基于用户行为分析的文献阅读价值评估方法*

尹盼盼，崔荣一†
（延边大学 计算机科学与技术学科 智能信息处理研究室，吉林 延吉 133002）

摘　要：综合文献自身客观价值和相对于查询用户的主观价值，提出了基于用户行为分析的文献阅读价值评估方法。首先，分析用户行为并构建下载文献标题的潜在语义空间，计算文献与下载文献的标题语义相关性，计算文献摘要、关键词中出现下载文献标题词的频次，进而计算其与用户行为的相关性；其次，采用期刊影响因子、论文加权被引频次、时间因子等指标，定量评估文献的自身价值；最后，综合主客观因素定量评估的结果，评估文献的阅读价值。实验表明，提出的文献阅读价值评估方法比基于单一方面因素评估文献阅读价值的传统方法更加合理、有效。

关键词：用户行为分析；文献阅读价值；潜在语义索引；文献分析；文献评价
中图分类号：TP391.4　　**文献标志码**：A　　**文章编号**：1001-3695(2012)04-1285-04
doi:10.3969/j.issn.1001-3695.2012.04.023

Evaluation of literature reading value based on user behavior

YIN Pan-pan, CUI Rong-yi†
(*Intelligent Information Processing Laboratory, Dept. of Computer Science & Technology, Yanbian University, Yanji Jilin* 133002, *China*)

Abstract: This paper proposed a literature evaluation method based on analysis of user behavior by synthesizing objective self-values of literature and subjective values relative to the query users. Firstly, in order to compute the relevance between literature and user behaviors, it analyzed users' behaviors and constructed latent semantic space of titles of downloaded literatures, computed the semantic relevance between titles of literatures in archives and downloaded literatures, and furthermore, it computed the frequencies of title-words in downloaded literatures which appeared in keywords and abstracts of literatures. Secondly, it evaluated self-values of literatures using impact factors, citation-weighted literatures and time factor. Finally, it evaluated reading values of literatures by synthesizing the results of objective and subjective factors. The experimental results show that the proposed evaluation method outperforms the method based on factors of single aspect.

Key words: user behavior analysis; literature reading value; latent semantic indexing; literature analysis; literature evaluation

0　引言

科学技术的高速发展促使科学研究日益专门化和综合化，作为创新主体的科技工作者对快速获取有用的信息和知识的需求更为迫切。由于大量知识存在于科技文献之中，因此科技文献的准备成为开展科研工作所必需的步骤和重要基础。然而在现实情况中，很多学者，特别是刚刚从事科研工作的研究者，在准备文献方面由于缺乏经验，缺少必要的指导，往往感到无从下手且耗费时间。而目前的电子检索系统只是将根据关键词搜索出的文献进行简单的罗列，或者只是按个别指标（如被引次数、发表时间等）进行简单排序，并不能为检索者提供有价值的信息和高效用的服务。于是，如何评估文献的阅读价值，如何向查询用户提供客观价值与主观需求相融合的知识服务并显著节约用户的时间，是一个亟待解决的、具有重要现实意义的课题。

文献[1,2]主要是模拟学者判断一篇文献的阅读价值的过程，综合分析一篇文章所在的期刊、文献作者、文献内容以及文献发表时间四个因素来评估一篇文献的阅读价值。但文献的阅读价值是一个相对的概念，它包含文献自身价值（公认的客观性评价）和相对于查询用户的价值（个体需求）主客观两方面的因素，它的高低有无与读者、文献主题相关性等有很大的关系。

本文从文献的自身价值和查询用户对文献的个体需求主客观两方面出发，全面衡量一篇文献的阅读价值，对反映用户研究背景和关注点的用户行为进行了分析，提出了文献相对于查询用户的主观价值以及文献自身客观价值的定量评估模型，并根据两方面因素的评价结果建立模型，全面评估文献的阅读价值。实验表明，本文提出的文献阅读价值评估方法比基于单一方面因素评估文献阅读价值的传统方法更加合理、有效。

1　用户行为分析

1.1　用户下载行为

信息和网络技术的发展，使得信息资源的数字化和电子化成为了可能，从而使得电子文献服务逐渐成为当代文献服务的重要组成部分。大量的电子期刊数据库给科技工作者开展科研准备工作提供了很大的便利，如中国知网（CNKI）、维普中文

收稿日期：2011-10-09；**修回日期**：2011-11-18　　**基金项目**：吉林省科技厅资助项目(20050703-1)

作者简介：尹盼盼(1987-)，女，山东聊城人，硕士研究生，主要研究方向为文本检索、模式识别；崔荣一(1962-)，男(朝鲜族)(通信作者)，吉林延吉人，教授，硕导，博士，CCF 会员，主要研究方向为模式识别、智能计算(cuirongyi@ybu.edu.cn).

科技期刊数据库、万方数据的 iLib 系统等信息服务数据库产品。

用户在使用这些文献检索系统时，系统会记录用户的查询信息以及返回的相关文档和点击信息，将这些信息称为查询行为[3]。用户提交查询后，检索系统会将用户输入的检索信息、选择浏览的文献信息以及用户下载的文献信息写入查询日志；与此同时，日志会记录用户下载文献的题名、来源数据库、下载页数、下载时间等信息。这些信息即为用户的下载行为。用户下载的文献往往是用户在接下来开展科研工作中需要进一步阅读的文献，这些文献的主题往往与用户的研究背景一致，并能反映用户的兴趣点与关注点。

1.2 文献与用户行为相关性的定量评估

科技文献的标题部分很精简地反映了文章的核心内容，同时与文档主题内容不相关的描述成分很少，是对一篇文献重要内容的概括[4]，因此文献标题中出现的关键词可以作为本篇文献的主题词，文献检索系统日志中记录的用户下载文献的标题信息能够反映用户的研究背景以及用户感兴趣的主题内容。

本文采用向量空间模型(vector space model，VSM)[5]来表示系统日志中记录的每一篇文献的标题信息。在 VSM 中，一篇文献的标题看成是由相互独立的若干词条(term)组成的集合，即 $TD_j=(td_{1j},td_{2j},\cdots,td_{mj})'$。对于每一个词条 td_{ij}，根据其在文献标题中隐含的语义及重要程度赋予一定的权值，文献标题的特征向量为 $(wd_{1j},wd_{2j},\cdots,wd_{mj})'$。利用权重评价函数 $tf\times idf$ 定义每个特征项的权重为

$$wd_{ij}=\frac{tf_{ij}\times\log(N/n_i)}{\sqrt{\sum_{i=1}^{m}(tf_{ij})^2\times[\log(N/n_i)]^2}} \tag{1}$$

其中：tf_{ij} 表示项 td_i 在文献标题 TD_j 中的文档内频数；idf_i 采用 $idf_i=\log(N/n_i)$ 表示项的反比文档频数；N 表示系统日志中记录的文献标题的总数目；n_i 表示项的文档频数。系统日志中记录的所有下载文献的标题表示为词项—文档矩阵：

$$A=[TD_1,TD_2,\cdots,TD_N]=\begin{bmatrix}wd_{11}&wd_{12}&\cdots&wd_{1N}\\wd_{21}&wd_{22}&\cdots&wd_{2N}\\\vdots&\vdots&&\vdots\\wd_{m1}&wd_{m2}&\cdots&wd_{mN}\end{bmatrix} \tag{2}$$

其中：第 i 行表示第 i 个词项 td_i 对各文献标题描述的贡献，第 j 列表示第 j 个文献标题 TD_j 中各词项的贡献。

本文采用潜在语义索引方法构造下载文献标题的词项—文档矩阵的潜在语义空间[6]，在用户行为潜在语义空间中计算文献库中各文献标题与下载文献标题的语义相关性。利用奇异值(SVD)分解方法，对 A 进行分解：

$$A=U\Sigma V^{\mathrm{T}} \tag{3}$$

其中：r 是矩阵 A 的秩；矩阵 U 的各列向量 $u_i(i=1,2,\cdots,r)$ 是矩阵 AA^{T} 的对应于特征值 σ_i^2 的左奇异向量；矩阵 V 的各列 $v_j(j=1,2,\cdots,r)$ 向量是矩阵 AA^{T} 的对应于特征值 σ_i^2 的右奇异向量，$\sigma_i(i=1,2,\cdots,r)$ 称为矩阵 A 的奇异值。当矩阵 A 的奇异值 $\sigma_i(i=1,2,\cdots,r)$ 衰减很快时，可以忽略一部分奇异值，得到矩阵 A 的近似表示。设选取 k，使式(4)成立：

$$\varphi=\frac{\sum_{i=1}^{k}\sigma_i^2}{\sum_{i=1}^{r}\sigma_i^2}>T\quad 1\leqslant k\leqslant r,0<T<1 \tag{4}$$

其中：T 为设定的阈值，反映高特征值占全体特征值的百分比，本文中取 $T=90\%$。由此得到 A 在最小平方误差意义下的 k-秩最佳近似：$A_k=U_k\Sigma_k V_k^{\mathrm{T}}$。其中：$k$ 是词项和文档近似表示空间的维数；U 和 V 删除相应的列得到 U_k 和 V_k；Σ 删除相应的行和列得到 Σ_k。

构建下载文献标题潜在语义空间后，通过投影 $P_k=U_kU_k^{\mathrm{T}}\in R^{m\times m}$，将原始文本向量投影到 k 维潜在语义空间中。文献库中各文献标题的原始文本向量 T_i 通过投影矩阵在潜在语义空间中表示为：$T_1=P_kT_i$。文献库中第 i 篇文献标题 T_i 与潜在语义空间中第 j 篇下载文献标题 TD_j 的相关性计算为

$$\mathrm{sim}(T_i,TD_j)=\frac{(P_kT_i)^{\mathrm{T}}(P_kTD_j)}{\|P_kT_i\|\cdot\|P_kTD_j\|}=\frac{(T_1)^{\mathrm{T}}(P_kTD_j)}{\|T_1\|\cdot\|P_kTD_j\|} \tag{5}$$

系统日志中记录的下载文献的总数目为 N，文献库中第 i 篇文献的标题与每篇下载文献标题的语义相关性如下所示：

$$\mathrm{sim}(T_i)=\begin{pmatrix}\mathrm{sim}(T_i,TD_1)\\\vdots\\\mathrm{sim}(T_i,TD_j)\\\vdots\\\mathrm{sim}(T_i,TD_N)\end{pmatrix} \tag{6}$$

文献的摘要精练地反映了整篇文档的主要内容，是对文献全文内容的概括，文献的关键词则是以词组形式体现了文档的关键内容[7]。文献的摘要和关键词能够正确地反映文档的主题内容，而且处理文档整篇内容的代价太高，对于大型文档更不可以采用这种整篇处理的方法。下载文献标题中各词项在文献库中文献摘要、关键词中的出现次数可以在一定程度上反映所下载文献与文献库中文献在主题内容上的相关性。每篇下载文献的标题词在文献库中第 i 篇文献摘要、关键词中出现的次数如下所示：

$$\mathrm{num}(K_A_i)=\begin{pmatrix}\mathrm{num}(K_A_i,tD_1)\\\vdots\\\mathrm{num}(K_A_i,tD_j)\\\vdots\\\mathrm{num}(K_A_i,tD_N)\end{pmatrix} \tag{7}$$

其中：$\mathrm{num}(K_A_i,tD_j)$ 表示第 j 篇下载文献标题中的所有词项在文献库中第 i 篇文献关键词、摘要中出现的总频次。

文献库中第 i 篇文献相对于用户行为的相关性定量评估 V_i_reader 如式(8)所示：

$$V_i_\mathrm{reader}=\mathrm{sim}(T_i)\times\mathrm{num}(K_A_i) \tag{8}$$

2 文献自身价值

2.1 期刊水平

文献的期刊水平是一项综合评估论文发表期刊的学术水平的指标。本文采用文献所在期刊的影响因子来评价文献的期刊水平。影响因子是 Garfield 于 1972 年提出，指某期刊前两年发表的论文在统计当年的平均被引用次数，即该期刊前两年发表论文在统计当年被引证次数/该期刊前两年发表的论文总数[8]。期刊论文的影响因子较好地量化了该期刊在学术研究方面的学术影响力以及学术思想传统的深度和广度，能够比较公平地评价各类期刊。影响因子越大，表明论文平均被引证的次数越多，期刊在学科发展中的作用越大、质量越高。

2.2　文献被引频次

文献的影响力主要体现在被引用上。被引频次的高低反映论文被同行重视的程度以及论文在科学发展和学术交流中所起的作用和影响力,也间接反映了该文的学术水平和价值[9]。一般来说,论文被引用得多,说明它在该专业领域内具有一定的影响力和权威性,尤其是被高质量的论文引用更是说明了论文本身的水平。

本文的文献被引频次用论文加权被引频次(F)来反映,论文加权被引频次不仅与论文被引用的次数有关,还与引用论文的期刊影响因子有重要关系。一篇论文被 k 种期刊所引用,引用期刊的影响因子值分别为 $F_{i1},F_{i2},\cdots,F_{ik}$,引用次数分别为 $n_1,n_2,\cdots,n_k$,则论文加权被引频次为

$$F_0=\sum_{j=1}^{k}n_iF_{ij}=n_1F_{i1}+n_2F_{i2}+\cdots+n_kF_{ik} \tag{9}$$

发表不久的新文献,其被引用次数可能较低,甚至出现零被引的情况,因此需要对式(9)所得的论文加权被引频次进行修正,修正后的论文加权被引频次如式(10)所示,其中 $\alpha=1$ 为修正值:

$$F=\alpha+F_0 \tag{10}$$

2.3　文献时间因子

用户在使用文献检索系统查找文献时,用户提交查询后,系统将文献库中与用户查询相关的文献返回给用户,并按照某一指标顺序(如相关度、时间等)对返回结果排序,这时用户更倾向于查看最新发表的文献。新近发表的文献,其被引次数可能较低,但是它可能反映了在某方面研究的新进展,有很强的阅读价值。

为突出这样新近发表文献的阅读价值,本文采用文献[10]中引入的时间因子,即

$$D(p)=(t-\min_{k\in P}\{T(k)\})/(\max_{k\in P}\{T(k)\}-\min_{k\in P}\{T(k)\}+1) \tag{11}$$

其中:T 是所有文献的发表年份构成的 $n\times1$ 的年限矩阵,n 为文献的总数;$D(p)$ 是文献 p 的时间因子;t 是文献 p 发表的年份;P 是所有文献组成的集合。

2.4　文献自身价值定量评估

文献的自身价值是从客观角度对文献进行的定量评价,本文主要采用文献所在期刊的影响因子(IF)、文献被引情况(F)、文献年限(D)这三个指标对文献的自身价值进行定量评估。定量评估的模型如式(12)所示:

$$V_i_\text{self}=IF\times F\times D(p) \tag{12}$$

3　文献阅读价值的综合评估

知识服务,就是从用户的角度出发,为用户提供解决问题的基于知识的深层次的服务[11]。本文以知识服务思想为指导,综合文献与用户行为的相关性以及文献自身价值两方面的定量评估结果,对文献阅读价值进行全面的评估。综合评估方法如式(13)所示,β 取 0.6:

$$\text{Value}_i=\beta\times V_i_\text{reader}+(1-\beta)\times V_i_\text{self} \tag{13}$$

4　实验结果及分析

4.1　实验过程

1)本实验主要是为了验证综合文献主客观两方面因素对文献阅读价值进行定量评估方法的合理性和有效性,实验将本文方法与现有文献检索系统基于单一方面因素判断文献阅读价值的排序结果进行比较。

2)在中国知网上输入“潜在语义分析”进行主题查询,返回结果中选定“计算机软件及计算机应用”学科领域,系统共返回 112 篇期刊论文,收集这 112 篇文献的标题、摘要、关键词信息,建立本地文献库;进入中国知网中“我的 CNKI”,查看系统日志中记录的本人一段时间内所下载文献的标题信息并保存到本地,用做用户下载行为的分析数据。

3)统计文献库中各文献所在期刊的影响因子,根据文献的被引频次及本文引证文献所在期刊的影响因子计算文献的加权被引频次,统计各文献的发表时间,计算文献对应的时间因子。文献库中部分文献对应期刊影响因子、论文被引频次、加权被引频次、发表时间、时间因子五个指标计算的结果如表 1 所示。

表 1　文献库中部分文献对应的五个指标

文献序号	文献标题	所在期刊	期刊影响因子	文献被引频次	文献加权被引频次	发表时间	时间因子
61	潜在语义分析中词汇—文本矩阵奇异值分解的并行实现	计算机应用与软件	0.515 0	0	1	2009/02	0.818 2
62	基于衰减词共现图的多文档摘要研究	小型微型计算机系统	0.460 0	0	1	2009/01	0.818 2
63	一种基于聚类分析的 BBS 主题发现算法研究	安徽师范大学学报	0.422 0	2	1.803 0	2009/01	0.818 2
64	利用信息检索和概率有限自动机的程序理解	浙江大学学报	1.025 0	0	1	2008/12	0.727 3
65	基于潜在语义分析的 BBS 主题发现算法研究	电脑知识域技术	0.162 0	0	1	2008/29	0.727 3
66	基于潜在语义分析的自动词法学习技术	情报理论与实践	1.500 0	0	1	2008/05	0.727 3
67	基于潜在语义分析的中文文本聚类方法研究	科技情报开发与经济	0.317 0	2	1	2008/27	0.727 3

4)对文本库中的文献数据进行分词处理。本文的分词软件采用中国科学院开源的中文分词工具 ICT-CLAS[12]。对文档库中的下载文献标题信息以及各文献进行分词,选择出现的名词、动词、形容词、区别词、字符串作为标志文本特征的词项。

5)计算各词项在文本中的出现权重,用 SVM 表示文献信息以及下载文献的标题信息,构造所有下载文献标题信息的词项—文档矩阵,利用 MATLAB 中的 SVD 工具包对词项—文档矩阵进行奇异值分解,构造用户行为的潜在语义空间。

6)对文献库中的每一篇文献,将其标题信息投影到用户行为潜在语义空间中,计算该文献标题与各下载文献标题的语义相关性,统计各下载文献标题词在文献库中各文献的摘要、关键词中出现的频次,计算文档库中各文献相对于查询用户的主观价值。

7)根据文献所在期刊影响因子、文献加权被引频次、文献时间因子计算文档库中各文献的自身价值。

8)综合文献自身价值和查询用户对文献的个体需求的定量评估结果,计算文献库中各文献的文献阅读价值,结果降序排列。

4.2　结果与分析

实验 1　验证查询用户对文献个体需求定量评估模型的合理性。依次读取文献库中各个文献,计算该文献与所有已下载文献标题的语义相关性,统计各下载文献标题词在该文献摘要、关键词中出现的频次,结合语义相关性和频次计算文献与用户行为的相关性。文献库中所有文献处理完后,对计算结果降序排列。表 2 为评估结果排在前十位的文献序号及其对应

相关性。

表 2 实验 1 的结果

文献序号	文献标题	与用户行为的相关性
68	Deep web 站点查询界面的潜在语义分析	118.229 9
107	中文信息的潜在语义分析	104.846 6
71	一种基于潜在语义分析和直推式谱图算法的文本分类方法 LSASGT	102.641 7
75	基于分块潜在语义的场景分类方法	96.071 9
98	基于潜在语义分析的汉语问答系统答案提取	93.708 6
36	基于潜在语义分析和改进的 HS-SVM 的文本分类模型研究	92.492 5
4	一种基于 LDA 的潜在语义区划分及 Web 文档聚类算法	91.345 0
106	潜在语义分析在中文信息处理中的应用	83.944 3
67	基于潜在语义分析的中文文本聚类方法研究	83.048 2
46	基于潜在语义分析的网页文本分类研究	82.269 8

CNKI 文献检索只能根据文献库中文献与用户输入查询条件的相关性这一单一因素返回查询结果,这样的过程具有很大的盲目性,而且用户通过查询条件并不能很好地描述其真正想要的文献,如 CNKI 的返回结果中就有很多有关“情感分析”“概率潜在语义分析”的文献,这些文献或许并不是用户真正想要查询的文献。

对用户的行为进行分析,将与用户行为相关性较大的文献返回给用户,能够针对用户的兴趣点、关注点返回相关文献。实验结果验证了本文提出的使用者对文献个人需求定量评估模型的合理性和有效性。

实验 2 验证文献自身价值定量评估模型的合理性。依次处理文献库中各个文献,采用该文献所在期刊的影响因子、论文加权被引频次、文献时间因子三个指标计算该文献的自身价值,对计算结果降序排列。表 3 为文献自身价值排在前十位的文献序号及其对应的文献自身价值定量评估的结果。

表 3 实验 2 的结果

序号	文献所在期刊	期刊影响因子	1 + 加权被引频次	发表时间	文献自身价值结果
98	计算机学报	1.796	7.560	2006	7.406 1
95	计算机研究与发展	1.117	7.057	2006	5.016 2
104	中文信息学报	1.130	5.553	2005	2.852 2
60	计算机工程与设计	0.789	1.604	2009	2.384 0
90	东北师大学报	0.699	5.088	2007	2.263 2
111	计算机工程	0.492	7.414	2004	2.081 2
94	情报科学	2.416	1.000	2007	1.537 5
58	小型微型计算机系统	0.460	3.862	2009	1.453 5
71	电子学报	0.686	2.744	2009	1.369 1
26	情报理论与实践	1.500	1.000	2010	1.363 6

由表 3 中数据可以很明显地看出,排在前十位的文献其所在期刊均为相应领域的核心期刊,而这些期刊上发表的文献都通过了相关领域专家严格的评议和质量审查,具有较高的自身价值。表 3 中数据验证了本文提出的文献自身价值评估模型的合理性。

实验 3 验证文献阅读价值综合评估模型的合理性。运用实验 1 和 2 中所得定量评估的结果计算文献的阅读价值,对计算结果降序排列。表 4 为文献阅读价值排在前十位的文献的序号、标题、所在期刊及其对应的文献阅读价值评估的结果。

观察表 4 中的数据发现,其与表 2 中实验 1 的数据很相似,其中文献 75 和 98 因为文献自身价值的影响,顺序发生了调整,文献 46 受到其文献自身价值较低的影响,文献的阅读价值有所下降。由表 4 数据可得出,本文文献阅读价值评估方法很好地结合了其相对查询用户的价值以及文献自身价值主客观两方面的因素,既能够返回给用户其感兴趣的文献,也同时能返回给用户学术水平较高的文献,验证了本文提出的文献阅读价值综合评估模型的合理性和有效性。

表 4 实验 3 的结果

序号	文献标题	文献所在期刊	文献阅读价值
68	Deep Web 站点查询界面的潜在语义分析	Journal of Southeast University(English Edition)	71.062 9
107	中文信息的潜在语义分析	华南理工大学学报	63.179 2
71	一种基于潜在语义分析和直推式谱图算法的文本分类方法 LSASGT	电子学报	62.132 6
98	基于潜在语义分析的汉语问答系统答案提取	计算机学报	59.187 6
75	基于分块潜在语义的场景分类方法	计算机应用	57.970 2
36	基于潜在语义分析和改进的 HS-SVM 的文本分类模型研究	图书情报工作	55.929 3
4	一种基于 LDA 的潜在语义区划分及 Web 文档聚类算法	中文信息学报	55.258 9
106	潜在语义分析在中文信息处理中的应用	计算机工程与应用	50.831 3
67	基于潜在语义分析的中文文本聚类方法研究	科技情报开发与经济	49.921 1
2	改进的概率潜在语义分析下的文本聚类算法	计算机应用	49.614 8

5 结束语

文献的阅读价值是一个相对的概念,它包含文献自身价值和查询用户对其是否认同主客观两方面因素。本文克服了现有文献阅读价值评估方法中只依靠单一方面因素进行评估的不足,提出了综合文献自身客观价值和相对于查询用户的主观价值两方面的因素,全面衡量一篇文献的阅读价值的方法。本文对用户使用文献检索系统的行为进行了分析,建立了基于潜在语义索引技术定量评估查询用户对文献个人需求的模型,建立了根据文献期刊水平、被引频次、发表年限等因素定量评估文献自身价值的模型。实验证明,本文提出的文献阅读价值评估方法比传统的基于单一方面因素评估文献阅读价值的方法更加合理、有效。

本文只根据用户的文献下载行为对用户行为进行了初步分析,对各种用户行为信息的深入挖掘是需要进一步研究的问题。另外,本文方法在评价文献自身价值时,只采用了三种定量指标进行评估,文献内容的重要性不仅体现在这些定量指标上,还受文献主题前沿性的影响,对影响文献自身价值的文献内容主题前沿性的挖掘分析也是今后进一步需要研究的问题。

参考文献:

[1] ZHANG Guang-qian, LIU Xin. Study on the method of ranking scientific papers[C]//Proc of International Conference on e-Business and e-Government. Washington DC: IEEE Computer Society, 2010: 3062-3066.

[2] ZHANG Guang-qian, LIU Xin. Research on recommendation of scientific papers based on reading value[C]//Proc of International Conference on Engineering and Business Management. Irine, California: Scientific Research Publishing, 2010: 4412-4417.

[3] 蒋在帆, 王斌. 基于用户行为分析的个人信息检索研究[J]. 中文信息学报, 2011, 25(1): 9-14, 122.

[4] KO Y, PARK J, SEO J. Automatic text categorization using the importance of sentences[C]//Proc of the 19th International Conference on Computational Linguistics. Stoudsburg, PA: ACL, 2002: 1-7.

(下转第 1331 页)

$$N_3 = P^{\mathrm{T}}(A_d + BMK_2), L_1 = (E_2 + E_3MK_2)^{\mathrm{T}}$$

对不等式(17)分别左乘和右乘矩阵 $\mathrm{diag}\{P^{-\mathrm{T}}\ P^{-\mathrm{T}}\ I\ I\ I\ I\}$, $\mathrm{diag}\{P^{-1}\ \ P^{-1}\ I\ I\ I\ I\}$,并令 $X^{\mathrm{T}} = P^{-1}$, $W = KP^{-1}$, $\bar{Q} = P^{-\mathrm{T}}QP^{-1}$,则矩阵不等式(17)与线性矩阵不等式(11)等价。又因为 X 可逆,所以控制器增益 $K_1 = YX^{-\mathrm{T}}$, $K_2 = ZX^{-\mathrm{T}}$,进而鲁棒 H_∞ 容错控制器为 $u(t) = YX^{-\mathrm{T}}x(t) + ZX^{-\mathrm{T}}(t - d(t))$。定理得证。

推论 1　对于系统式(1)中所有允许的不确定性和故障矩阵 M,由定理 1 可知,若以下的优化问题

$$\min_{\gamma}(X, Y, Z, \varepsilon_i, \gamma)\ \text{s. t. LMIs (8)\&(9)}$$

有一个最小值 $\gamma_{\min}$ 和一组可行解 (X^*, Y^*, Z^*),则 $u(t) = Y^*X^{*-\mathrm{T}}x(t) + Z^*X^{*-\mathrm{T}}(t - d(t))$ 是系统式(1)的鲁棒容错最优 H_∞ 控制器,相应的最小衰减干扰指标是 $\gamma_{\min}$。

3　仿真设计

考虑不确定线性时滞广义系统式(1),其参数如下:

$$E = \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix}, A = \begin{bmatrix} -0.9 & -0.3 \\ -0.6 & -0.8 \end{bmatrix}, A_d = \begin{bmatrix} 0.3 & 0.4 \\ 0.2 & -0.5 \end{bmatrix}, B = \begin{bmatrix} 0.3 & 0.4 \\ 0.2 & -0.5 \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, G = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}, H = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.1 \end{bmatrix}, E_a = \begin{bmatrix} 0.2 & 0.2 \\ 0.3 & 0.1 \end{bmatrix}$$

$$E_d = \begin{bmatrix} 0.2 & 0.1 \\ 0.3 & 0.1 \end{bmatrix}, E_b = \begin{bmatrix} 0.1 & 0.3 \\ 0.2 & 0.1 \end{bmatrix}, \gamma = 0.8, \eta = 0.4$$

考虑执行器故障矩阵 $M = \mathrm{diag}[0.8\quad 0.5]$,利用 MATLAB 中的 LMI 工具箱,求解可得

$$X = \begin{bmatrix} 0.2816 & 0.0482 \\ 0.0482 & 0.0682 \end{bmatrix}, \bar{Q} = \begin{bmatrix} 1.3426 & 0.0072 \\ 0.0072 & 1.5251 \end{bmatrix}$$

进一步求得

$$K_1 = YX^{-1} = \begin{bmatrix} -0.4746 & -34.3840 \\ -16.7746 & 56.8511 \end{bmatrix}$$

$$K_2 = ZX^{-1} = \begin{bmatrix} 0.0777 & 0.3769 \\ -0.6053 & -0.1220 \end{bmatrix}$$

当系统初始条件为 $x(0) = [1, 1]^{\mathrm{T}}$,对执行器在故障 $M = \mathrm{diag}(0.8, 0.5)$ 情形下,其状态 x_1 和 x_2 的零输入响应曲线分别如图 1、2 所示。图中,曲线 1 为执行器正常时系统的零输入响应,曲线 2 为执行器发生 M 故障时状态 x_1 和 x_2 的零输入响应。从仿真结果可以看出,系统在含有执行器故障时仍能是渐近稳定的,说明本文提出的方法是有效的。

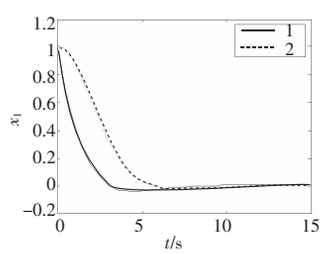

图 1　状态 x_1 的零输入响应

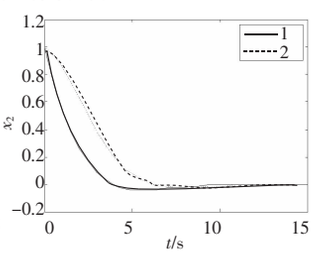

图 2　状态 x_2 的零输入响应

4　结束语

本文基于 Lyapunov 稳定性理论,研究了一类广义时滞不确定系统在有记忆状态反馈作用下的鲁棒 H_∞ 容错控制问题,同时考虑了系统对外部干扰信号抑制能力的性能指标,给出了在执行器增益故障情况下系统仍能渐近稳定的充分条件。仿真结果表明,该方法计算简便,可以方便求得鲁棒容错控制器,且控制器有很好的鲁棒性和容错性。

参考文献:

[1] 陈跃鹏,周祖德. 广义系统的鲁棒控制与容错控制[M]. 北京:科学出版社,2010.

[2] 陈跃鹏,周祖德,刘焕彬,等. 离散广义系统具有完整性的鲁棒二次稳定[J]. 控制理论与应用,2006, 23(1):14-18.

[3] CHEN Yue-peng, ZHANG Qing-ling. Integrity controller design for descriptor systems[J]. International Journal of Information and Systems Sciences, 2005, 1(3-4): 355-363.

[4] YAO Li-na, WANG Hong. Design of a fault diagnosis scheme for a class of singular nonlinear systems[J]. Journal of Control Theory and Applications, 2008, 6(2): 122-126.

[5] 张爱清,方华京. 时滞不确定广义系统的鲁棒 H_∞ 容错控制[J]. 自动化技术与应用,2007,26(5):16-19.

[6] 陶洪峰. 胡寿松. 模糊广义时滞系统的鲁棒集成容错控制器设计[J]. 南京航空航天大学学报,2009,41(3):397-402.

[7] 张显,段广仁. 离散广义线性时滞系统的有记忆状态反馈 H_∞ 容错控制[J]. 黑龙江大学自然科学学报,2008, 25(5):561-568.

[8] QIU Ji-qing, GUO Yu-ming, ZHAO Yan-chun, *et al*. Fault-tolerant control for a class of nonlinear uncertain singular systems[C]//Proc of International Conference on Web Information System and Mining. Washington DC: IEEE Computer Society, 2010: 391-395.

[9] XU Sheng-yan, DOOREN P V, STEFAN R, *et al*. Robust stability and stabilization for singular systems with state delay and parameter uncertainty[J]. IEEE Trans on Automatic Control, 2002, 47(7): 1122-1128.

[10] DAI L Y. Singular control systems [M]. New York: Springer-Verlag, 1989: 332.

[11] ZHOU Ke-min, KHARQONEKAR P P. Robust stabilization of linear systems with norm-bounded time varying uncertainty[J]. Systems and Control Letters, 1998, 10(1): 17-20.

[12] XIE Li-hua, FU Ming-yue, De SOUZA C E. H_∞ control and quadratic stabilization of systems with parameter uncertainty via output feedback [J]. IEEE Trans on Automatic Control, 1992, 37(8): 1253-1256.

[13] 舒伟仁,张庆灵. 时滞广义系统的状态反馈 H_∞ 控制[J]. 系统工程与电子技术,2004,26(12):1871-1875.

(上接第 1288 页)

[5] GEORGE T, VICKY P. A generalized vector space model for text retrieval based on semantic relatedness[C]//Proc of the 12th Conference of the European Chapter of the Association for Computational Linguistics: Student Workshop. NJ: Association for Computational Linguistics, 2009: 70-78.

[6] DEERWESTER S, DUMAIS S T, FURNAS G W, *et al*. Indexing by latent semanic analysis[J]. Journal of the American Society for Information Science, 1990, 41(6): 391-407.

[7] 王俊英. 基于科技文献的中文文本分类算法研究[D]. 秦皇岛:燕山大学,2007.

[8] 吴勤. 基于引证强度的学术论文质量评价方法研究[J]. 情报学报,2007,26(4):522-526.

[9] 胡德华,常小婉. 开放存取期刊论文质量和影响力的评价研究[J]. 图书情报工作,2008,52(2):61-64.

[10] 张光前,刘欣,冯永琴. 基于阅读价值的科技文献排序方法研究[J]. 情报学报,2009,28(6):177-241.

[11] 颜端武. 面向知识服务的智能推荐系统研究[D]. 大连:大连理工大学,2010.

[12] 中国科学院中文分词工具(ICTC LAS)[EB/OL]. http://ictclas.org/.