

# 基于标签信息特征相似性的协同过滤个性化推荐

何 明 要凯升 杨 芑 张久伶  
(北京工业大学信息学部 北京 100124)

**摘 要** 标签推荐系统旨在利用标签数据为用户提供个性化推荐。已有的基于标签的推荐方法往往忽视了用户和资源本身的特征,而且在相似性度量时仅针对项目相似性或用户相似性进行计算,并未充分考虑二者之间的有效融合,推荐结果的准确性较低。为了解决上述问题,将标签信息融入到结合用户相似性和项目相似性的协同过滤中,提出融合标签特征与相似性的协同过滤个性化推荐方法。该方法在充分考虑用户、项目以及标签信息的基础上,利用二维矩阵来定义用户-标签以及标签-项目之间的行为。构建用户和项目的标签特征表示,通过基于标签特征的相似性度量方法计算用户相似性和项目相似性。基于用户标签行为和用户与项目的相似性线性组合来预测用户对项目的偏好值,并根据预测偏好值排序,生成最终的推荐列表。在 Last.fm 数据集上的实验结果表明,该方法能够提高推荐的准确度,满足用户的个性化需求。

**关键词** 协同过滤,标签,推荐系统,相似性计算

**中图法分类号** TP391 **文献标识码** A

## Collaborative Filtering Personalized Recommendation Based on Similarity of Tag Information Feature

HE Ming YAO Kai-sheng YANG Peng ZHANG Jiu-ling

(Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China)

**Abstract** Tag recommendation systems are aimed to provide personalized recommendation using tag data for users. Previous tag based recommendation methods usually neglect the characteristics of users and items, and similarity measures are unconsidered fully incorporating effectively both user similarity and item similarity, which leads to deviation of recommendation results. To address this issue, this paper proposed the collaborative filtering recommendation method of combining tag features and similarity for personalized recommendation. Two-dimensional matrix is used to define actions among user-tag and tag-item based on integrating information among users, tags and items. Tag features representation is constructed, and user similarity and item similarity are calculated by similarity measure method based on tag features. The user preferences for items are predicted by their tag behaviors and linear combination of similarity of users and items, and the recommended list is generated according to the rank of preferences. The experimental results on Last.fm show that the proposed method can improve recommendation accuracy and satisfy the requirement for users.

**Keywords** Collaborative filtering, Tag, Recommendation systems, Similarity computation

## 1 引言

作为互联网时代的标志性技术,Web2.0 不断推动着互联网技术的变革。2017 年 6 月,在 Code 大会上发布的互联网趋势报告“Internet Trends”显示,全球互联网用户数已达到 34 亿,用户生成内容 UGC 使得互联网上的信息量爆增。在信息大数据时代,用户的个性化需求不断提高,面对海量数据信息,如何帮助用户有效获取满足其自身需要的信息,并有力改善“信息过载”问题,是数据科研工作者所面临的主要研究挑战之一。推荐系统<sup>[1-2]</sup>(Recommender Systems, RS)作为有效缓解该方法的方法<sup>[3]</sup>,通过分析用户的历史行为数据,为用户推荐个性化的内容。在信息智能时代,推荐系统已经成为互联网以及数据服务公司的核心技术模块之一,对推进推荐系统技术的发展具有重要的应用意义。

协同过滤<sup>[4-5]</sup>(Collaborative Filtering, CF)是推荐系统中应用最为广泛的推荐技术之一<sup>[6]</sup>,其基本思想是基于用户(Users)对项目(Items)的评分或其他行为模式,分析用户的喜好,进而为目标用户提供个性化的推荐,且不需要项目的显式特征表示。然而,在海量个性化需求的驱动下,协同过滤技术仍然面临一些挑战:1)冷启动问题<sup>[7]</sup>,即当新用户或项目出现时,由于缺乏它们的偏好信息而无法生成推荐;2)数据稀疏性问题,即当评分数据比较稀疏时,根据传统计算方法很难找到相似用户,导致推荐质量下降。出现上述问题的主要原因是数据不够充分。为了提供有效的推荐,还需要合适且容易获取的更多数据来丰富用户或项目的特征表示方式。

标签<sup>[8]</sup>(Tags)通常用来表示用户或者项目的属性,也可以用来表示用户对物品的理解和看法。其作为 Web 2.0 时代在社会化网络中的重要应用,体现了用户对资源的理解,既

本文受国家自然科学基金项目(91646201,91546111),北京市教委科研计划一般项目(KM201710005023)资助。

何 明(1975—),男,博士,副教授,主要研究方向为推荐系统、机器学习, E-mail: heming@bjut.edu.cn; 要凯升(1994—),男,硕士生,主要研究方向为推荐系统、数据挖掘; 杨 芑(1994—),男,硕士生,主要研究方向为数据挖掘; 张久伶(1990—),男,硕士生,主要研究方向为机器学习。

表达了信息资源的主要特征,又涵盖了用户与资源之间以及用户与用户之间的关系,兼具内容与关联的特征<sup>[9]</sup>。在Web2.0时代,可以更好地实现对各类信息的归类和处理。Delicious 共享书签、Flickr 共享照片、CiteUlike 共享学术文献等都应用了社会标签来描述和共享资源。将标签作为推荐的数据来源,充分利用自发标签直接反映用户兴趣和资源内容的特点,便有可能开发出同时具备内容过滤和协同过滤优越性的推荐技术,提高推荐系统的准确性和交互性<sup>[10]</sup>。Zhang等<sup>[11-13]</sup>介绍了基于“用户-项目-标签”三元组的推荐算法,将标签信息作为一个重要特征应用到推荐算法中,并以三分图的形式来描述用户、项目与标签三者之间的关系,最终实现个性化推荐。刘建勋等<sup>[14]</sup>提出了一种基于主题模型的方法,首先建立 Mashup 与 Web APIs 的描述文档以及 Mashup 与 APIs 之间的组合关系模型,然后寻找与待推荐标签 Mashup 的描述文档主题分布相似的 Web APIs,进而进行 Mashup 标签的自动推荐。张斌等<sup>[15]</sup>提出了一种基于概率的标签推荐方法,其进一步提升了标签推荐的质量。李锡荣等<sup>[16]</sup>通过考查近邻权重以及近邻标签权重,提出了一种一般性的软近邻投票框架。Yang等<sup>[17]</sup>从机器学习的视角,提出了一种为社会化推荐系统自动推荐标签的算法。孔欣欣等<sup>[18]</sup>提出了一种基于标签权重评分的推荐系统模型,其通过标签权重评分来更加准确地获取用户的评价和需求。Jomsri等<sup>[19]</sup>提出了基于标签的推荐系统架构,利用标签集来表达用户的偏好,并应用此偏好为用户推荐适合的研究论文。实验结果表明,用户自定义的标签能用于表达每个个体用户的偏好,提高了推荐的准确度。郭彩云等<sup>[20]</sup>提出了一种改进的基于标签的协同过滤算法,将用户评分融入到用户对标签权重的计算中,考虑用户不同兴趣程度的项目对推荐结果的影响,并使用指数渐进遗忘函数和时间窗口相结合的方法来捕捉用户兴趣的变化。李慧等<sup>[21]</sup>综合了用户的资源标签与标签概率模型,提出了一种个性化标签推荐方法。

由此可见,基于标签的推荐系统通过使用用户标注信息,更好地针对资源以及用户自身的特征信息为用户做出个性化推荐,提高推荐质量。然而,现有的标签系统中的推荐方法通常仅利用了用户-标签-资源之间的关系,忽视了用户和资源自身的特征;而且,在相似性度量方面,已有的方法仅仅针对项目相似性或用户相似性进行计算,并未充分考虑二者的有效结合,使得推荐结果的准确性出现偏差。目前,在基于标签的推荐过程中同时考虑项目相似性和用户相似性的研究还较少。因此,本文探索如何利用标签这一新型特征表示方式和相似性计算来提高个性化推荐系统的推荐准确度。本文的主要贡献和创新性如下:

1)首次将标签特征信息融入到结合用户相似性和项目相似性的协同过滤中来预测用户偏好,提出了基于用户-标签信息的协同过滤(UBTCF)和基于项目-标签信息的协同过滤(IBTCF);

2)基于用户-标签特征和项目-标签特征表示,提出了一种新的结合用户相似性和项目相似性的协同过滤个性化推荐算法(TFSCF);

3)通过仿真实验验证了方法的有效性,实验结果表明,将标签特征与相似性集成到协同过滤推荐中可以提高推荐的准确性。

## 2 传统协同过滤推荐算法

### 2.1 用户-项目评分矩阵

传统的协同过滤推荐算法中,首先将用户对项目的评分信息以  $m \times n$  阶的用户-项目评分的矩阵  $R(m, n)$  来表示,如表1所列。其中,  $m$  表示用户数,  $n$  表示项目数,  $R_{i,j}$  表示用户  $i$  对项目  $j$  的评分值。然后,计算目标用户或项目与其他用户或项目之间的相似度,找出与目标用户或项目最相似的“最近邻”集合作为目标用户的最近邻。最后,根据最近邻集合中的实际评分矩阵预测目标用户对未评分项目的评分。选取预测评分最高的  $N$  个项目作为推荐结果反馈给目标用户。

表1 用户-项目评分矩阵  $R(m, n)$

	Item <sub>1</sub>	...	Item <sub>j</sub>	...	Item <sub>n</sub>
User <sub>1</sub>	R <sub>1,1</sub>		R <sub>1,j</sub>		R <sub>1,n</sub>
⋮	⋮	⋮	⋮	⋮	⋮
User <sub>i</sub>	R <sub>i,1</sub>		R <sub>i,j</sub>		R <sub>i,n</sub>
⋮	⋮	⋮	⋮	⋮	⋮
User <sub>m</sub>	R <sub>m,1</sub>	...	R <sub>m,j</sub>	...	R <sub>m,n</sub>

### 2.2 相似性度量方法

相似性计算是协同过滤推荐算法中的关键技术之一,常用的相似性度量方法有3种,即余弦相似性、相关相似性和修正的余弦相似性。

1)余弦相似性。每个用户的所有评分数据被看作是  $n$  维项目空间中的一个向量,  $n$  为项目数量。如果用户没有对某个项目进行评分,则默认其评分为0。两个用户间的相似性可以看作是两个用户评分向量之间夹角的余弦值,余弦值越大,则用户间的相似性越高。假设两个用户  $x$  和  $y$  的评分向量分别为  $u$  和  $v$ ,则  $x$  和  $y$  之间的余弦相似性为:

$$\text{sim}(x, y) = \cos(u, v) = \frac{\vec{u} \times \vec{v}}{\|\vec{u}\| \|\vec{v}\|} \quad (1)$$

2)相关相似性,又称为 Person 相关性。假设用户  $u$  和用户  $v$  的共同评分项集合为  $R_w, R_u$  和  $R_v$ ,分别表示用户  $u$  和用户  $v$  对项目  $i$  的评分,则用户  $u$  和用户  $v$  的相关相似性为:

$$\text{sim}(u, v) = \frac{\sum_{i \in R_w} (R_{u,i} - \bar{R}_u) \times (R_{v,i} - \bar{R}_v)}{\sqrt{\sum_{i \in R_w} (R_{u,i} - \bar{R}_u)^2} \times \sqrt{\sum_{i \in R_w} (R_{v,i} - \bar{R}_v)^2}} \quad (2)$$

3)修正的余弦相似性。余弦相似性度量方法忽略了不同用户的评分尺度问题。修正的余弦相似性度量方法减去了用户对项目的平均评分。设用户  $x$  与用户  $y$  共同评分过的集合用  $R_{xy}$  表示,  $R_x$  和  $R_y$  分别表示用户  $x$  和用户  $y$  的评分集合,则用户  $x$  和用户  $y$  的修正余弦相似性为:

$$\text{sim}(x, y) = \frac{\sum_{i \in R_{xy}} (R_{x,i} - \bar{R}_x) \times (R_{y,i} - \bar{R}_y)}{\sqrt{\sum_{i \in R_x} (R_{x,i} - \bar{R}_x)^2} \times \sqrt{\sum_{i \in R_y} (R_{y,i} - \bar{R}_y)^2}} \quad (3)$$

### 2.3 评分预测

通常采用  $K$  近邻方法进行评分预测,即选择与目标用户最相似的  $K$  个用户作为最近邻集合来进行计算。假设集合  $U$  表示目标用户  $u$  的最近邻集合,则用户  $u$  对项目  $i$  的预测评分值可以用如下公式表示:

$$P(u, i) = \bar{R}_u + \frac{\sum_{u_k \in U} \text{sim}(u, u_k) \times (R_{u_k, i} - \bar{R}_{u_k})}{\sum_{u_k \in U} \text{sim}(u, u_k)} \quad (4)$$

其中,  $R_{u_k, i}$  表示  $u_k$  对项目  $i$  的非空评分,  $\bar{R}_{u_k}$  表示  $u_k$  在与  $u$  的

共同评分项集合中的平均评分,  $\bar{R}_u$  则表示  $u$  在所有的项目上的平均评分。

### 3 基于标签的协同过滤个性化推荐方法

随着用户以及项目数量的不断增加,用户-项目评分矩阵  $R(m,n)$  越来越大,同时由于用户很少给予项目评分,导致用户-项目评分矩阵较为稀疏,不能很好地从中提取用户的偏好特征。近年来,随着人们对社会化标签研究的不断深入,标签标注系统可以较好地解决这一问题,更好地实现信息分类,通过挖掘用户使用标签标记项目的历史行为数据,更好地针对项目以及用户自身的特征信息为用户做出个性化推荐,提高推荐质量。

#### 3.1 基于用户的协同过滤推荐方法

基于用户的协同过滤方法首先是通过查找与目标用户有相似兴趣的其他用户,然后根据相似用户的项目选择记录为目标用户推荐可能喜欢且没有选择过的项目。如图 1 所示,  $User_1$  在  $Product_2$  上有过行为记录,  $User_2$  在  $Product_1, Product_2, Product_3$  以及  $Product_4$  上均有行为记录,  $User_2$  在  $Product_2$  上有过行为记录,  $User_3$  在  $Product_2$  以及  $Product_3$  上有行为记录。根据用户的行为数据可以看出,相对于  $User_2$  来说,  $User_1$  与  $User_3$  的喜好的相似度更高,因此,当为  $User_3$  推荐物品时,根据  $User_1$  的行为记录,为  $User_3$  推荐  $Product_1$  以及  $Product_4$ 。

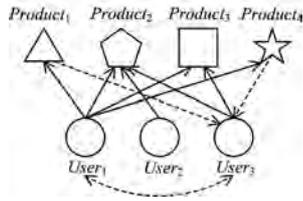


图 1 基于用户的协同过滤

由于对标签的应用越来越广泛,本文将标签信息融入到协同过滤推荐中,提出了 UBTCF。通过结合用户的标签行为数据和协同过滤方法,寻找并确定目标用户的最近邻,并通过最近邻的行为记录为目标用户进行推荐。

#### 3.2 基于项目的协同过滤方法

传统的基于项目的协同过滤方法是通过所有用户的历史偏好数据来确定项目之间的相似性,然后为用户推荐与他们之前喜欢的项目相似的项目,如图 2 所示。

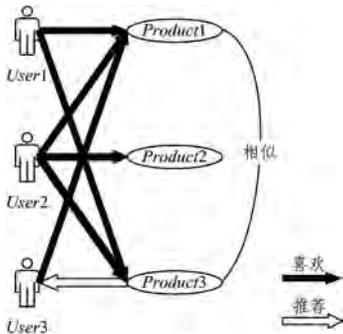


图 2 基于项目的协同过滤

在图 2 中,  $User_1$  喜欢  $Product_1$  和  $Product_3$ ,  $User_2$  喜欢项  $Product_1, Product_2$  以及  $Product_3$ ,  $User_3$  在  $Product_1$  上有行为记录。而  $Product_1$  与  $Product_3$  相似,因此当为  $User_3$  做

推荐时,推荐系统会推荐  $Product_3$ 。

但是这种方法在考虑项目之间的相似性时,忽略了项目本身的性质以及属性。基于此,本文将标签信息与协同过滤推荐相结合,提出了 IBTCF。通过结合项目的标签信息和协同过滤方法,确定项目之间的相似性,并为目标用户推荐与其所喜好项目相似度较高的项目。

#### 3.3 用户-标签-项目三分图

首先,本文尝试将标签信息引入协同过滤推荐方法中,构建用户-标签-项目三维关系,并从中获取用户、标签以及项目之间的关系,如图 3 所示。图 3 中,  $User_i$  表示用户,  $Tag_j$  表示标签,  $Item_k$  表示项目。如果用户  $User_i$  使用了标签  $Tag_j$ ,则在用户  $User_i$  和标签  $Tag_j$  之间使用一条连线进行连接,边上的值为用户  $User_i$  使用标签  $Tag_j$  的次数  $N$ ,表示用户  $User_i$  使用  $N$  次标签  $Tag_j$  用于标记项目。如果项目  $Item_k$  被标签  $Tag_j$  标记过,则在项目  $Item_k$  和标签  $Tag_j$  之间使用一条连线进行连接,边上的值为项目  $Item_k$  被标签  $Tag_j$  标记的次数  $N$ ,表示项目  $Item_k$  被标签  $Tag_j$  标记了  $N$  次。

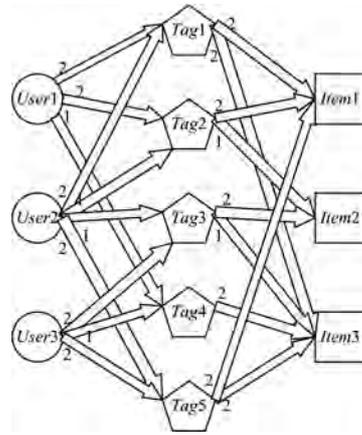


图 3 用户-标签-项目三维关系

#### 3.4 基于标签的关系映射

当得到用户-标签-项目三者之间的关系后,以标签为媒介,通过关系映射将用户-项目-标签构成的三维关系转化为用户-标签  $user\_tag$  以及标签-项目  $tag\_source$  两个二维标签行为矩阵,进而对两个矩阵分别进行处理。

其中,  $user\_tag$  用于表示用户  $u_i$  使用标签  $t_j$  的情况。

$$user\_tag_{i \times j} = \begin{bmatrix} u_1 t_1 & \cdots & u_1 t_j \\ \vdots & \ddots & \vdots \\ u_i t_1 & \cdots & u_i t_j \end{bmatrix} \quad (5)$$

$tag\_source$  用于表示项目  $s_k$  被标签  $t_j$  标记的情况。

$$tag\_source_{j \times k} = \begin{bmatrix} t_1 s_1 & \cdots & t_1 s_k \\ \vdots & \ddots & \vdots \\ t_j s_1 & \cdots & t_j s_k \end{bmatrix} \quad (6)$$

#### 3.5 基于标签的用户偏好计算

在基于标签的推荐协同过滤个性化推荐中,数据可以表示为一个三元组:

$$T = (U, L, S)$$

其中,  $U$  为用户集合,  $U = \{u_1, u_2, \dots, u_i, \dots, u_X\}$ ,  $X$  为用户总数,  $u_i$  为第  $i$  个用户,  $i = 1, 2, \dots, X$ ;  $L$  为标签集合,  $L = \{l_1, l_2, \dots, l_j, \dots, l_Y\}$ ,  $Y$  为标签总数,  $l_j$  为第  $j$  个标签,  $j = 1, 2, \dots, Y$ ;  $S$  为项目集合,  $S = \{s_1, s_2, \dots, s_k, \dots, s_Z\}$ ,  $Z$  为项目总数,  $s_k$  为第  $k$  个项目,  $k = 1, 2, \dots, Z$ 。

## 1) 用户及项目的标签特征表示

对于某个目标用户,需要根据用户使用标签的记录来获取用户基于标签的兴趣特征,可以根据式(7)获得:

$$\vec{F}_{ul} = \left( \frac{q_{ul}}{q_{ul}} \log \frac{X}{q_{lu}}, \dots, \frac{q_{ul}}{q_{ul}} \log \frac{X}{q_{lu}}, \dots, \frac{q_{ul}}{q_{ul}} \log \frac{X}{q_{lu}} \right) \quad (7)$$

其中,  $q_{ul}$  表示用户  $u$  使用标签  $l_j$  的次数,  $q_{ul}$  表示用户  $u$  使用标签的总数,  $X$  表示用户总数,  $q_{lu}$  表示使用标签  $l_j$  的用户数,

$\frac{q_{ul}}{q_{ul}}$  表示用户  $u$  使用标签  $l_j$  的频率,  $\log \frac{X}{q_{lu}}$  表示标签  $l_j$  在用

户所有标签中的重要程度,  $\frac{q_{ul}}{q_{ul}} \log \frac{X}{q_{lu}}$  表示标签  $l_j$  对于用户  $u$  的重要度。

由于标签在整个推荐过程中起到了一个关联的作用,最终目标还是确定用户与项目之间的关系,因此还需要考虑各个项目基于标签的特征表示,可以根据式(8)获得:

$$\vec{F}_{sl} = \left( \frac{q_{sl}}{q_{sl}} \log \frac{Z}{q_{ls}}, \dots, \frac{q_{sl}}{q_{sl}} \log \frac{Z}{q_{ls}}, \dots, \frac{q_{sl}}{q_{sl}} \log \frac{Z}{q_{ls}} \right) \quad (8)$$

其中,  $q_{sl}$  表示项目  $s$  被标记的标签总数,  $q_{sl}$  表示项目  $s$  上贴有的标签总数,  $Z$  表示项目总数量,  $q_{ls}$  表示被标签  $l_j$  标记过的项目总数,

$\frac{q_{sl}}{q_{sl}}$  表示标签  $l_j$  标记该项目  $s$  的频率,  $\log \frac{Z}{q_{ls}}$  表示

在项目  $s$  使用的所有标签中, 标签  $l_j$  对于项目  $s$  的重要度,

$\frac{q_{sl}}{q_{sl}} \log \frac{Z}{q_{ls}}$  表示标签  $l_j$  对于项目  $s$  的重要度。

## 2) 用户对项目的偏好计算

在确定了用户兴趣特征以及标签的特征表示之后,将二者结合起来,则可以得到用户对各个项目的偏好。

将用户  $u_i$  对项目  $s_k$  的偏好度记为:

$$F_{u_i s_k} = \vec{F}_{u_i l} \cdot \vec{F}_{sl} = \sum_{l=1}^Y F_{u_i l} \times F_{l s_k} \quad (9)$$

其中,  $F_{u_i s_k}$  表示用户  $u_i$  对项目  $s_k$  的偏好程度,  $u_i \in U, j=1, 2, \dots, X; s_k \in S, k=1, 2, \dots, Z$ 。

基于用户对项目的偏好程度构建用户-项目偏好矩阵, 记为:

$$F_{X \times Z} = \begin{pmatrix} F_{u_1 s_1} & \dots & F_{u_1 s_k} & \dots & F_{u_1 s_z} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ F_{u_i s_1} & \dots & F_{u_i s_k} & \dots & F_{u_i s_z} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ F_{u_x s_1} & \dots & F_{u_x s_k} & \dots & F_{u_x s_z} \end{pmatrix} \quad (10)$$

该矩阵可以反映出各个用户对各个项目的偏好程度。

## 3.6 基于标签的相似性计算

## 3.6.1 基于标签的用户相似性计算

用户相似性<sup>[22]</sup>表示的是两个不同用户间历史行为的相似程度。通过挖掘用户的历史行为数据,查询用户使用过的标签,寻找到与目标用户最近的邻居,从而将用户最近邻的偏好项目推荐给目标用户。

用户的特征信息都可以通过式(11)表示:

$$\vec{U}_i = (f_{i1}, f_{i2}, \dots, f_{ij}, \dots, f_{iY}) \quad (11)$$

其中,  $i=1, 2, \dots, S; j=1, 2, \dots, Y; f_{ij}$  表示用户  $u_i$  使用标签  $l_j$  的次数。

所有用户的行为信息可以通过用户行为表示矩阵  $U_{i \times i}$  表示:

$$U_{i \times i} = \begin{bmatrix} f_{i1} & \dots & f_{ii} \\ \vdots & \ddots & \vdots \\ f_{i1} & \dots & f_{ii} \end{bmatrix} \quad (12)$$

本文采用上文介绍的皮尔森相关相似性来计算各用户间的相似性。将用户评分替换成标签的使用次数。设用户  $U_a$  和用户  $U_b$  共同使用过的标签集用  $L_{ab}$  表示, 则用户  $U_a$  和用户  $U_b$  之间的相似性  $sim(U_a, U_b)$  可通过式(13)计算得到:

$$sim(U_a, U_b) = \frac{\sum_{i \in L_{ab}} (R_{a,i} - \bar{R}_a) \times (R_{b,i} - \bar{R}_b)}{\sqrt{\sum_{i \in L_{ab}} (R_{a,i} - \bar{R}_a)^2} \times \sqrt{\sum_{i \in L_{ab}} (R_{b,i} - \bar{R}_b)^2}} \quad (13)$$

其中,  $R_{a,i}$  和  $R_{b,i}$  分别表示用户  $U_a$  和用户  $U_b$  使用标签  $i$  的次数,  $\bar{R}_a$  和  $\bar{R}_b$  分别表示用户  $U_a$  和用户  $U_b$  使用标签的平均次数。

通过用户间的相似性计算,可以构建式(14)所示的用户相似性矩阵  $U_{XX}$ :

$$U_{XX} = \begin{bmatrix} 1 & \dots & U_{1b} & \dots & U_{1X} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ U_{a1} & & 1 & & U_{aX} \\ \vdots & & \vdots & \ddots & \vdots \\ U_{X1} & \dots & U_{Xb} & \dots & U_{XX} \end{bmatrix} \quad (14)$$

其中,  $a=1, 2, \dots, X, b=1, 2, \dots, X, U_{ab}$  表示用户  $U_a$  和用户  $U_b$  的相似性。

## 3.6.2 基于标签的项目相似性计算

项目的相似性表示的是不同项目之间的相似程度。通过比较各个项目上所标记的标签来确定各个项目之间的相似程度,根据用户已选项目,为用户推荐与其相似的项目。

项目  $S_k$  的特征信息可以通过式(15)表示:

$$\vec{S}_k = (f_{k1}, f_{k2}, \dots, f_{kj}, \dots, f_{kY}) \quad (15)$$

其中,  $k=1, 2, \dots, S; j=1, 2, \dots, Y; f_{kj}$  表示标签  $l_j$  标记项目  $s_k$  的次数。

所有项目的特征信息可以通过项目特征表示矩阵  $S_{k \times k}$  表示:

$$S_{k \times k} = \begin{bmatrix} f_{k1} & \dots & f_{kk} \\ \vdots & \ddots & \vdots \\ f_{k1} & \dots & f_{kk} \end{bmatrix} \quad (16)$$

本文采用上文介绍的皮尔森相关相似性来计算各个项目的相似性,将用户评分替换成标签的使用次数。设项目  $S_a$  和项目  $S_b$  共同使用过的标签集用  $L_{ab}$  表示, 则项目  $S_a$  和项目  $S_b$  之间的相似性  $sim(S_a, S_b)$  为:

$$sim(S_a, S_b) = \frac{\sum_{i \in L_{ab}} (R_{a,i} - \bar{R}_a) \times (R_{b,i} - \bar{R}_b)}{\sqrt{\sum_{i \in L_{ab}} (R_{a,i} - \bar{R}_a)^2} \times \sqrt{\sum_{i \in L_{ab}} (R_{b,i} - \bar{R}_b)^2}} \quad (17)$$

其中,  $R_{a,i}$  和  $R_{b,i}$  分别表示项目  $S_a$  和项目  $S_b$  上贴有标签  $i$  的次数,  $\bar{R}_a$  和  $\bar{R}_b$  分别表示项目  $S_a$  和项目  $S_b$  贴有标签的平均次数。

通过项目间的相似性计算,构建如式(18)所示的项目相似性矩阵  $I_{ZZ}$ :

$$I_{ZZ} = \begin{bmatrix} 1 & \dots & S_{1b} & \dots & S_{1Z} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ S_{a1} & & 1 & & S_{aZ} \\ \vdots & & \vdots & \ddots & \vdots \\ S_{Z1} & \dots & S_{Zb} & \dots & S_{ZZ} \end{bmatrix} \quad (18)$$

其中,  $a=1,2,\dots,Z; b=1,2,\dots,Z; S_{a,b}$  表示项目  $S_a$  和项目  $S_b$  的相似性。

### 3.7 用户偏好预测

根据各项目的特征以及用户间行为的相似性,可以通过寻找最近邻的偏好物品预测目标用户对各项目的偏好程度。根据用户相似性得到的用户预测偏好值可用  $PU_{us}$  表示:

$$PU_{us} = \sum_{k=1}^Z U_{u_i, u_j} \times P_{u_s, s_k} \quad (19)$$

其中,  $P_{u_s, s_k}$  表示用户对历史使用项目的偏好程度,  $U_{u_i, u_j}$  表示用户  $u_i$  和用户  $u_j$  的相似性。

根据各项目之间的相似性以及用户的历史行为,可以计算目标用户对各项目的偏好程度。根据项目相似性得到的用户预测偏好值可用  $PS_{us}$  表示:

$$PS_{us} = \sum_{k=1}^Z P_{u_s, s_k} \times S_{s_i, s_j} \quad (20)$$

其中,  $P_{u_s, s_k}$  表示用户对历史使用项目的偏好程度,  $S_{s_i, s_j}$  表示项目  $s_k$  和项目  $s_i$  的相似性。

为了更好地观察结合标签信息的用户相似性和项目相似性在推荐算法中所产生的影响,定义权重  $\alpha$  和  $\beta$ ,并将二者进行线性组合,即  $\alpha+\beta=1$ 。组合后的用户偏好预测计算公式如下:

$$pre = \alpha \times PU_{us} + \beta \times PS_{us} \quad (21)$$

其基本原理如图 4 所示。



图 4 结合用户相似性和项目相似性预测用户偏好的示意图

## 4 算法设计

传统的协同过滤推荐算法在计算最近邻或者项目相似性时,局限于用户对项目的评分,并未考虑用户和项目的其他相关信息。本文将标签信息融入协同过滤推荐算法中,将构建的标签信息三维关系映射成二维矩阵,并同时考虑了用户相似性与项目相似性对预测用户偏好的影响,从而提高推荐的准确度。

### 4.1 算法描述

首先,根据用户行为确定用户、标签以及项目三者之间的关系,构建用户-标签-项目三维关系;然后,将该关系表示转化为两个二维矩阵,分析用户  $User_i$  与标签  $Tag_j$ 、标签  $Tag_j$  与项目  $Item_k$  之间的关系,从而获得用户对项目的偏好程度;接着,计算用户相似程度和项目相似度,并将二者相结合用于用户偏好的预测中;最后,为用户生成 Top-N 推荐。算法的具体描述如下。

**算法 1** 基于标签信息特征相似性的协同过滤算法

输入:用户-标签-项目数据集,Top-N 的推荐项目数 N

输出:目标用户  $user_i$  的 Top-N 推荐项目集

Step1 建立用户-标签-项目关系表示

根据用户使用标签标记项目的历史行为,确定用户、标签以及项目三者之间的关系,构建用户-标签-项目三维关系。

Step2 构建用户-标签及标签-项目矩阵

基于用户-标签-项目三维关系,分析用户的标签行为,构建用户-标签矩阵  $user\_tag$ ;分析标签-项目的标注信息,构建项目-标签矩阵  $tag\_source$ 。

Step3 构建用户偏好矩阵

通过式(7)计算用户  $User_i$  的兴趣特征表示,由式(8)计算项目  $Item_k$  的标签特征表示。在此基础上,通过式(9)计算用户对项目的偏好度,进而构建用户对项目的偏好矩阵。

Step4 计算相似性

通过用户特征信息,根据式(13)计算用户间的相似度,并构建相应的用户相似性矩阵;通过项目的特征信息,由式(17)计算出项目间的相似度,同时构建相应的项目相似矩阵。

Step5 计算预测偏好值

基于用户使用标签标记项目的行为,通过用户相似性矩阵,寻找与用户  $User_i$  具有相似行为的其他用户  $User_k$ ,并根据相似用户的偏好计算目标用户  $User_i$  对各项目的预测偏好值;通过项目相似性矩阵,寻找用户  $User_i$  已评价项目与其他项目之间的相似程度,计算用户  $User_i$  对相似项目的预测偏好值。

Step6 确定权重

考虑到用户相似性和项目相似性均会对预测用户  $User_i$  的偏好产生影响,需要通过交叉验证来确定二者对预测结果产生影响所占的比重。在此基础上,计算用户  $User_i$  对项目最终的预测偏好值。

Step7 推荐生成

将用户  $User_i$  对各项目偏好的最终预测结果从大到小排序,根据项目数 N,生成 Top-N 推荐。

算法流程图如图 5 所示。



图 5 算法流程图

### 4.2 算法分析

假设有  $m$  个用户、 $n$  个标签、 $k$  个项目,数据规模为  $s$ 。Step1 及 Step2 遍历了所有的数据,计算用户-标签使用行为及项目-标签标注信息,时间复杂度为  $O(n^3)$ ;Step3 计算用户对项目的偏好度,进而构建用户-项目偏好矩阵,时间复杂度为  $O(n^2)$ ;Step4 分别计算用户相似性以及项目相似性,时间复杂度为  $O(n^2)$ ;Step5 计算用户对项目的偏好值,时间复杂度为  $O(n^2)$ ;Step7 对  $N$  个项目进行排序,寻找预测偏好值最大的  $N$  个项目,时间复杂度为  $O(n \log n)$ 。因此,综合分析算法的整个过程,为用户生成推荐结果的时间复杂度为  $O(n)$ ,总时间复杂度为  $O(n^3)$ 。

## 5 实验比较与分析

本实验的目的主要包括以下几个方面:

1) 根据本文提出的基于标签的协同过滤个性化推荐方法, 计算出用户相似以及项目相似的最佳权重  $\alpha$  和  $\beta$ 。

2) 考虑到推荐个数  $N$  对准确度的影响, 通过实验确定使准确度最高的  $N$  值。

3) 比较在不同  $Item$  数量下本文的方法与基于用户相似的协同过滤算法以及文献[20]方法的准确度, 验证本文提出的基于标签的协同过滤个性化推荐方法能有效提高准确度。

4) 通过交叉验证的方法来验证本文所提算法的有效性, 并观察随机选取数据对推荐准确率的影响。

### 5.1 实验设置

本文所用的实验数据来自于第五届推荐系统大会上公布的 *Last.fm* 数据集。 *Last.fm* 是一个音乐网站<sup>[23]</sup>, 允许用户使用标签为音乐和歌手做标记。为了避免根据数据集生成的矩阵过于稀疏, 本文对数据集进行了筛选, 挑选出至少使用过标签标记歌曲 10 次的用户及其相应的标签和歌曲被标记的情况构成实验所用数据集。该数据集共有 168437 条用户行为记录, 包括 11982 个项目、1493 个标签以及 1260 个用户。实验中, 将数据集随机划分为 10 份, 其中 9 份作为训练集, 1 份作为测试集。根据训练集可以确定用户对各个项目的偏好, 测试集用来计算算法的精确度等指标。

### 5.2 方法比较

本文采用以下 6 种方法作为基线系统进行评估和比较分析。

1) UBCF<sup>[24-27]</sup>: 基于用户的协同过滤算法。

2) IBCF<sup>[28-29]</sup>: 基于项目的协同过滤算法。

3) UBTCF: 本文提出的结合用户标签行为数据的协同过滤方法, 寻找并确定最近邻, 并通过最近邻的行为为目标用户进行推荐。

4) IBTCF: 本文提出的将标签信息融入到基于项目的协同过滤方法中, 确定各项目之间的相似性, 然后为用户推荐与其喜欢过的项目相似的项目。

5) ITCF<sup>[20]</sup>: 该方法将用户评分融入到用户对标签权重的计算中, 考虑用户不同兴趣程度的项目对推荐结果的影响。

6) TFSCF: 该方法为本文提出的方法。通过一定的权重值将用户相似性和项目相似性相结合, 从而为目标用户做出推荐。

### 5.3 评估标准

准确性指标是推荐系统中最重要的指标。最常使用的准确性指标包括准确率(Precision)和召回率(Recall)。准确率和召回率是信息检索和统计学分类领域的两个常用度量值, 用来评价结果的质量。其中, 准确率是检索出的相关文档数与检索出的总文档数的比率, 衡量的是检索系统的查准率; 召回率是指检索出的相关文档数与文档库中所有的相关文档数的比率, 衡量的是检索系统的查全率。假设  $D(s)$  是根据用户在训练集中的行为数据得出的推荐列表,  $T(s)$  是根据用户在测试集中的行为数据得出的推荐列表, 则推荐结果的准确率可由式(22)计算得到:

$$Precision = \frac{\sum_{i \in U} |D(s) \cap T(s)|}{\sum_{i \in U} |D(s)|} \quad (22)$$

推荐结果的召回率可由式(23)计算得到:

$$Recall = \frac{\sum_{i \in U} |D(s) \cap T(s)|}{\sum_{i \in U} |T(s)|} \quad (23)$$

Precision 和 Recall 是两个相互制约的指标, 因此需要综合考虑, 最常见的方法是 F-Measure (又称为 F-Score)。F-Measure 是 Precision 和 Recall 的加权调和平均:

$$F_\alpha = \frac{(\alpha^2 + 1) \times Precision \times Recall}{\alpha^2 \times Precision + Recall} \quad (24)$$

其中,  $\alpha$  是参数, 当  $\alpha=1$  时, 即为最常见的 F1-Measure。

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (25)$$

### 5.4 实验结果及分析

为了验证和比较本文所提方法的有效性, 进行以下几组实验。

1) 实验一: 该组实验主要用于确定用户相似性所占权重值  $\alpha$  以及项目相似性所占权重值  $\beta$ 。针对 3.4 节中提出的基于用户相似性以及项目相似性的用户偏好预测方法, 同时考虑式(21)中权重值  $\alpha$  和  $\beta$  对用户偏好预测的影响, 本实验取  $\alpha$  为 0.1~0.9, 每次间隔 0.1,  $\beta$  取  $1-\alpha$ , 对应于 0.9 到 0.1。依次观察  $\alpha$  以及  $\beta$  的变化对准确度的影响, 最终选出最佳的  $\alpha$  和  $\beta$  取值。实验结果如图 6 所示。

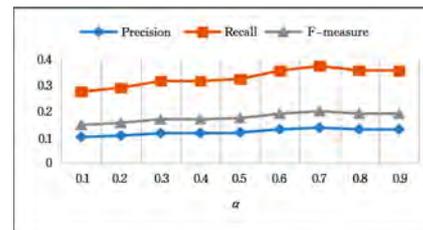


图 6 权重值  $\alpha$  和  $\beta$  对准确度的影响

由图 6 可以看出, Precision 和 Recall 以及 F-Measure 随着所取  $\alpha$  的增大即  $\beta$  的降低而升高, 当  $\alpha$  取值为 0.7, 对应  $\beta$  取值为 0.3 时, 推荐算法的精度最高, 性能最好;  $\alpha$  的值继续增大时, 3 种性能指标再次降低。因此, 本文的实验中,  $\alpha$  取值为 0.7,  $\beta$  取值为 0.3。

2) 实验二: 在进行 Top-N 推荐时,  $N$  的数量对推荐的结果同样具有很大的影响, 因此该组实验主要比较本文的 TF-SCF 方法在不同  $N$  值下 Precision, Recall 以及 F-measure 的变化情况。实验中,  $N$  的数量由 5 到 30 依次递增, 间隔为 5。实验结果如图 7 所示。

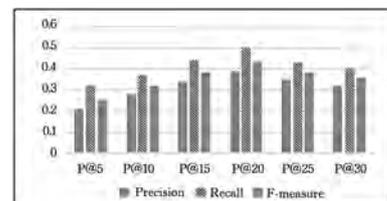


图 7 不同  $N$  数量下, TFSCF 推荐性能的变化

由图 7 可以看出, 当推荐个数为 5 时, Precision 为 0.20, Recall 为 0.33, F-measure 值为 0.25;  $N$  为 10 时, Precision 为 0.28; 随着推荐数量的不断增加, 推荐算法的效果不断升高, 直到  $N$  的数量等于 20 时, 推荐准确度为 0.38, 达到峰值, 当  $N$  值继续增加时, 推荐效果随着推荐数量的增加而降低。因此, 当  $N$  的数量为 20 时, 推荐效果最好。

3) 实验三: 由于在实验中, Item 的个数对推荐的结果具

有较大的影响,因此该组实验主要比较 TFSCF 方法、基于用户的协同过滤算法 UBCF、基于项目的协同过滤算法 IBCF、UBTCF 算法、IBTCF 算法以及文献[20]提出的 ITCF 算法在 Item 取 300,500,800,1300 等不同值时 *Precision*,*Recall* 以及 *F-measure* 的变化。实验结果如图 8—图 10 所示。

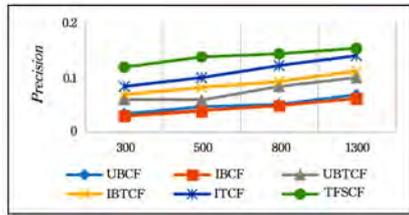


图 8 不同 Item 数量下,*Precision* 的变化情况

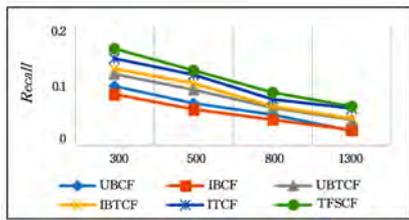


图 9 不同 Item 数量下,*Recall* 的变化情况

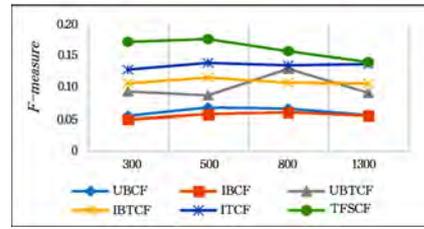


图 10 不同 Item 数量下,*F-measure* 的变化情况

由图 8—图 10 可以看出,随着 Item 个数的逐渐增加,各方法的 *Precision* 均逐渐提高,推荐效果越来越好。而 *Recall* 随着 Item 个数的增加逐渐降低。本文所提方法的 *Precision*,*Recall* 和 *F-measure* 均高于其他方法,说明其确实能够提高推荐的性能。

4)实验四:为了进一步验证本文提出推荐算法的推荐准确率不会受到数据选取随机性的影响,该组实验进行了数据的交叉验证。将数据集随机地分成 10 份,其中 9 份作为训练集,1 份作为测试集;得出实验结果之后,将未做过测试集的 1 份数据作为测试集,剩余 9 份作为训练集。观察 10 次结果中 *Precision*,*Recall* 以及 *F-measure* 的变化情况,实验结果如表 2 所列。

表 2 交叉验证中的相关性能指标

	300@1	300@2	300@3	300@4	300@5	300@6	300@7	300@8	300@9	300@10
<i>Precision</i>	0.11938	0.11938	0.11627	0.13278	0.12895	0.11268	0.12656	0.11507	0.11938	0.12344
<i>Recall</i>	0.32807	0.32807	0.31953	0.36489	0.35437	0.30966	0.34780	0.31623	0.32807	0.33925
<i>F-measure</i>	0.17506	0.17506	0.17050	0.19470	0.18909	0.16523	0.18558	0.16874	0.17506	0.18102

由表 2 可以看出, *Precision*,*Recall* 以及 *F-measure* 的波动范围在 0~0.02 之间,处于可以接受的范围,说明本文提出的基于用户相似性以及项目相似性的用户偏好预测算法的推荐效果并没有因数据选择的随机性而受到太大的影响。

**结束语** 标签作为 Web 2.0 时代在社会化网络中的重要应用,体现了用户对资源的理解,用户可以根据自己的兴趣偏好为物品等资源打上标签。本文根据自由标注的社会化标签以及协同过滤的思想,提出一种针对用户个体的个性化需求推荐算法。该算法使用标签体现资源特征以及用户兴趣偏好,并充分考虑用户的历史行为数据,首次提出将项目相似性以及用户相似性结合,使得推荐结果更加符合实际情况。通过实验和比较,验证了本文提出的方法能够提高推荐的准确性,在各种性能指标上优于其他方法。

未来我们将继续对本文中的算法进行优化,重点可以分析数据的增量式更新以及用户推荐的实时性、标签的相似性以及时效性等问题。

参 考 文 献

[1] ADOMAVICIUS G, TUZHILIN A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions[C]// Proceedings of the IEEE Transactions Knowledge and Data Engineering, 2005, 17(6): 734-749.  
 [2] KATARYA R, VERMA O P. Privacy-Preserving and Secure Recommender System Enhance with K-NN and Social Tagging [C]// 2017 IEEE 4th International Conference on Cyber Security and Cloud Computing (CSCloud). New York, 2017: 52-57.  
 [3] HUANG L W, LI D Y. A review of information recommendation in social media[J]. CAAI Trans. on Intelligent Systems, 2012, 7(1): 1-8.

[4] SU X, KHOSHGOFTAAR T M. A survey of collaborative filtering techniques [J]. Advances in Artificial Intelligence, 2009, 2009(12): 4.  
 [5] ORTEGA F, HERNANDO A, BOBADILLA J, et al. Recommending items to group of users using Matrix Factorization based Collaborative Filtering [J]. Information Sciences, 2016, 345(C): 313-324.  
 [6] LIN J, SUGIYAMA K, KAN M Y, et al. Addressing cold-start in app recommendation: latent user models constructed from twitter followers[C]// 36th International ACM SIGIR Conference on Research and Development in Information Retrieval. Dublin, Ireland, 2013: 283-292.  
 [7] WANG L C, MENG X W, ZHANG Y J. Context-Aware recommender systems: A survey of the state-of-the-art and possible extensions[J]. Journal of Software, 2012, 23(1): 1-20.  
 [8] SYMEONIDIS P. ClustHOSVD: Item Recommendation by Combining Semantically Enhanced Tag Clustering With Tensor HOSVD[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2016, 46(9): 1240-1251.  
 [9] MISTRY O, SEN S. Tag recommendation for social bookmarking: Probabilistic approaches [J]. Multiagent and Grid Systems, 2012, 8(2): 143-163.  
 [10] 于洪, 李俊华. 一种解决新项目冷启动问题的推荐算法[J]. 软件学报, 2015, 26(6): 1395-1408.  
 [11] ZHANG Z K, LIU C, ZHANG Y C, et al. Solving the cold-start problem in recommender systems with social tags [J]. EPL (Europhysics Letters), 2010, 92(2): 28002.  
 [12] ZHANG Z K, ZHOU T, ZHANG Y C. Tag-Aware recommender systems: A state-of-the-art survey [J]. Journal of Computer Science and Technology, 2011, 26(5): 767-777.

- [13] ZHANG Z K, ZHOU T, ZHANG Y C. Personalized recommendation via integrated diffusion on user-item-tag tripartite graphs [J]. *Physica A: Statistical Mechanics and its Applications*, 2010, 389(1):179-186.
- [14] 刘建勋, 石敏, 周栋, 等. 基于主题模型的 Mashup 标签推荐方法 [J]. *计算机学报*, 2017, 40(2):520-534.
- [15] 李锡荣, 许洁萍, 薛盛博, 等. 基于软近邻投票的图像标签相关性计算 [J]. *计算机学报*, 2014, 37(6):1365-1371.
- [16] 张斌, 张引, 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.
- [17] YANG S, LU Z, GILES C L. Automatic tag recommendation algorithms for social recommender systems [J]. *ACM Transactions on the Web*, 2011, 5(1):1-31.
- [18] 孔欣欣, 苏本昌, 王宏志, 等. 基于标签权重评分的推荐模型及算法研究 [J]. *计算机学报*, 2017, 40(6):1440-1452.
- [19] JOMSRI P, SANGUANSINTUKUL S, CHOOCHAIWATTANA W. A framework for tag-based research paper recommender system: An IR approach [C] // *Proceedings of the 2010 IEEE 24th Int'l Conf. on Advanced Information Networking and Applications Workshops*. 2010:103-108.
- [20] 郭彩云, 王会进. 改进的基于标签的协同过滤算法 [J]. *计算机工程与应用*, 2016, 52(8):56-61, 147.
- [21] 李慧, 马小平, 胡云, 等. 融合主题与语言模型的个性化标签推荐方法研究 [J]. *计算机科学*, 2015, 42(8):70-74.
- [22] AR Y, BOSTANCI E. A genetic algorithm solution to the collaborative filtering problem [J]. *Expert Systems with Applications*, 2016, 61:122-128.
- [23] 李瑞敏, 林鸿飞, 闫俊. 基于用户-标签-项目语义挖掘的个性化音乐推荐 [J]. *计算机研究与发展*, 2014, 51(10):2270-2276.
- [24] BREESE J S, HECKERMAN D, KADIE C. Empirical analysis of predictive algorithms for collaborative filtering [C] // *Proc. of the 14th Conference on Uncertainty in Artificial Intelligence*. Madison, Wisconsin, USA, 1998:43-52.
- [25] HERLOCKER J L, KONSTAN J A, BORCHERS A, et al. An algorithmic framework for performing collaborative filtering [C] // *22th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. Berkeley, CA, USA, 1999:230-237.
- [26] JIN R, CHAI J Y, SI L. An automatic weighting scheme for collaborative filtering [C] // *27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. Sheffield, UK, 2004:337-344.
- [27] RESNICK P, IACOVOU N, SUCHAK M, et al. An open architecture for collaborative filtering of netnews [C] // *1994 ACM Conference on Computer Supported Cooperative Work*. Chapel Hill, NC, USA, 1994:175-186.
- [28] SARWAR B, KARYPIS G, KONSTAN J, et al. Item-based collaborative filtering recommendation algorithms [C] // *10th International Conference on World Wide Web*. Hong Kong, China, 2001:285-295.
- [29] DESHPANDE M, KARYPIS G. Item-based top-n recommendation algorithms [J]. *ACM Transactions on Information System*, 2004, 22(1):143-177.

(上接第 401 页)

**结束语** 本文针对传统协同过滤算法中用户相似性计算不准确和推荐准确率偏低的问题, 将时间特征、评分差异、项目属性综合考虑到相似性的计算方法中, 然后再将改进的相似性计算方法融入到基于用户的协同过滤算法中, 提出了一种基于评分差异和用户兴趣相似度的协同过滤算法。在计算用户相似性时, 考虑到用户的兴趣随时间变化, 把时间因素融合到基于改进信息熵的相似性计算方法中, 评分项目属性也影响着用户的兴趣相似性, 因此综合两部分的相似性计算方法得出一种新的用户相似性计算方法。实验结果表明, 该相似性计算方法更能准确地找出目标用户的邻居集用户, 有效地提高了推荐系统的准确率。

用户兴趣可能随时间和情景而发生改变, 因此, 在下一步的工作中, 将建立情景模型以反映用户兴趣变化的因素, 并将其加入到用户相似性的计算方法中, 以更真实地描述用户兴趣和查找用户的最近邻居。

### 参考文献

- [1] CAI Y, LEUNG H, LI Q, et al. Typicality-based collaborative filtering recommendation [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2014, 26(3):766-779.
- [2] KALELI C. An entropy-based neighbor selection approach for collaborative filtering [J]. *Knowledge-Based Systems*, 2014, 56(C):273-280.
- [3] 汪静, 印鉴, 郑利荣, 等. 基于共同评分和相似性权重的协同过滤推荐算法 [J]. *计算机科学*, 2010, 37(2):99-104.
- [4] BOBADILLA J S, ORTEGA F, HERNANDO A, et al. A collaborative filtering approach to mitigate the new user cold start problem [J]. *Knowledge-Based Systems*, 2012, 26:225-238.
- [5] JIA D, ZHANG F, LIU S. A robust collaborative filtering recommendation algorithm based on multidimensional trust model [J]. *Journal of Software*, 2013, 8(1):11-18.
- [6] JU C, XU C. A new collaborative recommendation approach based on users clustering using artificial bee colony algorithm [J]. *The Scientific World Journal*, 2013, 2013(3):869658.
- [7] 陈志敏, 李志强. 基于用户特征和项目属性的协同过滤推荐算法 [J]. *计算机应用*, 2011, 31(7):1748-1750.
- [8] HUANG M, SUN L, DU W. Collaborative filtering recommendation algorithm based on item attributes [C] // *2014 15th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*. IEEE, 2014:101-106.
- [9] ZHANG J, PENG Q, SUN S, et al. Collaborative filtering recommendation algorithm based on user preference derived from item domain features [J]. *Physica A: Statistical Mechanics and its Applications*, 2014, 396(2):66-76.
- [10] PIAO C H, ZHAO J, ZHENG L J. Research on entropy-based collaborative filtering algorithm and personalized recommendation in e-commerce [J]. *Service Oriented Computing and Applications*, 2009, 3(2):147-157.