



融合用户信任度的概率矩阵分解群组推荐算法

宋玉龙, 马文明, 刘彤彤

(烟台大学 计算机与控制工程学院, 山东 烟台 264005)

摘要: 利用推荐系统进行群组推荐时, 群组成员之间的交互关系对推荐结果有很大影响, 但传统的群组推荐算法较少考虑用户信任度的重要性, 致使社交关系信息不能得到充分利用。在群组融合时考虑群组内用户间的交互关系, 提出一种基于用户信任度和概率矩阵的群组推荐算法。在获取用户信任度数据后, 使用概率矩阵分解(PMF)算法补全信任度矩阵并进行归一化处理, 得到相似度矩阵, 同时在后验概率计算过程中加入用户间的信任度因素, 通过极大化后验概率获得预测评分。在此基础上, 对群组中用户的权重进行归一化处理, 使用基于用户交互关系的权重策略融合群组成员偏好, 得到最终的推荐结果。在 Epinions 和 FilmTrust 数据集上的实验结果表明, 该算法可使融合结果更具群组特性, 同时提高推荐结果的可靠性和可解释性, 且均方根误差和命中率均优于 PMF、NeuMF、RippleNet 等对比算法。

关键词: 群组推荐; 偏好融合; 概率矩阵分解; 用户信任度; 协同过滤

开放科学(资源服务)标志码(OSID):



中文引用格式: 宋玉龙, 马文明, 刘彤彤. 融合用户信任度的概率矩阵分解群组推荐算法[J]. 计算机工程, 2022, 48(1): 105-111.

英文引用格式: SONG Y L, MA W M, LIU T T. Group recommendation algorithm incorporating user trust with probability matrix factorization[J]. Computer Engineering, 2022, 48(1): 105-111.

Group Recommendation Algorithm Incorporating User Trust with Probability Matrix Factorization

SONG Yulong, MA Wenming, LIU Tongtong

(School of Computer and Control Engineering, Yantai University, Yantai, Shandong 264005, China)

[Abstract] When using the recommendation system for group recommendation, interactions between group members have a great impact on recommendation results, but traditional group recommendation algorithms consider little the importance of user trust, and fail to make full use of social relationship information. Based on probability matrix and user trust, this paper proposes a group recommendation algorithm that considers interactions between users in a group. After obtaining data of user trust, the Probability Matrix Factorization(PMF) algorithm is used to complete the trust matrix, which is then normalized to obtain the similarity matrix. The factor of trust between users is added in the posterior probability calculation process to obtain the prediction score by maximizing the posterior probability. On this basis, the weights of users in the group are normalized, and a weight strategy based on user interactions is used to fuse the preferences of group members to obtain the recommendation result. The experimental results on Epinions dataset and FilmTrust dataset show that the proposed algorithm can make the fusion results to have more group characteristics, and improve the reliability as well as interpretability of the recommendation results. The algorithm displays a lower root mean square error and higher hit rate than PMF, NeuMF, RippleNet and other comparison algorithms.

[Key words] group recommendation; preference fusion; Probability Matrix Factorization(PMF); user trust; Collaborative Filtering(CF)

DOI: 10.19678/j.issn.1000-3428.0059526

0 概述

互联网技术的快速发展和媒体渠道日趋多样化

使人们能够更便捷地获取信息,但同时也面临信息过载的问题,用户难以从海量数据中筛选出自己真正感兴趣的部分^[1]。推荐系统能够实现对海量信息

基金项目: 国家自然科学基金(61602399)。

作者简介: 宋玉龙(1993—),男,硕士研究生,主研方向为机器学习、推荐系统;马文明(通信作者),副教授;刘彤彤,硕士研究生。

收稿日期: 2020-09-15 **修回日期:** 2020-11-22 **E-mail:** mwmytu@126.com

的处理与分析^[2],得到最契合需求的信息,为用户提供个性化的推荐结果^[3-4]。传统推荐系统的受众往往是单个用户,而目前以多个用户组成的群体为单位的个性化推荐需求越来越多(如为一个旅游团的人推荐旅游景区或者给公司团队建设的员工推荐娱乐场地^[5]),因此,需要将传统推荐系统从针对单个用户拓展为适用于群组用户^[6],即实现群组推荐。

群组推荐基于对单个用户偏好的拟合以及群组融合策略。概率矩阵分解(Probability Matrix Factorization, PMF)算法用于拟合单个用户偏好,其在矩阵分解的基础上加入了用户和项目特征向量的先验信息^[7],但只考虑用户与项目的评分数据^[8],忽略了用户之间的交互性,而用户之间的交互关系往往会对用户的偏好产生较大影响。在完成单个用户的偏好拟合后,群组推荐中的另一个问题是群组成员的偏好融合,这要求群组推荐系统将群组内的单个成员偏好融合为可以表现所有组内成员的整体偏好^[9]。进行群组推荐时往往需要考虑公平性和用户满意度等问题,既要尽量满足组内成员个人的偏好,又要体现群组的整体偏好,使推荐结果满足整个群组的需求^[10],然而现有群组融合策略也较少考虑组内成员间的交互关系。

为使融合结果更具群组特性,提高推荐结果的可靠性和可解释性,本文提出一种融合用户信任度的群组推荐算法,考虑群组内用户交互关系进行群组融合。在获取用户信任度数据后,训练得到组内每个成员的信任向量和被信任向量,并以被信任向量均值为基准向量,将每个成员的信任向量与其做点积运算得到对应权重,通过归一化后的权重评分求和得到群组评分,以此生成推荐列表。

1 相关工作

1.1 群组融合

群组融合方法通常可分为偏好融合与推荐结果融合两类,如图1所示。偏好融合是指在获取组内所有成员偏好模型后,将成员偏好融合为群组偏好,利用群组偏好进行群组推荐^[11]。推荐结果融合是指获得组内所有成员的推荐结果或评分后,采用融合策略将所有推荐结果或评分合并为群组结果或者评分^[12]进行群组推荐。

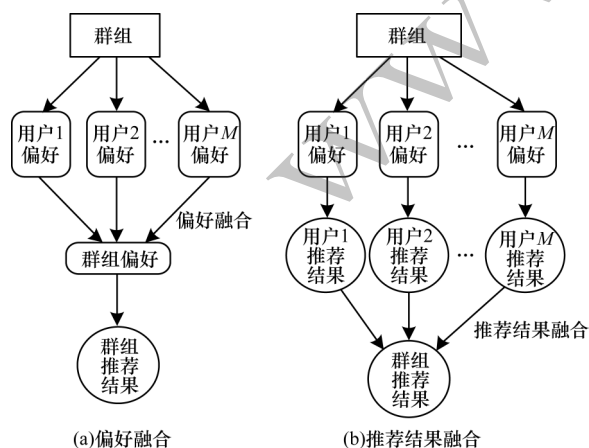


图1 群组融合策略

Fig.1 Group fusion strategy

推荐结果融合又可分为列表融合与评分融合。列表融合是指将每个用户的推荐列表合并成一个群组推荐列表,评分融合则是指将组内用户的评分融合成群组评分。在现有的融合策略中,均值策略与最小痛苦策略应用最为广泛。

均值策略将群组内所有成员对某个项目的评分取平均值作为该群组对该项目的评分,如式(1)所示:

$$p^G = \left\{ \text{avg}(d^u) : \exists d^u \in \bigcup_{u \in G} p^u \right\} \quad (1)$$

最小痛苦策略可规避组内有成员对某个项目特别厌恶,但该项目的平均分却很高而进入推荐列表的情况。该策略取组内用户对项目的评分的最小值作为该群组对该项目的评分,如式(2)所示:

$$p^G = \left\{ \min(d^u) : \exists d^u \in \bigcup_{u \in G} p^u \right\} \quad (2)$$

此外,研究者还提出了最受尊敬者策略、最大愉悦策略等融合策略,这些策略在不同的情境下会产生不同的效果。

1.2 概率矩阵分解

在推荐系统中,协同过滤(Collaborative Filtering, CF)是一种被普遍使用的推荐技术,其利用兴趣相投、喜好相似的群体的偏好来推荐用户感兴趣的项目与信息,通过个人所给予信息的反馈记录实现过滤,从而帮助其他人筛选信息^[13-15]。可以根据是否采用机器学习的建模思想将协同过滤技术划分为基于内容和基于模型两类。在基于模型的协同过滤技术中,矩阵分解算法以高准确率和良好可伸缩性受到广泛关注^[16]。

基于矩阵分解的协同过滤将用户和项目的特征转化为潜在向量,通过计算用户或项目之间潜在向量的相关性进行推荐^[17]。在使用矩阵分解算法^[18]进行评分预测时,通常将用户与项目表示为 $M \times N$ 的矩阵(其中 M 和 N 分别表示用户和项目的数量),矩阵中的元素对应用户对项目的评分。该矩阵通常很稀疏,故需要补全其中缺失的评分。矩阵分解算法将矩阵 $R_{M \times N}$ 分解为2个维度更低的矩阵的乘积 $U_{K \times M}^T V_{K \times N}$ (其中 K 表示潜在向量的维度),通过不断学习迭代使 $U_{K \times M}^T V_{K \times N}$ 逼近评分矩阵 $R_{M \times N}$,同时得到未评分项目的预测评分。

SALAKHUTDINOV于2007年提出概率矩阵分解算法,其假设用户潜在向量、项目潜在向量及两者的内积均服从高斯分布,通过极大化它们的后验概率来获得更为准确的潜在向量。然而,PMF算法忽略了用户之间的交互性,而用户之间的信任度等交互关系往往会对用户的偏好产生较大影响。例如,用户U特别喜欢电影A,通常也会对其好友介绍电影A并鼓励他们观看,而朋友也更愿意接受所信任的人的推荐^[19]。由于传统的推荐算法没有考虑这样的情况,因此不能充分利用大量社交信息。

2 融合用户相似度的概率矩阵分解群组推荐

2.1 群组推荐框架

本文提出的群组推荐模型框架如图2所示, 具体流程如下:

1) 获取用户信任度数据, 使用经典概率矩阵分解算法补全信任度矩阵 T , 获得用户信任度特征向量 L 和 R 。

2) 对信任度矩阵每一行使用 SoftMax 进行归一化, 得到相似度矩阵 F 。

3) 获取用户-项目评分数据, 使用联合相似度的概率矩阵分解算法进行补全, 得到用户对未评分项目的预测评分。

4) 使用融合信任度的权重策略合并群组内所有成员的评分, 获取整个群组对项目的评分, 并根据评分生成推荐列表。

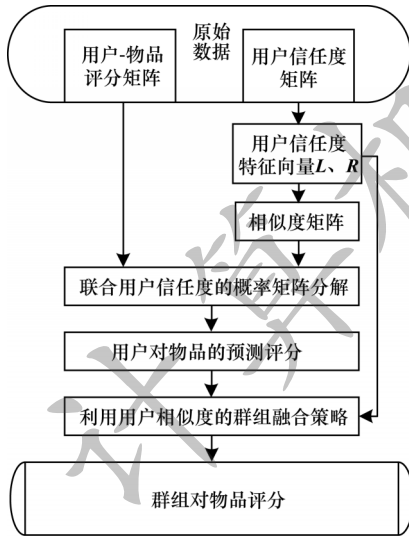


图2 群组推荐模型框架

Fig.2 Framework of group recommendation model

2.2 用户相似度矩阵构建

本文使用 Epinions 数据集, 其中包含若干用户间的信任关系与用户对项目的评分。构建用户相似度矩阵需要所有两两用户间的信任关系。因此, 首先使用概率矩阵分解算法对用户信任度矩阵进行补全。

假设有 M 个用户, 以 T_{lr} 表示用户 l 对用户 r 的信任度, 构成一个 $M \times M$ 的信任度矩阵。将目标信任度矩阵分解为 2 个维度更低的矩阵的乘积 $L_{K \times M}^T R_{K \times M}$, 其中 K 为潜在向量维度, L, R 表示用户信任度隐特征向量。

假设用户信任度 T_{lr} 由用户 l 的潜在向量和用户 r 的潜在向量的内积来决定, 且该信任度服从高斯分布, 即:

$$T_{lr} \sim N(L_l^T R_r, \sigma^2) \quad (3)$$

则观察到的信任度矩阵的条件概率为:

$$p(T|L, R, \sigma^2) \sim \prod_{l=1}^M \prod_{r=1}^M N(L_l^T R_r, \sigma^2)^{I_{lr}} \quad (4)$$

其中: I_{lr} 为指示函数, 如果用户 l 对用户 r 存在信任数据, 则为 1, 否则为 0。

再假设用户潜在特征向量都服从均值为 0 的高斯先验分布, 即:

$$p(L|\sigma_L^2) \sim \prod_{l=1}^M N(L_l|0, \sigma_L^2 \mathbf{I}) \quad (5)$$

$$p(R|\sigma_R^2) \sim \prod_{r=1}^M N(R_r|0, \sigma_R^2 \mathbf{I}) \quad (6)$$

其中: \mathbf{I} 表示一个对角阵。则 L 和 R 的后验概率为:

$$p(L, R|T, \sigma_L^2, \sigma_R^2, \sigma^2) \propto p(T|L, R, \sigma^2) p(L|\sigma_L^2) p(R|\sigma_R^2) \quad (7)$$

两边取对数得到:

$$\begin{aligned} \ln p(L, R|T, \sigma_L^2, \sigma_R^2, \sigma^2) = & -\frac{1}{2\sigma^2} \sum_{l=1}^M \sum_{r=1}^M I_{lr} (T_{lr} - L_l^T R_r)^2 - \\ & -\frac{1}{2\sigma_L^2} \sum_{l=1}^M L_l^T L_l - \frac{1}{2\sigma_R^2} \sum_{r=1}^M R_r^T R_r - \frac{1}{2} \left(\sum_{l=1}^M \sum_{r=1}^M I_{lr} \right) \ln \sigma^2 - \\ & -\frac{1}{2} MK \ln \sigma_L^2 - \frac{1}{2} MK \ln \sigma_R^2 + C \end{aligned} \quad (8)$$

其中: C 为无关常数。

通过最小化以下目标函数来最大化后验概率:

$$\begin{aligned} L = & \frac{1}{2} \sum_{l=1}^M \sum_{r=1}^M I_{lr} (T_{lr} - L_l^T R_r)^2 + \\ & + \frac{\lambda_L}{2} \sum_{l=1}^M \|L_l\|_{\text{Fro}}^2 + \frac{\lambda_R}{2} \sum_{r=1}^M \|R_r\|_{\text{Fro}}^2 \end{aligned} \quad (9)$$

其中: $\lambda_L = \sigma^2 / \sigma_L^2$; $\lambda_R = \sigma^2 / \sigma_R^2$ 。

等式 L 分别对 L_l 和 R_r 求偏导得:

$$\partial E / \partial L_l = (T_{lr} - L_l^T R_r) R_r - \lambda_L L_l \quad (10)$$

$$\partial E / \partial R_r = (T_{lr} - L_l^T R_r) L_l - \lambda_R R_r \quad (11)$$

然后使用随机梯度下降更新 L_l 和 R_r , 直到收敛或达到最大迭代次数。

训练完成后可获得每个用户的 2 个特征向量 L 和 R , 将任意用户 l 的左向量 L_l 与另一个用户 r 的右向量 R_r 点乘可获得 l 对 r 的信任度, 即:

$$T_{lr} = \langle L_l, R_r \rangle \quad (12)$$

首先计算每对用户之间的信任度矩阵 T , T_{lr} 表示用户 l 对用户 r 的信任度。矩阵的第 l 行数据表示第 l 个用户对其他用户信任度。此时对每行数据进行 SoftMax 操作, 将每行信任度总和归一, 获得相似度矩阵 F :

$$F_{lr} = e^{T_{lr}} / \sum_{r=1}^M e^{T_{lr}} \quad (13)$$

其中, F_{lr} 表示用户 l 与用户 r 的相似度。

2.3 融入用户信任度的 FPMF 算法

假设有 M 个用户和 N 个项目, 将每个用户与项目的评分作为一个 $M \times N$ 矩阵 $R_{M \times N}$, $R_{i,j}$ 表示用户 i 对项目 j 的评分。再假设 R 服从于均值为 $U_i^T V_j$, 方差为

σ_R^2 的高斯分布,其概率分布为:

$$p(\mathbf{R}|\mathbf{U}, \mathbf{V}, \sigma_R^2) = \prod_{i=1}^M \prod_{j=1}^N N(\mathbf{U}_i^T \mathbf{V}_j, \sigma_R^2) \quad (14)$$

由相似度矩阵可知,用户的特征向量与相似度矩阵 \mathbf{F} 中其他用户特征向量乘以权重求和后应相等,即:

$$\mathbf{U}_i = \sum_{t=1}^M F_{i,t} \mathbf{U}_t \quad (15)$$

其中: $F_{i,t}$ 表示用户 i 与用户 t 的相似度。则用户特征矩阵 \mathbf{U} 的高斯先验分布如下:

$$p(\mathbf{U}|\mathbf{F}, \sigma_U^2, \sigma_F^2) \propto p(\mathbf{U}|\sigma_U^2) p(\mathbf{U}|\mathbf{F}, \sigma_F^2) = \prod_{i=1}^M N(\mathbf{U}_i|0, \sigma_U^2 \mathbf{I}) \times \prod_{i=1}^M N\left(\mathbf{U}_i \sum_{t=1}^M F_{i,t} \mathbf{U}_t, \sigma_F^2 \mathbf{I}\right) \quad (16)$$

再假设项目特征向量 \mathbf{V} 也服从高斯分布,即:

$$p(\mathbf{V}|\sigma_V^2) = \prod_{j=1}^N N(\mathbf{V}_j|0, \sigma_V^2 \mathbf{I}) \quad (17)$$

则可得 \mathbf{U} 和 \mathbf{V} 的后验概率分布为:

$$p(\mathbf{U}, \mathbf{V}|\mathbf{R}, \mathbf{F}, \sigma_R^2, \sigma_U^2, \sigma_V^2) \propto p(\mathbf{R}|\mathbf{U}, \mathbf{V}, \sigma_R^2) p(\mathbf{U}|\mathbf{F}, \sigma_U^2, \sigma_F^2) p(\mathbf{V}|\sigma_V^2) \quad (18)$$

两边取对数可得:

$$\begin{aligned} \ln p(\mathbf{U}, \mathbf{V}|\mathbf{R}, \mathbf{F}, \sigma_R^2, \sigma_U^2, \sigma_V^2) = & -\frac{1}{2\sigma_R^2} \sum_{i=1}^M \sum_{j=1}^N I_{i,j}^R (\mathbf{R}_{i,j} - \mathbf{U}_i^T \mathbf{V}_j)^2 - \\ & \frac{1}{2\sigma_U^2} \sum_{i=1}^M \left(\left(\mathbf{U}_i - \sum_{t=1}^M F_{i,t} \mathbf{U}_t \right)^T \left(\mathbf{U}_i - \sum_{t=1}^M F_{i,t} \mathbf{U}_t \right) \right) - \\ & \frac{1}{2\sigma_U^2} \sum_{i=1}^M \mathbf{U}_i^T \mathbf{U}_i - \frac{1}{2\sigma_V^2} \sum_{j=1}^N \mathbf{V}_j^T \mathbf{V}_j - \frac{1}{2} \left(\sum_{i=1}^M \sum_{j=1}^N I_{i,j}^R \right) \ln \sigma_R^2 - \\ & \frac{1}{2} M K \ln \sigma_U^2 - \frac{1}{2} N K \ln \sigma_V^2 + C \end{aligned} \quad (19)$$

通过最小化以下目标函数来最大化后验概率:

$$\begin{aligned} L = & \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^N I_{i,j}^R (\mathbf{R}_{i,j} - \mathbf{U}_i^T \mathbf{V}_j)^2 + \frac{\lambda_U}{2} \sum_{i=1}^M \|\mathbf{U}_i\|_{\text{Fro}}^2 + \\ & \frac{\lambda_V}{2} \sum_{j=1}^N \|\mathbf{V}_j\|_{\text{Fro}}^2 + \frac{\lambda_F}{2} \sum_{i=1}^M \left\| \mathbf{U}_i - \sum_{t=1}^M F_{i,t} \mathbf{U}_t \right\|_{\text{Fro}}^2 \end{aligned} \quad (20)$$

其中: $\lambda_U = \sigma_R^2 / \sigma_U^2$; $\lambda_V = \sigma_R^2 / \sigma_V^2$; $\lambda_F = \sigma_R^2 / \sigma_F^2$ 。

使用梯度下降方式对 \mathbf{U}_i 和 \mathbf{V}_j 进行更新,直到收敛或达到最大迭代次数。

2.4 融合策略

在群组融合方面,本文提出一种新的基于用户交互的融合策略,这将再次利用之前获得的用户信任网络,具体流程如下:

获取所有用户的信任度隐特征向量 \mathbf{R} ,求得 \mathbf{R} 的平均值 \mathbf{R}_{mean} ,即:

$$\mathbf{R}_{\text{mean}} = \frac{1}{M} \sum_{i=1}^M \mathbf{R}_i \quad (21)$$

将每个用户的左向量 \mathbf{L} 与 \mathbf{R}_{mean} 点乘,获得该用户的权重,即:

$$w_i = \langle \mathbf{L}_i, \mathbf{R}_{\text{mean}} \rangle \quad (22)$$

在群组推荐时,将该群组中用户的权重使用 SoftMax 函数进行归一化,即:

$$w_i = e^{w_i} / \sum_{j=1}^M e^{w_j} \quad (23)$$

使用权重策略将组内用户评分合并成群组评分:

$$R(G, j) = \sum_{i \in G} w_i R_{i,j} \quad (24)$$

其中: $R(G, j)$ 表示群组 G 对项目 j 的评分。

3 实验结果与分析

3.1 实验数据集

Epinions 数据集是从一个在线商品网站收集的多图数据集,其中包含了多种关系,如评论者对另一个评论者的态度(信任/不信任),以及评论者对商品的评分,共包含 664 824 条评分数据和 487 183 条信任度数据。随机将评分数据中的 80% 作为训练集,其余的 20% 作为测试集。

本文同时还在 FilmTrust 数据集上进行了验证。FilmTrust 为 2011 年从网站 FilmTrust 完整抓取下来的数据集。该数据集由两部分组成:用户-物品评分和用户间信任度关系,其中评分包含 35 497 条数据,信任关系包含 1 853 条数据。

3.2 评估指标

由于数据集中没有确切的群组及群组对项目的实际评分,本文实验将对同一项目具有相同评分且不小于 5 个用户的单位作为一个群组,则该相同评分可看作群组对项目的实际评分,这将得到 N 个群组及每个群组对一个项目的确切评分。实验使用以下 2 种评估指标:

第 1 个评估指标是均方根误差(RMSE),计算公式如下:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i,j} (\mathbf{R}_{i,j} - \hat{\mathbf{R}}_{i,j})^2} \quad (25)$$

在第 2 个评估指标计算方式中,将群组评分为 5 分的项目作为该群组的应推荐项目,随机选取 $M-1$ 个项目与该项目一起预测群组评分并根据评分高低进行排序,计算应推荐项目在前 K 个项目出现的频率并作为命中率(HR@K)。

$$\text{HR@K} = \frac{1}{N} \sum_i |\text{Rank}_i \leq K| \quad (26)$$

其中: Rank_i 表示项目 i 在该 K 个项目中的排序名次。

本次实验选择的 M 值为 20。

3.3 实验过程及分析

3.3.1 λ_U 和 λ_V 的选择

式(20)中的 λ_U 和 λ_V 也可以起到正则化系数的作用,相较于 λ_F 取值较小,通常在 0.000 1 到 0.1 之间。

图3展示了在其他参数不变的情况下,参数 λ_u 和 λ_v 取0.000 1至0.01时对RMSE和HR@5的影响。可以看出,当 λ_u 和 λ_v 都取值在0.001左右时,RMSE

最小且HR@5最大,即此时的实验效果最好。因此,在后续实验中将采用0.001作为 λ_u 和 λ_v 的取值。

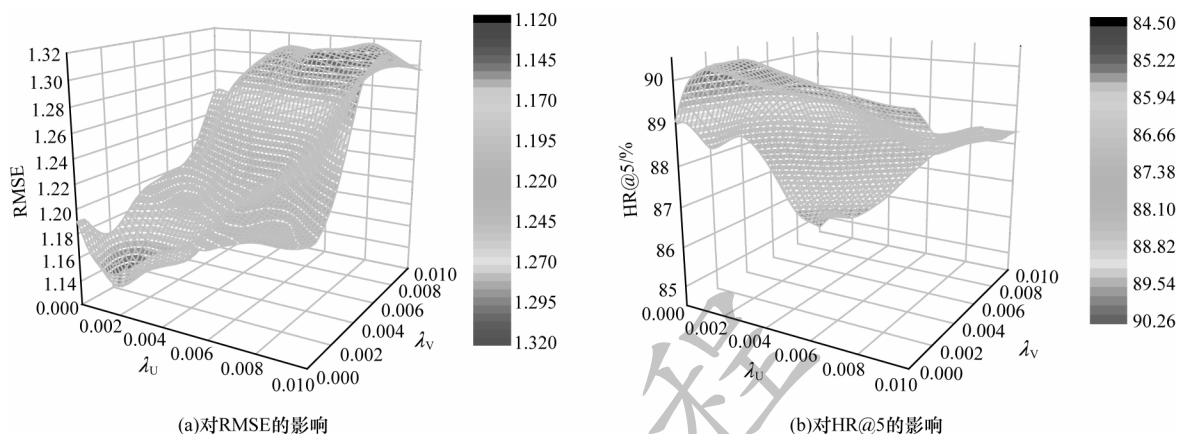


图3 参数 λ_u 和 λ_v 对推荐性能的影响

Fig.3 Influence of parameters λ_u and λ_v to recommended performance

3.3.2 λ_F 的选择

本文算法相较于经典PMF算法改进之处在于后验推导公式中加入了 $\frac{\lambda_F}{2} \sum_{i=1}^M \left\| U_i - \sum_{t=1}^M F_{k,t} U_t \right\|_{Fro}^2$ 部分,当 λ_F 取0时,将退化为普通PMF算法。

图4展示了当 λ_F 取值在0.1至20之间时推荐结果的RMSE和HR@5变化情况。从中可以看出,随着 λ_F 的增大,RMSE和HR@5均呈现一定的变化。RMSE逐渐减小后慢慢开始回升,而HR@5在逐渐提高后开始回落。在此后的对比实验中,将采用10作为 λ_F 的取值。

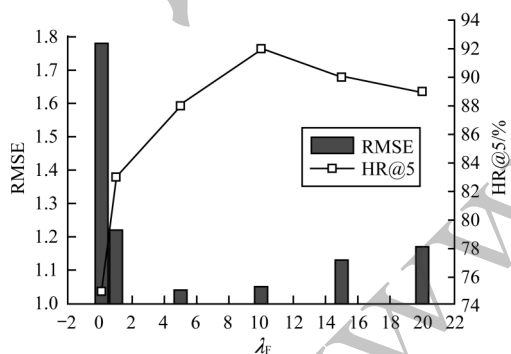


图4 不同 λ_F 取值下的RMSE和HR@5

Fig.4 RMSE and HR@5 with different λ_F

3.3.3 对比算法

在本文实验中,选择使用以下对比算法:

1)经典矩阵分解(Matrix Factorization, MF)算法。以下算法在协同过滤方法中“共现矩阵”的基础上,加入了隐向量的概念,加强了模型处理稀疏矩阵的能力,具有更好的扩展性与灵活性。但是与协同过滤一样,不方便加入用户、项目和上下文相关特征,使得矩阵分解丧失了利用很多有效信息的

机会。

2)经典PMF算法。相比矩阵分解,PMF算法可以更好地应对数据稀疏问题。

3)多层感知机(Multi-Layer Perceptron, MLP)算法。通过多层神经元拟合用户对项目的评分。

4)NeuMF(Neural Matrix Factorization)算法。该算法结合了广义矩阵分解和多层感知机,可以同时提取低维和高维特征^[20],但由于该模型也是基于协同过滤的思想构造的,因此并没有引入更多其他类型的特征。

5)RippleNet算法。该算法将知识图谱作为推荐系统的辅助信息来源,利用实体关系三元组分析用户的偏好倾向并推理出哪些新的实体项可能是该用户可能喜欢的。其中知识图谱指的是由类似(阿甘正传,电影-导演,罗伯特·泽米吉斯)的事实三元组构成。模型输入为:用户 u 和用户的历史纪录 $V_{u,i}$,以及项目 v ;模型输出:用户点击/选择/喜欢该项目的概率,由于该算法旨在输出用户喜欢项目的概率,在本次实验中改为回归方法以计算评分^[21]。

6)SIGR(Social Influence-based Group Recommender)算法。该算法将注意力机制和一种二分图嵌入模型BGEM作为基本块。采用了注意力机制来学习每个用户的社交影响,并将其应用于不同的群组中。为准确捕捉用户社交影响,SIGR设计了一种新的深度社交影响学习框架,来利用并整合用户的全局/局部社交网络结构信息^[22]。

为控制各方法的数据集与输出类型相同,将MLP、NeuMF的输出由原本的分类改为回归并计算评分,每层神经元个数分别为64,32,16,8。

3.3.4 实验结果及分析

图5展示了随着迭代次数的增加,FPMF方法与各对比算法使用均值融合策略时的RMSE收敛情况。

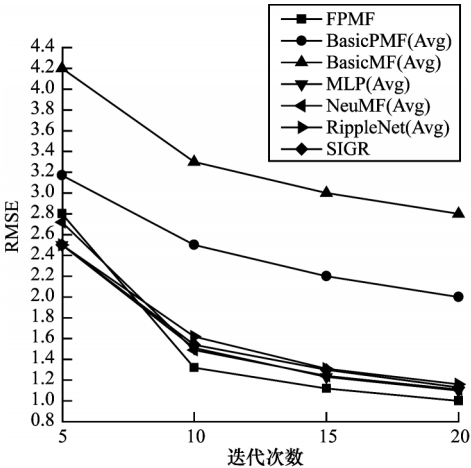


图5 各算法训练时的RMSE收敛情况
Fig.5 RMSE convergence of each algorithm when training

在图6中,其他对比方法分别使用了均值策略和最小痛苦策略作为群组融合方法。以下为各方法的超参数调整达到最优后的RMSE效果。

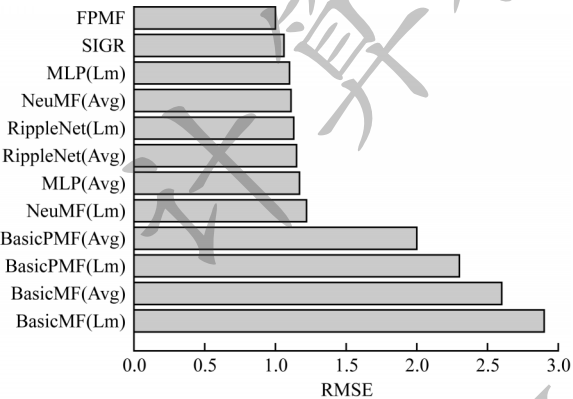


图6 各算法超参数调整达到最优的RMSE
Fig.6 RMSE of each algorithm when the super parameters are optimal

表1和表2分别展示了Epinions和FilmTrust数据集下各方法的HR@K。可以看出,原本在RMSE效果明显占优的MLP方法和NeuMF方法使用回归方法计算评分后进行排序的HR@K不再具有优势,而FPMF方法在该指标下的性能高于其他方法。

同时由表2可以看出,FPMF算法在FilmTrust数据集上的表现略逊于在Epinions上的表现,较为合理的解释是FilmTrust数据集中的用户信任度关系相比于Epinions较为稀疏,FPMF算法精度较为依赖信任度数据,失去信任度数据后的算法性能开始向基础PMF靠拢。

表1 Epinions数据集上不同K值下的命中率
Table 1 Hit rate with different K on Epinions dataset %

推荐算法	HR@1	HR@2	HR@3	HR@5
FPMF	44	71	85	91
PMF(Avg)	17	30	52	65
PMF(Lm)	23	34	51	65
SIGR	39	62	77	85
RippleNet(Avg)	32	49	58	77
RippleNet(Lm)	28	44	51	69
MF(Avg)	15	21	40	60
MF(Lm)	16	25	38	55
MLP(Avg)	8	15	22	40
MLP(Lm)	8	12	23	35
NeuMF(Avg)	7	13	18	43
NeuMF(Lm)	8	17	20	39

表2 FilmTrust数据集上不同K值下的命中率
Table 2 Hit rate with different K on FilmTrust dataset %

推荐算法	HR@1	HR@2	HR@3	HR@5
FPMF	35	65	79	84
PMF(Avg)	19	31	52	61
PMF(Lm)	22	29	56	66
SIGR	32	55	64	78
RippleNet(Avg)	31	48	55	73
RippleNet(Lm)	29	41	51	66
MF(Avg)	17	20	39	61
MF(Lm)	14	18	44	59
MLP(Avg)	10	19	25	42
MLP(Lm)	11	14	21	39
NeuMF(Avg)	8	19	29	36
NeuMF(Lm)	9	12	25	32

3.3.5 模型可解释性

在Epinions数据集的测试结果中,小组成员393号受组内其他成员信任。表3所示数据为该组对不同项目的群组评分以及组内成员的个体评分,由于393号成员的被信任程度较高,因此令其在群组偏好融合过程中具有较高的权重,从表中也可以看出,群组评分结果与393号成员个体评分具有较高的相关度,这进一步说明了模型的可解释性。

表3 群组评分及单个组员评分案例
Table 3 Example of ratings of group and individual member

用户ID	项目ID及群组评分			
	1033(3.8)	266(2.1)	9658(2.5)	660(4.4)
393	4.1	2.3	2.2	4.6
175	3.2	4.6	3.8	4.2
2411	1.9	3.9	3.2	3.8
611	3.1	1.2	1.7	3.2
398	4.7	2.8	1.6	1.9
1855	2.9	4.1	3.2	2.5
884	4.4	3.8	2.4	4.9
3875	3.5	3.2	1.4	3.9
4348	2.3	1.3	3.6	3.5
1197	2.7	3.1	4.1	2.7

4 结束语

针对当前群组推荐算法较少考虑成员交互关系的问题,本文提出融合用户信任度的概率矩阵分解群组推荐算法。对概率矩阵分解算法进行改进,在后验概率计算部分引入用户间的信任度,充分挖掘单用户的历史偏好。同时,在群组融合部分引入用户的信任度隐特征向量来计算评分权重,充分利用群组的社交关系特性。实验结果表明,本文方法能够有效提升群组推荐性能。下一步将在概率矩阵分解中融合信任度数据本身的置信程度以及用户的评分习惯等因素,进一步提升本文算法的推荐性能。

参考文献

- [1] YU W H, ZHANG H D, HE X N, et al. Aesthetic-based clothing recommendation[C]//Proceedings of 2018 World Wide Web Conference. New York, USA: ACM Press, 2018: 649-658.
- [2] CHAO Y W, AHMED A, BEUTEL A, et al. Recurrent recommender networks[C]//Proceedings of the 10th ACM International Conference on Web Search & Data Mining. New York, USA: ACM Press, 2017: 495-503.
- [3] COVINGTON P, ADAMS J, SARGIN E. Deep neural networks for YouTube recommendations[C]//Proceedings of 2016 ACM Conference on Recommender Systems. New York, USA: ACM Press, 2016: 191-198.
- [4] LI X M, XU G Q, TANG M H. Community detection for multi-layer social network based on local random walk[J]. Journal of Visual Communication and Image Representation, 2018, 57: 91-98.
- [5] ARDISSONO L, GOY A, PETRONE G, et al. INTRIGUE: personalized recommendation of tourist attractions for desktop and hand held devices[J]. Applied Artificial Intelligence, 2003, 17(8/9): 687-714.
- [6] 张玉洁, 杜雨露, 孟祥武. 组推荐系统及其应用研究[J]. 计算机学报, 2016, 39(4): 745-764.
ZHANG Y J, DU Y L, MENG X W. Research on group recommendation system and its application[J]. Chinese Journal of Computers, 2016, 39(4): 745-764. (in Chinese)
- [7] SALAKHUTDINOV R. Probabilistic matrix factorization[C]//Proceedings of International Conference on Neural Information Processing Systems. [S. l.]: Curran Associates, Inc., 2007: 1257-1264.
- [8] 康雁, 李涛, 李浩, 等. 融合知识图谱与协同过滤的推荐模型[J]. 计算机工程, 2020, 46(12): 73-79, 87.
KANG Y, LI T, LI H, et al. Recommendation model fusing with knowledge graph and collaborative filtering[J]. Computer Engineering, 2020, 46(12): 73-79, 87. (in Chinese)
- [9] ROY S B, AMER-YAHIA S, LA A C, et al. Space efficiency in group recommendation[J]. The VLDB Journal, 2010, 19(6): 877-900.
- [10] WANG W, ZHANG G, LU J. Member contribution-based group recommender system[J]. Decision Support Systems, 2016(87): 80-93.
- [11] SHI J, WU B, LIN X. A latent group model for group recommendation[C]//Proceedings of 2015 IEEE International Conference on Mobile Services. Washington D. C., USA: IEEE Press, 2015: 233-238.
- [12] LI X M, XU G Q, JIAO L T, et al. Multi-layer network community detection model based on attributes and social interaction intensity[J]. Computers and Electrical Engineering, 2019, 77: 300-313.
- [13] YANG B, LEI Y, LIU J, et al. Social collaborative filtering by trust[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(8): 1633-1647.
- [14] SINGH M, MECHROTRA M. Impact of biclustering on the performance of biclustering based collaborative filtering[J]. Expert System with Applications, 2018, 113: 443-456.
- [15] CHEN H, YIN H, WANG W, et al. PME: projected metric embedding on heterogeneous networks for link prediction[C]//Proceedings of the 24th ACM SIGKDD International Conference. New York, USA: ACM Press, 2018: 1-5.
- [16] CHEN S, PENG Y. Matrix factorization for recommendation with explicit and implicit feedback[J]. Knowledge-Based Systems, 2018, 158: 109-117.
- [17] CHEN R, HUA Q Y, WANG B, et al. A novel social recommendation method fusing user's social status and homophily based on matrix factorization technique[J]. IEEE Access, 2019, 7: 18783-18798.
- [18] 王根生, 潘方正. 融合多元异构信息的矩阵分解推荐算法[J]. 小型微型计算机系统, 2020, 41(7): 1406-1412.
WANG G S, PAN F Z. Matrix factorization recommendation algorithms based on multiple heterogeneous information fusion[J]. Journal of Chinese Computer Systems, 2020, 41(7): 1406-1412. (in Chinese)
- [19] LAI C H, LIU D R, LIN C S. Novel personal and group-based trust models in collaborative filtering for document recommendation[J]. Information Sciences, 2013, 239(1): 31-49.
- [20] HE X N, LIAO L Z, ZHANG H W, et al. Neural collaborative filtering[C]//Proceedings of the 26th International Conference on World Wide Web. New York, USA: ACM Press, 2017: 173-182.
- [21] WANG H W, ZHANG F Z, WANG J L, et al. RippleNet: propagating user preferences on the knowledge graph for recommender systems[C]//Proceedings of the 27th ACM International Conference on Information and Knowledge Management. New York, USA: ACM Press, 2018: 417-426.
- [22] YIN H, WANG Q, ZHANG K, et al. Social influence-based group representation learning for group recommendation[C]//Proceedings of the 35th IEEE International Conference on Data Engineering. Washington D. C., USA: IEEE Press, 2019: 1-5.

编辑 金胡考