

DOI: 10.11992/tis.201802011

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20180416.1325.008.html>

融合协同过滤与用户偏好的旅游组推荐方法

陈君同, 古天龙, 常亮, 宾辰忠, 梁聪

(桂林电子科技大学 广西可信软件重点实验室, 广西 桂林 541004)

摘 要:近年来, 组推荐系统已经逐渐成为旅游推荐领域的研究热点之一。传统的推荐系统面临的数据稀疏性问题和在组推荐系统中同样存在。基于评分的推荐系统中, 可以把组推荐系统分为对单个用户的偏好预测和对组内成员预测结果的融合两个阶段。为提高推荐的效果, 提出了一种融合协同过滤与用户偏好的旅游组推荐方法, 它考虑了用户的预测评分和组推荐结果的准确性。在协同过滤中通过加入相似性影响因子和关联性因子进行预测评分, 然后在均值策略和最小痛苦策略的基础上, 提出了满意度平衡策略, 该策略考虑了组内成员的局部满意度和整体满意度。实验表明, 所提出的方法提高了推荐的准确率。

关键词:组推荐; 旅游推荐; 数据稀疏性; 协同过滤; 用户偏好; 均值策略; 最小痛苦策略

中图分类号: TP391 **文献标志码:** A **文章编号:** 1673-4785(2018)06-0999-07

中文引用格式: 陈君同, 古天龙, 常亮, 等. 融合协同过滤与用户偏好的旅游组推荐方法[J]. 智能系统学报, 2018, 13(6): 999-1005.

英文引用格式: CHEN Juntong, GU Tianlong, CHANG Liang, et al. A tourist group recommendation method combining collaborative filtering and user preferences[J]. CAAI transactions on intelligent systems, 2018, 13(6): 999-1005.

A tourist group recommendation method combining collaborative filtering and user preferences

CHEN Juntong, GU Tianlong, CHANG Liang, BIN Chenzhong, LIANG Cong

(Guangxi Key Laboratory of Trusted Software, Guilin University of Electronic Technology, Guilin 541004, China)

Abstract: In recent years, the group recommendation system has gained much attention in the field of tourism recommendation. The problem of data sparsity faced by the traditional recommendation system also exists in the group recommendation system. In the scoring-based recommendation system, the group recommendation system can be divided into two stages: preference prediction for individual users and aggregation of the forecast results of group members. To improve the effect of recommendation, a tourist group recommendation approach is proposed that incorporates collaborative filtering and users' preferences. It considers the accuracy of user's predictive scores and the group recommendation result. In the collaborative filtering, the predictive score is calculated by adding the similarity impact factor and the relevancy factor. Based on the average strategy and the least misery strategy, a satisfaction balance strategy is proposed, which considers both of the partial satisfaction and whole satisfaction of the group members. A series of conducted experiments show that the proposed method yields more accurate recommendations.

Keywords: group recommendation; tourism recommendation; data sparsity; collaborative filtering; user's preference; average strategy; least misery strategy

随着信息技术和互联网的发展, 网络正成为人们规划旅游的重要信息来源^[1]。人们逐渐从信

息匮乏的时代走入了大数据的时代, 在海量数据的背景下, 如何快速找到对用户最有价值的信息, 显得越来越重要, 推荐系统便应运而生^[2]。

以往的推荐系统主要关注于单个用户, 在电视节目^[3]、音乐、电影、新闻等方面取得了很好的效果, 但是对于旅游领域还没有给出完善的推荐

收稿日期: 2018-02-07. 网络出版日期: 2018-04-16.

基金项目: 国家自然科学基金项目(61572146, U1501252, U1711263); 广西创新驱动重大专项项目(AA17202024); 广西自然科学基金项目(2016GXNSFDA380006).

通信作者: 宾辰忠. E-mail: binchenzhong@guet.edu.cn.

方案^[4]。首先,和电影的数据相比,旅游评分的数据难以获得;其次,用户的评分矩阵比较稀疏;最后,旅游通常是多个用户以群组的形式参与的,因此,结合所有成员偏好的组推荐系统将成为旅游推荐领域的一个研究热点^[5]。基于协同过滤的组推荐系统中,在对单个用户的预测阶段,使用传统的协同过滤算法计算用户(项目)之间的相似度时,没有考虑用户共同评分的项目数和评分关联程度对相似性造成的影响,例如,两个兴趣差别较大的游客,可能同时感兴趣的景点比较少,当用户的共同评分比较少时,传统的协同过滤算法便无法准确地度量用户之间相似性^[6];在群组成员预测结果的融合阶段,效果比较好的偏好融合策略中有均值策略和最小痛苦策略^[7],均值策略把成员对项目的平均评分作为群组的得分,然而没有考虑少数成员的不满意度,最小痛苦策略选择成员对项目最小的评分作为群组的得分,却忽略了多数成员的偏好。

本文在组推荐过程中,首先改进了协同过滤算法,它结合相似性影响因子和关联性因子,以解决旅游推荐中面临的数据稀疏性问题;其次在组偏好建模阶段,提出了一种新的融合策略——满意度平衡策略,它同时考虑了组内成员的局部满意度和整体满意度;最后,通过在旅游数据集上的实验分析,验证了所改进的方法能够有效地提高推荐的质量。

1 相关工作

1.1 协同过滤算法

协同过滤算法是电子商务推荐领域中一种最为成功的推荐算法^[8]。它不需要用户主动提供个人需求信息,而是根据他们已有的评分记录,获得用户的潜在偏好。这种推荐算法能否取得良好的效果,很大程度上取决于用户的评分数据。

1.2 组推荐关键技术

组推荐关键技术包括融合方法和融合策略。融合方法分为模型融合和推荐融合。模型融合根据群组成员的用户偏好模型融合成群组偏好模型,然后基于群组偏好模型生成组推荐;推荐融合先根据传统算法获得每个用户的预测评分后,再根据预测评分进行融合,也可以融合推荐项目列表得到群组的推荐列表^[5,7]。两种融合方法各有自己的不足:模型融合易受到评分稀疏性的影响,推荐融合忽略了群组成员之间的交互^[5]。

组推荐系统中常用的融合策略有均值策略、最小痛苦策略、最开心策略^[9]。文献^[10]通过一系列实验评估,指出乘法策略、均值策略、最小痛苦

策略以及痛苦避免均值策略较好。文献^[7]通过对组推荐系统的文献研究分析,发现使用最普遍的策略有均值策略、痛苦避免均值策略及最小痛苦策略,但这些策略对于不同特征的群组适用性也不尽相同。

2 基于协同过滤的组推荐

协同过滤算法作为一种基本的方法常常被应用于组推荐系统中。基于协同过滤的组推荐包括4个阶段:相似性度量、选择邻居、预测评分、确定推荐项目^[11]。

首先通过评分矩阵中用户的评分计算两两用户(项目)的相似度,然后根据K近邻的方法预测当前用户对未知项目的评分,最后通过融合策略结合所有组内成员的偏好生成组推荐列表。图1为本文中用到的基于协同过滤的组推荐整体框架。

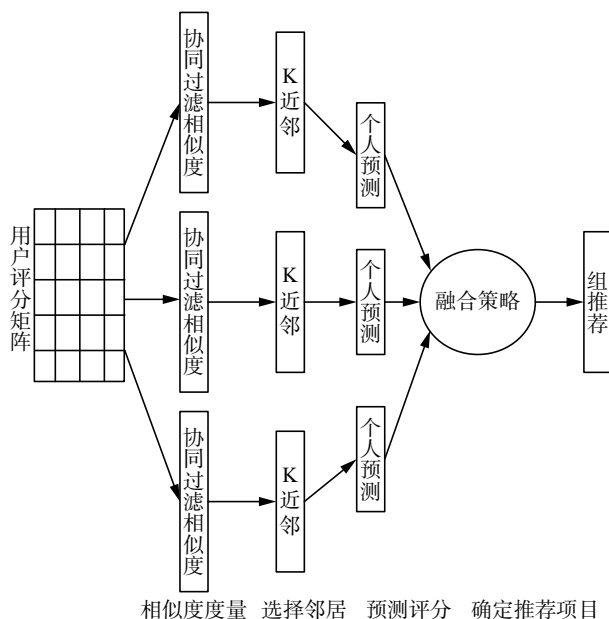


图1 基于协同过滤的组推荐整体框架

Fig. 1 The whole framework of group recommendation based on collaborative filtering

其中,协同过滤算法可以分为基于用户的最近邻推荐(user-based nearest neighbor recommendation)和基于项目的最近邻推荐(item-based nearest neighbor recommendation)。

2.1 基于用户的最近邻推荐

基于用户的最近邻推荐假设当前用户会喜欢与之有相似偏好的用户喜欢的项目。目前比较常用的相似度计算方法有余弦相似度(cosine similarity)、皮尔森相关相似度(Pearson correlation coefficient)等^[12-13]。本文主要选用相似度定义如式(1)所示。

$$\text{sim}(m, n) = \frac{\sum_{i \in I_{mn}} R_{m,i} \cdot R_{n,i}}{\sqrt{\sum_{i \in I_m} R_{m,i}^2} \sqrt{\sum_{i \in I_n} R_{n,i}^2}} \quad (1)$$

式中: $\text{sim}(m, n)$ 表示用户 m 和 n 的余弦的相似度, $R_{m,i}$ 表示用户 m 对项目 i 的评分, $R_{n,i}$ 表示用户 n 对项目 i 的评分, I_m 表示用户 m 评分的项目集合, I_n 表示用户 n 评分的项目集合, I_{mn} 表示用户 m 和用户 n 共同评分的项目集合。

找出与当前用户 m 有相似偏好的 K 个用户集合 KNN_m , 则用户 m 对项目 i 的预测评分为

$$p_{m,i} = \bar{R}_m + \frac{\sum_{n \in \text{KNN}_m} \text{sim}(m, n) \times (R_{n,i} - \bar{R}_n)}{\sum_{n \in \text{KNN}_m} |\text{sim}(m, n)|} \quad (2)$$

式中 \bar{R}_m 、 \bar{R}_n 分别表示用户 m 和用户 n 的平均评分。

2.2 基于项目的最近邻推荐

基于项目的最近邻利用用户对项目的评分来计算相似度。本文选用的相似度如式 (3) 所示。

$$\text{sim}(i, j) = \frac{\sum_{m \in R_{ij}} (R_{m,i} - \bar{R}_i)(R_{m,j} - \bar{R}_j)}{\sqrt{\sum_{m \in R_{ij}} (R_{m,i} - \bar{R}_i)^2} \sqrt{\sum_{m \in R_{ij}} (R_{m,j} - \bar{R}_j)^2}} \quad (3)$$

式中: $\text{sim}(i, j)$ 表示项目 i 和 j 的皮尔森相似度^[14-15], \bar{R}_i 、 \bar{R}_j 表示项目 i 和项目 j 的平均得分, R_{ij} 表示对项目 i 和 j 都有过评分的用户集合。

找出要预测的项目 i 的 K 个相似项目集合 KNN_i , 则用户 m 对项目 i 的预测值为

$$p_{m,i} = \bar{R}_i + \frac{\sum_{j \in \text{KNN}_i} \text{sim}(i, j) \times (R_{m,j} - \bar{R}_j)}{\sum_{j \in \text{KNN}_i} |\text{sim}(i, j)|} \quad (4)$$

2.3 数据稀疏性问题及算法改进

相似度是协同过滤中的重要指标, 决定着预测评分的好坏, 同时由于在旅游领域, 存在用户自身的综合因素使得旅游推荐不同于以往电子商务领域的推荐。例如, 用户可以每月看一场电影, 但却很少有人每年旅行很多次。因此在旅游推荐中, 不可避免地会存在数据的稀疏性问题。传统的相似度计算方法在评分数据丰富的情况下可以给出很好的效果, 但是在旅游推荐中, 传统的推荐方法在计算用户 (项目) 间的相似度时, 可能会忽略用户评分数据的稀疏程度对相似度计算结果造成的影响。如表 1 所示 (0 表示没有评分), 以基于用户的相似度为例:

1) 用户 A 和用户 B 共同评分的景点数量多于用户 C 和 D 共同评分的景点数量, 所以用户 A 和用户 B 的相似度应该比用户 C 和用户 D 的相似度

更高。然而, 用户 A 和 B 的余弦相似度计算结果为 0.838 1, 用户 C 和 D 的相似度结果为 0.910 4 (余弦相似度的范围为 [0, 1])。显然, 传统的相似度不能很好地计算他们之间的相似度。为了解决这一问题, 本文利用用户对景点共同评分的数目与用户对景点评分总数目的关系来调整用户之间的相似性, 本文称之为相似性影响因子 h_1 , 定义为

$$h_1 = 2 \sqrt{\frac{I_{mn}}{I_m \cup I_n}} - 1 \quad (5)$$

式中: I_m 为用户 m 评分的景点集合, I_n 为用户 n 评分的景点集合。用户 m 和 n 的共同评分景点数 I_{mn} 在其全部评分景点 $I_m \cup I_n$ 中占的比例越大, 则相似性影响因子 h_1 就越大, 相似度的值可能越大; 反之 h_1 越小, 相似度的值可能越小。

表 1 用户评分矩阵
Table 1 User rating matrix

用户	龙脊梯田	银子岩	七星景区	漓江
A	5	5	3	5
B	5	0	3	5
C	2	5	0	0
D	0	5	1	0
E	5	2	0	0
F	3	4	0	0

2) 用户 E 和用户 F 根据余弦相似度计算结果为 0.854 2, 表示他们的偏好很相似; 根据皮尔森和修正余弦相似度计算结果为 -1, 表示他们的偏好完全相反 (相似度的范围为 [-1, 1])。实际上他们的相似度不应该完全相反, 也不应该很相似, 而是介于两者之间。随着用户和评分矩阵规模的增加, 类似这种情况也会影响推荐结果的准确性。为解决此类问题, 本文利用用户评分的关联性来调整用户之间的相似性, 用户的共同评分向量越接近, 相似度的值可能越大, 反之相似度的值可能越小。本文称之为关联性因子 h_2 , 定义为

$$h_2 = 1 - \sqrt{1 - \left(\frac{\sum_{i \in I_{mn}} 2R_{m,i} \cdot R_{n,i}}{\sum_{i \in I_m} (R_{m,i}^2 + R_{n,i}^2)} \right)^2} \quad (6)$$

相似度的计算是协同过滤中最重要的一步。旅游领域面临的数据稀疏性问题, 使得原有的相似度方法很难准确度量用户之间的相似性。这是因为在用户评分数据稀疏的情况下, 传统的方法主要考虑用户共同评分之间的相似性, 却忽略了用户在个别项目上的评分是相似的而在其他项目上不一定相似的现象, 用户只有在比较多的项目上评分相似时, 他们的偏好才能认为是相似的; 此外, 传统的相似度无法准确区分某些相似度相

同但偏好差别很大的用户之间的相似性。

本文提出的相似性影响因子 h_1 和关联性因子 h_2 全面考虑了用户共同评分项和评分关联性对相似度量度的影响,可以有效缓解由于数据稀疏性问题导致的相似度计算不够准确的问题。改进的

$$P_{m,i} = \bar{R}_m + \frac{\sum_{n \in \text{KNN}_m} \left[h_1 \times \left(\partial \times \frac{\sum_{i \in I_m} R_{m,i} \cdot R_{n,i}}{\sqrt{\sum_{i \in I_m} R_{m,i}^2} \sqrt{\sum_{i \in I_n} R_{n,i}^2}} + (1 - \partial) \times h_2 \right) \times (R_{n,i} - \bar{R}_n) \right]}{\sum_{n \in \text{KNN}_m} \left[h_1 \times \left(\partial \times \frac{\sum_{i \in I_m} R_{m,i} \cdot R_{n,i}}{\sqrt{\sum_{i \in I_m} R_{m,i}^2} \sqrt{\sum_{i \in I_n} R_{n,i}^2}} + (1 - \partial) \times h_2 \right) \right]} \quad (8)$$

同理,基于项目相似度的情况类似。

2.4 修正的偏好融合策略

在个人用户对项目预测评分的基础上,融合策略可以将组内成员的偏好进行融合,从而得到整个组对于项目的得分,根据得分的大小生成最终的推荐列表。由于群组中成员的偏好可能相同,也可能不同,此时单个成员的偏好不能作为整个组的偏好,因此如何获取群组成员的共同偏好,缓解各成员间的偏好冲突,也是组推荐要解决的问题^[7]。目前效果比较好的偏好融合策略有均值策略(average)和最小痛苦策略(least misery)。均值策略选择成员对项目评分的平均值作为群组对于项目的得分,可以表示为

$$R_{g,i} = \text{avg}(R_{m,i} : m \in g) \quad (9)$$

式中: g 表示群组, $R_{g,i}$ 表示群组 g 对于项目 i 的得分。

最小痛苦策略选择成员对项目最小的评分作为群组对于项目的得分,可以表示为

$$R_{g,i} = \min(R_{m,i} : m \in g) \quad (10)$$

均值策略只考虑群组成员的平均偏好程度,但可能忽略了少数成员的不满意度;最小痛苦策略根据组内成员对项目的最小评分做参考,却可能会忽略多数人的感受。如表2所示,根据均值策略,景点1和景点3对于包含用户1、2、3和用户4的群组是等价的,但对于景点1而言,均值策略显然没有考虑用户4的感受,相比较来说,选择景点3可能比景点1更好一些;根据最小痛苦策略,景点3、4和景点5对于群组是等价的,相对于景点3而言,组内成员对景点4或景点5更感兴趣,此时少数人可能服从多数,组成员很可能会优先选择景点5,最小痛苦策略只考虑了成员的最小满意度却忽略了大多数人的偏好。

为此,本文考虑了以上两种策略的不足之处,定义一种修正的方法来平衡组成员局部满意度与整体满意度之间的关系,本文称之为满意度平衡策略

相似度 $\text{sim}(m,n)_h$ 如式(7)所示:

$$\text{sim}(m,n)_h = h_1 \times (\partial \times \text{sim}(m,n) + (1 - \partial) \times h_2) \quad (7)$$

式中: ∂ 表示参数, $\partial \in [0, 1]$, $\text{sim}(m,n)$ 表示传统的相似度,将式(7)代入式(2)即可得到用户对景点的预测评分,如式(8)所示。

(satisfaction balance), 群组 g 对景点 i 的得分用满意度平衡策略表示为 $\text{balance}_{g,i}$,定义如式(11)所示。

$$\text{balance}_{g,i} = \text{avg}(R_{m,i} : m \in g) + \frac{1}{S} \text{avg}(R_{m,i} : m \in g) * \min(R_{m,i} : m \in g) \quad (11)$$

式中: S 表示群组中成员的数目。根据定义,在表2中可以看出,按照满意度平衡策略计算的结果更能反映群组对景点的总体兴趣,根据修正的融合策略得到的组推荐列表为景点5、4、3、2、1。

表2 不同融合策略示例

Table 2 Different aggregation strategies examples

用户	景点1	景点2	景点3	景点4	景点5
用户1	5	5	4	5	4
用户2	5	3	4	5	5
用户3	5	5	4	4	5
用户4	1	5	4	4	5
均值策略	4	4.50	4	4.50	4.75
最小痛苦策略	1	3	4	4	4
满意度平衡策略	5	7.88	8	9	9.50

3 实验方案设计

3.1 数据获取

目前旅游推荐领域还没有公开实验数据集,国内外学术研究使用的数据主要来自旅游网站或问卷的方式。采用调查问卷的方式获取数据可能会存在一些不可避免的缺点:如果设计的问题太多往往会让用户失去耐心,太少反而不会获取足够的信息;可能会涉及到用户隐私问题;用户还可能会存在回忆误差。本文数据集由携程网(<http://www.ctrip.com/>)爬取桂林市的景点信息整理获得,包括用户ID,景点ID及用户对景点的评分。最初获取的数据集包括18 354个用户对255个景

点的 48 473 条评价,评分范围为 1~5 分。

3.2 数据预处理

为了方便实验,对获取的数据进行以下处理:

删除用户重复的评分记录;删除没有用户评分的景点;删除评分记录小于 3 条的用户。最终得到的旅游数据集如表 3 所示。

表 3 实验数据

Table 3 Experimental dataset

用户总数	景点总数	评分总数	稀疏等级
3 667	230	23 628	0.972

3.3 电影数据集

实验中除了旅游数据集之外,还采用了目前在衡量推荐算法中常用的由美国 Minnesota 大学 Grouplens 项目提供的 Movielens 数据集。如表 4 所示,该数据集由 943 个用户对 1 682 部电影的 100 000 条评价组成,每个用户的评分记录不少于 20 条,评分范围为 1~5 分。

表 4 实验数据

Table 4 Experimental dataset

用户总数	电影总数	评分总数	稀疏等级
943	1 682	100 000	0.937

3.4 评价标准

平均绝对误差^[16](mean absolute error, MAE)和均方根误差^[8](root-mean-square error, RMSE)是衡量预测评分和真实评分之间相近程度的评价标准,本文使用 MAE 和 RMSE 对个人预测结果的准确性进行检验。如式(12)、(13)所示。

$$MAE = \frac{1}{N} \sum_{c=1}^N |P_c - R_c| \quad (12)$$

$$RMSE = \sqrt{\frac{\sum_{c=1}^N (P_c - R_c)^2}{N}} \quad (13)$$

式中: P_c 和 R_c 分别表示用户对测试集中项目的预测评分和真实评分, N 为评分数目。

归一化折损累计增益 (normalized discounted cumulative gain, nDCG) 是信息检索中一种衡量推荐列表准确率的评价指标^[5,17],本文使用 nDCG 对组推荐的结果进行检验。设 q_1, q_2, \dots, q_k 为群组推荐项目的排名列表,则用户 u 的 DCG 和 nDCG 定义如式(14)、(15)所示:

$$DCG_k^u = r_{uq_1} + \sum_{c=2}^k \frac{r_{uq_c}}{\log_2(c)} \quad (14)$$

$$nDCG_k^u = \frac{DCG_k^u}{IDCG_k^u} \quad (15)$$

式中: r_{uq_c} 表示用户 u 对群组推荐结果中排在第 c 个位置项目的预测(或真实)评分, $IDCG_k^u$ 为用户 u 对

于群组的 k 个推荐项目最大的 DCG 值,计算出群组中各个用户的 nDCG 值,然后取它们的平均值作为组推荐列表的 nDCG。nDCG 是介于 0~1 之间的数, nDCG 的数值越大,表明推荐列表中项目的排序越准确,推荐准确度越高。

4 实验结果

4.1 传统的协同过滤和改进的方法比较

图 2 和图 3 分别给出了 2.1 节中介绍的基于用户的协同过滤 (UBCF) 和 2.2 节中介绍的基于项目的协同过滤 (IBCF)、通过引入相似性影响因子与关联性因子改进的基于用户的协同过滤 (imp-UBCF) 和改进的基于项目的协同过滤 (imp-IBCF) 的折线图。随机选择旅游数据集中 80% 的评分记录作为训练集, 20% 的评分记录作为测试集, 经多次实验, 取 $\theta=0$, 观察邻居数目 K 从 5~30 每次增加 5 时, 各个推荐算法的性能。图 2、3 中可以看出, 在相同邻居数目的情况下, 无论是使用基于用户的协同过滤还是基于项目的协同过滤预测评分, 基于改进方法的 MAE 和 RMSE 均低于传统的方法, 表明改进的方法在计算用户或项目之间的相似性时可以起到更好的效果, 进而提高用户对项目预测评分的准确性。这是因为传统的相似度方法 (比如基于用户的相似度) 在评分矩阵稀疏的情况下, 无法准确度量用户共同评分的景点数量与用户之间相似度的关系, 而且在一定程度上也忽略了用户对于不同景点的兴趣差异。

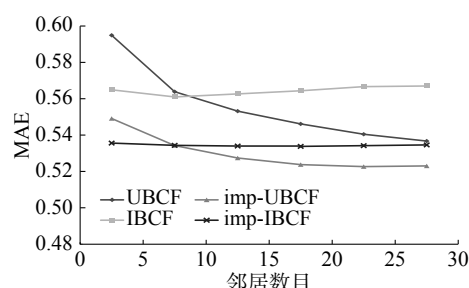


图 2 不同推荐方法下的 MAE

Fig. 2 MAE for different recommended methods

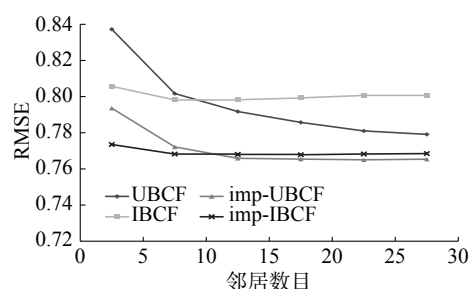


图 3 不同推荐方法下的 RMSE

Fig. 3 RMSE for different recommended methods

4.2 改进的协同过滤方法检验

为了验证改进的协同过滤方法在群组中的有效性,随机选择旅游数据集中的3人、4人、6人组成3类群组,重复进行10次实验,取邻居数目 $K=30$ 。本节使用基于用户的协同过滤和改进的基于用户的协同过滤方法对每一个群组中所有用户的预测评分进行检验,取10次实验的平均值为最后结果,如图4所示,群组中整体的MAE低于传统的方法,表明改进的方法在给群组中所有成员对项目进行预测评分时,计算结果更加准确。

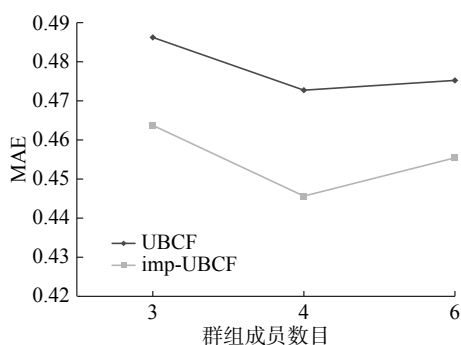


图4 不同推荐方法下的MAE

Fig. 4 MAE for different recommended methods

4.3 满意度平衡策略检验

4.3.1 偏好融合策略

在组偏好融合阶段,常见的融合策略有均值策略(average)、最小痛苦策略(least misery)、乘法策略(multiplicative)^[10]、最开心策略(most pleasure)^[5]、痛苦避免均值策略(average without misery)^[1]。其中,乘法策略通过将成员对项目的评分做乘法得到的结果作为整个组的得分,然后根据得分按照大小排序生成推荐列表;最开心策略选择成员对项目的最高评分作为群组的得分;痛苦避免均值策略在排除单个成员评分低于某个阈值的项目后再根据成员对项目的平均评分作为群组的得分。

4.3.2 偏好融合策略对比与分析

在使用改进的协同过滤方法对群组成员预测评分的基础上,本节实验在旅游数据集上比较了本文提出的满意度平衡策略(satisfaction balance)与常用的融合策略的实验结果。由图5可以看出,在群组的规模不同时,修正的偏好融合策略(satisfaction balance)表现的效果比较好。这是因为least misery是用少数成员的意见决定整个组的选择,推荐的项目可能不是群组中所有用户最喜欢的;average在给群组推荐项目时,只考虑了所有成员的整体满意度,却忽略了组内不同成员的感受;multiplicative是用所有成员的评分乘积作为群组结果,当某些成员之间的偏好差别较大

时,群组的得分也可能比较高;most pleasure只考虑了成员的最大满意度;而average without misery虽然排除了成员评分较小的项目,但在其他项目的推荐上仍然存在因少数成员的相对不满意度使得群组得分较高的情况;而本文修正的融合策略考虑了用户的整体满意度和局部满意度之间的关系,使得推荐的项目能够更好地反应出整个群组的偏好,推荐结果更加准确。

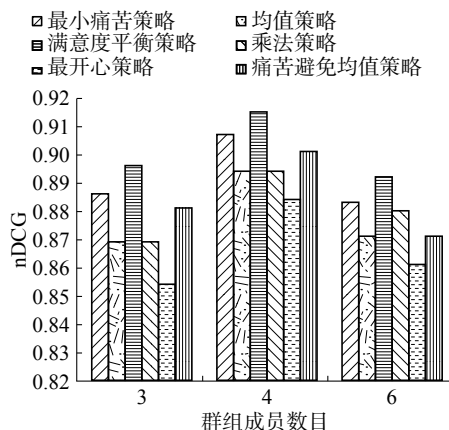


图5 不同融合策略下的nDCG

Fig. 5 nDCG for different aggregation strategies

4.4 不同数据集下的实验结果

图6为在Movielens数据集上,UBCF、IBCF、imp-UBCF和imp-IBCF的折线图。可以看出,imp-IBCF在Movielens数据集上提高的效果最为明显,和UBCF、IBCF相比,分别提高了5%和6.7%,同样验证了本文所提方法的有效性。

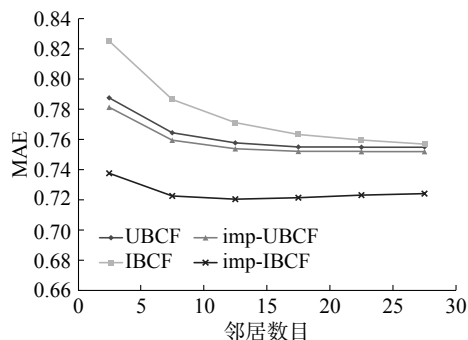


图6 不同推荐方法下的MAE

Fig. 6 MAE for different recommended methods

5 结束语

本文针对组推荐系统在旅游推荐领域面临的数据稀疏性问题,提出一种改进的协同过滤方法来提高对单个用户预测评分的准确性;然后针对现有的融合策略——均值策略和最小痛苦策略忽略了局部满意度或整体满意度的问题进行了修正,修正的偏好融合策略同时考虑了两者的不足之处。通过不同的实验证明,本文提出的方法在

一定程度上提高了预测的准确性和推荐的准确率。未来的工作是收集旅游领域的文本信息,对用户进行广度的偏好分析,进一步提高推荐的效率。

参考文献:

- [1] DE PESSEMIER T, DHONDT J, MARTENS L. Hybrid group recommendations for a travel service[J]. Multimedia tools and applications, 2017, 76(2): 2787-2811.
- [2] 雷震, 文益民, 王志强, 等. 基于影响力控制的热传导算法[J]. 智能系统学报, 2016, 11(3): 328-335.
LEI Zhen, WEN Yimin, WANG Zhiqiang, et al. Heat conduction controlled by the influence of users and items[J]. CAAI transactions on intelligent systems, 2016, 11(3): 328-335.
- [3] YU Zhiwen, ZHOU Xingshe, HAO Yanbin, et al. TV program recommendation for multiple viewers based on user profile merging[J]. User modeling and user-adapted interaction, 2006, 16(1): 63-82.
- [4] 常亮, 曹玉婷, 孙文平, 等. 旅游推荐系统研究综述[J]. 计算机科学, 2017, 44(10): 1-6.
CHANG Liang, CAO Yuting, SUN Wenping, et al. Review on tourism recommendation system[J]. Computer science, 2017, 44(10): 1-6.
- [5] 胡川, 孟祥武, 张玉洁, 等. 一种改进的偏好融合组推荐方法[J]. 软件学报, doi: 10.13328/j.cnki.jos.005288.
HU Chuan, MENG Xiangwu, ZHANG Yujie, et al. An enhanced group recommendation method based on preference aggregation[J]. Journal of software, doi: 10.13328/j.cnki.jos.005288.
- [6] 吴毅涛, 张兴明, 王兴茂, 等. 基于用户模糊相似度的协同过滤算法[J]. 通信学报, 2016, 37(1): 198-206.
WU Yitao, ZHANG Xingming, WANG Xingmao, et al. User fuzzy similarity-based collaborative filtering recommendation algorithm[J]. Journal on communications, 2016, 37(1): 198-206.
- [7] 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 Yujie, DU Yulu, MENG Xiangwu. Research on group recommender systems and their applications[J]. Chinese journal of computers, 2016, 39(4): 745-764.
- [8] 李慧, 马小平, 胡云, 等. 融合上下文信息的社会网络推荐系统[J]. 智能系统学报, 2015, 10(2): 293-300.
LI Hui, MA Xiaoping, HU Yun, et al. Social network recommendation system mixing context information[J]. CAAI transactions on intelligent systems, 2015, 10(2): 293-300.
- [9] MASTHOFF J. Group modeling: selecting a sequence of television items to suit a group of viewers[J]. User modeling and user-adapted interaction, 2004, 14(1): 37-85.
- [10] MASTHOFF J. Group recommender systems: combining individual models[M]//RICCI F, ROKACH L, SHAPIRA B, et al. Recommender Systems Handbook. Boston, USA: Springer, 2011: 677-702.
- [11] BOBADILLA J, ORTEGA F, HERNANDO A, et al. Recommender systems survey[J]. Knowledge-based systems, 2013, 46: 109-132.
- [12] LI Wenqiang, XU Hongji, JI Mingyang, et al. A hierarchy weighting similarity measure to improve user-based collaborative filtering algorithm[C]//Proceedings of the 20162nd IEEE International Conference on Computer and Communications. Chengdu, China, 2016: 843-846.
- [13] CHEN Xuegang, XIA Mingna, CHENG Jieren, et al. Trend prediction of internet public opinion based on collaborative filtering[C]//Proceedings of the 201612th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery. Changsha, China, 2016: 583-588.
- [14] 邓爱林, 朱扬勇, 施伯乐. 基于项目评分预测的协同过滤推荐算法[J]. 软件学报, 2003, 14(9): 1621-1628.
DENG Ailin, ZHU Yangyong, SHI Baile. A collaborative filtering recommendation algorithm based on item rating prediction[J]. Journal of software, 2003, 14(9): 1621-1628.
- [15] LI Dongsheng, CHEN Chao, LV Qin, et al. An algorithm for efficient privacy-preserving item-based collaborative filtering[J]. Future generation computer systems, 2016, 55: 311-320.
- [16] CHRISTENSEN I, SCHIAFFINO S, ARMENTANO M. Social group recommendation in the tourism domain[J]. Journal of intelligent information systems, 2016, 47(2): 209-231.
- [17] BALTRUNAS L, MAKCINSKAS T, RICCI F. Group recommendations with rank aggregation and collaborative filtering[C]//Proceedings of the 4th ACM Conference on Recommender Systems. Barcelona, Spain, 2010: 119-126.

作者简介:



陈君同,男,1989年生,硕士研究生,主要研究方向为智能推荐系统。



古天龙,男,1964年生,教授,博士生导师,博士,主要研究方向为形式化方法、知识工程与符号推理、协议工程与移动计算、可信泛在网络、嵌入式系统。主持国家863计划项目、国家自然科学基金、国防预研重点项目、国防预研基金等30余项。发表学术论文130余篇,其中被SCI/EI检索60余篇,出版学术著作3部。



常亮,男,1980年生,教授,博士,中国计算机学会高级会员。主要研究方向为数据与知识工程、形式化方法、智能系统。主持并完成多项科研项目,其中国家自然科学基金项目1项、广西自然科学基金项目1项,发表学术论文70余篇,其中被SCI/EI检索60余篇。