

基于用户兴趣和项目分类的协同过滤推荐算法

丛洪杰¹, 龚安¹, 李华昱¹, 帅训波²

(1. 中国石油大学(华东) 计算机与通信工程学院, 山东 青岛 266580;
2. 中国石油勘探开发研究院 计算机应用技术研究, 北京 100083)

摘要:随着用户产生的评分数据的稀疏性越来越高,传统的协同过滤算法在计算用户相似性和项目相似性的过程中暴露出弊端,导致推荐质量急剧下降。针对用户评分矩阵数据稀疏性高、推荐精度低等问题,首先采用用户兴趣分布对评分矩阵预测填充,以降低数据的稀疏性,然后在寻找最近邻居的过程中,提出基于项目分类的修正余弦相似性的度量方法,利用用户评分在项目类别内的偏离程度来改进修正余弦相似性,以寻找更加准确的 K 近邻。在MovieLens 1m数据集上进行了实验,结果表明,在用户评分数据稀疏性较高的情况下,该方法可以有效地改善传统的余弦相似性,修正余弦相似性度量方法所存在的问题,显著提高了推荐质量。

关键词:协同过滤;修正余弦相似性;用户兴趣分布;项目分类

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2018)11-0085-04

doi:10.3969/j.issn.1673-629X.2018.11.019

A Collaborative Filtering Recommendation Algorithm Based on User Interest and Item Classification

CONG Hong-jie¹, GONG An¹, LI Hua-yu¹, SHUAI Xun-bo²

(1. School of Computer & Communication Engineering, China University of Petroleum (East China),
Qingdao 266580, China;
2. Institute of Computer Application Technology, China Petroleum Exploration and
Development Research Institute, Beijing 100083, China)

Abstract: As the sparsity of the score data generated by users becomes higher and higher, the traditional collaborative filtering algorithm exposes defects in the calculation of user similarity and project similarity, leading to a sharp decline in recommendation quality. Aiming at the problems of high sparsity and low recommendation accuracy of user score matrix data, we firstly use the user interest distribution to predict and fill the score matrix, so as to reduce the sparsity of data. Secondly, we propose a measure method of modified cosine similarity based on item attribute classification in the process of finding nearest neighbors, and use the degree of deviation of user rating in the item attribute category for a more accurate K -nearest neighbor. The experiment is carried out in the MovieLens 1m datasets, which shows that in the case of high sparse of user score data, the proposed method can effectively improve the traditional cosine similarity, correct the problems existing in the cosine similarity measurement method, and remarkably improve the recommended quality.

Key words: collaborative filtering; modified cosine similarity; user interest distribution; item classification

0 引言

互联网和移动互联网的蓬勃发展,带来了信息快速增长,在海量数据中获取对自己有价值的信息更加困难,个性化推荐算法成为解决这一问题最有效的技术之一^[1]。在个性化推荐算法中,协同过滤推荐在工业界和学术界都取得了很大的成功^[2]。

随着大数据时代的来临,协同过滤的弊端也越发突出,包括数据稀疏性、冷启动、可扩展性差、小规模数据离线处理算法无法应用到大规模数据处理等^[3]。针对这些问题,学者们提出了许多解决办法。邓爱林等^[4]为减小数据稀疏性带来的推荐精度低的问题,提出了一种基于项目评分预测填充空缺值的方法,然后

收稿日期:2017-12-06

修回日期:2018-04-12

网络出版时间:2018-06-29

基金项目:国家油气重大专项(2017ZX05013-001)

作者简介:丛洪杰(1989-),男,硕士研究生,研究方向为数据挖掘、个性化推荐;龚安,副教授,通讯作者,CCF会员(62929M),研究方向为大数据智能处理;李华昱,博士,副教授,CCF会员(19453M),研究方向为语义Web与数据集成。

网络出版地址: <http://cnki.net/kcms/detail/61.1450.TP.20180629.1703.022.html>

采用一种新颖的相似性度量获得目标用户更加准确的最近邻居,使得推荐质量有所提升。Melville P 等^[5]在缓解评分数据的稀疏性方面,则采用基于内容的方法预填充用户评分矩阵中的未评分项,从而提高推荐精度。

韦素云等^[6]考虑了项目类别和兴趣度等因素,提出一种改进的协同过滤推荐算法。该算法首先计算项目之间的类别距,然后结合项目类别信息,考虑不同项目之间的相关程度,构造出项目类别相似性矩阵,衡量项目间相似性的标准采用改进的条件概率的方法。孟祥武等^[7]提出大数据时代的推荐系统,充分利用丰富的用户反馈、社会化网络等信息进一步提高推荐系统的性能和用户满意度。国琳等^[8]提出通过构建和分析用户兴趣分布曲线以发现兴趣领域专家,并甄别状态不正常的伪专家。

王立才等^[9]提出上下文感知的推荐系统,利用上下文信息,改善推荐系统的推荐精确度和用户满意度,从面向过程的角度论述了上下文感知推荐系统的研究进展和难点。刘平峰等^[10]提出用户兴趣图谱,建立兴趣领域本体,集成兴趣图谱的动态性,实现用户兴趣匹配与定位,进而提高推荐系统的精度。Pessemier T D 等^[11]提出个性化的混合推荐模型,该模型更多考虑了用户的偏好、约束限制和用户反馈等因素,使得推荐个人旅行线路更加符合用户兴趣。Oh J 等^[12]提出一种个性化流行趋势匹配算法,计算目标用户的个性化趋势,匹配用户兴趣偏好,提升预测评分,产生更加精准的推荐列表。

以上所述研究虽然对推荐算法做出了改进,并且取得了不错的效果,但对于项目类别和用户兴趣没有加以充分挖掘利用。对此,文中算法引入用户兴趣分布预测填充评分矩阵空缺值,缓解数据稀疏性,改进相似性度量方法计算项目相似性,以减小预测评分的误差。

1 项目分类

1.1 评分矩阵

基于项目的协同过滤算法中,计算项目的相似性是关键步骤。项目类别数据和用户评分数据为文中推荐算法的基础数据,由此可构建项目类别分布矩阵、用户兴趣分布矩阵和用户评分矩阵。

(1)项目类别分布矩阵。

项目类别分布矩阵由 $n \times k$ 的矩阵 $C(n,k)$ 表示,如表 1 所示,其中 n 为项目的个数, k 为项目类别的个数。

(2)用户兴趣分布矩阵。

用户兴趣分布矩阵由 $m \times k$ 矩阵 $D(m,k)$ 表示,见

表 2,其中 m 为用户的数目, k 为用户兴趣种类数。

表 1 项目类别分布矩阵 $C(n,k)$

	Class ₁	...	Class _j	...	Class _k
Items ₁	C _{1,1}	...	C _{1,j}	...	C _{1,k}
⋮	⋮	⋮	⋮	⋮	⋮
Items _i	C _{i,1}	...	C _{i,j}	...	C _{i,k}
⋮	⋮	⋮	⋮	⋮	⋮
Items _n	C _{n,1}	...	C _{n,j}	...	C _{n,k}

表 2 用户兴趣分布矩阵 $D(m,k)$

	Class ₁	...	Class _j	...	Class _k
Users ₁	D _{1,1}	...	D _{1,j}	...	D _{1,k}
⋮	⋮	⋮	⋮	⋮	⋮
Users _i	D _{i,1}	...	D _{i,j}	...	D _{i,k}
⋮	⋮	⋮	⋮	⋮	⋮
Users _m	D _{m,1}	...	D _{m,j}	...	D _{m,k}

(3)用户评分矩阵。

用户评分矩阵由 $m \times n$ 的矩阵 $R(m,n)$ 表示,见表 3,其中用户数目用 m 表示,项目数目用 n 表示, R_{ij} 表示用户 i 对项目 j 的评分。

表 3 用户评分矩阵 $R(m,n)$

	Items ₁	...	Items _j	...	Items _n
Users ₁	R _{1,1}	...	R _{1,j}	...	R _{1,n}
⋮	⋮	⋮	⋮	⋮	⋮
Users _i	R _{i,1}	...	R _{i,j}	...	R _{i,n}
⋮	⋮	⋮	⋮	⋮	⋮
Users _m	R _{m,1}	...	R _{m,j}	...	R _{m,n}

1.2 相似性度量方法

传统的基于邻域的协同过滤算法,度量用户或项目间的相似性是关键的一步。以下是计算项目之间相似性主要采用的三种方法:

(1)余弦相似性(cosine similarity):把项目看作 m 维用户的空间向量。设向量 i 和向量 j 分别代表用户对项目 i 和项目 j 的评分,则项目 i 和项目 j 之间的相似性 $\text{sim}(i,j)$ 定义如下:

$$\text{sim}(i,j)=\frac{\sum_{u \in U} R_{u,i} R_{u,j}}{\sqrt{\sum_{u \in U} R_{u,i}^2} \sqrt{\sum_{u \in U} R_{u,j}^2}}$$

(1)

其中, $R_{u,i}$ 、 $R_{u,j}$ 分别为用户 u 对项目 i 和项目 j 的评分; U 表示用户对项目 i 和项目 j 都有评分的用户集合。

(2)皮尔森系数(Pearson correlation)。

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

其中, U 表示对项目 i 、 j 有共同评分的用户集合;

\bar{R}_i 和 \bar{R}_j 分别为对项目 i 、 j 都有评分的用户对项目 i 和项目 j 评分的均值。

(3) 修正余弦相似性 (adjusted cosine similarity)。

$$\text{sim}(i, j) = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_u)(R_{u,j} - \bar{R}_u)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_u)^2} \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}_u)^2}} \quad (3)$$

其中, \bar{R}_u 为用户 u 已评分项目的平均值

2 基于项目分类的修正余弦相似性

用户规模和项目数目的增加,使得评分矩阵的数据稀疏性逐渐增加,从而导致预测评分精度低。目前解决办法就是对用户未评分项进行预填充,固定缺省值是最简单且常用的方法之一,该方法可以有效地提高系统的推荐精度。但实际生活中评分矩阵中的缺省值显然不可能都一样,因此填充固定缺省值的方法没有从根本上解决数据稀疏性问题。文中以用户兴趣分布评分矩阵求相似用户,对空缺值进行预测填充,减小用户评分数据的稀疏性。

传统的修正余弦相似性的度量方法是对用户的评分去中心化,通过减去用户的平均评分,改善了用户评分尺度不同的问题。评分尺度问题主要由用户评分的标准不同而导致,比如评分区间为 1~5 分时,用户对于项目评分 A 可能是 3 分以上喜欢,而 B 则 4 分以上才为喜欢。但是上述方法仅仅考虑了用户评分尺度问题,对于项目的类别用户的评分标准也是有所不同的,比如用户对动作类或科幻类的电影更加偏好,那么用户会普遍给这一类别的电影更高的评分。文中提出对项目进行分类,计算用户在项目的每种类别中的评分标准,以此来改进计算项目相似性的过程,以寻找更加准确的 K 近邻。

2.1 用户兴趣分布填充空缺值

基于项目的协同过滤,主要由计算项目的相似邻居,预测目标用户对项目的评分和生成推荐列表等过程组成,其中计算项目的相似邻居需要用户对项目 i 、 j 同时具有评分。求用户对项目 i 和对项目 j 分别有评分的并集公式表示为:

$$U = U_i \cup U_j \quad (4)$$

其中, U 表示所有用户的集合; U_i 和 U_j 分别表示对项目 i 和项目 j 有过评分的用户集合。

对用户集合中分别对项目 i 、 j 未评分的项进行填充预测评分,定义 $P\{p_{u,i}, p_{u,j}, \dots\}$ 。

预测填充策略,本阶段采用用户兴趣分布矩阵作为求最近邻居的输入数据,通过基于用户的协同过滤算法进行预测评分。

$$P_{u,i} = \frac{\sum_{v \in N(i)} \text{sim}_D(u, v) r_{v,i}}{\sum_{v \in N(i)} \text{sim}_D(u, v)} \quad (5)$$

其中, $P_{u,i}$ 为预测评分; $\text{sim}_D(u, v)$ 为用户 u 和用户 v 通过用户兴趣分布矩阵求得的相似度; $r_{v,i}$ 为相似用户对项目 i 的评分。

2.2 项目分类修正余弦相似性

$$\text{sim}(i, j) = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_{u_{c(i)}})(R_{u,j} - \bar{R}_{u_{c(j)}})}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_{u_{c(i)}})^2} \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}_{u_{c(j)}})^2}} \quad (6)$$

其中, $\bar{R}_{u_{c(i)}}$ 为项目 i 所属类别中用户 u 在该类别的平均评分, $\bar{R}_{u_{c(j)}}$ 同理。

根据项目类别,可以把项目分成不同的兴趣类别,比如 Action、Adventure、Fantasy、Horror 等;用户的兴趣取向也被分为上述类别,这样用户对于项目的评分转化为用户对于兴趣取向的评分,可以通过项目评分矩阵 $\mathbf{R}(m, n)$ 和项目类别列表,计算得到用户兴趣分布矩阵 $\mathbf{D}(m, k)$ 。对于项目的属性类别,可以通过数据集中项目类别标签来生成项目类别矩阵 $\mathbf{C}(n, k)$ 。由此可以获取 $\bar{R}_{u_{c(i)}}$:

$$\bar{R}_{u_{c(i)}} = \frac{\mathbf{D}(m, k) \cdot \mathbf{C}(n, k)^T}{|\mathbf{C}(k)|} \quad (7)$$

其中, $\bar{R}_{u_{c(i)}}$ 为用户 u 对项目 k 所属兴趣类别的平均评分。一个项目所属类别可以有多个,例如,电影 Toy Story 类别属性为 Animation、Children's、Comedy,则矩阵 $\mathbf{C}(n, k)$ 相应的属性值上标记为 1。 $|\mathbf{C}(k)|$ 为项目 k 所属兴趣类别向量的模(也就是所属类别的累加值)。

2.3 算法描述

输入:用户的项目评分矩阵 \mathbf{R} , 项目类别信息,项目集合 Items, 目标用户集合 U , 项目邻居数目 K , 推荐列表长度 N

输出:目标用户 u 对项目的预测评分,且生成目标用户 u 的推荐列表

(1) 根据数据集中项目分类信息构建项目类别分布矩阵 $\mathbf{C}(n, k)$;

(2) 根据项目分类信息和用户项目评分矩阵 \mathbf{R} , 计算用户兴趣分布矩阵 $\mathbf{D}(m, k)$;

(3) 由上述两步的结果计算出用户对项目所属兴

趣类别的平均评分 $\overline{R_{u,(k)}}$;

(4) 求用户对项目 i 和项目 j 分别有评分的用户的并集;

(5) 对并集中用户对项目 i 或项目 j 未评分项,采用用户兴趣分布的协同过滤算法进行预填充值;

(6) 构建用户项目评分矩阵,求得项目 i 和项目 j 有着共同评分的用户集 U ;

(7) 利用基于项目分类的修正余弦相似性来计算项目之间的相似度,生成相似度最大的 k 个项目邻居;

(8) 通过求得的相似邻居和相似度值来预测目标用户对项目的评分;

(9) 通过对预测评分的排序,推荐 top N 项目给目标用户。

3 实验结果与分析

3.1 实验数据集

实验数据采用明尼苏达大学 GroupLens Research 项目小组的 MovieLens 数据集,规模大小为 MovieLens 1 m。用户评分数据包含约 100 万条,其中用户数量为 6 040,电影数量为 3 952。评分标准为 1~5 的整数,数字越大,用户越喜欢,反之,则对该电影不感兴趣。数据的稀疏等级为 $1 - \frac{1\ 000\ 209}{3\ 952 \times 6\ 040} = 0.958\ 098$ 。把数据随机生成 8 份,选择一份作为测试集,其余为训练集,进行交叉实验。在该数据集中,电影分为 18 种类别,每一部电影至少属于一种或多种类别。

3.2 评价指标

实验采用平均绝对误差 (mean absolute error, MAE) 来评价算法的推荐精度^[13]。假定目标用户的预测评分集为 $P\{p_1, p_2, \dots, p_n\}$, 与其对应测试集中的真实评分集为 $Q\{q_1, q_2, \dots, q_n\}$, 则 MAE 表示为:

$$MAE = \sum_{i=1}^N |p_i - q_i| / N \quad (8)$$

3.3 实验方案

首先对数据进行预处理,构建用户的兴趣分布数据以及项目分类数据,然后使用 Java 编程语言对推荐算法进行实现,分别对传统的余弦相似性、修正余弦、改进的修正余弦进行实现,对目标用户和电影进行预测评分,评估推荐精度;最后对比文中算法与其他两种算法的推荐精度。

3.4 实验结果分析

为了验证算法的有效性,在使用相同数据集规模及验证方法上,对传统协同过滤算法中使用的余弦相似性、修正余弦相似性的相似性度量方法与文中提出的改进修正余弦相似性进行实验对比。最近邻居的规模从 10 到 100 递增,实验结果如图 1 所示。

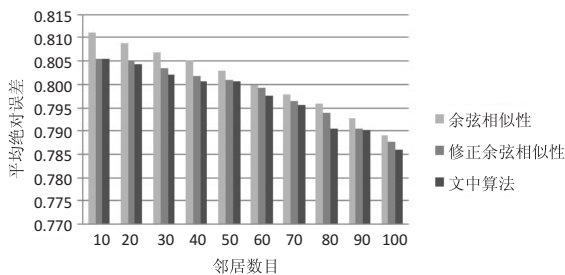


图 1 推荐算法的 MAE 值对比

从图 1 可以看出,文中提出的方法随着最近邻居的增加 MAE 值逐渐减小,且低于传统协同过滤方法。因为文中方法考虑了项目的类别以及对于用户兴趣分布的因素,在此基础上计算的用户的最近邻居相似度更加符合现实,更加准确,而通过更加准确的最近邻居用户群体,可以获得更加符合目标用户的预测评分。所以该算法可以有效提升推荐质量,为用户提供满意的个性化推荐列表。

4 结束语

针对协同过滤算法所面临的数据稀疏性、推荐精度低等^[14]问题,采用了基于用户兴趣分布方法初步填充评分矩阵中的未评分项;在计算项目相似性阶段,提出一种基于项目分类改进修正余弦相似性度量方法,即由用户平均评分去中心化的方法改为对用户兴趣类别平均分去中心化,求出最近邻居并生成推荐列表。实验结果表明,该方法具有较小的 MAE,在一定程度上提高了推荐质量。

参考文献:

- [1] 弗朗西斯科·里奇. 推荐系统: 技术、评估及高效算法 [M]. 北京: 机械工业出版社, 2015.
- [2] BREESE J S, HECKERMAN D, KADIE C. Empirical analysis of predictive algorithms for collaborative filtering [C]// Proceedings of the fourteenth conference on uncertainty in artificial intelligence. Madison, Wisconsin: Morgan Kaufmann Publishers Inc., 1998: 43-52.
- [3] HUANG Z, CHEN H, ZENG D. Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering [J]. ACM Transactions on Information Systems, 2004, 22(1): 116-142.
- [4] 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.
- [5] MELVILLE P, MOONEY R J, NAGARAJAN R. Content-boosted collaborative filtering for improved recommendations [C]// Eighteenth national conference on artificial intelligence. Edmonton, Canada: [s. n.], 2002: 187-192.
- [6] 韦素云, 业宁, 吉林根, 等. 基于项目类别和兴趣度的协同过滤推荐算法 [J]. 南京大学学报: 自然科学版, 2013, 49

3 结束语

针对危险源原因分析中存在的分析主要依赖人的参与和传统关联规则挖掘算法中候选集过多影响挖掘效率的问题,提出了一种基于混合蚁群关联规则挖掘的危险源原因分析算法 HA-MACR。该算法通过引入粒子群优化蚁群初始信息素浓度,根据频繁项集来确定,避免了蚁群算法初始时的盲目性,有效提高了危险源原因分析的效率与质量。实验结果表明,该算法充分利用了蚁群优化算法的寻优能力和正反馈性,将其应用于对危险源原因的分析过程中,通过挖掘到的最大频繁项集产生质量较好的关联规则,进而得到分析结果,不仅提高了关联规则的挖掘效率,同时也能有效提高危险源原因分析的准确性和执行效率。

参考文献:

[1] GROTH K, WANG Chengdong, MOSLEH A. Hybrid causal methodology and software platform for probabilistic risk assessment and safety monitoring of socio-technical systems [J]. Reliability Engineering & System Safety, 2010, 95(12): 1276-1285.

[2] PURTON L, CLOTHIER R, KOUROUSIS K. Assessment of technical airworthiness in military aviation; implementation and further advancement of the bow-tie model [J]. Procedia Engineering, 2014, 80: 529-544.

[3] 赵阳, 吴廖丹. 一种自底向上的最大频繁项集挖掘方法 [J]. 计算机技术与发展, 2017, 27(8): 57-60.

[4] 周发超, 王志坚, 叶枫, 等. 关联规则挖掘算法 Apriori 的研究改进 [J]. 计算机科学与探索, 2015, 9(9): 1075-1083.

[5] 郭进伟, 皮建勇. 一种基于 FP-growth 的并行 SON 算法的实现 [J]. 微型机与应用, 2014, 33(8): 60-63.

[6] DU Jiaoling, ZHANG Xiangli, ZHANG Hongmei, et al. Research and improvement of apriori algorithm [C]//Proceedings 2nd international workshop on intelligent systems and

applications. Dalian, China: IEEE, 2010.

[7] AGRAWAL J, AGRAWAL S, SINGHAI A, et al. SET-PSO-based approach for mining positive and negative association rules [J]. Knowledge and Information Systems, 2015, 45(2): 453-471.

[8] ZHOU Zhiping, ZHANG Daowen, SUN Ziwen, et al. An adaptive hybrid PSO and GSA algorithm for association rules mining [C]//International conference on cloud computing and security. [s. l.]: Springer International Publishing, 2015: 469-479.

[9] 罗茜. 遗传模拟退火算法挖掘关联规则的应用 [D]. 广州: 中山大学, 2010.

[10] 黄红星, 王秀丽, 黄习培. 挖掘最大频繁项集的改进蚁群算法 [J]. 计算机工程与应用, 2011, 47(13): 161-165.

[11] 陈昌敏, 谢维成, 范颂颂. 自适应和最大最小蚁群算法的物流车辆路径优化比较 [J]. 西华大学学报: 自然科学版, 2011, 30(3): 5-8.

[12] 许凯波, 鲁海燕, 程毕芸, 等. 求解 TSP 的改进信息素二次更新与局部优化蚁群算法 [J]. 计算机应用, 2017, 37(6): 1686-1691.

[13] LAI M, TONG X. A metaheuristic method for vehicle routing problem based on improved ant colony optimization and Tabu search [J]. Journal of Industrial & Management Optimization, 2017, 8(2): 469-484.

[14] 杨程, 范强, 王涛, 等. 基于多维特征的开源项目个性化推荐方法 [J]. 软件学报, 2017, 28(6): 1357-1372.

[15] SARIEFF N B, BUNIYAMIN N. Comparative study of genetic algorithm and ant colony optimization algorithm performances for robot path planning in global static environments of different complexities [C]//IEEE international symposium on computational intelligence in robotics and automation. Daejeon, South Korea: IEEE, 2016: 132-137.

[16] 周 晓, 葛洪伟, 苏树智. 基于信息素的自适应连续域混合蚁群算法 [J]. 计算机工程与应用, 2017, 53(6): 156-161.

(上接第 88 页)

(2): 142-149.

[7] 孟祥武, 纪威宇, 张玉洁. 大数据环境下的推荐系统 [J]. 北京邮电大学学报, 2015, 38(2): 1-15.

[8] 国琳, 左万利. 基于兴趣图谱的用户兴趣分布分析及专家发现 [J]. 电子学报, 2015, 43(8): 1561-1567.

[9] 王立才, 孟祥武, 张玉洁. 上下文感知推荐系统 [J]. 软件学报, 2012, 23(1): 1-20.

[10] 刘平峰, 朱孔真, 杨柳, 等. 基于用户兴趣图谱的个性化推荐系统设计 [J]. 武汉理工大学学报: 信息与管理工程版, 2014, 36(3): 341-344.

[11] PESSEMIER T D, DHONDT J, MARTENS L. Hybrid group recommendations for a travel service [J]. Multimedia Tools

& Applications, 2017, 76(2): 2787-2811.

[12] OH J, PARK S, YU H, et al. Novel recommendation based on personal popularity tendency [C]//Proceedings of the 2011 IEEE 11th international conference on data mining. [s. l.]: IEEE, 2011: 507-516.

[13] LAI Siwei, XIANG Liang, DIAO Rui, et al. Hybrid recommendation models for binary user preference prediction problem [J]. Journal of Machine Learning Research, 2012, 18: 137-151.

[14] ADOMAVICIUS G, KWON Y O. Improving aggregate recommendation diversity using ranking-based techniques [J]. IEEE Transactions on Knowledge & Data Engineering, 2012, 24(5): 896-911.