

DOI: 10.7652/xjtuxb201610018

一种利用多群组智慧的协同推荐算法

郑修猛, 陈福才, 吴奇, 朱宇航, 黄瑞阳

(国家数字交换系统工程技术研究中心, 450002, 郑州)

摘要: 针对当前基于社会网络的推荐系统大多数采用一般的启发式方法,存在节点复杂路径选择和信任弱传递现象导致推荐精确度不高的问题,以及针对推荐系统固有的冷启动问题,提出了一种利用多群组智慧的协同推荐算法。该算法首先根据用户的社会属性和社会信任关系信息进行群组划分,将用户分为多个不同的群组;然后分析群组中用户的社会活动和社会关系等,建立一种利用多群组的评分预测模型,并利用群组评分预测新用户的评分。该算法通过对社会网络进行深层次的群组挖掘,利用多群组智慧可以有效提高推荐效果,利用群组评分可改善对冷启动用户的推荐。仿真实验表明,该算法相比传统的协同推荐算法在效果评分上提高了约 0.2,相比其他社会化推荐算法进一步提高了约 0.02,并有效解决了冷启动问题。

关键词: 协同推荐;社会网络;群组智慧;冷启动

中图分类号: TN311.1 **文献标志码:** A **文章编号:** 0253-987X(2016)10-0118-07

A Collaborative Filtering Recommendation Algorithm Using Multiple Groups Intelligence

ZHENG Xiumeng, CHEN Fucui, WU Qi, ZHU Yuhang, HUANG Ruiyang

(National Digital Switching System Engineering & Technological R&D Center, Zhengzhou 450002, China)

Abstract: A collaborative filtering recommendation algorithm using the multiple groups intelligence is proposed to address the problems that most of the current recommendation systems that are based-on social networks use the general heuristic methods and have drawbacks of choice of complex paths and weak transferring of trust phenomenon which leads to low recommendation precision, and there exists the inherent cold start problem in recommendation systems. The proposed algorithm divides users into several different groups from their social attribution and social trust relationship information. Then predictive models are built based on multi-group through analyzing the user's social activities and social relationships in the groups and the evaluated group scores are used to predict ratings of new users. The algorithm uses a deep group mining in a social network, the multi group intelligence to effectively improve the effect of the recommendation, and the evaluated group score to improve the recommendation of the cold start users. Simulation results and comparisons with the traditional collaborative recommendation algorithm and other social recommendation algorithms show that the recommendation effects of the proposed algorithm improve by about 0.2 and 0.22, respectively, and that the problem of cold start is effectively solved.

Keywords: collaborative filtering recommendation; social network; group intelligence; cold start

收稿日期: 2016-01-13。 作者简介: 郑修猛(1989—),男,硕士生;陈福才(通信作者),男,研究员。 基金项目: 国家自然科学基金资助项目(61171108);国家重点基础研究发展计划资助项目(2012CB315901,2012CB315905);国家科技支撑计划资助项目(2014BAH30B01)。

网络出版时间: 2016-07-22 网络出版地址: <http://www.cnki.net/kcms/detail/61.1069.T.20160722.1953.004.html>

传统的协同过滤推荐系统普遍存在推荐精确度不高和冷启动问题。将社会网络中的人与人之间的信任关系应用到推荐系统中使得推荐结果更加真实,能够提高推荐的精确度和缓解冷启动问题。

在相关研究中,研究者通过纳入社会化上下文信息等,提出了社会化推荐算法^[1]。其中在传统的协同推荐算法基础上,结合社会网络中的社会信任关系,研究者提出了基于信任的推荐算法^[2]。早期的研究大数集中在信任网络中的信任搜索策略和信任传递等方面,相继出现了 Tidal Trust 算法^[3]、Mole Trust 算法^[4]和 Trust Walker 算法^[5]等推荐算法,但是这些算法存在复杂路径选择和信任弱传递等问题。后来,研究者将矩阵分解模型引入到基于信任的推荐系统中,出现了 Social MF 算法^[6]、Trust MF 算法^[7]等推荐算法,但是此类研究大多数从用户或者资源对象的角度进行,却很少基于社会网络内部结构来对推荐进行分析。随着社会化网络理论的成熟,基于社会网络的社团发现技术开始在传统的协同推荐算法中使用。在相关研究中,文献^[8]通过用户聚类手段等,删除群组外一些不活跃用户,降低了待分析用户的规模,但是对于边缘冷启动用户的推荐问题无法解决;Liu 等人通过将社会网络中的特征向量加权来促进群组的贡献,并在社团中进行用户评分的规范化处理^[9];Li Yu 等通过社团特征预测用户的偏好,然后对用户-项目矩阵进行填充等^[10]。但是,以上基于群组发现的推荐算法均未考虑网络中的节点同时属于多个群组的情况。在真实的社会网络里,1 个用户可以属于多个不同类型的群组,同时这些群组大部分是重叠的(如图 1 所示)。用户在不同群组中的归属度是不同的,反过来不同社会群组又对特定的个体在项目的选择上也会产生不同的影响。

因此,本文考虑用户的多群组现象,在社会群组推荐系统的基础上,通过引入社团发现中的层叠社团理论和社会计算中的群体智慧相关概念,提出了

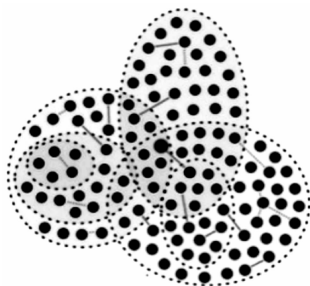


图 1 重叠群组示例图

利用多群组智慧的协同推荐算法(MGICF)。本文基于用户的社会属性、社会行为和社会信任网络等,采用 2 种不同的群组划分策略产生多个聚合的模型,结合群组中用户偏好信息,建立预测模型。与现有方法相比,本文方法可以有效提高推荐效果,利用群组评分可解决冷启动问题。

1 社会群组的相关概念

1.1 社会群组描述

在推荐系统中,用户集合为 $\{u_1, u_2, \dots, u_n\}$,项目集合为 $\{I_1, I_2, \dots, I_m\}$,评分矩阵为 R , R 中元素 R_{uj} 代表用户 u 对项目 j 的评分,取值范围为 $[1, 5]$ 。本文将用户分成不同的群组 $\{G_1, G_2, \dots, G_N\}$,群组中的用户社会网络关系数据可用图模型表示,每个点表示 1 个用户,每条边表示用户之间的社会关系,目标用户和群组中的邻居用户关系示意图如图 2 所示。图 2 中群组中的某个用户 u_1 属于 N 个不同的群组,在群组 G_1 中有 6 个用户,其他 5 个用户和 u_1 之间存在着必然的社会联系,本文用 $S(u, v)$ 表示群组中用户 u 和用户 v 之间的社会关系。

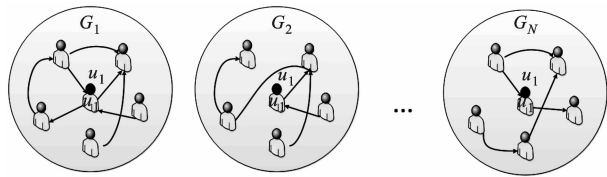


图 2 目标用户和群组中的邻居用户关系示意图

1.2 群组中社会关系定义

群组中用户之间的社会关系可以是真实的社会网络关系,也可以用用户共同参与的社会活动等表示。因此,本文定义了群组中的 3 种社会关系。

定义 1 将用户是否存在共同评分定义为第 1 种社会关系,用 S_c 表示。若群组中的用户 u, v 之间有共同评分时, $S_c(u, v) = 1$, 反之 $S_c(u, v) = 0$, S_c 是对称无向的。

定义 2 将用户在真实的社会网络中的社会信任关系定义为第 2 种社会关系,用 S_T 表示。若群组中用户 u 信任用户 v , 则 $S_T(u, v) = 1$; 若用户 v 对用户 u 不存在信任关系则 $S_T(u, v) = 0$, S_T 是有向的。

定义 3 将用户之间的评分相似度定义为第 3 种社会关系,用 S_s 表示。 $S_s(u, v) = \text{sim}(u, v)$, S_s 是无向的。传统协同推荐算法通过修正的余弦相似度计算用户之间的评分相似度,如下式所示

$$\text{sim}(u, v) = \frac{\sum_{i \in I_u \cap I_v} (R_{ui} - \bar{R}_u)(R_{vi} - \bar{R}_v)}{(\sum_{i \in I_u} (R_{ui} - \bar{R}_u)^2)^{1/2} (\sum_{i \in I_v} (R_{vi} - \bar{R}_v)^2)^{1/2}} \quad (1)$$

式中: I_u 和 I_v 分别为用户 u 、 v 评分过的项目集合; $\bar{R}_u$ 和 $\bar{R}_v$ 分别为用户 u 、 v 的平均评分。

1.3 群组中社会关系分析

一般认为用户在群组中的影响力越大,对群组的评分影响力越大。文献[11]在分析真实信任关系数据集时得出:一个用户的跟从者越少,其更容易选择信任其所信任的用户。因此,可以有这样的解释:如果其他用户对用户 i 信任度越大,表示用户 i 具有更鲜明的偏好,不易受到群组中其他用户和整个群组的影响。由群组中的 3 种社会关系,本文又定义了用户信任度、群组归属度和群组贡献度如下。

定义 4 用户信任度,表示群组中用户在信任网络中的重要程度,用 D_T 表示。

定义 5 群组归属度,表示用户归属于不同群组的程度,用 D_B 表示。

定义 6 群组贡献度,表示推荐系统中社会群组对用户评分预测的贡献程度,用 D_C 表示。

用户的信任度越大,用户的偏好越明显;用户对群组的归属度越大,表明用户对群组的信赖程度越大。因此,本文假定群组贡献度与用户的信任度成反比,与群组归属度是等价的,则群组的贡献度为

$$D_C \propto \frac{1}{D_T}; \quad D_C = D_B \quad (2)$$

本文利用定义 2 的第 2 种社会关系,利用 PageRank 算法来计算用户在群组中的信任度

$$D_T(u) = \frac{1-d}{M} + d \left(\sum_{i=1}^n \frac{D_T(i)}{C(D_T(i))} \right) \quad (3)$$

式中: M 为群组中的所有用户数; n 为信任 u 的用户的总数; $C(D_T(i))$ 代表 i 信任用户数; d 为损失参数,据经验默认为 $d=0.85$ 。

于是,本文定义群组的贡献度为

$$D_C(u) \approx \frac{1}{D_T(u) + \alpha} + \beta \quad (4)$$

式中: $D_T(u)$ 为用户 u 在群组中的信任度。例如,当用户在群组信任度为 1 时,表示用户偏好明确,群组对其贡献度为 0,当用户与其他用户不存在社会关系时,信任度为 0,该用户的评分可设定为群组评分,因此群组对其影响为 100%。点 $(0, 1)$ 和 $(1, 0)$ 代入求解 α 和 β 。拟合曲线如图 3 所示。

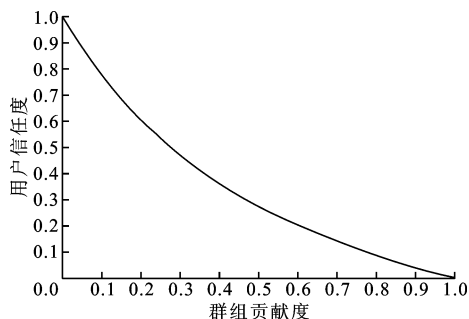


图 3 群组贡献度和用户信任度反比例关系拟合曲线

2 利用多群组智慧的推荐算法模型

2.1 社会群组划分策略

本文根据用户的显性特征和隐性特征分别对用户群组进行划分,提出了 2 种社会群组划分策略:基于人口统计的群组划分策略和基于信任矩阵分解的群组划分策略。

(1) 基于人口统计的群组划分策略。在 MovieLens 数据集中除了包含用户-项目的评分信息,还有用户属性信息(年龄、性别和职业等),以此对用户进行群组划分,这是一种最为简便的划分策略。一般具有相同属性的用户兴趣爱好越相似,比如女性观众对韩剧的喜好程度比男性高,男性更热衷于电子产品。利用用户的社会属性进行群组划分,1 个用户具有多个社会属性,用户会同时被划分到多个群组中。用户群组中与其他用户又发生必然的社会关系,通过分析群组中的用户之间的关系,建立多群组预测模型。

(2) 基于信任矩阵分解的群组划分策略。在 Cias 和 Epinions 数据集中不但包含了评分信息,还包含已有的社会信任关系信息。用户之间的社会信任关系构成的连接矩阵是规模比较大的非负矩阵。针对复杂网络中的社团发现算法有很多,目前代表性的算法有 GN 算法^[12]和 Newman 快速算法^[13]等。但是,以上算法都是对网络进行单一的社团划分,网络中的节点被划分到多个完全独立的社团中,而在真实网络中,1 个节点很可能属于多个不同的社团,根据社团发现中重叠社团相关理论,可以利用矩阵的特征向量来发现网络社团重叠结构。因此,本文采用文献[14]中非负矩阵分解的方法对用户进行群组划分。

假设拥有 n 个节点的信任网络的邻接矩阵为 $A \in \mathbf{R}^{n \times n}$,则非负矩阵分解方法的策略为:通过寻找最

大近似原始网络数据 $\mathbf{A}$ 的 2 个低秩因子矩阵 $\mathbf{W}$ 和 $\mathbf{H}$ 来实现社区发现。分解后得到的基向量矩阵 $\mathbf{W}$ 表示网络降维后的群组特征,具有稀疏性和线性无关性,而归属矩阵 $\mathbf{H}$ 则表示相应节点与群组的隶属程度。于是,就把用户划分为 K 个不同的群组,归属矩阵 $\mathbf{H}$ 可以转换为用户的群组的归属度 U_{belong} 。在非负矩阵分解过程中,一般采用欧几里德距离最小化方式,优化的目标函数 $O(E)$ 为

$$\min_{\mathbf{W}, \mathbf{H}} O(E) = \min_{\mathbf{W}, \mathbf{H}} \|\mathbf{A} - \mathbf{WH}\|_F^2 \quad (5)$$

s. t. $\mathbf{W} \geq 0, \mathbf{H} \geq 0$

式中: $\|\cdot\|_F$ 为 Frobenius 范数,用来度量目标函数的逼近程度; $\mathbf{W} \in \mathbf{R}^{n \times K}$ 和 $\mathbf{H} \in \mathbf{R}^{K \times n}$ 分别是分解之后得到的关于模式节点的基矩阵和归属矩阵, n 表示网络中的节点个数, K 表示相关模式节点子空间的聚类个数,反映了信任网络中存在的群组的个数。

2.2 利用多群组的目标用户评分预测

在对目标用户 u 进行评分预测时,本文在传统的基于用户的协同推荐算法的基础上,考虑群组内相邻用户的影响以及群组对用户贡献度,通过加权求和预测用户的评分

$$\hat{R}_{ui} = \bar{R}_u + \sum_{n=1}^N \left(\frac{\sum_{v \in X(u, i, n)} S_S(u, v) R_{vi} - \bar{R}_v}{\sum_{v \in X(u, i, n)} |S_S(u, v)|} \cdot \frac{D_C(n)}{\sum_{i=1}^n D_C(i)} \right) \quad (6)$$

式中: N 为用户 u 所属的群组的个数; $\bar{R}_u$ 为用户 u 的平均得分; $D_C(n)$ 为群组 n 对用户 u 评分的贡献度。利用传统的协同推荐算法中基于用户的预测算法,从群组 n 中找出与目标用户 u 最相似且对项目 i 评分过的 k 个用户 v 作为邻居集合 $X(u, i, n)$, 计算 k 个邻居对项目 i 评分的加权平均值,作为用户 u 对项目 i 的评分预测值。

2.3 新用户的评分预测问题

当用户的邻居用户过少或者为新用户时,本文利用群组推荐系统的相关理论,为新用户进行评分预测。群组推荐系统从为特定用户的个性化推荐扩展到为整个群组的推荐,当然群组推荐系统不仅可以为群组推荐,也可以为单个用户推荐,并能解决多准则推荐问题和新用户引起的冷启动问题^[15]。在群组推荐系统研究中,研究者们提出了多种不同的评分策略,如平均满意度、最小失望度、最大满意度等合并特定用户的推荐列表,使推荐结果让群组中的所有人一致满意。

本文考虑群组评分对用户评分预测的影响,设 $G_R(n, i)$ 为群组 n 的群组评分,当用户 u 为新用户或者邻居用户数小于 k 时,可以根据用户 u 的所属社会群组评分 $G_R(n, i)$ 代替式(6)基于邻居相似度的评分

$$\hat{R}_{ui} = \sum_{n=1}^N \left(G_R(n, i) \cdot \frac{D_C(n)_n}{\sum_{i=1}^n D_C(i)} \right) \quad (7)$$

基于社会群组推荐算法的目标是使群组评分尽最大可能拟合群组所有成员的偏好,群组图中节点的链接数越多,表示用户在群组的权重越高,对群组评分时的影响就越大。因此,本文提出基于用户权重的群组评分策略。

首先通过用户之间的第 1 种社会关系,可以计算用户 u 在群组评分时的权重

$$W(u) = \frac{\sum_{j=1}^M S_C(u, j)_{uj}}{\sum_{i, j \in G} S_C(i, j)_{ij}} \quad (8)$$

式(8)中, M 表示群组 G 中所有用户的数量,通过计算用户 u 存在的社会关系与群组 G 中所有社会关系的比值,可得群组的最终评分为

$$G_{Rj} = \sum_{u=1}^N W(u) R_{uj} \quad (9)$$

基于人口统计学的群组划分伪代码程序如下。

输入 用户属性信息 $U_A(a_1, a_2, \dots, a_n)$ 、评分 R_{ui} 、信任 T 。

输出 评分预测值 $\hat{R}_{ui}$ 。

(1) for $j \leftarrow 1$ to n do

(2) $G_{Rj} \leftarrow \text{predict}(W(u), R_{uj})$

(3) end

(4) for $u \leftarrow 1$ to n do

(5) $D_T(u) \leftarrow \text{PageRank}(D_T(i))$

(6) end

(7) $D_C(u) \leftarrow \frac{1}{D_T(u) + \alpha} + \beta$

(8) $\hat{R}_{ui} \leftarrow \text{predict}(N, S_S, D_C(u), G_{Ri})$

基于非负矩阵分解的群组划分伪代码如下。

输入 用户属性信息 $U_A(a_1, a_2, \dots, a_n)$ 、评分 R_{ui} 、信任 T 。

输出 评分预测值 $\hat{R}_{ui}$ 。

(1) 设置目标函数 $\min_{\mathbf{W}, \mathbf{H}} \|\mathbf{A} - \mathbf{WH}\|_F^2$

(2) for $j \leftarrow 1$ to n do

(3) $G_{Rj} \leftarrow \text{predict}(W(u), R_{uj})$;

- (4) end
- (5) for $u \leftarrow 1$ to n do
- (6) $D_C(k) = D_B(k)$
- (7) end
- (8) $\hat{R}_u \leftarrow \text{predict}(N, S_S, D_C(u), G_{R_i})$

2.4 算法复杂度分析

按照本文设计的 2 种群组划分策略计算基于多群组智慧的协同推荐算法的复杂度。基于人口统计的群组划分策略,用户节点为 n ,用户的社会属性有 g 种,需要划分为 g 个群组,但主要仍是计算两两用户之间的相似度,计算复杂度为 $O(n^2)$ 。在不同群组的贡献度求解时,还需要计算用户的信任度,在每个群组中构建的隐性信任矩阵上进行 PageRank 算法的计算复杂度也约为 $O(n^2)$,因此算法的复杂度为 $O(n^2)$ 。当然,用户的信任度也可以离线计算。基于信任矩阵分解的群组划分策略中,除了需要计算两两用户之间的相似度,还需要对信任矩阵进行分解,其中信任关系图中边数为 m ,非负矩阵分解的算法复杂度除了矩阵乘法运算之外,还取决于迭代次数 l ,其算法复杂度约为 $O(lmk^2)$,综合两者,算法的复杂度为 $O(n^2) + O(lmk^2)$ 。

3 仿真实验

为了评估本文所提 MGICF 算法的推荐效果,采用推荐系统中常用的 MovieLens、Ciaos 和 Epinions 数据集进行测试实验,并与现有的推荐算法进行对比实验。

3.1 实验数据介绍

MovieLens 数据集中除了用户项目评分信息外,还包含电影及用户的属性信息。在用户属性文档中包括年龄、性别、职业等相关信息,在项目属性文档中包括电影的类型。在 Ciaos 和 Epinions 数据集中除了包含评分信息还包含用户之间的信任关系。数据集统计情况如表 1 所示。

表 1 实验数据集节点数量和关系数量统计情况

数据集	用户数	项目数	评分	信任关系数
MovieLens	943	1 682	100 000	0
Epinions	40 163	139 738	664 824	487 183
Ciaos	7 375	99 746	278 483	111 781

3.2 算法评价指标

在实验过程中,可利用推荐系统中 2 个常用的误差度量参数——平均绝对偏差 M_{AE} 和均方根误差

R_{MSE} ,当 M_{AE} 或者 R_{MSE} 越小时,表明预测结果越精确。因此,本文采用 M_{AE} 和 R_{MSE} 作为推荐性能评价指标,计算公式如下

$$M_{AE} = \frac{\sum_{R_{ij} \in S} |\hat{R}_{ij} - R_{ij}|}{|S|} \quad (10)$$

$$R_{MSE} = \left(\frac{1}{|S|} \sum_{R_{ij} \in S} (\hat{R}_{ij} - R_{ij})^2 \right)^{1/2} \quad (11)$$

式中: R_{ij} 是测试数据中的真实评分; $\hat{R}_{ij}$ 是通过推荐算法预测的评分; S 表示测试集。

3.3 实验结果及分析

首先根据基于人口统计的群组划分策略,在 MovieLens 数据集上进行推荐算法的性能测试实验。将传统的基于用户的推荐算法(User-based)、按照用户社会属性划分的为单一群组内的协同推荐算法(Group by Gender, Group by Age, Group by Occupation)和多群组智慧的推荐算法 MGICF(G1+G2+G3)的推荐效果进行比较。

根据 MovieLens 数据集,首先按性别分为 2 个群组,男性为第 1 组,女性为第 2 组;年龄按照年龄段分为 5 个群组,将低于 21 年龄组设为第 1 组,将 21 到 30 年龄组设为第 2 组,将 31 到 40 年龄组设为第 3 组,将 41 到 50 年龄组设为第 4 组,将大于 50 年龄组设为第 5 组;按照职业细分为 21 个群组,这些职业分别为管理员、艺术家、医生、教育家、工程师、艺人、经理、按摩师、家庭主妇、律师、图书管理员、市场投资者、农民、程序员、退休人员、销售员、科学家、学生、技术员、作家、其他,将这些职业按照 1 至 21 进行分组,划分后群组中用户数统计分布图如图 4 所示。

由于 MovieLens 数据集中不包含用户之间的直接信任关系,实验过程中需要重新定义用户之间的隐性信任关系。假设用户 a 和用户 b 存在共同的评分,则 a 和 b 之间存在信任关系;否则 a 和 b 不存在信任关系。其中,当用户 a 评论过的项目数大于或等于用户 b 评论过的项目数, b 对 a 的信任值为 1;反之,则 a 对 b 的信任值为 0。用户之间的隐性信任关系如下

$$\left. \begin{aligned} \text{trust}(b \rightarrow a) &= 1, & I_a \cap I_b \neq \emptyset; & |I_a \cap I_b| \geq \varphi \\ & & |I_a| \geq |I_b| \\ 0, & & \text{其他} \\ \text{trust}(a \rightarrow b) &= 1, & I_a \cap I_b \neq \emptyset; & |I_a \cap I_b| \geq \varphi \\ & & |I_a| \leq |I_b| \end{aligned} \right\} \quad (12)$$

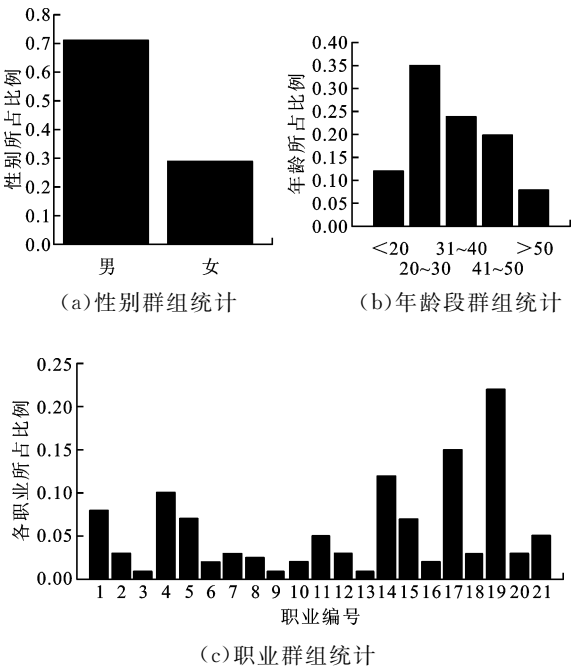


图 4 不同群组数据统计

在每个群组中利用 PageRank 算法进行用户信任度 $D_T(u)$ 计算,然后计算群组对用户的评分预测过程中的贡献度 $D_C(n)$ 。

图 5 为多群组智慧推荐算法与其他几种算法的比较。由图 5 实验结果得知,基于人口统计的群组划分策略,可以在一定程度上提高推荐系统的准确度,但整体结果并不是十分明显。分析原因如下:①由于在社会属性划分时,用户的显性社会特征捕捉不全面,加上现有数据集的局限性,可利用的社会属性仅有 3 种;②由于在用户的评分过程,用户的性别、年龄、职业等特征一定程度上已经反映在评分结果中。

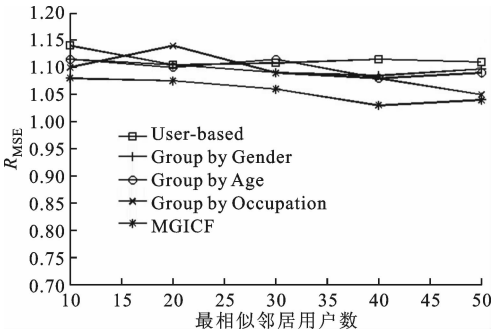


图 5 基于用户、单一群组和多群组智慧的推荐算法效果比较

根据第 2 种基于信任矩阵分解的群组划分策略,分别在 Ciao 和 Epinions 数据集上进行测试。将本文算法的测试结果与传统的基于近邻用户的协同推荐算法 UserKNN、基本的矩阵分解算法 Bi-

asedMF 以及 2 种基于信任矩阵分解算法 SocialMF 和 TrustMF 进行比较,如表 2 所示。

表 2 5 种不同算法在 2 个数据集上的实验结果对比

算法	M_{AE}		R_{MSE}	
	Ciaos	Epinions	Ciaos	Epinions
UserKNN	0.899	0.918	1.183	1.207
BiasedMF	0.795	0.817	1.090	1.056
SocialMF	0.749	0.822	0.981	1.067
TrustMF	0.742	0.819	0.983	1.095
MGICF	0.733	0.816	0.974	1.066

在此之前,需要提前对群组个数 K 参数进行预估计,也就是非负矩阵分解迭代运算中 H 矩阵行数,这里采用的谱分析的方法。由第 1 种社会关系构成邻接矩阵,将邻接矩阵作为基矩阵进行估计,社团数可以由该邻接矩阵带的谱分析进行预估计,本文进行相关实验后将 K 取值 10。

图 6 为本文 MGICF 算法与现有几种算法的对比。由图 6 实验结果可知,基于多群组智慧的推荐效果在 Ciao 和 Epinions 数据集上均得到一些提升,相比传统的基于邻居用户的协同推荐算法,提升最为明显。但可能由于用户之间的信任关系数据依然存在稀疏性和弱社区效应问题,相比其他基于信任推荐算法,效果有所提升,但是幅度不是十分明显。不同数据集上提升效果对比,在 Ciao 数据集上面比 Epinions 数据集上面的效果更为明显,这是由于不同数据集中的信任关系强度和边数不同。

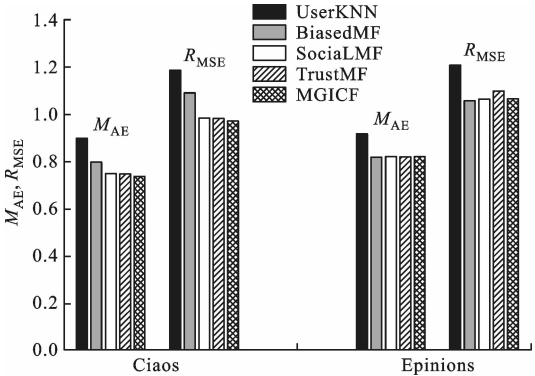


图 6 本文 MGICF 与现有算法比较

综上,本文利用多群组智慧的协同推荐算法有如下优点:①可以根据多重标准进行聚合,避免了传统协同推荐系统中默认单一群组,利用不同的兴趣群组协同进行目标用户评分预测,提高预测精确度;②利用群组评分对新用户的评分预测,解决了推荐系统的冷启动问题;③多群组智慧非常类似于并行化处理思想,为后续的并行化算法提供借鉴。

4 结 论

本文提出了一种利用多群组智慧的协同推荐算法,以及基于人口统计和基于信任矩阵分解的2种群组划分策略。本文算法通过将用户划分在不同的群组中,分析群组中的社会关系进行综合评分,能有效提高推荐精确度。此外,考虑群组评分,可以解决对冷启动用户进行推荐的问题。本文提出的多群组智慧的推荐算法均是基于用户群组的,并未考虑项目也可以属于多群组现象;并且随着时间变化用户兴趣和群组结构都会发生变化,考虑时间因素的影响也是下一步的研究方向。

参考文献:

- [1] 孟祥武,刘树栋,张玉洁,等. 社会化推荐系统研究[J]. 软件学报, 2015, 26(6): 1356-1372.
MENG Xiangwu, LIU Shudong, ZHANG Yujie. Research on social recommender systems [J]. Journal of Software, 2015, 26(6): 1356-1372.
- [2] MASSA P, BHATTACHARJEE B. Using trust in recommender systems: an experimental analysis [C]// Proceedings of the 2nd International Conference on Trust Management. Berlin, Germany: Springer-Verlag, 2004, : 221-235.
- [3] GOLBECK J. Personalizing applications through integration of inferred trust values in semantic web-based social networks [C]// Proceedings of Workshop on Semantic Network Analysis. Tilburg, Netherlands: Sun SITE Central Europe CEUR-WS, 2005: 15-28.
- [4] MASSA P, AVESANI P. Trust metrics on controversial users: balancing between tyranny of the majority and echo chambers [J]. Semantic Web and International System, 2007, 3(1): 39-64.
- [5] JAMALI M, ESTER M. Trustwalker: a random walk model for combining trust-based and item-based recommendation [C]// Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA: ACM, 2009: 397-406.
- [6] YANG B, LEI Y, LIU D, et al. Social collaborative filtering by trust [C]// Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence. Washington, D. C., USA: IJCAI, 2013: 2747-2753.
- [7] 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. New York, USA: ACM, 2010: 135-142.
- [8] LING Y, GUO D, FEI C, et al. User-based clustering with top-n recommendation on cold-start problem [C]// 2013 Third International Conference on Intelligent System Design and Engineering Applications. Piscataway, NJ, USA: IEEE, 2013: 1585-1589.
- [9] 刘继, 邓贵仕. 基于加权谱分析的用户网络社团协作推荐方法 [J]. 大连理工大学学报, 2010, 50(3): 438-443.
LIU J, DENG Gui-shi. A collaborative recommendation method based on user network community with weighted spectral analysis [J]. Journal of Dalian University of Technology, 2010, 50(3): 438-443.
- [10] YU L. Using ontology to enhance collaborative recommendation based on community [C]// Proceedings of the 9th International Conference on Web-Age Information Management. Piscataway, NJ, USA: IEEE, 2008: 45-49.
- [11] TANG J, GAO H, LIU H. Mtrust: discerning multifaceted trust in a connected world [C]// Proceedings of the Fifth ACM International Conference on Web Search and Data Mining. New York, USA: ACM, 2012: 93-102.
- [12] GIRVAN M, NEWMAN M E J. Community structure in social and biological networks [J]. Proceedings of the National Academy of Sciences of the United States of America, 2002, 99(12): 7821-7826.
- [13] NEWMAN M E. Fast algorithm for detecting community structure in networks [J]. Physical Review: E Statistical Nonlinear & Soft Matter Physics, 2004, 69(6): 066133.
- [14] 常振超, 陈鸿昶, 刘阳, 等. 基于联合矩阵分解的节点多属性网络社团检测 [J]. 物理学报, 2015, 64(21): 218901.
CHANG Zhenchao, CHEN Hongchang, LIU Yang, et al. Community detection based on joint matrix factorization in networks with node attributes [J]. Acta Physica Sinica, 2015, 64(21): 218901.
- [15] MASTHOFF J. Group recommender systems: combining individual models [M]// RICCI F, ROKACH L, SHAPIRA B, et al. Recommender Systems Handbook. Berlin, Germany: Springer-Verlag, 2011: 677-702.

(编辑 刘杨)