

# VRE 中基于内容过滤的论文推荐算法\*

李永<sup>1</sup>, 徐德智<sup>1</sup>, 张勇<sup>2</sup>, 邢春晓<sup>2</sup>

(1. 中南大学 信息科学与工程学院, 长沙 410083 2. 清华大学 信息技术研究院, 北京 100084)

**摘要:** 针对虚拟研究环境中的重要资源论文, 提出了基于内容过滤的推荐算法, 即根据研究者兴趣实现个性化服务, 推荐所需论文。该算法采用矢量空间模型作为用户兴趣和资源描述模型, 使用余弦相似度计算资源推荐度, 基于效率考虑, 利用朴素贝叶斯分类算法减小搜索空间。实验表明, 推荐效果和效率得到了明显改善。

**关键词:** 虚拟研究环境; 个性化; 矢量空间模型; 余弦相似度; 朴素贝叶斯

中图分类号: TP393

文献标志码: A

文章编号: 1001-3695(2007)09-0058-03

## Content-based paper recommending algorithm in VRE

LI Yong<sup>1</sup>, XU De-zhi<sup>1</sup>, ZHANG Yong<sup>2</sup>, XING Chun-xiao<sup>2</sup>

(1. School of Information Science & Engineering, Central South University, Changsha 410083, China 2. Research Institute of Information Technology, Tsinghua University, Beijing 100084, China)

**Abstract:** This paper proposed a content-based recommending algorithm to recommend papers in VRE for researchers. The algorithm utilized VSM (vector space model) to represent users' interests and resources, calculated recommended degree by cosine similarity. To improve efficiency, used naive Bayes classify method to reduce the searching space. Experimental results demonstrate that this approach can produce better accuracy and performance.

**Key words:** VRE (virtual research environment); personalized services; VSM; cosine similarity; naive Bayes

随着我国国家科技创新体系计划<sup>[1]</sup>的实施, 科研机构的发展拥有了更多的机遇, 也遇到了严峻的挑战。未来国家创新体系建设要着重发挥科研机构和大学在科技创新与人才培养方面的核心作用。因此, 如何利用计算机技术在各个科研机构中以及科研机构之间建立一个可以更好地实现研究资源(包括设备、材料、数据、人力等)共享、加强研究人员的科研协作、培养科技创新人才的研究环境, 是有效促进国家科技创新体系计划实施必须解决的重要问题。

英国的 JISC (Joint Information Systems Committee)<sup>[2]</sup>是支持科研和教育的一个基金组织。为了帮助科学家进行研究, 加强科研机构之间的合作, JISC 提出了虚拟研究环境(VRE)的概念<sup>[3]</sup>; VRE 的概念不仅仅是某个研究领域的一个 VRE, 而是要定义相关的标准并建立一个通用的框架, 让其他人使用这个框架并在上面逐渐加入具体领域的 VRE 应用。<sup>[4]</sup>“JISC 支持的众多项目中有 15 个是关于 VRE 的, 如 GROWL<sup>[4]</sup>、Sakai<sup>[5]</sup>以及文献 [6, 7] 介绍了一些建立 VRE 使用的 Web services、grid 和工作流等技术。我国以网络为基础的科学活动环境综合试验平台 CROWN<sup>[8]</sup> (China research and development environment over wider-area network) 就是面向国内各个领域的科学家的一个典型虚拟研究环境。

VRE 中, 随着用户数量不断增多, 研究资源日益增长, 研究人员查找和发现自己感兴趣的资源成为一件非常耗时、耗力的事情, 会出现信息过载和信息迷向问题。因此, 需要在 VRE 中引入个性化服务, 通过研究不同用户兴趣, 主动为用户推荐

最需要的研究资源, 从而更好地解决 VRE 中资源日益庞大与用户寻找有用资源费时费力之间的矛盾。VRE 中的每个研究者都会阅读大量论文, 论文资源无疑是一种重要的研究资源。所以, 需要在 VRE 中进行论文的推荐。

近年来, 应用到各种个性化推荐系统的推荐算法主要基于内容过滤和协同过滤<sup>[9-12]</sup>。协同过滤算法的基本思想是根据用户兴趣的相似性来推荐资源, 把与当前用户相似的其他用户意见提供给当前用户。但是协作过滤依赖于大量用户对大量资源的打分数据, 而且在推荐用户没有评价的资源与新资源方面存在局限性。VRE 中允许任意数量的用户建立任意大小规模的研究项目, 往往会出现很多新建的项目无法获得大量的用户打分数据。对于论文, 可以获得其完整的内容描述, 这样更能从内容本质上推荐给用户真正感兴趣的资源。现阶段, 国内外的各种会议、期刊不断刊登出大量新的论文, 研究者必须要在其领域跟踪前沿技术, 而基于内容过滤同样可以推荐新资源给用户。文献 [13] 中提到的图书推荐系统也正是使用基于内容过滤的推荐算法来推荐图书。因此, 基于内容过滤的推荐算法更加适合在 VRE 推荐论文资源。

## 1 基于内容过滤的推荐算法

### 1.1 用户兴趣和论文资源的模型表达

VRE 中的论文资源本质上是一种文档。文档的表示模型中最典型的是 VSM<sup>[14]</sup>。对于论文资源, 直接用 VSM 把每篇

收稿日期: 2006-07-28; 修返日期: 2006-10-12 基金项目: 国家自然科学基金资助项目(60473078)

作者简介: 李永(1982-), 男, 湖南汨罗人, 硕士研究生, 主要研究方向为个性化服务和虚拟研究环境(liyongjie1982@gmail.com); 徐德智(1963-), 男, 湖南湘潭人, 教授, 主要研究方向为 Web 数据处理、XML、个性化服务; 张勇(1973-), 男, 河南商丘人, 副研究员, 主要研究方向为数据库、e-science 和地理信息系统; 邢春晓(1968-), 男, 河南南阳人, 研究员, 主要研究方向为数字图书馆、电子政务、数字娱乐、个性化服务。

论文表示成一个向量  $V_d = (w_1 f_1) (w_2 f_2) \dots (w_n f_n)$ 。 $V_d$  记录了论文中每个关键词  $w$  和单词出现的频率  $f$ 。为了保证后面向量余弦计算的精度,不进行特征选取,直接用单词作为关键词)。

对于用户兴趣,主要通过用户填写的记录兴趣的文件、用户自己上传的论文资源以及用户要求推荐论文资源时给出的查询词组三种途径获得能反映用户兴趣的关键词。同样用 VSM 把每个用户的兴趣表示成一个向量  $V_u = (w_1 f_1) (w_2 f_2) \dots (w_n f_n)$ 。 $V_u$  记录了能反映用户兴趣的关键词  $w$  以及关键词的权重  $v$ 。

$$v = k_1 f_1 + k_2 f_2 + k_3 f_3 \tag{1}$$

其中  $f_1$  表示  $w$  在该用户记录兴趣的文件中的出现次数; $f_2$  表示  $w$  在其上传文件中的出现次数; $f_3$  表示  $w$  在其查询词中的出现次数; $k_1$ 、 $k_2$  和  $k_3$  表示对这三个频率赋予的权值。一个词在用户记录兴趣的文件和查询词中一般都只会出现一次。 $f_1$  和  $f_3$  都等于 1,所以式(1)简化为

$$v = k_1 + k_2 f_2 + k_3 \tag{2}$$

为了与论文中特征词出现频率的数量级保持一致,以保证相似性计算的精度,根据实验效果, $k_1$ 、 $k_2$  和  $k_3$  分别取值为 2、0.1 和 5。

1.2 相似度计算方法

论文资源的模型和用户兴趣的模型均用 VSM 表示,可以把资源对用户的推荐度计算转换为向量间的余弦相似度的计算。

用户兴趣的 VSM 模型  $V_u$  的关键词向量为( $w$  表示关键词,总共  $m$  个) $u = (w_1 \mu_1, w_2 \mu_2, \dots, w_m \mu_m)$ 。论文资源  $i$  的 VSM 模型  $V_d$  的关键词向量为( $w$  表示关键词,总共  $n$  个) $d_i = (w_1 \mu_1, w_2 \mu_2, \dots, w_n \mu_n)$ 。将两个特征词向量合并后得到本次相似度计算的合并向量  $D$ ( $w$  表示合并后的关键词,总共  $k$  个) $D = u \cup d_i = (w_1 \mu_1, w_2 \mu_2, w_3 \mu_3, \dots, w_k \mu_k)$ 。其中,由于  $u$  和  $d_i$  中可能存在相同的关键词,合并后关键词向量中的关键词个数为  $k$ ,满足  $\max(m, n) \leq k \leq m + n$ 。

根据  $D$  就可以得到这次相似度计算时用户的相似度计算向量为  $D_u = (v_1 \mu_1, v_2 \mu_2, v_3 \mu_3, \dots, v_k \mu_k)$ , $D_u$  中的  $v$  表示  $D$  中对应的关键词在用户模型中的权重。如果这个词不存在,则取 0(为了减小陡变,进行了平滑处理,根据实验效果,取值为 0.1)。同样得到论文的相似度计算向量为  $D_i = (f_1 f_2 f_3 \dots f_k)$ , $D_i$  中的  $f$  表示  $D$  中对应的关键词在论文中的频率。如果这个词不存在,则取 0。

用余弦相似度来计算  $D_u$  与  $D_i$  之间的相似度  $\text{sim}(D_u, D_i)$ ,即为将论文资源  $i$  推荐给用户的推荐度:

$$\text{rec}(u, i) = \left( \sum_{j=1}^k v_j \times f_j \right) / \sqrt{\left( \sum_{j=1}^k v_j^2 \right) \left( \sum_{j=1}^k f_j^2 \right)} \tag{3}$$

通过计算每篇论文与用户的相似度,可以获得推荐度的有序排列,然后把推荐度最大的  $N$  篇论文推荐给用户。算法 1 给出了基于内容过滤的推荐算法的描述。

算法 1 基于内容过滤的推荐算法  
输入:用户 VSM 模型  $V_u$ ,  $M$  篇论文资源的 VSM 模型  $V_d(i = 1, 2, \dots, M)$   
输出:用户的 top- $N$  推荐集  
step 1 从用户的 VSM 模型  $V_u$  获得它的关键词向量  $u$ :  
 $u = (w_1 \mu_1, w_2 \mu_2, \dots, w_m \mu_m)$   
step 2 从论文  $i$  的 VSM 模型  $V_d$  获得它的关键词向量  $d_i$ :

$d_i = (w_1 \mu_1, w_2 \mu_2, \dots, w_n \mu_n)$   
step 3 合并两个关键词向量  $u$  和  $d_i$ ,获得本次相似度计算的合并向量  $D$ :  
 $D = (w_1 \mu_1, w_2 \mu_2, w_3 \mu_3, \dots, w_k \mu_k), \max(m, n) \leq k \leq m + n$   
step 4 依据  $D$ ,计算用户和论文资源在这次相似度计算中各自的向量值  $D_u$  和  $D_i$ :  
 $D_u = (v_1 \mu_1, v_2 \mu_2, v_3 \mu_3, \dots, v_k \mu_k)$   
 $D_i = (f_1 f_2 f_3 \dots f_k)$   
step 5 使用余弦相似度来计算论文资源  $i$  推荐给用户的推荐度:  
 $\text{rec}(u, i) = \left( \sum_{j=1}^k v_j \times f_j \right) / \sqrt{\left( \sum_{j=1}^k v_j^2 \right) \left( \sum_{j=1}^k f_j^2 \right)}$   
step 6 重复 step 2,计算每一篇论文对于该用户的推荐度  
step 7 将所有论文资源按推荐度的大小排列。其中推荐度最大的  $N$  个资源将作为用户的推荐集

1.3 算法 1 的实验结果分析

实验使用由六个不同的研究者用户,每人上传 50 篇共 300 篇的论文作为待推荐的论文资源,然后对其中每个用户进行 top-50 的推荐。如果推荐的论文属于该用户上传的论文,则表示推荐成功(假设用户上传的论文资源肯定是用户感兴趣的);反之,则表示推荐失败。

实验选择 300 篇论文作为待推荐资源,主要是为了从一定的角度评价推荐效果。在实际为用户推荐论文资源的过程中,只会将其他用户上传的 250 篇论文作为待推荐的论文资源,即不会重新推荐用户自己上传的论文资源。

采用信息检索领域广泛使用的查准率(precision)和误判率(fallout)<sup>[15]</sup>来评价实验效果,它们的定义如下:

$$\text{precision} = \text{推荐成功的论文数} / \text{被推荐的全部论文数} \tag{4}$$

$$\text{fallout} = \text{推荐失败的论文数} / \text{其他用户上传的论文数} \tag{5}$$

根据实验结果,表 1 给出了对六个用户进行推荐时的查准率和误判率的比较。

表 1 算法 1 的推荐效果

用户	查准率/%	误判率/%	用户	查准率/%	误判率/%
U <sub>1</sub>	54	9.2	U <sub>4</sub>	34	13.2
U <sub>2</sub>	80	4	U <sub>5</sub>	14	17.2
U <sub>3</sub>	14	17.2	U <sub>6</sub>	36	12.8

由表 1 可以看出,六个用户之间的推荐效果相差比较明显。用户 2 的推荐效果最好,用户 3 和 5 的推荐效果最差。经分析发现,用户 2 提供的记录兴趣文件比其他用户更加详细;用户 1 和 2 在推荐时都提供了查询关键词,这样就能更反映用户自身的兴趣,推荐效果更好。因此,必须对用户的兴趣模型进行充分学习,进一步真实、完整地反映用户兴趣,才能对用户提供更加有效的推荐。

2 改进后的推荐算法

在实际的 VRE 中,当遇到一个大规模甚至是超大规模的研究项目时,参与的研究者很多。论文资源也会很多,每次要对同一个用户在线计算所有论文的推荐度,时间复杂度太高。所以,必须在保证推荐效果的前提下,尽量减小搜索空间,即减少待推荐的论文资源数量,提高推荐效率。

当 VRE 中一个大规模的研究项目成立时,用户提供的论文不会是散乱的,一般来讲都已分成若干类别。即使没有分类,也可以用已有的各种聚类算法对这些论文进行聚类。最终,可以得到一个已经分好类的论文集合(实验中,论文资源在用户上传时已经分成若干类别)。

通过实验发现,对同一个用户推荐的论文大多数都是属于相同的两、三个类别,这些类别恰好基本概括了用户的研究方向,或非常相近。如果能首先找到这两、三个类别,然后仅仅计算这些类别中的论文资源对于该用户的推荐度,这样就可以大大减小搜索空间,降低时间复杂度,提高推荐效率。

## 2.1 基于朴素贝叶斯的分类

本文利用实现简单且分类效果很好的朴素贝叶斯分类算法<sup>[16]</sup>来计算用户兴趣的 VSM 中关键词属于不同论文资源类别的概率,以获得用户兴趣属于该类别的概率。假定每个关键词对给定的论文资源类别的影响独立于其他关键词,那么用户兴趣属于类别的概率等于  $u = (w_1, w_2, \dots, w_m)$  中关键词  $w$  属于类别  $c$  的概率的乘积。

在计算推荐度时,为了保证计算精度,没有对关键词进行特征选取。但是,在利用朴素贝叶斯分类算法时,太多没有代表性的词实际上是一种噪声数据,影响分类效果。所以必须进行特征选取,去除那些表现力不强的词,筛选出针对该类的特征词集合。实验中,选用互信息量<sup>[17]</sup>作为特征词的选取方法。

## 2.2 利用朴素贝叶斯分类的推荐算法

将朴素贝叶斯分类算法引入基于内容过滤的推荐算法中,提出一种改进算法。首先需要预先处理各个论文资源类别,对于每个类别中的论文,获得其 VSM 模型,然后将该类别中所有论文的 VSM 模型合并,并计算每个关键词属于这个类别的概率:

$$P(w_k | c_j) = [1 + \sum_{i=1}^{|D|} N(w_k, d_i)] / [ |V| + \sum_{s=1}^{|V|} \sum_{i=1}^{|D|} N(w_s, d_i) ] \quad (6)$$

其中  $P(w_k | c_j)$  为关键词  $w_k$  在类别  $c_j$  中出现的概率;  $|D|$  为该类别的论文数;  $N(w_k, d_i)$  为  $w_k$  在类别  $c_j$  的论文  $d_i$  中出现的频率;  $|V|$  为类别  $c_j$  中总的词数;  $\sum_{s=1}^{|V|} \sum_{i=1}^{|D|} N(w_s, d_i)$  为该类别所有关键词的词频之和。这样就得到该类别的 VSM 模型向量:  $V_{c_j} = (w_1, p_1)(w_2, p_2) \dots (w_{n_1}, p_{n_1})$ 。  $V_c$  记录了类别中每个关键词  $w$  和该词属于该类别概率  $p$ 。

计算每个关键词  $w$  和类别  $c$  的互信息量:

$$I(w_k, c_j) = \log_2 [P(w_k | c_j) / P(w_k)] \quad (7)$$

其中  $P(w_k)$  为关键词  $w_k$  在所有论文中的出现概率,其计算公式与  $(w_k | c_j)$  的计算公式相同,只是  $|D|$  表示全体论文数。

将关键词的互信息量按大小排列,选择比例为  $\alpha$  (根据实验效果  $\alpha = 0.019$ ) 的互信息量较大的词作为该类别的特征词。经过特征选取,排除了那些对分类没有太多意义的关键词,得到一个新的 VSM 模型。其中每个特征词都较能反映该类别的特征,就将它作为该类别的 VSM 模型:  $V_{c_j} = (w_1, p_1)(w_2, p_2) \dots (w_{n_2}, p_{n_2})$ ;  $n_2 = \alpha n_1$ 。然后,再计算用户兴趣属于各类别的概率,获得用户兴趣最可能属于的两个类别(根据实验,取两个类别综合效果最好)。最后,计算这两个类别中的论文资源对该用户的推荐度,把前  $N$  个资源作为 top- $N$  推荐集。算法 2 给出了改进后的基于内容过滤的推荐算法的描述。

算法 2 利用朴素贝叶斯分类的基于内容过滤推荐算法

输入: 用户 VSM 模型  $V_u$ ,  $L$  个论文资源类别的 VSM 模型  $V_{c_j} (j = 1, 2, \dots, L)$ ,  $M$  篇论文资源的 VSM 模型  $V_{d_i} (i = 1, 2, \dots, M)$ 。

输出: 用户的 top- $N$  推荐集

step 1 对用户的 VSM 模型  $V_u$  获得它的特征词向量  $u$ :

$$u = (w_1, w_2, \dots, w_m)$$

step 2 使用朴素贝叶斯分类法来计算用户兴趣属于论文资源类别  $j$  的概率:

$$P(c_j | u) = P(c_j) \prod_{k=1}^m P(w_k | c_j)$$

step 3 重复 step 2, 计算用户属于每个论文资源类别的概率, 选取得到用户兴趣最可能属于的论文资源类别  $c_{\max_1}$  和  $c_{\max_2}$

step 4 对论文资源类别  $c_{\max_1}$  和  $c_{\max_2}$  中的论文资源  $i$ , 从其 VSM 模型  $V_{d_i}$  中获得它的关键词向量  $d_i$ :

$$d_i = (w_1, w_2, \dots, w_n)$$

step 5 合并两个关键词向量  $u$  和  $d_i$ , 获得该次相似度计算的合并向量  $D$ :

$$D = (w_1, w_2, w_3, \dots, w_k), \max(m, n) \leq k \leq m + n$$

step 6 依据  $D$  计算用户和论文资源在这次相似度计算中各自的向量值  $D_u$  和  $D_i$ :

$$D_u = (v_1, v_2, v_3, \dots, v_k)$$

$$D_i = (f_1, f_2, f_3, \dots, f_k)$$

step 7 使用余弦相似度来计算论文资源  $i$  推荐给用户的推荐度:

$$\text{rec}(u, i) = (\sum_{j=1}^k v_j \times f_j) / \sqrt{(\sum_{j=1}^k v_j^2)(\sum_{j=1}^k f_j^2)}$$

step 8 重复 step 7, 计算论文资源类别  $c_{\max_1}$  和  $c_{\max_2}$  中的每一篇论文对于该用户的推荐度

step 9 按推荐度的大小排列。其中推荐度最大的  $N$  个资源将作为用户的推荐集

## 2.3 算法 2 的实验结果分析

根据算法 2 的实验结果, 同样对六个用户进行推荐效果的比较, 如表 2 所示。

表 2 算法 2 的推荐效果

用户	查准率/%	误判率/%	用户	查准率/%	误判率/%
U <sub>1</sub>	70	6	U <sub>4</sub>	58	8.4
U <sub>2</sub>	86	2.8	U <sub>5</sub>	56	8.8
U <sub>3</sub>	64	7.2	U <sub>6</sub>	60	8

对比两个算法的推荐效果, 如图 1 和图 2 所示。

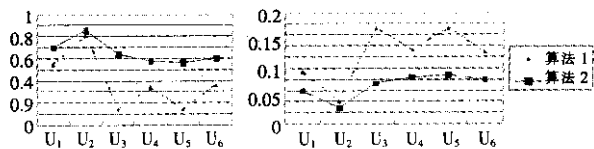


图 1 算法 1 和算法 2

的 precision 的对比

图 2 算法 1 和算法 2

的 fallout 的对比

由图 1 和 2 可以看出, 利用了朴素贝叶斯分类算法后的推荐效果明显比原算法的推荐效果好。六个用户的查准率和召回率提高, 误判率都有所下降, 而且六个用户之间的推荐效果相差也降低了。分析发现, 主要是因为每个用户提交的论文都是自己感兴趣的研究方向, 往往都集中在两类别中, 而本文的实验就正好找到这两个类别, 大大减小了待推荐资源中不应推荐资源的比例, 所以推荐效果更加明显、有效。

## 3 结束语

本文针对虚拟研究环境中研究资源日益庞大与用户寻找有用资源费时费力之间的矛盾, 提出了将个性化服务引入虚拟研究环境中, 并针对虚拟研究环境中重要的研究资源论文, 设计了利用朴素贝叶斯分类算法提高推荐效率的基于内容过滤的推荐算法。根据 VRE 中不同用户的兴趣, 推荐符合其兴趣的论文, 促进了研究人员更好地利用研究资源, 更好地进行研究合作。

未来的工作包括两方面: 一方面, 怎样更 (下转第 89 页)

向的算法获得明文的用户密码,而是依然采用正向的 MD5 单向散列值,从本地的存储用户库获得相关的身份信息;采用 MD5 单向散列值的密码,比较数据包中需要验证的字段。这种方法并不增加服务器端的负荷,却使网络的安全性得到了大幅度提高。Radius 协议只是服务器针对发起呼叫方用户身份的认证,而本文的认证机制还支持被叫方对发起呼叫的用户身份的验证。这是通过定义了专门用于鉴别的头域来实现的。这个头域不仅包括用户名和密码,还包括用户的 ENUM 号码,增加了攻击者假冒身份的难度。

5) 同时在认证服务器与 SIP 服务器之间传递的信息包含服务器生成的随机数。这个随机数会有一个有效期,可以防止重放攻击。

5 总结

本文介绍了引入 ENUM 技术实现的 VoIP 认证系统的关键问题研究和技术方案,该方案利用 E. 164 号码独特的分配特性把 E. 164 电话号码引入到 VoIP 认证系统中,使得 VoIP 系统的身份认证与传统的 PSTN 电话网络中的身份认证一样,即通过电话号码来认证。同时本方案也可以作为 VoIP 网络 and 传统电信网络相互融合的一种可行性方案。该方案的认证方式可以适用于发起呼叫端认证与被动呼叫端认证相结合的方式,具有良好的认证可靠性。本文利用全球统一的 E. 164 号码,用于构建适合于全球范围的有序、可靠的 VoIP

通信,对于基于此实现其他可信任 VoIP 通信业务,具有一定的参考意义。

参考文献:

[ 1 ] 李晓东, 阎保平. 基于 DNS 的 ENUM 技术及其应用研究[ J ]. 微电子学与计算机, 2003, 20( 1 ): 46-49, 53.

[ 2 ] 王峰, 钱华林. ENUM 系统以及解析服务的研究[ J ]. 计算机应用研究, 2003, 20( 7 ): 41-45.

[ 3 ] 封涛, 阎保平, 朱晓莲. 基于 Handle 名字解析系统的 ENUM 技术解决方案[ J ]. 计算机应用研究, 2006, 23( 10 ): 223-225.

[ 4 ] KUHM D R, WALSH T J, FRIES S. Security considerations for voice over IP systems[ R ]. [ S. l. ] : National Institute of Standard and Technology( NIST ), 2005.

[ 5 ] ENDLER D, GHOSAL D, JAFARI R. VoIP security and privacy threat taxonomy public release 1.0[ R ]. [ S. l. ] : Voice over IP Security Alliance( VOIPSA ), 2005.

[ 6 ] FALTSTROM P, MEALLING M. The E. 164 to uniform resource identifiers( URI ) dynamic delegation discovery system( DDDS ) application( ENUM ) [ S ]. [ S. l. ] : IETF, 2004.

[ 7 ] PETERSON J. RFC3764, Enumservice registration for SIP addresses-of-record[ S ]. [ S. l. ] : IETF, 2004.

[ 8 ] MEALLING M, DANIEL R. RFC2915. The naming authority pointer( NAPTR ) DNS resource record[ S ]. [ S. l. ] : IETF, 2000.

[ 9 ] FOSTER M, McGARRY T, YU J. RFC3482, Number portability in the global switched telephone network( GSTN ) : an overview[ S ]. [ S. l. ] : IETF, 2003.

(上接第 60 页) 有效地学习用户兴趣并提高系统的运行效率和推荐效果; 另一方面, 在虚拟研究环境中, 用户是专业化程度较高的研究人员, 资源多样化, 包括设备、材料、数据、人力、论文、实验数据文档、软件等, 怎样针对这一个特殊的环境, 依据环境自身的特点, 为用户提供更好的个性化服务。

参考文献:

[ 1 ] 中华人民共和国国务院. 国家中长期科学和技术发展规划纲要[ EB/OL ]. 2006-02-09. [http://www.gov.cn/jrzq/2006-02/09/content\\_183787\\_7.htm](http://www.gov.cn/jrzq/2006-02/09/content_183787_7.htm).

[ 2 ] JISC. The joint information systems committee[ EB/OL ]. 2003-08-28. <http://www.jisc.ac.uk/>.

[ 3 ] JISC. Virtual research environments programme[ EB/OL ]. 2006-03-17. [http://www.jisc.ac.uk/programme\\_vre.html](http://www.jisc.ac.uk/programme_vre.html).

[ 4 ] CCLRC Daresbury Laboratory. GROWL home. [ EB/OL ]. 2006-03-16. <http://tyne.dl.ac.uk/GROWL/>.

[ 5 ] Sakai Community. Sakai Project[ EB/OL ]. 2006-04-20. <http://www.sakaiproject.org/>.

[ 6 ] SIM Y W, WANG C, CARR L A, *et al.* A Web/grid services approach for a virtual research environment implementation[ DB/OL ]. 2005-09. <http://eprints.ecs.soton.ac.uk/11162/>.

[ 7 ] SIPOS G, NEMETH C, LEWIS G J, *et al.* Executing workflow-based grid applications with the collaborative P-GRADE portal[ DB/OL ]. 2005-09. <http://eprints.ecs.soton.ac.uk/11162/>.

[ 8 ] 北京航空航天大学计算机新技术研究所. CROWN portal[ EB/OL ]. 2004-01-01. <http://www.crown.org.cn/>.

[ 9 ] RESNICK, IACOVOU N, SUCHAK M, *et al.* GroupLens : an open

architecture for collaborative filtering of netnews[ C ]//Proc of CSCW. Chapel Hill, North Carolina : ACM Press, 1994 : 175-186.

[ 10 ] KARYPIS G. Evaluation of item-based Top-N recommendation algorithms[ C ]//Proc of CIKM 2001. Atlanta : ACM Press, 2001 : 247-254.

[ 11 ] ZENG Chun, XING Chun-xiao, ZHOU Li-zhu, *et al.* Similarity measure and instance selection for collaborative filtering[ J ]. International Journal of Electronic Commerce, 2004, 4( 8 ): 115-129.

[ 12 ] ADOMAVICIUS G, TUZHILIN A. Toward the next generation of recommender systems : a survey of the state-of-the-art and possible extensions[ J ]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17( 6 ): 734-749.

[ 13 ] MOONEY R J, ROY L. Content-based book recommending using learning for text categorization[ C ]//Proc of the 5th ACM Conference on Digital Libraries. San Antonio [ s. n. ], 2000.

[ 14 ] SALTON G, WONG A, YANG C S. A vector space model for automatic indexing[ J ]. Information Retrieval and Language Processing, 1975, 18( 11 ): 613-620.

[ 15 ] SALTON G, LESK M E. Computer evaluation of indexing and text processing[ J ]. Journal of the Association for Computing Machinery, 1968, 15( 1 ): 8-36.

[ 16 ] LANGLEY P, IBA W, THOMPSON K. An analysis of bayesian classifiers[ C ]//Proc of the 10th National Conference on Artificial Intelligence. Menlo Park : AAAI Press, 1992 : 223-228.

[ 17 ] VIOLA P, WELLS W M. Alignment by maximization of mutual Information[ C ]//Proc of the 5th International Conference on Computer Vision. Boston, MA [ s. n. ], 1995 : 16-23.