

# 基于标准标签的用户兴趣模型研究

杨 晶<sup>1</sup>, 成卫青<sup>1</sup>, 郭常忠<sup>2</sup>

(1. 南京邮电大学 计算机学院, 江苏 南京 210003;  
2. 烟台大学 数学与信息科学学院, 山东 烟台 264005)

**摘 要:**信息大爆炸的网络时代,个性化推荐是解决信息“超负载”的有效办法。用户兴趣模型是个性化推荐的核心,关系着整个推荐系统的推荐质量。标签一直被用于资源分类,在个性化推荐方面却很少使用。文中采取向量空间模型的建模方法,利用个性化标签描述用户兴趣,并提出一套简洁有效的标签标准化方法—基于属性共现率的标签标准化以及基于聚类的标签标准化方法对用户的自定义标签进行标准化。该模型能有效降低用户兴趣模型的向量维数,避免分析标签语义的复杂过程,且能够从用户的角度贴切地表达用户兴趣。实验结果表明该模型有助于提高个性化推荐的服务质量。

**关键词:**个性化推荐;用户兴趣模型;向量空间模型;标准标签

**中图分类号:**TP39

**文献标识码:**A

**文章编号:**1673-629X(2013)10-0208-04

doi:10.3969/j.issn.1673-629X.2013.10.052

## Research on User Interest Model Based on Standard Tag

YANG Jing<sup>1</sup>, CHENG Wei-qing<sup>1</sup>, GUO Chang-zhong<sup>2</sup>

(1. College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210003, China;  
2. College of Mathematics and Information Science, Yantai University, Yantai 264005, China)

**Abstract:** Faced to the Internet age of information explosion, the personalized recommendation is an effective way to solve the “information overload”. User interest model as the core of personalized recommendation determines the quality of the recommendation system. Tags have been used for the classification of resources; however, they are seldom used in personalized recommendation. In this paper, vector space model is used in modeling, where personalized tags are used to describe user interests. A set of simple and effective methods are proposed to standardize user's custom tags, including a standardization method based on attribute co-occurrence frequency and a standardization method based on clustering. Thus, the vector dimension of the user interest model can be reduced effectively, avoiding complex tag semantic analysis, as well as being able to aptly express user's interests from their point of view. The experimental results show that the proposed user interest model can help to improve the quality of personalized recommendation.

**Key words:** personalized recommendation; user interest model; vector space model; standard tag

## 0 引言

个性化推荐系统能有效地缓解网站的信息的“超负载”现象<sup>[1-2]</sup>。个性化推荐研究利用用户在系统中预设的参数<sup>[3]</sup>或者网页的访问记录作为数据源,通过数据挖掘<sup>[4]</sup>等办法获取用户的兴趣偏好。标签常被用于资源分类,可以表现用户的兴趣爱好,但是语义模糊问题使得标签在个性化推荐中的使用并不多。文中采用向量空间模型表示法建立基于标签的用户兴趣模型,提出一套简洁有效的标签标准化方法将用户自定义标签标准化,标准化过程结束以后利用标准标签作

为向量空间模型表示法的关键词,进而完成用户兴趣建模。

## 1 基于标准化标签的用户兴趣建模

个性化推荐的成效很大程度上依赖于用户兴趣模型,兴趣模型的好坏直接关系着系统的推荐质量。目前比较常用的用户模型表示法有:主题表示法、关键词列表的表示法、基于神经网络的表示法、基于本体论(Ontology)的表示法和基于向量空间模型的表示法<sup>[5]</sup>。

收稿日期:2012-12-25

修回日期:2013-03-28

网络出版时间:2013-07-24

**基金项目:**国家自然科学基金资助项目(61170322, 71171117);软件开发环境国家重点实验室开放课题(SKLSDE-2011KF-0X);江苏省自然科学基金资助项目(BK2010524)

**作者简介:**杨 晶(1987-),女,硕士生,CCF会员,研究方向为个性化推荐;成卫青,副教授,工学博士,CCF会员,研究方向为网络测量。

**网络出版地址:**http://www.cnki.net/kcms/detail/61.1450.TP.20130724.1005.035.html

向量空间模型 (Vector Space Model, VSM) 表示法是目前为止使用较多且效果较好的特征表示法<sup>[6]</sup>, 最早用在 SMART 信息检索系统中, 目前已成为自然语言处理中常用的模型<sup>[7]</sup>。它把用户的兴趣模型表示成一个  $n$  维的特征向量  $\{(T_1, W_1), (T_2, W_2), \dots, (T_n, W_n)\}$ , 用以表示用户感兴趣的方面以及对这一方面的感兴趣程度。其中,  $T_i$  表示兴趣特征项;  $W_i$  是特征项  $T_i$  的权重。

定义两个建立模型需要用到的概念:

定义 1 标准标签: 人为制定且语义公认明确的标签, 各标签之间的相关度为 0 或很小。由标准标签组成的集合称为标准标签库。

定义 2 词语共现率: 在若干字符串中出现同一个词的概率, 以下简称为共现率。

文中提出一套标签标准化方法将用户的自定义标签映射到标准标签, 利用基于向量空间模型的表示法建立用户兴趣模型。建模方法如下:

Step1: 将用户的自定义标签进行标签标准化, 与系统可理解的标准标签建立映射关系。

Step2: 利用 TF-IDF 方法计算标准标签的权重。

Step3: 用向量空间模型表示法表示模型, 用标准标签作为兴趣关键词, 结合 Step2 中计算出的标签权值表示用户的兴趣模型。

## 1.1 标签标准化

标签标准化能够有效地缓解语义模糊、歧义等问题。文中为了提高标签标准化效率, 提出一组标签标准化方法——基于属性共现率的标签标准化和基于聚类的标签标准化。二者各有所长, 可以在较大程度弥补彼此的不足。标签标准化过程如图 1 所示。

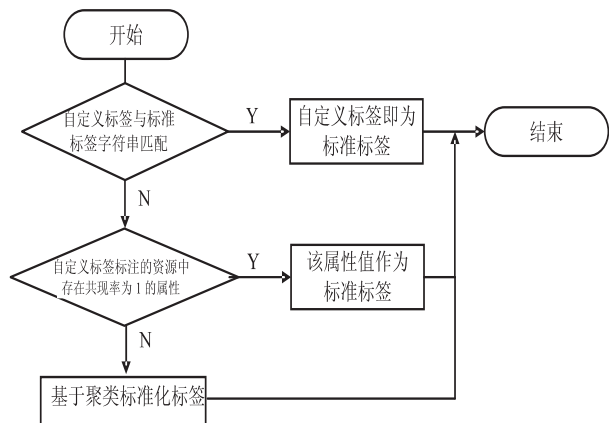


图 1 标签标准化过程

### 1.1.1 基于属性共现率的标签标准化

基于属性共现率的标签标准化利用被某一标签标注的资源集中资源属性的共性来表示这一标签的真正含义, 属性值可以作为该自定义标签的标准标签的条件是, 该属性值在资源集中的共现率为 100%。

基于属性共现率的标签标准化不仅可以在一定程度上减少标签的总量、解决用户标注的偏差问题, 同时也避免了基于聚类的标签标准化过程的一些缺陷:

(1) 避免两个含义相同或相近的标签在标注完全不同资源的情况下不能有效聚类的弊端。

(2) 减少需要聚类的标签数量。

设待标准化标签为  $t$ ,  $t$  标注的资源集合  $P_t = \{p_1, p_2, \dots, p_l\}$ , 对集合  $P_t$  中的资源做共性分析, 若某一属性值共现率为 1 则将该属性值作为  $t$  的标准标签。若该属性值不在标准标签库中, 则将其存入标准标签词库以扩充词库。若待标准化标签  $t$  经过此方法标准化失败, 则利用下一方法——基于聚类的标签标准化对  $t$  进行标准化。

### 1.1.2 基于聚类的标签标准化

基于聚类<sup>[8]</sup>的标签标准化是标签标准化的第三步, 其主要作用是将第二步标准化失败的自定义标签进行聚类以使全部自定义标签得以标准化。同时, 聚集同一类型的标签可以更准确地反应用户在一些资源上的兴趣度<sup>[4]</sup>。关于标签聚类, 文献[9]是利用 LSA 和 SVD 方法对标签-资源矩阵进行分析, 指出标签之间存在的相似性, 并利用 SOM 方法对标签进行聚类。文献[10]则是运用马尔可夫聚类算法对标签共现网络进行分析, 进而实现对标签的聚类。以上聚类方法虽有一定的效果, 但由于算法本身的复杂度高, 且使用以上算法时需要事先对标签做格式处理, 加大了标签聚类的难度, 也延长了聚类时间。涉及文中标签聚类的定义如下:

定义 3 标签  $t_i$  和  $t_j$  的互信息  $I(t_i; t_j)$  以及熵  $H(t_i)$  定义如下:

$$I(t_i; t_j) = p(t_i, t_j) \log \frac{p(t_i, t_j)}{p(t_i)p(t_j)} \quad (1)$$

$$H(t_i) = -p(t_i) \log p(t_i) \quad (2)$$

其中,  $p(t_i, t_j)$  是指  $t_i$  和  $t_j$  标记在同一资源上的概率, 即:  $p(t_i, t_j) = \frac{|P_{t_i} \cap P_{t_j}|}{|\bigcup_{k=1}^n P_{t_k}|}$  ( $P_{t_i}$  和  $P_{t_j}$  分别指标签  $t_i$  和  $t_j$  标注的资源集合,  $n$  为标签总数, 资源总数大于 1);  $p(t_i)$ ,  $p(t_j)$  分别代表标签  $t_i$ ,  $t_j$  标记资源的概率, 即:  $p(t_i) = \frac{|P_{t_i}|}{|\bigcup_{k=1}^n P_{t_k}|}$ ,  $p(t_j) = \frac{|P_{t_j}|}{|\bigcup_{k=1}^n P_{t_k}|}$ 。

当  $t_i$  和  $t_j$  相关时,  $I(t_i; t_j) > 0$ ; 当  $t_i$  和  $t_j$  完全不相干时,  $I(t_i; t_j) = 0$ 。

定义 4 标签  $t_i$  和  $t_j$  的相关度  $r_{ij}$  定义如下:

$$r_{ij} = \frac{I(t_i; t_j)}{\frac{H(t_i) + H(t_j)}{2}} \quad (3)$$

当  $t_i$  和  $t_j$  标注的资源完全相同时,  $r_{ij} = 1$ , 标签  $t_i$  和  $t_j$  完全相关; 当  $t_i$  和  $t_j$  标注的资源完全不同时,  $r_{ij} = 0$ , 标签  $t_i$  和  $t_j$  完全不相相关; 其他情况下  $0 < r_{ij} < 1$ 。

定义 5 标签相关矩阵  $R$  定义为:

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix} = (r_{ij})_{m \times n} \quad (4)$$

其中,  $n$  为标准标签个数;  $m$  为经过方法一和方法二标准化失败的用户自定义标签 (以下简称自定义标签) 个数;  $r_{ij}$  为标签相关度, 见式 (3)。

文中利用模糊聚类<sup>[11]</sup> 的思想提出基于标准化的标签聚类。过程描述如下: 首先, 把标准标签作为聚类中心, 计算自定义标签与聚类中心的相关度, 将自定义标签聚集到相关度最大的聚类中心所属的类中。直到剩下的自定义标签与聚类中心的相关度均小于一个阈值  $\delta$  (文中相关度阈值  $\delta = 0.21$ ) 时, 停止聚类。将剩下的自定义标签互相聚类, 并将新的聚类中心存入到标准标签词库。再重新计算自定义标签与标准标签的相关度, 调整自定义标签的聚类, 得到自定义标签的标准标签。

基于标准化的标签聚类算法:

设标准标签集合  $T_s = \{t_{s1}, t_{s2}, \dots, t_{sn}\}$ ; 自定义标签集合  $T_p = \{t_{p1}, t_{p2}, \dots, t_{pm}\}$ ; 阈值  $\delta$ , 当标签间的相关度小于  $\delta$  时, 认为标签之间不相关。

a) 将  $T_s$  中的标签作为聚类中心, 按照式 (3) 计算集合  $T_s$  和  $T_p$  中所有元素之间的相关度。

b) 利用式 (4) 计算标签相关度矩阵  $R$ 。

c) 选出  $R$  中最大的元素  $r_{ij}$ , 若其大于等于阈值  $\delta$ , 将标签  $t_{pi}$  聚集到以标签  $t_{sj}$  为中心的类中, 删除矩阵  $R$  的第  $i$  行, 得到  $(m-1) \times n$  维矩阵  $R_{(1)}$ 。

d) 从矩阵  $R_{(1)}$  开始, 重复执行过程 c), 直到矩阵中最大的元素小于阈值  $\delta$ , 聚类停止, 此时得到的类聚集结果为  $C_{(1)} = \{c_1, c_2, \dots, c_n\}$ 。

e) 将剩余的与聚类中心相关度小于  $\delta$  的自定义标签  $T_{PL} = \{t_{pk}, t_{pq}, \dots, t_{pw}\}$  聚类, 集合  $T_{PL}$  的聚类计算方法如下:

① 以  $T_{PL}$  自身作为聚类中心, 按照式 (3) 计算集合  $T_{PL}$  中所有标签之间的相关度, 然后执行步骤 b) 计算相关矩阵  $R_L$ 。

② 找出主对角线以上 (或以下) 的最大元素  $r_{sl}$ , 以标记资源的数量较大的标签  $t_{ps}$  作为聚类中心, 将标记资源数量较小的标签  $t_{pl}$  聚集到以  $t_{ps}$  为中心的类中, 同时删除矩阵  $R_L$  的第  $l$  行和第  $l$  列。

③ 从矩阵  $R_L$  开始, 重复执行 ②, 直到矩阵  $R_L$  中最

大的元素小于阈值  $\delta$ , 聚类停止, 此时得到的类聚集结果为  $C_{(2)} = \{c_{n+1}, c_{n+2}, \dots, c_{n+h}\}$ 。

f) 调整聚类结果: 令  $T_{ps} = T_p - T_{PL}$ , 以  $C_{(2)}$  作为新的聚类中心, 按照式 (3) 计算集合  $T_{ps}$  中元素与聚类中心  $C_{(2)}$  的相关度, 调整集合  $T_{ps}$  中的元素聚类结果。

最后获得的聚类结果为:  $C = C_{(1)} + C_{(2)} = \{c_1, c_2, \dots, c_n, c_{n+1}, \dots, c_{n+h}\}$ 。聚类结束。

## 1.2 标签权重的计算

计算标签  $t$  的权重的方法有很多, 文中使用 TF-IDF<sup>[7]</sup> 方法加权方案。

$$w(t) = \frac{tf_i \times \log(N/n + 0.01)}{\sqrt{\sum_{i \in D_i} [tf_i \times \log(N/n + 0.01)]^2}} \quad (5)$$

其中,  $tf_i$  表示用户  $u_i$  使用标签  $t$  的次数;  $N$  为用户集合  $U$  中用户的总数;  $n$  为用户集合中使用过标签  $t$  的用户数;  $D_i$  表示用户  $u_i$  的标签集合。式 (5) 的取值范围为  $[0, 1]$ 。

用户的兴趣模型表示为:  $\{(t_1, w(t_1)), (t_2, w(t_2)), \dots, (t_n, w(t_n))\}$ 。

## 2 实验设计与结果分析

### 2.1 实验数据

豆瓣 (www.douban.com) 是一个 web2.0 网站, 所有的内容、分类、筛选、排序都是由用户产生或决定, 是个性化推荐研究方向采用的实验数据源之一<sup>[12]</sup>。文中以 HTMLParser 为基础架构, 通过研究豆瓣电影站点 (<http://movie.douban.com/>) 的结构, 创建该站点数据的工程。该工程功能如下: 首先通过分析豆瓣的群组用户获取一系列的用户 id, 然后以用户 id 为参数, 通过豆瓣公开的 API 获取用户的标签以及标签标注的电影集合。

文中选取两个数据集合  $D_1$  和  $D_2$  作为实验对象, 为了保证实验的结果不受多余因素的影响, 每个数据集都包含 10 个用户及每个用户标注的 100 个电影, 且均为从豆瓣数据集中随机选取的, 不同的只是实验数据的筛选条件。

数据集  $D_1$  中的用户及电影需满足以下条件:

- (1) 每个用户标注的电影总数大于 100。
- (2) 用户  $u_i$  标注的电影  $Item_{ij}$  使用的标签数大于等于 5。
- (3) 电影  $Item_{ij}$  至少被豆瓣数据集中的 100 个人标记过。

(4) 按照电影被标注的时间顺序从每个用户的标注电影集合中分别选取前 100 部电影。

数据集  $D_2$  包含不同于数据集  $D_1$  的 10 个用户, 用户及电影需满足以下条件:



- (1) 每个用户标注的电影总数大于 100。
- (2) 按照电影被标注的时间顺序从每个用户的标注电影集合中分别选取前 100 个电影。

2.2 实验方法及实验结果

文中以豆瓣电影的热门标签作为标准标签的数据源,经过人为筛选后建立标准标签库。设用户  $u$  的兴趣模型为  $\{(t_{u1},w(t_{u1})), (t_{u2},w(t_{u2})),\cdots, (t_{um},w(t_{um}))\}$ ,模型中的标准标签集合  $T_u = \{t_{u1},t_{u2},\cdots,t_{um}\}$ ;标注电影  $p$  的标准标签集合为  $T_p = \{t_{p1},t_{p2},\cdots,t_{pm}\}$ ;支持数  $\text{Sup} = |T_p \cap T_u|$ ,即每个电影的标准标签与用户模型中标准标签的重合数。为了检验模型的有效性以及电影的标注次数与模型推荐质量之间的关系,实验过程中分别选取了数据集  $D_1$  和  $D_2$  中电影的 top10 以及 top30 标签两组数据做对比,进而检验标注次数对模型质量的影响。由于实验环境限制,仅用查全率( $R$ )来衡量模型的质量。

查全率( $R$ ) = 
$$\frac{\text{信息源中利用模型向用户推荐的电影数}}{\text{信息源中用户标注过的全部电影数}}$$
由查全率公式可知,该值越接近 1,代表模型质量越好;反之,越接近 0,代表模型质量越差。

- 实验方法如下:
- (1) 对数据集  $D_1$  和  $D_2$  中每个用户,分别选择前 60 个电影做训练集。在标准标签库的基础上,利用文中的建模方法对  $D_1$  和  $D_2$  中 20 个用户分别建立用户兴趣模型。
- (2) 用每个用户余下的 40 个电影做测试集,分别选取每个电影的 top10 和 top30 (若不足,则全选) 标准标签来进行仿真实验。

在数据集  $D_1$  中,当  $\text{Sup} \geq 6$  时,此电影推荐给用户;在数据集  $D_2$  中,当  $\text{Sup} \geq 3$  时,则此电影推荐给用户。实验结果如表 1 所示。

表 1 平均查全率

选取的标签数	$D_1$ 平均查全率	$D_2$ 平均查全率
Top10	0.868	0.775
Top30	0.889	0.792

2.3 实验结果分析

实验结果表明,文中提出的基于标准标签的用户兴趣模型具有可行性。从数据集  $D_2$  的实验结果中可以看出,查全率并不是很高,分析结果发现, $D_2$  中某个电影被用户标注的次数很少,甚至只被标注过一次。导致利用  $\text{Sup}$  值进行过滤时,将此电影过滤掉进而无法推荐给用户,可以通过调整  $\text{Sup}$  值的大小来改善这种情况;另外,通过对  $D_1$  和  $D_2$  的实验结果对比可以看出,一个电影被标注的次数越多,越能体现这个电影本身的特征,查全率也就越高,进而推荐的准确率和个性

化推荐质量也就越高。

3 结束语

随着 web3.0 的应用,标签技术也不断发展,利用标签技术提供良好的个性化服务越来越受到学者们的关注,然而标签的语义问题一直是标签技术发展的瓶颈。文中提出的一组标签标准化方法——基于属性共现率的标签标准化和基于聚类的标签标准化有效地避免了复杂的语义处理,且两个标准化方法互相弥补、互相协作共同完成对用户自定义标签的标准化。

文中基于向量空间模型建立用户的兴趣模型,利用标准标签作为兴趣项,根据传统的权值计算 TF-IDF 方法计算每个兴趣项的权重,进而完成用户兴趣建模。实验结果表明,文中提出的标准化方法为建立准确的用户兴趣模型打下了坚实的基础,用户兴趣模型效果较好。

当然,文中还存在着一些有待继续研究的问题,例如, $\text{Sup}$  值作为文中衡量推荐的标准简单易懂,但是同时也存在考虑因素不全等问题,因此,研究适用于该模型的推荐算法势在必行;另外,标签是用来显式地表示用户的兴趣爱好,且一旦用户建立标签,就说明用户已经对其标注的内容形成一种长期的稳定兴趣。然而用户的兴趣是复杂多变的,只根据用户的稳定兴趣进行推荐是不全面的,如何分析用户的行为挖掘其隐性的兴趣爱好,并将用户行为产生的即时兴趣与稳定兴趣结合起来建立用户的兴趣模型,以及研究适用于该模型的推荐算法将是下一步的研究重点。

参考文献:

[1] Lee Jong-Seok,Jun Chi-Hyuck, Lee Jaewook, et al. Classification-based collaborative filtering using market basket data [J]. Expert Systems with Applications, 2005, 29 (3) : 700 - 704.

[2] Ahn A J. A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem [J]. Information Sciences, 2008, 178 (1) : 37-51.

[3] Miller B N, Albert I, Lam S K, et al. MovieLens unplugged: Experiences with an occasionally connected recommendation system [C]//Proc of Int'l Conf on Intelligent User Interfaces. New York, USA: [s. n. ], 2003.

[4] Xia Xiufeng, Zhang Shu, Li Xiaoming. A personalized recommendation model based on social tags [C]//Proc of DBTA. [s. l. ]: [s. n. ], 2010: 1-5.

[5] 林霜梅,汪更生,陈弈秋. 个性化推荐系统中的用户建模及特征选择[J]. 计算机工程, 2007, 33 (17) : 196-198.

[6] 赵银春,付关友,朱征宇. 基于 Web 浏览内容和行为相结合的用户兴趣挖掘 [J]. 计算机工程, 2005, 31 (12) : 93-94.

(下转第 215 页)

本次实验的隧道并不存在异常事件。

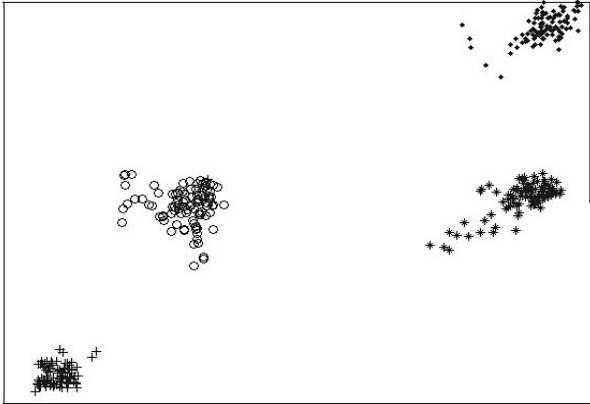


图 3 ECRT 优化算法输出

3.2 算法性能分析

为了验证该算法的性能,笔者在相同环境下,分别使用关联规则算法和文中算法同时对上文实验数据进行了比对,其算法误差精度如图 4 所示,可见文中算法具有更低的误差,尤其是在大量数据的情况下,其精确度较高。

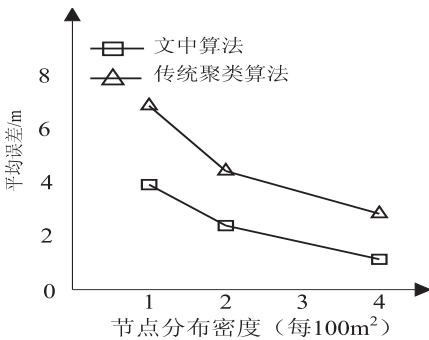


图 4 两种算法精确度情况

4 结束语

文中根据隧道实时监测事件的需要,在原有的关联规则算法分析的基础上,提出了一种新的算法,用实验验证了改算法该有效性,并通过与关联规则算法的比对,得出该算法的性能更高,为该算法的广泛使用奠

定了基础。

参考文献:

[1] 吴 斌,郑 毅,傅伟鹏,等.一种基于群体智能的客户行为分析算法[J]. 计算机学报,2003,26(8):913-918.

[2] 颜宏文,马 瑞,晏弼成.基于信息熵构造判定树的数据挖掘算法的设计与实现[J]. 计算机工程与应用,2003,40(23):180-182.

[3] 翁怀荣,张洪伟,钟 响,等.基于改进的蚁群算法的聚类分析及其在 HRM 中的应用[J]. 计算机应用,2008,25(8):1908-1912.

[4] 王雪松,石 琦,高 珍.基于视频数据的城市隧道交通运行特征与安全研究[J]. 中国安全科学学报,2011(8):129-137.

[5] 周悦来,谭建豪.基于网格和信息熵的多密度聚类算法[J]. 计算机系统应用,2011,20(10):189-192.

[6] Lin Chengru, Liu Kenhao, Chen Ming-syan. Dual Clustering: Integrating Data Clustering over Optimization and Constraint Domains[J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(5):628-637.

[7] Zhang Xueping, Wang Jiayao, Wu Fang. A Novel Spatial Clustering with Obstacles Constraints Based on Genetic Algorithms and K-Medoids[J]. International Journal of Computer Science and Network Security, 2006, 6(10):605-610.

[8] Huang Ming, Bian Fuling. A Grid and Density Based Fast Spatial Clustering Algorithm[C]//Proc. of International Conference on Artificial Intelligence and Computational Intelligence. San Sebastian, Spain: [ s. n. ], 2009:260-263.

[9] Hu Caiping, Qin Xiaolin. A novel spatial cluster algorithm with sampling[M]. Heidelberg:Springer-Verlag, 2007:216-225.

[10] 毛德梅,丁瑞国.对数据挖掘中关联规则算法的比较研究[J]. 皖西学院学报,2006,22(5):27-30.

[11] 李 芸,李青山.数据挖掘中关联规则挖掘方法的研究及应用[D]. 西安:西安电子科技大学,2007.

[12] 刘义安,羊 斌.关联规则挖掘中对 Apriori 算法的一种改进研究[J]. 计算机应用,2007,27(2):418-420.

(上接第 211 页)

[7] 宗成庆.统计自然语言处理[M]. 北京:清华大学出版社, 2011:341-342.

[8] 罗 可,蔡碧野,吴一帆,等.数据挖掘中聚类研究[J]. 计算机工程与应用,2003,39(20):182-184.

[9] Choy S O, Lui A K. Web information retrieval in collaborative tagging systems[C]//Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence. Washington, DC: IEEE Computer Society, 2006:352-355.

[10] Baldassarri A, Cattuto C, Loreto V, et al. Ranking and community detection in undirected networks[EB/OL]. [ 2008-10-05 ]. <http://www.tagora-project.eu/wp-content/2007/04/talk-servedio-folkfrank.pdf>.

[11] 边肇祺,张学工.模式识别[M].第2版.北京:清华大学出版社,2000.

[12] 田莹颖.基于社会化标签系统的个性化信息推荐探讨[J]. 图书情报工作,2010,54(1):50-53.

基于标准标签的用户兴趣模型研究

作者：[杨晶](#)，[成卫青](#)，[郭常忠](#)，[YANG Jing](#)，[CHENG Wei-qing](#)，[GUO Chang-zhong](#)

作者单位：[杨晶, 成卫青, YANG Jing, CHENG Wei-qing\(南京邮电大学 计算机学院, 江苏 南京, 210003\)](#)  
[, 郭常忠, GUO Chang-zhong\(烟台大学 数学与信息科学学院, 山东 烟台, 264005\)](#)

刊名：[计算机技术与发展](#)

ISTIC

英文刊名：[Computer Technology and Development](#)

年，卷(期)：

2013(10)

本文链接：[http://d.g.wanfangdata.com.cn/Periodical\\_wjtz201310052.aspx](http://d.g.wanfangdata.com.cn/Periodical_wjtz201310052.aspx)