文提出一种基于评分填充与信任信息的混合推荐算法,首先使用基于用户与项目的实值受限玻尔兹曼机算法对原始评分矩阵进行评分填充,以缓解评分稀疏性问题,之后使用基于信任信息的WSO算法对待评分进行预测。实验结果表明,本文算法在Epinions以及Ciao数据集上显示出较高的推荐准确度。本文主要针对数据稀疏性问题进行探索,在数据处理过程中去除了冷启动用户及项目对推荐结果所造成的影响,在下一步的工作中将对冷启动问题进行研究。

参考文献 (References)

- [1] 杜倩. 基于聚类的加权Slope One推荐技术研究[D]. 北京:北京工业大学, 2016: 66. (DU Q. Research on weighted Slope One recommendation technology based on clustering [D]. Beijing: Beijing University of Technology, 2016: 66.)
- [2] LIU Y, LIU D, XIE H, et al. A research on the improved Slope One algorithm for collaborative filtering[J]. International Journal of Computing Science and Mathematics, 2016, 7(3): 245-253.
- [3] 刘方婷. 融入时间因素和信任机制的矩阵分解推荐算法[D]. 秦皇岛:燕山大学, 2017: 60. (LIU F T. The matrix decomposition recommendation algorithm combined by time factor and trust mechanism[D]. Qinhuangdao: Yanshan University, 2017: 60.)
- [4] 何洁月, 马贝. 利用社交关系的实值条件受限玻尔兹曼机协同过滤推荐算[J]. 计算机学报, 2016, 39(1): 183-195. (HE J Y, MA B. Based on real-valued conditional restricted Boltzmann machine and social network for collaborative filtering[J]. Chinese Journal of Computers, 2016, 39(1): 183-195.)
- [5] ZHAO X, ZHU C, CHENG L. Coupled Bayesian matrix factorization in recommender systems [C]// Proceedings of the 2017 IEEE International Conference on Data Science and Advanced Analytics. Piscataway: IEEE, 2017: 522-528.
- [6] SALAKHUTDINOV R, MNIH A, HINTON G. Restricted Boltzmann machines for collaborative filtering [C]// Proceedings of the 24th International Conference on Machine Learning. New York: ACM, 2007: 791-798.
- [7] MNIH A, SALAKHUTDINOV R R. Probabilistic matrix factorization [C]// Proceedings of 20th International Conference on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc., 2007: 1257-1264.
- [8] GEORGIEV K, NAKOV P. A non-IID framework for collaborative filtering with restricted Boltzmann machines [C]// Proceedings of the 30th International Conference on Machine Learning. New York: JMLR.org, 2013: III-1148-III-1156
- [9] 霍淑华. 基于协同过滤的评分预测推荐算法研究[D]. 北京:北京工业大学, 2015: 67. (HUO S H. Research on collaborative filtering rating prediction algorithm [D]. Beijing: Beijing University of Technology, 2015: 67.)
- [10] HINTON G. A practical guide to training restricted Boltzmann machine[D]. Toronto: University of Toronto, 2010: 20.
- [11] YAO W, HE J, HUANG G, et al. Modeling dual role preferences for trust-aware recommendation [C]// Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2014: 975-978.
- [12] 张琦, 柳玲, 文俊浩. 一种基于领域信任及不信任的奇异值分解推荐算法[J]. 计算机科学, 2019, 46(10): 27-31. (ZHANG Q, LIU L, WEN J H. Recommendation algorithm with field trust and distrust based on SVD [J]. Computer Science, 2019, 46(10): 27-31.)
- [13] WANG M, MA J. A novel recommendation approach based on users weighted trust relations and the rating similarities [J]. Soft Computing, 2016, 20(10): 3981-3990.
- [14] 陈文俊, 倪静. 社会网络环境下基于信任传递的推荐模型研究[J]. 计算机应用研究, 2020, 37(5): 1364-1367. (CHEN W J, NI J. Research on recommendation model based on trust transitivity in social network environment [J]. Application Research of Computers, 2020, 37(5): 1364-1367.)
- [15] 潘一腾, 何发智, 于海平. 一种基于信任关系隐含相似度的社会化推荐算法[J]. 计算机学报, 2018, 41(1): 65-81. (PAN Y T, HE F Z, YU H P. Social recommendation algorithm based on implicit similarity in trust [J]. Chinese Journal of Computers, 2018, 41(1): 65-81.)
- [16] GONG B Y, XUE Y Z. Improved SVD algorithm based on Slope One [C]// 2018 CCDC: Proceedings of the 30th Chinese Control and Decision Conference. Piscataway: IEEE, 2018: 1002-1006.
- [17] LIU X, OUYANG Y, RONG W, et al. Item category aware conditional restricted Boltzmann machine based recommendation [C]// Proceedings of the 2015 International Conference on Neural Information Processing, LNCS 9490. Cham: Springer, 2015: 609-616.

This work is partially supported by the National Natural Science Foundation of China (61772249).

SHEN Xueli, born in 1969, M. S., professor. His research interests include information security, recommendation systems.

LI Zijian, born in 1994, M. S. candidate. His research interests include recommendation systems.

HE Chenhao, born in 1992, M. S. His research interests include deep learning, recommendation systems