

# 基于标签聚类与项目主题的协同过滤推荐算法

李昊阳 符云清

(重庆大学软件学院 重庆 401331)

**摘 要** 传统基于项目的协同过滤算法在计算项目相似度时仅依靠评分数据,未考虑项目的自身特征。社会化标注的出现使得标签能在一定程度上反映项目特征,但标签具有语义模糊的特点,因此直接将标签纳入协同过滤算法存在一定问题。为解决上述问题,提出一种改进的基于项目的协同过滤推荐算法。该算法对标签进行聚类并生成主题标签簇,根据项目标注情况计算项目与主题间的相关度并生成项目-主题相关度矩阵,同时将其与项目-评分矩阵相结合来计算项目间的相似度,采用协同过滤完成对目标项目的评分预测,以实现个性化推荐。在 Movielens 数据集上的实验结果表明,该算法能够解决标签的语义模糊问题并提升推荐质量。

**关键词** 社会化标注,标签聚类,项目主题,协同过滤,个性化推荐

中图分类号 TP301 文献标识码 A DOI 10.11896/j.issn.1002-137X.2018.04.041

## Collaborative Filtering Recommendation Algorithm Based on Tag Clustering and Item Topic

LI Hao-yang FU Yun-qing

(School of Software, Chongqing University, Chongqing 401331, China)

**Abstract** The traditional item-based collaborative filtering algorithm only focuses on the rating data without the characteristics of items when calculating the similarity between items. The appearance of social tagging can reflect the characteristics of items, but there are some semantic fuzziness problems while adding the social tags into the collaborative filtering algorithm directly. To solve the problems above, this paper put forward an improved item-based collaborative filtering recommendation algorithm. It clusters social tags to generate tag clusters which represent different topics, and calculates the relevance between items and topics to generate item-topics matrix according to the tagging results of items. The similarity between items is calculated by combining item-topics matrix with item-ratings matrix, the rating of target items are predicted through the collaborative filtering algorithm, and the personalized recommendation is realized. Experimental results on MovieLens dataset show that the proposed algorithm can eliminate the semantic fuzziness and improve the quality of recommendation.

**Keywords** Social tagging, Tag clustering, Item topics, Collaborative filtering, Personalized recommendation

随着信息技术的迅猛发展,信息过载已成为亟待解决的难题,推荐系统作为解决信息过载的主要手段,得到了广泛的关注与研究。它通过分析系统中显式或隐式的信息,根据用户的喜好为其提供商品和服务推荐,有效地解决了信息过载问题<sup>[1]</sup>。推荐算法作为推荐系统的核心部分,是决定推荐效果的重要前提,目前个性化推荐的方法主要有基于内容的推荐、基于知识的推荐、协同过滤推荐以及混合推荐等,协同过滤推荐算法是其中应用得最为成功的一种。

传统的协同过滤推荐算法分为两类,分别是基于用户的协同过滤和基于项目的协同过滤,二者皆是根据用户的历史评分数据来预测偏好并产生推荐<sup>[2-3]</sup>。协同过滤推荐技术的优势在于不需要考虑被推荐资源的内容,仅依靠评分矩阵进行个性化推荐。但随着系统中用户与资源数量的不断增加,该技术面临着诸多挑战,如仅依靠评分矩阵来计算用户或物品相似度,而未考虑用户本身的兴趣;在数据量庞大的情况下

评分矩阵过于稀疏,难以保证推荐准确性等问题。

随着 Web 2.0 时代的到来,用户使用互联网的方式发生了显著变化,网站社交性的增强使得用户更愿意分享他们的偏好与兴趣,社会化标注允许用户对互联网中的资源打上任意标签,从各个方面对资源进行描述<sup>[4]</sup>。具有代表性的网站有 Delicious、CiteULike、Last.fm 以及豆瓣等。用户进行资源标注时所使用到的标签可被看作用户与资源之间的纽带,它既表达了用户自身的兴趣爱好,又能反映资源的主题<sup>[5]</sup>,因此标签可以作为推荐系统中用户信息与资源信息的重要补充而应用于个性化推荐之中。

近年来,将标签纳入个性化推荐算法的研究成为了一大热点。文献[6]通过分析用户在标注资源与评分时所侧重的不同兴趣点来提高协同过滤算法中项目相似度的计算精度,以改善推荐效果。文献[7]根据用户对标签的使用情况建立用户偏好模型,利用标签对资源的标注情况计算资源间的相

到稿日期:2017-03-12 返修日期:2017-06-15

李昊阳(1991—),男,硕士生,CCF 会员,主要研究方向为数据挖掘、个性化推荐,E-mail:525193177@qq.com;符云清(1969—),男,博士,教授,主要研究方向为计算机体系结构、计算机网络与通信、现代远程教育等,E-mail:yqfu@cqu.edu.cn(通信作者)。

似性;然后根据用户的历史行为来预测其对未评价资源的偏好并产生推荐。文献[8]采用张量分解的形式对用户、资源和标签三者之间的潜在语义关联进行分析,进一步改善了标签推荐算法的效果。还有一些研究者将标签与协同过滤推荐技术相结合,提出了多种基于标签的推荐算法<sup>[9-10]</sup>。但是上述文献本质上均是通过标签语义完全匹配的方式来设计算法,未考虑标签中的语义模糊问题。

根据上述分析,本文提出一种基于标签聚类 and 项目主题的协同过滤算法(ICFT)。该算法将系统中的标签进行聚类,把聚类后的每个标签簇视为一个主题,通过计算标注该项目的标签与不同标签簇之间的相关程度来生成项目-主题相关度矩阵,然后将其与项目-评分矩阵相结合来计算项目间的综合相似度,产生目标项目的最近邻,进而实现项目的评分预测与推荐。

## 1 相关工作

### 1.1 基于项目的协同过滤算法

传统基于项目的协同过滤算法认为,用户对目标项目的评分与用户对该项目的相似项目评分接近,因此该算法的核心是计算项目间的相似度,主要步骤为:

- 1) 建立  $m$  行  $n$  列的用户-项目评分矩阵;
- 2) 计算目标项目  $i$  与其他项目  $j$  之间的相似度;
- 3) 将其他项目按照计算出的相似度由高到低进行排列,选取其中前  $K$  个作为目标项目的最近邻集合;
- 4) 计算用户  $u$  对目标项目  $i$  的最近邻集合评分的加权和,以此作为用户  $u$  对目标项目  $i$  的预测评分。

项目间相似度的计算是基于共同评分的用户集进行的,公式如下:

$$sim(i, j) = \frac{\sum_{u \in U_{i,j}} (r_{u,i} - \bar{r}_u)(r_{u,j} - \bar{r}_u)}{\sqrt{\sum_{u \in U_{i,j}} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{u \in U_{i,j}} (r_{u,j} - \bar{r}_u)^2}} \quad (1)$$

其中,  $U_{i,j}$  表示对项目  $i$  和  $j$  均进行过评分的用户集合。

用户  $u$  对于尚未评价的项目  $i$  的评分预测通过式(2)计算。

$$pred(u, i) = \frac{\sum_{j \in I_u} sim(i, j) * r_{u,j}}{\sum_{j \in I_u} sim(i, j)} \quad (2)$$

其中,  $I_u$  代表用户  $u$  评分过的项目集合,  $sim(i, j)$  代表物品  $i$  与  $j$  之间的相似度。

由于基于项目的协同过滤算法在计算项目相似度时是根据项目的共同评分集合来进行的,因此只有在两个项目的共同评分较多的情况下得出的项目相似度才较为准确。但在现实中,由于系统数据量庞大,评分数据存在着较大的稀疏性,两个项目的共同评分数据较少,导致有些与目标项目实际有较大差异的项目因计算得出的相似度较高而被纳入项目最近邻集合,同时,单纯地依靠评分计算项目相似度而未考虑项目自身特点的算法策略,也在一定程度上降低了最近邻选取的准确性,影响了推荐效果。

### 1.2 标签共现与复杂网络

标签的出现能够为推荐系统提供额外的信息,但是由于社会化标注允许用户给资源标注任意标签,使得标签具有一定的随意性,且存在同义词、多义词等语义模糊问题,如在与计算机科学相关的资源中单词“Java”代表一门编程语言,在

地理方面的资源中“Java”代表爪哇岛,而在食品相关的资源中“Java”代表某种咖啡的名称;同时,用户随意标注还会引起标签数据量过大的问题,导致标签数据中存在较多的信息噪声。因此在利用标签数据进行个性化推荐时,需要消除信息噪声以提高推荐结果的准确性。

共词分析法是一种内容分析方法,该方法通过对一些词语两两统计它们在同一篇文档中出现的次数来构建共词矩阵,通过相似性测量来反映这些词语之间的亲疏关系。文献[11]指出若两个标签被同时标注在某些资源上,则表明二者之间有某种语义联系。因此,标签之间的亲疏关系可以通过标签共现的形式体现。

在社会化标注过程中,用户、资源与标签之间形成了复杂的联系,其中不同标签之间的共现关系构成了复杂的标签网络,并且其具有复杂网络的一些重要特征<sup>[12]</sup>。复杂网络往往具有一定的社团结构,其社团内部的节点之间连接较为紧密,各社团之间的连接比较稀疏。因此可以利用社团划分的方式对标签共现网络中的标签进行聚类。

## 2 基于连边社团检测算法的标签聚类

在标签共现网络中,将每个标签视为一个节点,把标签之间的共现关系视为节点间的边,标签的共现次数视为边的权重。这样,所有节点与边构成一个加权无向网络  $G(V, E, W)$ ,其中  $V$  为网络中的节点集合,  $E$  为边的集合,  $W$  为边的权重。

利用复杂网络中社团检测的方法将联系紧密的标签划分到相同社团中。被划分到同一社团的标签往往具有较高的语义相关度,因而能够表达相似的主题。项目被不同主题的标签标注的情况可在一定程度上反映项目的主题,因此对其进行分析以实现对项目内容的补充。

在传统的社团检测算法中,每个节点只能被划分到一个社团中,但由于标签共现网络中存在多义标签,使得某些标签可能会分属不同的标签簇,因此本文使用 Ahn 等提出的连边社团检测算法<sup>[13]</sup>对标签共现网络进行社团划分。连边社团检测算法认为一个社团是一组紧密连接的边的集合,而不是通常定义的紧密相连的节点的集合。其优势在于一个节点能够与多条边相连,如果这些边隶属于不同的社团,那么该节点相应地隶属于不同社团,这使一些标签属于多个不同的标签簇。

如图 1 所示,“Java”标签若与“Coffee”“Cappuccino”“Starbucks”等标签共同出现则表示咖啡种类,若与“Island”“travel”“Indonesia”等标签一起出现则表示爪哇岛,而若与“Sun”“C++”“Programming”等标签共同出现则表示编程语言。连边社团检测算法能够有效地将多义词划分到不同的聚类簇中,从而有效解决标签语义模糊的问题。

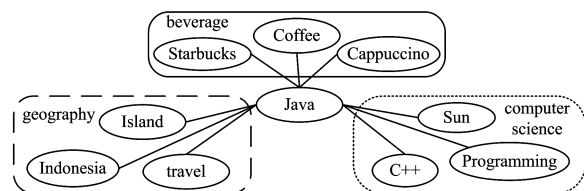


图 1 复杂网络中的多义标签

Fig. 1 Polysemous tags in complex network

该算法认为:在复杂网络中,社团的内部结构相比社团之间应该更为紧密。因此该算法的基本思想是:根据标签共现

网络中连边的凝聚过程得到整个网络的层次树结构,判断在聚类过程进行到哪一步时网络中各个社团的内部结构总体最为紧密,使得聚类效果达到最佳。其具体步骤如下:

1) 首先进行社团的初始化,将每条边看作一个社团,然后计算相邻社团的边相似度,选择相似度最大的社团进行合并,并利用树状图将每次合并记录下来;

2) 重复上述过程,直至整个网络凝聚成一个社团为止,此时得到整个网络的层次树状图;

3) 为了得到最佳社团划分结果,在网络层次树状图上计算每一步聚合完成后当前网络的划分密度,在划分密度达到最大时,网络的划分结果达到最优;

4) 将最优结果下连边的聚类还原为节点的聚类,即可得到标签聚类的最佳结果。

该算法的关键在于连边相似度的计算和划分密度的计算,其中连边相似度采用 Jaccard 相似度进行计算,公式如下:

$$sim(e_{ik}, e_{jk}) = \frac{|n_+(i) \cap n_+(j)|}{|n_+(i) \cup n_+(j)|} \quad (3)$$

其中,  $e_{ik}$  和  $e_{jk}$  表示具有公共节点  $K$  的两条邻居边,  $n_+(i)$  表示节点  $i$  和它所有邻居节点的集合,  $n_+(j)$  表示节点  $j$  和它所有邻居节点的集合。

划分密度是文献[13]引入的用于衡量连边社团检测算法何时达到最优的参数。假设一个包含  $m$  条边的网络被划分成  $c$  个社团  $\{P_1, P_2, \dots, P_c\}$ , 其中社团  $P_c$  包含  $m_c$  条连边和  $n_c$  个节点, 则定义该社团的归一化密度为:

$$D_c = \frac{m_c - (n_c - 1)}{n_c(n_c - 1)/2 - (n_c - 1)} \quad (4)$$

其中,  $n_c - 1$  是连接  $n_c$  个节点所需的最少连边数, 而  $n_c(n_c - 1)/2$  是  $n_c$  个节点之间最大可能的连边数, 因此社团划分密度  $D_c$  比较的是该社团内部的连边数与理论最大连边数, 该数值越大表明社团内部结构越紧密。整个网络的划分密度就是各个社团划分密度的加权和, 其公式如下:

$$D = \frac{1}{M} \sum_c m_c D_c = \frac{2}{M} \sum_c m_c \frac{m_c - (n_c - 1)}{(n_c - 2)(n_c - 1)} \quad (5)$$

可以看出, 划分密度  $D$  达到最大值时, 整个网络中各个社团内部的结构总体达到最紧密状态, 即社团划分达到最佳。

### 3 结合项目主题相似度与项目评分的协同过滤算法

通过上文可知, 标签簇能够代表不同的主题, 因此通过计算项目与标签簇之间的相关程度能够得到项目的主题, 进而可以根据项目的主题相关度计算项目间的主题相似度, 这在一定程度上弥补了传统协同过滤算法未考虑项目本身特点的缺陷; 同时, 通过纳入更多信息来计算项目间的相似度也能够一定程度上缓解评分数据稀疏的问题。

本文提出的算法框架如图 2 所示。

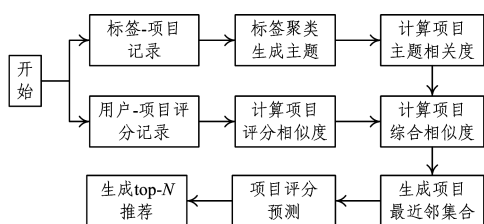


图 2 本文推荐算法的框架

Fig. 2 Framework of recommendation algorithm

### 3.1 项目主题相关度计算

本文中项目的主题通过项目的标注情况来反映。若标注项目的标签大部分来自某个标签簇, 则可以认为该项目在很大程度上与此标签簇所代表的主题相关。依次计算项目在每个主题上的相关度, 建立项目-主题相关度矩阵, 其相关度计算公式如下:

$$R_{ik} = \frac{\sum_{T_l \in C_k} count(R_i, T_l)}{\sum_{T_l \in T} count(R_i, T_l)} \quad (6)$$

其中,  $R_{ik}$  表示项目  $R_i$  与标签簇  $C_k$  的相关度,  $C_k$  表示第  $k$  个标签簇中的标签集合,  $T$  表示标注项目  $R_i$  的所有标签的集合,  $count(R_i, T_l)$  表示标签  $T_l$  标记项目  $R_i$  的次数。根据上述方法, 计算并生成如图 3 所示的项目-主题相关度矩阵, 其中  $N$  表示标签簇个数,  $M$  表示项目个数。

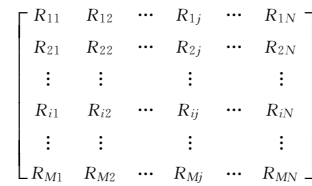


图 3 项目-主题相关度矩阵

Fig. 3 Item-topic relevance matrix

根据项目-主题相关度矩阵, 采用改进的余弦相似度公式计算项目之间的主题相似度, 公式如下:

$$sim(i, j) = \frac{\sum_{k=1}^N (R_{ik} - \bar{R}_i)(R_{jk} - \bar{R}_j)}{\sqrt{\sum_{k=1}^N (R_{ik} - \bar{R}_i)^2} \sqrt{\sum_{k=1}^N (R_{jk} - \bar{R}_j)^2}} \quad (7)$$

其中,  $\bar{R}_i$  与  $\bar{R}_j$  分别表示项目  $i$  和  $j$  与每个标签簇相关度的平均值。

### 3.2 结合项目主题相似度与项目评分的项目相似度计算方法

本文综合考虑项目之间的主题相似度与评分相似度, 提出了结合项目主题相似度与评分相似度的项目间相似度计算方法。以项目  $i$  与项目  $j$  为例, 本文计算方法如下:

- 1) 根据用户-项目评分矩阵, 利用式(1)计算项目  $i$  与项目  $j$  之间的评分相似度, 记作  $simR(i, j)$ ;
- 2) 根据项目-主题相关度矩阵, 利用式(7)计算项目  $i$  与项目  $j$  的主题相似度, 记作  $simT(i, j)$ ;
- 3) 将评分相似度与主题相似度相结合, 计算项目  $i$  与项目  $j$  之间的综合相似度, 用  $sim(i, j)$  表示:

$$sim(i, j) = \alpha simT(i, j) + (1 - \alpha) simR(i, j) \quad (8)$$

其中,  $\alpha$  为相似度调节系数, 用于调整  $simR(i, j)$  与  $simT(i, j)$  所占权重。

### 3.3 结合项目主题相似度与项目评分的协同过滤算法

设用户-项目评分矩阵为  $r_{m \times n}$ , 项目主题相关度矩阵为  $R_{M \times N}$ , 本文提出的结合项目主题相关度的协同过滤算法的具体描述如算法 1 所示。

#### 算法 1 结合项目主题相关度的协同过滤算法

输入: 用户-项目评分记录、标签-项目记录、目标项目的最近邻居数  $K$ 、目标用户  $u$

输出: 目标用户  $u$  的 top-N 推荐列表

- 步骤 1 根据标签-项目记录, 采用基于连边社团检测的标签聚类算法将标签聚类成不同标签簇, 代表不同主题;
- 步骤 2 根据式(6)计算项目主题相关度, 生成项目-主题相关度矩阵

$R_{M \times N}$ ;

- 步骤 3 从目标用户  $u$  未评分的项目集合  $I$  中将任意的一个项目  $i \in I$  作为目标项目,结合项目-评分矩阵  $r_{m \times n}$  与项目-主题相关度矩阵  $R_{M \times N}$ ,根据式(8)计算项目间相似度;
- 步骤 4 取相似度较高的前  $K$  个项目作为目标项目的最近邻,根据式(2)计算目标项目的预测评分值;
- 步骤 5 重复步骤 3-4,计算用户  $u$  所有未评分项目的预测评分,将评分较高的前  $N$  个项目形成 top-N 推荐列表反馈给用户。

### 4 实验结果与分析

#### 4.1 实验数据集

本实验采用的是 GroupLens 项目组开放的 Movielens 数据集<sup>[14]</sup>。该数据集来源于 Minnesota 大学开发的电影推荐系统 Movielens,用户可以对电影进行评分,分值最低为 1,最高为 5;同时用户也可以对自己观看过的电影标注标签。

该数据集包括用户评分数据集和用户-标签-电影数据集,其中包含 71567 个用户对 10681 部电影的 10000000 个评分记录和 95580 个标签记录,并且每个用户至少评价了 20 部电影。

在对标签进行合并大小写、单复数等处理后,选择标注次数大于 40 的 423 个标签,并选取被标注次数大于 20 的 1307 部电影,将所选数据的 80% 作为训练集、剩余 20% 作为测试集进行实验。

#### 4.2 评价标准

本实验采用平均绝对误差 MAE 作为评价指标。MAE 作为推荐系统中常用的评价指标之一,通过计算用户的预测评分与实际评分之间的偏差来度量预测的准确性,MAE 值越小则推荐质量越高。设测试集中共有  $N$  条数据,分别为  $\{r_1, r_2, \dots, r_n\}$ ,算法对这些数据的预测值为  $\{p_1, p_2, \dots, p_n\}$ ,则 MAE 的计算公式为:

$$MAE = \frac{\sum_{i=1}^N |p_i - r_i|}{N} \tag{9}$$

#### 4.3 实验结果及分析

实验首先针对参数的选取进行探讨,分析不同参数值对算法的影响,并在实验得出的最优参数的基础上将本文算法与其他算法进行对比。

##### 4.3.1 针对相似度调节系数的实验

根据本文描述的聚类方法,最终将标签聚成 10 个标签簇。为了探讨相似度调节系数  $\alpha$  对算法的影响,分别设定项目的最近邻为 20,40 和 60,将  $\alpha$  的值从 0.1 逐次递增至 0.9 进行实验。如图 4 所示,随着  $\alpha$  的不断增大,算法的 MAE 值先逐渐减小,然后逐渐增大,且 3 种最近邻数目下的 MAE 均在  $\alpha$  取 0.4 时取得最低值,由此可知在目标项目最近邻数目固定的情况下  $\alpha$  为 0.4 时能够产生较好的推荐结果。

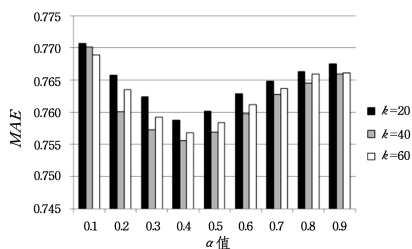


图 4 不同  $\alpha$  值下 MAE 的比较

Fig. 4 Comparison of MAE under different values of  $\alpha$

##### 4.3.2 针对目标项目最近邻的实验

由上述实验得知, $\alpha$  为 0.4 时能够取得较好的推荐效果,因此固定  $\alpha$  值为 0.4,将目标项目的最近邻数量由 10 逐次递增至 70 进行实验。如图 5 所示,随着最近邻数目的不断增加,算法的 MAE 值逐渐减小,推荐效果逐步提升,在项目最近邻为 30 时,MAE 值达到最小,推荐效果达到最佳;之后,随着目标项目最近邻的增大,MAE 值逐渐增大,推荐效果随之变差。因此在实验中选取目标项目的最近邻为 30 时能够得到最佳的推荐效果。

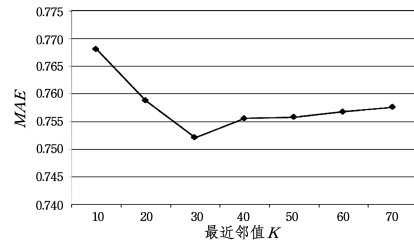


图 5 不同最近邻数下 MAE 比较

Fig. 5 Comparison of MAE under different values of nearest neighbor

##### 4.3.3 与其他算法的比较

为了验证本文算法的效果,选取基于项目的协同过滤算法(BCF)<sup>[15]</sup>和基于用户项目混合的协同过滤算法(Hybrid-CF)<sup>[16]</sup>与本文算法进行对比实验。文献<sup>[16]</sup>表明与传统的协同过滤相比,Hybrid-CF 算法能够提高预测精度并缓解数据稀疏性。设相似度调节系数  $\alpha$  为 0.4, $K$  值从 10 到 70 逐渐递增,实验结果如图 6 所示。

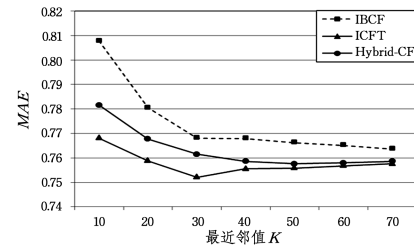


图 6 不同算法的 MAE 比较

Fig. 6 Comparison of MAE under different algorithms

可以发现,在任意的最近邻居数  $K$  值下,本文提出的算法比 IBCF 算法、Hybrid-CF 算法具有更低的 MAE 值,这表明本文提出的算法在实验中具有更好的推荐效果。

同时可以发现,在目标项目最近邻不断增大的情况下,本文提出的算法与两种对比算法之间的 MAE 差值逐步缩小,这是因为随着目标项目最近邻个数的逐渐增加,由标签带来的项目主题信息对于推荐算法的改善效果正在逐渐减弱。

**结束语** 传统基于项目的协同过滤算法在计算项目相似度时仅依靠评分信息,未考虑项目的自身特征。本文采用社会化标注中产生的标签数据作为项目信息的补充,利用标签共现网络与连边社团检测算法对标签进行聚类并生成主题标签簇,根据项目标注情况计算项目主题相关度,生成项目-主题相关度矩阵;然后利用项目-评分矩阵与项目-主题相关度矩阵计算项目的综合相似度,选择用户的未评分项目作为目标进行评分预测,从而产生个性化推荐;最后在 Movielens 数据集上对本文提出的算法进行实验。实验结果表明本文算法

能够改善推荐效果。

由于本文在分析项目主题相关性时仅考虑了标签的共现关系与标注频次,因此在后续的研究中将尝试在本文算法的基础上结合标签语义分析与标签重要程度来挖掘项目主题,更加准确地分析项目主题相似度,以提高推荐的质量。

### 参 考 文 献

- [1] BOBADILLA J, ORTEGA F, HERNANDO A, et al. Recommender systems survey[J]. Knowledge-based Systems, 2013, 46(1):109-132.
- [2] RESNICK P, IACOVU N, SUCHAK M, et al. GroupLens: an open architecture for collaborative filtering of netnews[C]//ACM Conference on Computer Supported Cooperative Work. ACM, 1994:175-186.
- [3] LINDEN G, SMITH B, YORK J. Amazon.com recommendations: item-to-item collaborative filtering [J]. IEEE Internet Computing, 2003, 7(1):76-80.
- [4] FENG W, WANG J. Incorporating heterogeneous information for personalized tag recommendation in social tagging systems [C]//ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2012:1276-1284.
- [5] ZHANG B, GAO Y, GAO K N, et al. Combining Relation and Content Analysis for Social Tagging Recommendation[J]. Journal of Software, 2012, 23(3):476-488. (in Chinese)  
张斌,高引,高克宁,等.融合关系与内容分析的社会标签推荐[J].软件学报,2012,23(3):476-488.
- [6] SONG H, LU P, ZHAO K. Improving item-based collaborative filtering recommendation system with tag [C]//International Conference on Artificial Intelligence, Management Science and Electronic Commerce. IEEE, 2011:2142-2145.
- [7] CAI Q, HAN D M, LI H S, et al. Personal Resource Recommendation Based on Tags and Collaborative Filtering[J]. Computer Science, 2014, 41(1):69-71. (in Chinese)  
蔡强,韩东梅,李海生,等.基于标签和协同过滤的个性化资源推荐[J].计算机科学,2014,41(1):69-71.
- [8] LI G, WANG S, LI Z Y, et al. Personalized Tag Recommendation Algorithm Based on Tensor Decomposition[J]. Computer Science, 2015, 41(4):30-35. (in Chinese)  
李贵,王爽,李征宇,等.基于张量分解的个性化标签推荐算法[J].计算机科学,2015,41(4):30-35.
- [9] WANG W P, WANG J H. Hybrid Recommendation Method Based on Tag and Collaborative Filtering[J]. Computer Engineering, 2011, 37(14):34-35. (in Chinese)  
王卫平,王金辉.基于Tag和协同过滤的混合推荐方法[J].计算机工程,2011,37(14):34-35.
- [10] KIM H N, JI A T, HA I, et al. Collaborative Filtering Based on Collaborative Tagging for Enhancing the Quality of Recommendation [J]. Electronic Commerce Research & Applications, 2010, 9(1):73-83.
- [11] HALPIN H, ROBU V, SHEPHERD H. The complex dynamics of collaborative tagging[C]//International Conference on World Wide Web(WWW 2007). Banff, Alberta, Canada, May. DBLP, 2007:211-220.
- [12] WU C, ZHOU B. Complex network analysis of tag as a social network[J]. Journal of Zhejiang University (Engineering Science), 2010, 4(11):2194-2197. (in Chinese)  
吴超,周波.基于复杂网络的社会化标签分析[J].浙江大学学报:工学版,2010,4(11):2194-2197.
- [13] AHN Y, BAGROW J, LEHMANN S. Link Communities Reveal Multiscale Complexity In networks [J]. Nature, 2010, 466(7307):761-764.
- [14] GroupLens[EB/OL]. <http://www.grouplens.org>.
- [15] SARWAR B, KARYPIS G, KONSTAN J, et al. Item-based collaborative filtering recommendation algorithms [C]//International Conference on World Wide Web. ACM, 2001:285-295.
- [16] JI H, LI J, REN C, et al. Hybrid collaborative filtering model for improved recommendation [C]//IEEE International Conference on Service Operations and Logistics, and Informatics. IEEE, 2013:142-145.
- [12] DAO T H, JEONG S R, AHN H. A novel recommendation model of location-based advertising: Context-Aware Collaborative Filtering using GA approach[J]. Expert Systems with Applications, 2012, 39(3):3731-3739.
- [13] LIU R S, YANG T C. Improving Recommendation Accuracy by Considering Electronic Word-of-Mouth and the Effects of Its Propagation Using Collective Matrix Factorization [C]//IEEE Datacom. IEEE, 2016.
- [14] ZHEN G L, ZHU F X, YAO X. Recommendation Rating Prediction Based on Attribute Boosting with Partial Sampling[J]. Chinese Journal of Computers, 2016, 39(8):1501-1514. (in Chinese)  
郑麟,朱福喜,姚杏.基于属性提升与局部采样的推荐评分预测[J].计算机学报,2016,39(8):1501-1514.
- [15] EL-MOEMEN S A, HASSAN T, SEWISY A A. A Context-Aware Recommender System for Personalized Places in Mobile Applications [J]. International Journal of Advanced Computer Science & Applications, 2016, 7(3):442-448.
- [16] CHAMPIRI Z D, SHAHAMIRI S R, SALIM S S B. A systematic review of scholar context-aware recommender systems [J]. Expert Systems with Applications, 2015, 42(3):1743-1758.

(上接第219页)