

# 基于二级向量描述的搜索引擎个性化服务模型<sup>\*</sup>

徐静秋 朱征宇 谭明红 任翔  
(重庆大学计算机学院 重庆 400044)

**摘要** 本文介绍了一种搜索引擎个性化服务模型。用二级向量进行文本特征提取和用户兴趣建模,关键词向量能快速定位用户的兴趣领域,而扩展词向量能准确反映用户在该领域上的兴趣偏好。当用户提交关键词时,本系统根据学习到的用户兴趣描述模型计算词间相关度,自动增加几个个性化扩展词提交给搜索引擎,实现不同用户键入相同关键词能返回不同信息的目的。实验结果充分表明本系统应用于搜索引擎个性化服务领域的有效性和实用性。

**关键词** 搜索引擎,个性化,二级向量描述,词间相关度

## A Search Engine Personalization Model Based on Double Vector Description

XU Jing-Qiu ZHU Zheng-Yu TAN Ming-Hong REN Xiang  
(College of Computer, Chongqing University, Chongqing 400044)

**Abstract** A novel personalized search engine model is proposed in this article. It describes user profiles and text feature with double vector. Keyword vector helps to rapidly find a user's interest domain. Expansion word vector can accurately express his preference in the domain. When a user submit query keywords, our system automatically computes the term-term associations according to user profiles. More personalized expansion words are determined using a search engine. Thereby different search results are returned to different users who input the same key words. The test results show the feasibility and applicability of the presented work for personalized information service of a search engine.

**Keywords** Search engine, Personalization, Double vector description, Term-term associations

### 1 目前搜索引擎存在的不足

随着 Internet 的发展,搜索引擎已经被越来越多的人使用,而优秀的搜索引擎也不断涌现,如:Google, Yahoo 和 Sogou 等,将网络资源进行了一定的组织,从而给人们的学习和生活带来了许多便利。搜索引擎的发展是一个不断探索的过程,如今已从第二代向第三代发展,其 Web 信息检索的性能更加优越。第三代搜索引擎是正在研究开发的智能搜索引擎,个性化要求是它的主要特色之一。然而,目前的搜索引擎仍然存在许多不足之处。

虽然现在许多网站和商家都声称向用户提供了个性化服务,但是大部分还只是形式上的,而用户需要的是真正内容上的个性化服务。大多数搜索引擎是不区分用户的。例如:用户输入关键词“邓亚萍”进行搜索,返回的网页集包括她工作、学习、家庭生活等诸多方面的信息。也就是说,不同用户输入相同的关键词进行查询时,输出的结果是完全一样的,而与用户本身的兴趣偏好没有任何关系。因此很多不相关的信息被推荐给用户,而真正感兴趣的信息排在了推荐列表的后面,大多数用户都不会访问搜索引擎返回的 20 页以后的网页<sup>[1]</sup>,导致推荐效率大大降低。

针对上述情况,本文提出基于二级向量描述的个性化推荐系统设计。当用户向搜索引擎提交查询关键词时,系统根据日常学习到的用户个人兴趣,计算关键词和扩展词间相关度,自动增加几个体现用户偏好的扩展词一起提交给搜索引擎,这样能过滤出需要的信息,从而可以提高搜索引擎的

检索效率。

### 2 基于用户兴趣描述的个性化推荐模型

#### 2.1 本模型的总体结构

如图 1 所示,本模型的特点是用户兴趣代理和推荐系统都在客户端,通过兴趣代理捕获用户兴趣,将用户兴趣描述(user profile)存放在用户兴趣库中。推荐系统的功能就是当用户键入关键词时,提取兴趣库和知识库的信息,通过词间相关性度量算法自动增加几个个性化的扩展词一同提交给搜索引擎,从而返回相关的搜索结果,有效可行地实现搜索引擎个性化服务的目的。

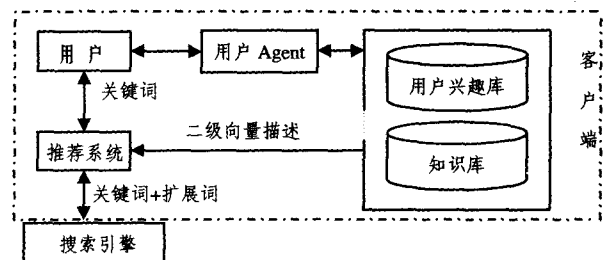


图 1 模型总体框架图

#### 2.2 用户兴趣描述模型(User Profile)

首先在客户端建立用户兴趣代理,当用户浏览的网页数量达到某个阈值时(如浏览 60 个网页),代理生成用户近期兴趣视图(Current Interest View)<sup>[2]</sup>。假设用户至多有  $M$  个兴

<sup>\*</sup> 高等学校博士学科点专项科研基金资助课题(20030611016)、重庆大学骨干教师计划基金项目(2003A33)。徐静秋 硕士生,研究方向为个性化服务和 Web 智能检索、数据挖掘;朱征宇 副教授,博士,研究方向为 Web 智能检索、电子商务和数据库技术。

趣领域,即兴趣子类,那么该视图就由  $M$  个兴趣子类组成,如军事、体育、财经等。

本模型的特点是用二级向量的形式描述用户对某一子类的兴趣。所谓二级向量就是用一组关键词向量和扩展词向量描述用户的子类兴趣。关键词向量能快速定位兴趣领域,而扩展词向量能准确反映用户在该领域上的兴趣偏好。例如有两个用户关于体育明星子类的关键词向量都是(邓亚萍),用户甲的扩展词向量为(学习、大学),用户乙的扩展词向量为(学子、活动、公益)。关键词向量快速地定位用户的兴趣领域是体育明星子类的邓亚萍,但通过扩展词向量可知用户甲关心她的读书求学情况,而用户乙则对她参与各种社会慈善活动感兴趣。用户兴趣建模就是通过扩展词向量来区分不同用户对同一对象所关心的不同侧面。这要求确定一个全局词典(关键词来自于该词典),可根据领域知识或专业词典得到,比如体育类的词汇就是由体育明星的名字和与体育专业相关的术语组成,这个词典较小,有利于快速定位兴趣领域。同时确定一个扩展词词典(扩展词来源于它),它包含一些与领域知识无关的普通词。这样无论是用户兴趣描述还是文本特征均用二级向量的形式表示。

### 2.3 个性化推荐算法

当用户向搜索引擎提交关键词时,推荐系统根据学习到的用户兴趣描述模型,自动产生几个个性化扩展词并提交给搜索引擎,这有助于准确定位搜索范围。这种算法既利用了当前搜索引擎的优点,如实时响应用户的查询请求、返回信息全面等,同时有效地弥补了搜索引擎不区分用户,未考虑用户兴趣偏好的不足。因此,即使用户输入的关键词相同,通过本系统提供的少量扩展词,也可以返回与用户兴趣相关的结果集,以实现搜索引擎的个性化服务。

#### 2.3.1 算法中的基本定义和理论基础

首先,以术语为行、文件为列做一个大矩阵。设一共有  $t$  行  $d$  列,矩阵名为  $X$ 。矩阵的元素为术语在文件中的出现频度。

给定矩阵  $X$ ,根据文[3]可知,基于  $X$  可以有三类同文件检索密切相关的问题:

- (1) 术语  $i$  和  $j$  有多相似? 即术语的类比和聚类问题;
- (2) 文件  $i$  和  $j$  有多相似? 即文件的类比和聚类问题;
- (3) 术语  $i$  和文件  $j$  有多相关? 即术语和文件的关联问题。

下面用矩阵  $X$  来进行这三类比较。

- (1) 比较两个术语

做“正向”乘法:  $X * X'$ ,其中  $X_{ij}$  代表术语  $i$  和  $j$  的相似程度。

- (2) 比较两个文件

做“逆向”乘法:  $X' * X$ ,其中  $X_{ij}$  代表文件  $i$  和  $j$  的相似程度。

- (3) 比较一个文件和一个术语

矩阵  $X$  本身。其中  $X_{ij}$  代表术语  $i$  和文件  $j$  的相关程度。

以上理论为定量地分析未出现在初始信息中的对象关系提供了依据,如通过术语-文档关联矩阵  $X$  可以定量地计算出词与词间的相关度、文档与文档间的相关度等。本文中的推荐算法就是应用矩阵  $X$  做“正向”乘法  $X * X'$  比较关键词与扩展词间的相关性,从而求出与关键词关联程度最高的几个扩展词。

### 2.3.2 特征词-文档关联矩阵 $X$ 的生成

当用户向搜索引擎输入关键词,如何确定能代表用户个人偏好同时与该关键词关联程度最高的几个扩展词,是算法设计中解决的关键问题。在用户兴趣建模部分,客户端代理将学习到的用户兴趣描述(user profile)存入兴趣库中,取用户最感兴趣的前  $M$  个子类,每个子类分别用一个关键词向量和一个扩展词向量表示。计算关键词与扩展词间的关系是以生成用户近期兴趣视图的训练网页集作为基础,这些训练网页是各兴趣子类、关键词向量和扩展词向量产生的依据。从训练网页中提取出同时包含至少一个关键词和一个扩展词的文档集,组成一个特征词-文档关联矩阵  $X$ ,如表 1。

表 1 特征词-文档关联矩阵  $X$

	166	231	282	307	375	405	434	437	540	544
邓亚萍	4	18	14	9	4	13	11	8	35	18
手术	0	4	2	5	0	0	0	7	0	4
学子	1	5	3	2	1	0	10	11	10	5
活动	0	0	2	0	3	18	2	0	25	0
眼镜	0	12	0	0	0	0	0	0	0	12
冠军	4	0	4	0	1	19	0	3	14	0
体育	2	0	0	0	0	25	0	3	14	0
激光	5	4	0	0	0	0	0	6	0	4
公益	0	0	0	0	0	11	0	0	31	0
慈善	0	0	0	0	3	19	0	0	18	0

表 1 是从某个关心邓亚萍参加社会公益活动的用户浏览的网页中生成的术语-文档关联矩阵。这个二维矩阵中第一行代表文档号,第一列代表特征词,每个单元格表示特征词在文档中出现的频次。这些特征词来源于该用户体育子类的二级向量描述。取权重最大的前  $m$  个关键词作为该用户的关键词向量描述,取权重在前  $n$  位的扩展词作为扩展词向量描述,在实验中为了体现系统对关心同一对象不同侧面的用户群的个性化推荐效果,体育子类的关键词向量中只取了“邓亚萍”一个关键词,即  $m=1$ ,扩展词向量取了权重在前 9 位的扩展词,即  $n=9$ 。参数  $m, n$  阈值的设置可以根据实际情况适当调整。由于篇幅有限,表 1、表 2、表 3 只显示了部分实验数据。

#### 2.3.3 特征词-文档关联矩阵 $X$ 的修正

计算关键词与扩展词间的相似度时,本文采用向量空间相似度计算方法<sup>[5]</sup>:

$$\sin(\omega_1, \omega_2) = \frac{\omega_1 \cdot \omega_2}{\|\omega_1\|_2 \cdot \|\omega_2\|_2} \quad (1)$$

表 2 修正的特征词-文档关联矩阵  $XN$

	166	231	282	307	375	405	434	437	540	544
邓亚萍	0.072	0.323	0.251	0.161	0.072	0.233	0.197	0.143	0.627	0.323
手术	0	0.157	0.078	0.196	0	0	0	0.274	0	0.157
学子	0.05	0.252	0.151	0.101	0.05	0	0.503	0.553	0.503	0.252
活动	0	0	0.064	0	0.097	0.579	0.064	0	0.804	0
眼镜	0	0.57	0	0	0	0	0	0	0	0.57
冠军	0.163	0	0.163	0	0.041	0.776	0	0.123	0.572	0
体育	0.069	0	0	0	0	0.866	0	0.104	0.485	0
激光	0.25	0.2	0	0	0	0	0	0.3	0	0.2
公益	0	0	0	0	0	0.334	0	0	0.942	0
慈善	0	0	0	0	0.114	0.721	0	0	0.683	0

式(1)是向量空间相似度计算公式,两个词语向量  $\omega_1$  和  $\omega_2$  被看作是向量空间中的两个向量。通过计算两个向量

的夹角的余弦来衡量相互之间的相似度,夹角越小,相似度越高。如表1所示,每一行表示一个特征词向量,分量值对应一个特征词在不同文档中的权重。

为了进行词语间相似度计算(行与行的比较),在算法中首先对矩阵  $X$  进行修正,将矩阵行按  $\| \cdot \|$  模归一化,得

到矩阵  $X_N$ ,如表2所示。

### 2.3.4 关键词与扩展词的相似度计算

计算出修正后的特征词-文档关联矩阵后,再对特征词进行类比,对矩阵  $X_N$  做正向乘法  $X_N * X_N'$ ,得到如下所示矩阵  $M = X_N * X_N'$  (见表3)。

表3 词间关系矩阵  $M$

	邓亚萍	手术	学子	活动	眼镜	冠军	体育	激光	公益	慈善
邓亚萍	1	0.474	0.747	0.675	0.529	0.613	0.526	0.392	0.669	0.605
手术	0.474	1	0.262	0.005	0.685	0.046	0.028	0.78	0	0
学子		0.262	1	0.452	0.287	0.391	0.305	0.28	0.474	0.35
活动		0.005	0.452	1	0	0.924	0.891	0	0.952	0.978
眼镜	0.529	0.685	0.287	0	1	0	0	0.599	0	0
冠军	0.613	0.046	0.391	0.924	0	1	0.973	0.078	0.799	0.955
体育	0.526	0.028	0.305	0.891	0	0.973	1	0.049	0.746	0.956
激光	0.392	0.78	0.28	0	0.599	0.078	0.049	1	0	0
公益		0	0.474	0.952	0	0.799	0.746	0	1	0.885
慈善	0.605	0	0.35	0.978	0	0.955	0.956	0	0.885	1

矩阵  $M$  中的每个单元格就反映训练文档集中词间相关度,本文主要关心的是关键词(代表某主题)与扩展词(描述内容细节)之间的联系。如表3所示,第2列数据是主要考察对象,它反映了关键词“邓亚萍”与这一组扩展词的关系。确定选取的扩展词数量也是实验的关键内容。如果扩展出的新的查询词项太多,会产生大量噪声<sup>[7]</sup>,对相似度评分造成极大的影响。因此在传统的查询扩展中,扩展的程度选择是一个重要而尚未很好解决的问题。为了避免过多的扩展词带来大量噪声,规定扩展词的数量最多不超过4个。但在实验中我们发现关键词并不一定与相似度最大的前4个扩展词的相关度都高,很多时候仅有1~3个扩展词与关键词的关联程度大。这里借用了求相对误差的公式<sup>[8]</sup>合理地选取扩展词:

$$E_r(x) = \frac{x - x^*}{x^*} \quad (2)$$

$$|E_r(x)| \leq \delta \quad (3)$$

公式(2)中  $x^*$  表示矩阵  $M$  中与关键词相似度最大的扩展词对应的相似度值,  $x$  表示其它扩展词与该关键词的相似度值。公式(3)中  $\delta$  参数表示  $x$  与  $x^*$  的相对差值,它反映了其它扩展词  $x$  相对于  $x^*$  与关键词的相关度,可以根据实验情况设置。除了选取相似度值最大的那个扩展词外,其它扩展词的取舍就依据公式(3),实验中取  $\delta = 10\%$ ,即  $x$  与  $x^*$  的相对差值不超过  $10\%$ ,都表明扩展词  $x$  与关键词有较强的关联度。

因此在选取扩展词时,除了选取矩阵  $M$  中相似度最大的值  $x^*$  对应的词外,还选取那些与  $x^*$  相对差值不超过  $10\%$  的扩展词,同时扩展词的总数不超过4个。如表3第2列特殊标记的三个单元格,表示经过公式(3)的运算,从关于邓亚萍参与慈善活动这一主题的训练网页中选定的扩展词对应的相似度值。可以得出结论:与关键词“邓亚萍”最相关的扩展词是“学子、活动、公益”。

### 3 实验结果及分析

假设有四个喜欢浏览邓亚萍网页的用户,由于每个人的喜好不同,他们往往关心的是邓亚萍生活的不同侧面,如有的喜欢看她发奋读书的故事、有的对她的家庭生活特别感兴趣等。能否依据每个用户的浏览历史,对一群提交相同关键词的用户进行个性化推荐是衡量本文算法优劣的标准。可以用信息检索领域中评估系统效果的两个标准:查全率(Recall)和查准率(Precision),作为评判算法执行效率的尺度。在Google中进行了测试,用“邓亚萍”作为查询词进行检索,总共返回了878张网页(测试时间2006年10月20日),把整个返回的网页集作为一个全集,也是后面叙述中所称的原集。从中选取四个具有代表性的主题,在原集中找出与每个主题相关的所有网页,表4第2列就表示原集中与每个主题内容相关的网页数量。

表4 新方法与原方法对各主题查全率的统计对比

主题	原集相关网页数	Top-N	查询词(关键词+扩展词)	原方法	新方法	原查全率	新查全率	提高倍数
公益活动	30	30	邓亚萍 学子 活动 公益	1	13	0.033333	0.433333	12
工作	71	80	邓亚萍 工作 运动员 奥运会 北京	22	47	0.309859	0.661971	1.13636
学习	79	80	邓亚萍 学习 大学	10	54	0.126582	0.683544	4.4
家庭	130	130	邓亚萍 孩子 法国 中国	18	89	0.138462	0.684615	3.94444
平均查全率						0.152059	0.615866	3.05018

(1) 查全率:返回的相关网页数与所有的相关网页数量的比值。

首先将原集中某一主题的所有网页分为训练集(training set)和测试集(testing set)两部分,训练集用来生成术语间关系矩阵  $M$ (如表3所示),得到反映用户兴趣偏好的一组扩展词。在增加了扩展词后进行再次搜索,从搜索结果中产生

Top-N 推荐集,将既在 Top-N 推荐集又在测试集中的网页数量与测试集中网页数量的比值作为查全率 Recall 的标准<sup>[9]</sup>,有

$$\text{Recall} = \frac{|\text{testing} \cap \text{Top-N 推荐集}|}{|\text{testing}|} \quad (4)$$

采用公式(4),希望达到的最理想的效果就是新推荐算法得到的 Top-N 推荐集能尽可能多地召回原集中所有相关网

页。为了对比原搜索和个性化搜索的效果,每个主题 Top-N 推荐集中 N 的取值等于或略大于原集中相关网页数,如表 4、表 6 中 Top-N 列所示。公式(4)中  $|testing \cap Top-N 推荐集|$  表示既在 Top-N 推荐集又在测试集中的网页数量,  $|testing|$  表示测试集中网页数量。

表 4 是两种搜索方法查全率的对比。原方法指仅用关键词进行搜索的结果,新方法表示添加了个性化扩展词后的搜索结果。第 5 列“原方法”表示关于某主题原方法 Top-N 推荐集召回的网页数,第 6 列“新方法”表示新方法 Top-N 推荐集召回的网页数。

实验表明,新方法搜索结果中与主题相关的网页普遍排名靠前。如关于邓亚萍参与公益活动这一主题的网页。原集中总共仅有 30 张,且排名靠后,前 30 张中仅有一张相关网页。但用新方法添加了几个扩展词后,前 30 张网页中召回了 13 张,较之原方法查全率提高了 12 倍。表 5 是新方法与原方法的网页排名情况对比,由此可见新推荐算法的效果是相当明显的。

表 5 新方法与原方法的网页排名对比

新排名	原排名	主题
1	540	邓亚萍中国奥运健儿公益第一人
2	361	邓亚萍资助学子恢复视力追踪
3	52	邓亚萍泉州光明行捐助仪式昨日举行
4	231	邓亚萍到天津资助天津寒门学子
5	603	邓亚萍 15 万泉州助学子恢复视力
6	27	邓亚萍山城献爱心
7	307	邓亚萍来津参加全国眼科年会公益活动
10	399	邓亚萍资助贫困近视学子进行激光手术
13	544	身在英伦读博士 心系津门自强儿
18	74	邓亚萍捐献奖金 帮助贫困学子治疗近视
20	758	亚萍将“行善”来金华
21	282	乒坛邓亚萍金华献爱心
26	804	为贫困学子摘掉眼镜

图 2 以图形的方式反映对比结果,每个主题左边紫色的柱状图表示原方法的查全率,右边红色的柱状图表示新方法的查全率。

(2) 查准率:返回检索结果中符合用户需要的网页数量与返回网页数量的比值:

表 6 新方法与原方法对各主题准确率的统计对比

主题	Top-N	查询词(关键词+扩展词)	原方法	新方法	原查准率	新查准率	提高倍数
公益活动	30	邓亚萍 学子 活动 公益	1	23	0.033333	0.76667	22
工作	80	邓亚萍 工作 运动员 奥运会 北京	22	65	0.275	0.8125	1.954545
学习	80	邓亚萍 学习 大学	10	64	0.125	0.8	5.4
家庭	130	邓亚萍 孩子 法国 中国	18	110	0.138462	0.84615	5.111111
平均查准率					0.142949	0.80633	4.640695

**总结与展望** 从目前的商业应用看,网络信息服务已经进入以用户为中心的更先进的理念竞争阶段,网络只有有了用户才能产生经济效益。

如何吸引更多用户的注意力、如何使用户更长时间地留在自己的网站、如何使用户满意等一系列问题,都要依靠提供个性化的信息服务才能解决。因此个性化搜索引擎产品必然具有很大的市场前景。本文对提出的个性化推荐模型进行了系统的分析和评述,实验结果表明它有效地实现了搜索引擎的个性化服务。由于它并没有改变搜索引擎本身的结构,合

$$Precision = \frac{\text{返回的相关网页数}}{|Top-N 推荐集|} \quad (5)$$

由于互联网上信息的动态更新,不同时间和使用不同查询词进行搜索,返回的结果集是不同的。实验表明在增加扩展词后,Google 返回的结果集与原集不同,出现了一些在原集中不存在但与查询主题相关的网页。因此,准确率表示为返回的 Top-N 推荐集中符合用户需要的网页数量与该推荐集网页总数 ( $|Top-N 推荐集|$ ) 的比值。根据公式(5)求出原方法与新方法所得推荐结果集的准确率,表 6 与图 3 是得到的对比结果。实验表明,添加了几个反映用户主题偏好的扩展词后,返回的文档集中与用户兴趣相关的网页明显增加,从而提高了搜索的准确率。表 6 第 4 列“原方法”表示原方法的 Top-N 推荐集中与主题相关的网页数,第 5 列“新方法”表示新方法的 Top-N 推荐集中与主题相关的网页数。从表 6 可以看出,新算法推荐的准确率比原方法平均提高了 4.64 倍,体现了搜索引擎个性化服务的效果。

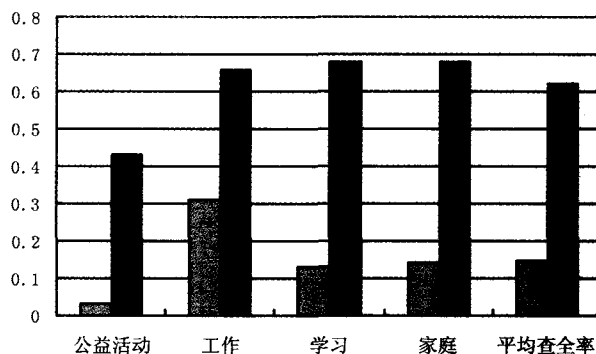


图 2 各主题查全率对比图

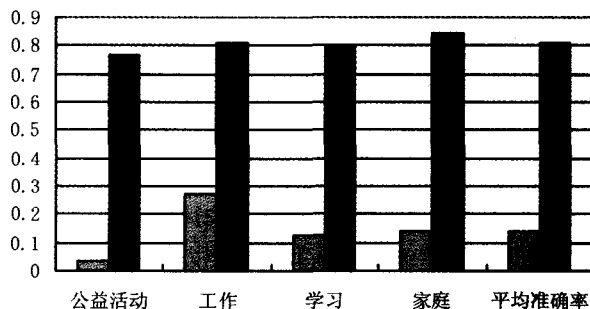


图 3 各主题准确率对比图

理地利用目前搜索引擎实时响应、返回信息全面等优点,因此本模型具有很大的应用前景。但还有一些工作有待解决,比如对搜索结果进一步过滤,推荐给用户高质量的信息等。

### 参考文献

1 Jansen B J, Spink A, Bateman J, et al. Real life information retrieval: a study of user queries on the web. SIGIR Forum, 1998, 32(1):5~17

# Web 服务器系统中基于反馈控制的比例延迟保证

潘文平 慕德俊 戴冠中 武航星

(西北工业大学自动化学院 西安 710072)

**摘要** 随着 Web 应用商业领域的广泛使用, Web 服务器系统需要在高负载下提供区分服务, 以满足用户的不同需求。为实现以延迟作为评价指标的区分服务, 本文在 Web 服务器系统的连接管理和请求处理两个层次建立了基于反馈控制的比例延迟保证模型。模型中的反馈控制器通过动态计算和调节不同类别客户占用的资源:(服务线程和数据库连接), 能保证高优先级的客户较快得到服务而不同类别客户的平均延迟比保持不变。为测试闭环系统的性能, 设计了两种分别服从均匀分布和重尾分布的动态负载。仿真结果表明, 即使并发客户连接的数目剧烈变化, 控制器作用下的服务器系统仍然能够达到较好的比例延迟保证, 可靠地为用户提供区分服务。

**关键词** 反馈控制, 比例区分服务, Web 服务器

## Feedback Control-based Proportional Delay Guarantees in Web Server Systems

PAN Wen-Ping MU De-Jun DAI Guan-Zhong WU Hang-Xing

(College of Automation, Northwestern Polytechnical University, Xi'an 710072)

**Abstract** As more business applications become Web enabled, Web server systems evolve to provide service differentiation to satisfy different requirements of clients under heavy load conditions. To implement service differentiation with respect to delays in the Web server system, in this paper, proportional delay guarantee models based on feedback control are constructed at connection management level and request processing level. Controllers in the models can dynamically calculate and adjust the amount of resources, the worker threads and the database connections, for different types of clients according to their proportional delay requirements. Two kinds of workloads, which follow uniform and heavy-tailed distributions respectively, are generated for simulation of the close-loop system. Experiment results indicate that, proportional delay guarantees can be achieved in the Web server system successfully even if the number of the concurrent clients is changed abruptly under different kinds of workloads.

**Keywords** Feedback control, Proportional differentiation service, Web server

### 1 Web 应用的延迟及相关工作

在 HTTP 协议之下, 客户通过 TCP 连接向 Web 服务器或 Web 应用服务器发送请求, 并接收服务器返回的响应。从客户的角度来看, 对于 Web 应用, 其端到端的延迟主要由以下几个方面构成: (1) 连接延迟。指 TCP 连接从建立到被服务线程或服务进程接受的时间间隔, 即已经完成的连接在 Accept Queue 中的排队时间。图 1(2) 显示了已建立的 TCP 连接进入 Accept Queue 的情形。当 accept() 被调用时, 此队列队首的连接出队并被分配给服务线程或服务进程。(2) 处理延迟。即服务器读取客户请求, 生成响应所需的时间。(3) 传输延迟。主要指在客户与服务器间通过三次握手建立 TCP 连接(如图 1<sup>[1]</sup>所示), 以及通过 TCP 连接传输请求和响应所需的时间。

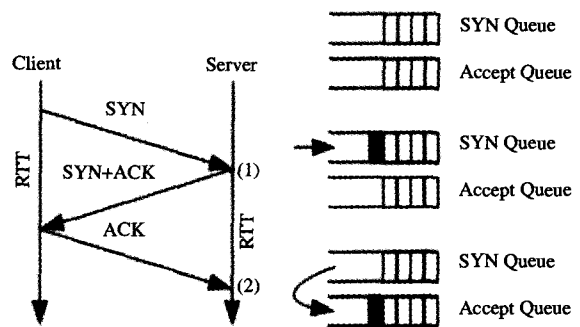


图 1 三次握手与连接延迟

为在服务器上实现以延迟作为评价指标的区分服务和性能保证, 近年来人们展开了大量工作。当发送请求的并发的客户数目多于 Web 服务器如 Apache 等设置的最大服务线程

潘文平 博士研究生, 主要研究方向为 Web QoS 控制、网络拥塞控制; 慕德俊 教授, 博士生导师, 主要研究领域为计算机网络、信息安全; 戴冠中 教授, 博士生导师, 主要研究领域为自动控制、复杂系统与复杂性理论等; 武航星 博士研究生, 主要研究方向为网络拥塞控制。

- 朱征宇, 裴仰军, 陈华月, 等. 个性化服务中用户近期兴趣视图的生成. 计算机工程与设计, 2005(4): 952~953
- Deerwester S, Dumais S T, Furnas G W, et al. Indexing by latent semantic analysis. Journal of the American Society for Information Science, 1990, 41(6): 391~407
- Salton G. Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer [ED/OL]. Addison-Wesley, <http://citeseer.nj.nec.com/content/26897/0>, 1989
- Zhang Yuejie, Zhang Tao, Chen Shijie. Research on Lucene-based English-Chinese Cross-Language Information Retrieval.

- Journal of Chinese Language and Computing, 15 (1): 25~32
- Wang S, Tanaka Y. Topic-oriented query expansion for web search. In: Proceedings of the 15th International Conference on World Wide Web, Edinburgh, Scotland, May 2006
- 张敏, 宋睿华, 马少平. 基于语义关系查询扩展的文档重构方法. 计算机学报, 2004(10): 1398
- 瞿瑞彩, 谢伟松. 数值分析, 天津, 天津大学出版社, 2001. 2~3
- 赵亮, 胡乃静, 张守志. 个性化推荐算法设计. 计算机研究与发展, 2008(8): 989~990