

基于网页链接与用户反馈的 PageRank 算法改进研究

曹姗姗 王 冲

(桂林电子科技大学计算机科学与工程学院 桂林 541004)

摘要 在网页链接结构的排序算法 PageRank 的基础上,提出了一种改进的 Bias PageRank(BPR)算法。为了提高用户对网页排序结果的满意度,该算法结合网页链接结构与用户反馈信息(点击率、最近一次点击时间等)进行综合分析,从而从网页设计与用户角度共同对网页 PR 值进行合理分配,以在一定程度上达到高质量网页尽量排序靠前、信息价值低的网页尽量下沉的目的。仿真实验表明,BPR 算法在一定程度上改善了排序效果,提升了用户信息检索满意度。

关键词 网页排序,链接结构,用户反馈,PageRank,算法改进

中图分类号 TP391.3 文献标识码 A DOI 10.11896/j.issn.1002-137X.2014.12.039

Improved PageRank Algorithm Based on Links and User Feedback

CAO Shan-shan WANG Chong

(College of Computer Science and Engineering, Guilin University of Electronic Technology, Guilin 541004, China)

Abstract Based on the PageRank algorithm, this paper proposed an improved PageRank algorithm named Bias PageRank, which not only takes the link structure between pages into consideration, but also users' feedback information, such as frequency of click and interval of recent click. Through comprehensively analyzing these information, BPR algorithm can make pages of high quality rise and pages of poor quality fall to a certain extent. Experiments indicate that BPR algorithm can improve the ranking result and users' satisfaction.

Keywords Page rank, Link structure, User feedback, PageRank, Algorithm improvement

1 引言

随着网络的高速发展,网民数量急剧增加,报告显示,截止 2013 年 6 月底,我国网民数量已达 5.91 亿^[1],同时互联网信息呈几何级数增长。面对海量信息和大规模网络用户,如何能让用户高效快捷地搜索到所需信息,成为网络搜索机制的首要问题。而网页排序技术正是解决这一问题的关键技术之一,因为该技术作为搜索引擎的核心部分,是评价搜索引擎质量高低的一个重要指标。

目前关于网页排序的算法丰富多样,主要包括基于网页内容分析的网页排序算法和基于链接结构分析的网页排序算法两大类。其中著名的搜索引擎 Google 所采用的网页排序算法是基于网页链接结构分析的 PageRank 算法,该算法的成功应用,证实了它所具有的实践应用价值和理论研究价值。

2 算法分析

2.1 PageRank 算法

在借鉴传统引文分析思想的基础上,1998 年斯坦福大学计算机学院研究生 Lawrence Page 和 Sergey Brin 提出了一种基于网络链接分析的算法^[2],即最早的传统 PageRank 算法。该算法利用网页之间的链接关系把整个网络结构关系表示为

Web 图,如图 1 所示。将网页 V 对网页 U 的链接看作是 V 对 U 的一次投票,那么网页 V 就需要把自己的一部分 PageRank 值分给 U,这就是 PageRank 算法的基本思想。显然,V 本身的权威值越高,U 获得的值就越高。同时,U 被链接的次数越多,其 PageRank 值也就越高。最后根据各个网页 PR 值的高低对搜索结果进行排序。

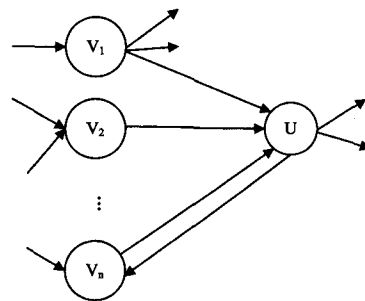


图 1 Web 图

PageRank 算法建立在随机网络冲浪模型的基础上,假设用户在 Web 上是随机浏览的,并且 Web 上所有网页都是可访问的,那么用户可以有两种选择:

- (1) 点击当前页上的一个链接;
- (2) 通过键入网址,随机跳转到一个页面进行访问。

用户点击一个链接的概率为 d ,那么直接跳转的概率就

到稿日期:2014-01-24 返修日期:2014-04-28 本文受教育部项目(2011YSD120030),广西教学改革项目(2011JGA043)资助。

曹姗姗(1988-),女,硕士生,主要研究方向为多媒体技术,E-mail:16672909@qq.com;王冲(1972-),男,副教授,硕士生导师,主要研究方向为信息检索技术、多媒体教育软件技术与应用理论。

是 $1-d$ 。这里的概率 d 又被称为阻尼系数,它的取值通常为 $0.85^{[3]}$ 。

根据以上原理,一个网页 PageRank 值的计算公式如下所示:

$$PR(U) = (1-d) + d \left[\frac{PR(V_1)}{Out(V_1)} + \dots + \frac{PR(V_n)}{Out(V_n)} \right] \quad (1)$$

其中, V_i 是指向 U 的网页, $\{V_1, V_2, \dots, V_n\}$ 是所有指向 U 的网页集合。 $Out(V_i)$ 是网页 V_i 链出的链接数,即出度。通过公式我们可以看出,PageRank 算法仅仅依赖于网页间的链接关系,所以该计算过程可以离线完成,这样就有效减少了在线查询时的计算量,从而大大提高了用户查询响应效率。但是 PageRank 算法仍存在一些不足,例如:

(1) 平均分配网页 PR 值,没有考虑到不同网页质量不同,所有被链接的网页都获得相同 PR 值。但实际上,用户往往倾向于高质量权威性网页,所以应对权威网页分配较高 PR 值;

(2) 偏向旧网页,旧网页由于在网络中存在时间长,被链接可能性更大。相对于新网页,旧网页的 PR 值就会偏高,而实际中用户对有价值的新信息需求也较大,用户倾向于具有高信息价值的网页,无论旧网页还是新网页;

(3) 忽略用户兴趣,PageRank 算法仅根据网页链接结构对网页进行排序,而网页间的链接是由网页设计者决定的。因此该算法仅从网页设计者角度出发,忽略了用户的角度。

2.2 相关改进算法

针对第一个不足, Xing 等提出了加权 PageRank 算法 (Weighted PageRank, WPR), 即利用网页的链入链接个数和链出链接个数,对网页链接权重分别添加基于网页入度和基于网页出度的权重因子^[9]。文献[6,7]所提出的算法均采用了 WPR 算法中的加权因子来对网页权威值进行分配。同时,文献[4,10]也利用网页入度和出度,提出了基于网页入出度比例的不均衡分配算法。虽然利用网页入度与出度在一定程度上能解决 PageRank 算法平均分配的问题,但仍存在偏向旧网页和忽略用户兴趣的问题。

针对第二个不足,通常改进的 PageRank 算法通过引入时间因子来提升新网页的 PR 值。文献[6,8,9,11]均通过引入一个与网页的权值呈反比的时间权值函数来提升新网页,使新网页上浮。但是该方法并未考虑到热门新网页与同时间发布的普通新网页的不同,使得两者获得相同程度的上浮,这显然是不合理的。同样,高质量的旧网页和同时间发布的普通旧网页获得相同程度的下沉,也是不合理的。

针对第三个不足,考虑到用户对网页的兴趣,文献[7,8]均提取了用户在网页上的停留时间作为用户反馈的兴趣因子。同时,由于网页点击信息也是用户兴趣的重要体现,它还可以体现该网页质量和价值,因此 Gyanendra Kumar 等通过提取网页链接的访问量,从用户角度对 PageRank 算法进行改进^[12]。但是这些算法均未改善 PageRank 算法偏向旧网页的问题。而文献[13]利用网页点击总量及用户最新点击时间,对网页进行了 PR 值不均等分配,该算法虽然对 PageRank 算法中的上述 3 个不足做出了改进,但对于最近偶尔被点击的旧网页来说,由于在网络中的时间长,其点击总量比最新热门网页高,它所获得的 PR 值就会远高于新网页,这也是不合理的。

因此,本文结合能够隐含反馈用户兴趣的网页点击量以及网页有关的时间因子,对 PageRank 算法进行改进,使得改进后的 PageRank 算法能够改善原算法的上述 3 个不足,从而在一定程度上提高网页 PR 值分配的合理性。

3 改进的 PageRank 算法

随机网络冲浪模型假定用户如果沿着网页中的链接进行跳转,那么从当前网页 V_i 跳转到其所有链出网页的概率都是相同的,即 $\frac{1}{Out(V_i)}$ 。但实际上,用户浏览网页时,往往偏好于一些高质量高价值的网页。因此,从网页设计者和用户角度出发,我们认为被链接次数多且点击率高的网页是高质量网页,同时最近被点击网页的 PR 值应高于相同条件下长时间未被点击网页的 PR 值。所以本文引入了偏好因子和网页热度因子来对 PageRank 算法进行改进。改进的 PageRank 算法称为 BPR (Bias PageRank) 算法,其计算公式如下:

$$BPR(U) = (1-d) + d \left[BPR(V_1) * B(V_1 \rightarrow U) + \dots + BPR(V_n) * B(V_n \rightarrow U) + \delta * H(U) \right] \quad (2)$$

其中,偏好因子 $B(V_i \rightarrow U)$ 是用户从当前网页 V_i 跳转到它所指向的网页 U 的概率,即链接 $V_i \rightarrow U$ 的权重。BPR 算法中的 $H(U)$ 代表了网页 U 在当前的热门度,是网页的热度 (Hot) 因子。越热门的网页,最近一次被点击的时间与当前时间的间隔就越小,所以该间隔成为衡量网页当前热门度的标准。而其中热度因子 $H(U)$ 的系数 δ 代表网页的热度在权威值分配中的重要程度。这里的偏好因子和热度因子均包含了对用户隐式反馈信息的分析。

所谓用户隐式反馈就是通过分析用户搜索的行为信息来得出用户对搜索结果的认同度,它是通过服务器对用户行为进行分析得出结果。所以隐式反馈不需要用户对搜索结果进行显式的信息反馈,从而减轻了用户的使用负担。

3.1 偏好因子

由于点击行为能够反映出用户的兴趣,是用户的一种隐式反馈,因此从网页设计者与用户角度出发,BPR 算法中的偏好因子 $B(V_i \rightarrow U)$ 在分析网页链接结构的同时,考虑了网页点击率 (Frequency of Click),从而使分配更合理,更好地满足用户需求。这里 $B(V_i \rightarrow U)$ 的值就是链接 $V_i \rightarrow U$ 的权重,其计算公式如下:

$$B(V_i \rightarrow U) = \frac{\alpha * FC(U) + 1}{\sum_{w \in F(V_i)} [\alpha * FC(W) + 1]} \quad (\alpha \geq 0) \quad (3)$$

其中, $F(V_i)$ 是网页 V_i 所有前向链接 (Forward Link) 的网页集,即网页 V_i 指向的网页的集合,该集合的元素个数就是网页 V_i 的出度。 α 是网页点击率相对于网页链接结构在权威值分配时所占的比重。 $\alpha=0$,代表不考虑网页点击率,此时 BPR 算法就演变成仅考虑网页链接结构的 PageRank 算法。而 α 越大,代表网页点击率越重要。 $FC(U)$ 是网页 U 的月点击率 (Frequency of Click),计算公式如下:

$$FC(U) = \frac{Click(U)}{T(U)} \quad (4)$$

其中, $T(U)$ 是与网页 U 被搜索引擎搜索到的次数有关的函数。由于当前很多网页设计不规范,无法从网页中获取与时间有关的信息,但一般搜索引擎服务器的搜索周期为一个月,因此搜索引擎搜索到网页的次数与该网页的存在时间成正比,可以用网页被搜索到的次数代替其存在时长^[11]。而

Click(U)是与网页U的点击量有关的函数,见式(5):

$$Click(U) = \ln(N+1) + \beta \quad (5)$$

式中, N 是网页 U 的点击量。为了补偿新网页, 这里添加了补偿因子 β , 取值为 1。为了平衡网页点击量与网页存在时间在网页权威值分配中的比重, 这里对网页被搜索引擎搜索到的次数 M 同样做取对数操作, 如下式:

$$T(U) = \ln(M+2) \quad (6)$$

综合式(5)和式(6)可将式(4)写为:

$$FC(U) = \frac{\ln(N+1) + \beta}{\ln(M+2)} \quad (7)$$

由式(7)可以看出, 网页的点击量 N 越高, 被搜索到的次数 M 越小(即网页存在时间越短), 该网页的点击率就越高。如此非热点的旧网页随着访问者越来越少, 会在一定程度上下沉, 而高质量新网页上浮, 从而满足用户需求。

通过引入偏好因子, 我们可以看出 BPR 算法能够从用户和网页设计者的角度, 依据信息价值的高低对网页分配不同的权威值, 同时能够合理对待新旧网页, 如果该网页信息价值高, 依据 BPR 算法, 它获得的 BPR 值就较高。从而使当前信息价值稍低的旧网页能够下沉, 当前价值高的网页上浮。

3.2 热度因子

考虑到对当前热门网页分配的权重应比一般网页大, BPR 算法加入了热度因子 $H(U)$ 。该因子取决于网页最近被浏览的时间与被搜索引擎搜索到的日期间隔, 以月为单位, 公式如下:

$$H(U) = \begin{cases} 0.1, & t(U) \leq 1 \\ \frac{0.1}{1 + \lambda * t(U)}, & t(U) > 1 \end{cases} \quad (8)$$

其中, $t(U)$ 是网页 U 最近一次被浏览时间与用户提交搜索请求时当前日期的时间间隔, 以月为单位。如果网页 U 在最近一个月内被点击浏览过, 热度因子就为 0.1, 代表该网页当前是热门网页。当最近一次点击发生在 $t(U)$ 个月前, 当前网页的热度因子则根据 $t(U)$ 的大小, 与其成反比减小。所以网页越久未被浏览, 它的热度因子就越小, 代表该网页当前并不热门。式中 λ 是衰减系数, 用于控制时间间隔的比重, 这里取值为 $\frac{1}{12}$ 。 $t(U)$ 的取值如下:

$$t(U) = \begin{cases} t_{now}(U) - t_{last}(U), & t_{last}(U) \neq Null \\ T(U), & t_{last}(U) = Null \end{cases} \quad (9)$$

式中, $t_{last}(U)$ 记录网页 U 最近被点击的时间。 $t_{now}(U)$ 是用户提交搜索要求时的当前时间。 $T(U)$ 是网页 U 被搜索引擎搜索到的次数, 与网页存在时间成正比, 与式(4)中 $T(U)$ 相同。若 $t_{last}(U) \neq Null$, 那么 $t(U)$ 就是最近一次点击距离当前时间的间隔。若 $t_{last}(U) = Null$, 那么 $t(U)$ 就是网页 U 的存在时间, 用网页被搜索到的次数 $T(U)$ 来表示。

通过添加网页热度因子, 对网页的 BPR 值做出修整, 使得当前热门网页能够获得较高的权重值。而网页热度在网页权重值中的重要性, 可通过式(2)中 δ 因子进行修整。

结合式(2)可以得出: BPR 算法的流程结构与原 PageRank 算法基本相同, 而且 BPR 算法也是收敛的。

4 实验仿真

4.1 主要实验步骤

我们使用了 Heritrix、Lucene、eclipse 等工具来搭建实验

环境, 以对 BPR 算法进行实验仿真, 验证其效果。实验主要步骤如下:

(1) 对 Web 数据的收集。实验抓取本校国际学院约 5000 个页面, 并生成一个用于记录网页链接结构信息的 Web 图;

(2) 采用 java 语言分别实现传统 PageRank 算法及 BPR 算法, 计算各网页的权威值。这里对 BPR 算法中的 α, β 因子均取 1, 为不使热度因子对网页权威值影响过大, 系数 δ 取 0.1;

(3) 对 eclipse 平台下的项目添加 Lucene 3.0.jar 包, 根据 Lucene 提供的框架对网页信息进行抽取及中文分词, 然后分别连同计算出的 PR 值和未添加热度因子的 BPR 值, 以 Document(field₁, ..., field_n) 的形式存储在 Lucene 的存储接口中, 建立索引;

(4) 通过比较查询得到的结果页面次序, 来分析 BPR 算法相较于传统 PageRank 算法改进的效果。

4.2 实验结果

为方便分析, 这里将搜索结果及其有关信息共同显示出来, 然后对两种算法得到的前 10 个排序结果进行对比分析。

图 2—图 4 中的搜索关键字均为“大学”。

次序	网页标题	PR值	点击率	被搜索次数/最近点击时间
1	韩国又石大学	1.0566705	2.3762667	29 2012-11-22
2	澳大利亚悉尼科技大学	0.7804651	2.2917982	29 2013-11-22
3	澳大利亚南澳大学	0.7804651	2.2918122	29 2013-12-22
4	澳大利亚考瑞大学	0.7804651	2.2921145	29 2013-11-22
5	澳大利亚海伯大学	0.7804651	2.2485306	29 2013-5-22
6	澳大利亚纽卡大学	0.7804651	2.2513242	29 2012-11-22
7	澳大利亚悉尼科技大学	0.7804651	3.265079	2 2013-12-2
8	澳大利亚南威尔士大学	0.7804651	2.4503798	29 2013-11-22
9	马来西亚林登大学招生简章	0.7804651	2.5530913	25 2013-4-22
10	韩国又石大学董事长、校长一行访问我校	0.5475961	2.4528325	21 2012-10-1

图 2 传统 PageRank 算法得到的前 10 个网页

次序	网页标题	BPR值	点击率	被搜索次数/最近点击时间
1	韩国又石大学	1.08804	2.3752667	29 2012-11-22
2	澳大利亚悉尼科技大学	0.8900542	3.265079	2 2013-12-2
3	马来西亚林登大学招生简章	0.7656808	2.5530913	25 2013-4-22
4	澳大利亚南威尔士大学	0.7478829	2.4503798	29 2013-11-22
5	澳大利亚悉尼科技大学	0.7214436	2.2917982	29 2013-11-22
6	澳大利亚考瑞大学	0.7204598	2.2921145	29 2013-11-22
7	澳大利亚南澳大学	0.72040823	2.2918122	29 2013-12-22
8	澳大利亚纽卡大学	0.7133904	2.2513242	29 2012-11-22
9	澳大利亚海伯大学	0.7129036	2.2485306	29 2013-5-22
10	德国德雷斯顿工业大学(名校·免学费)招生	0.6292358	4.666478	5 2013-12-3

图 3 未添加热度因子的 BPR 算法得到的前 10 个网页

图 2 为采用传统 PageRank 算法所得到的前 10 个结果页面。由于 PageRank 算法仅从设计者角度考虑了网页间的链接结构, 因此在图 2 中, 页面被链接的次数越多, PR 值越高, 排名越靠前。而图 3 采用 BRP 算法(此时还未加入热度因子), 则从网页设计者与用户角度综合考虑了网页的链接结构以及网页本身所隐含的用户反馈信息, 来对 BPR 值进行合理分配。

次序	网页标题	BPR值	点击率	被搜索次数/最近点击时间
1	韩国又石大学	1.09144	2.3762667	29 2012-11-22
2	澳大利亚悉尼科技大学	0.8905416	3.265079	2 2013-12-2
3	马来西亚林登大学招生简章	0.7697808	2.5530913	25 2013-4-22
4	澳大利亚南威尔士大学	0.753489	2.4503798	29 2013-11-22
5	澳大利亚悉尼科技大学	0.7271105	2.2917982	29 2013-11-22
6	澳大利亚南澳大学	0.7264082	2.2918122	29 2013-12-22
7	澳大利亚考瑞大学	0.7261253	2.2921145	29 2013-11-22
8	澳大利亚海伯大学	0.7171636	2.2485306	29 2013-5-22
9	澳大利亚纽卡大学	0.71679044	2.2513242	29 2012-11-22
10	德国德雷斯顿工业大学(名校·免学费)招生	0.6292358	4.666478	5 2013-12-3

图 4 添加热度因子的 BPR 算法得到的前 10 个网页

比较图 2 与图 3, 图 2 中排名第 1 的“韩国又石大学”被链

接的次数最多,同时由于该网页的点击率并不低,因此利用综合考虑的 BPR 算法,在图 3 中其排名并没有下降。同时,通过观察,图 2 中位于第 7 位的“澳大利亚名校-悉尼科技大学”网页 PR 值与排在其前面的各澳大利亚大学相同,说明它被链接的次数也较多。但该网页被搜索到的次数比前面各澳大利亚大学少,即存在时间较短,而且其点击率较高,这就体现出了用户偏好,因此采用 BPR 算法该网页排名提升至第 2 位,如图 3 所示。这里可以通过 α 因子调节网页链接结构与点击率间的比例,来调整网页上浮下沉的幅度。

比较图 2 与图 3,显然 BPR 算法体现出了对新旧网页合理分配权威值的效果,高价值新网页上浮,低价值旧网页下沉。

图 4 是采用添加了热度因子的 BPR 算法所搜索得到的前 10 个网页。相较于图 3,这里通过添加与网页最近被点击时间有关的热度因子,来对网页当前信息价值在一定程度上做出修正。

比较图 3 与图 4,我们可以看出,图 3 中排名第 7 位的“澳大利亚南澳大学”最近被点击的时间较近,体现了该网页的热度,因此添加了热度因子的 BPR 算法对该网页的位置进行修正后,其排名被提升至第 6 位,如图 4 所示。这里,热度因子在网页 BPR 值分配过程中所占的比重可以通过式(2)中的系数 δ 来调整。

同时,实验采用文献[6]中用户满意度评估的方法,满意评估公式 $S = \sum_{i=1}^n (n-i+1) * s_i$,其中 n 是网页总数,实验选取 n 为前 50 个网页, i 为 n 个网页中的第 i 个, s_i 为满意系数。网页分为 4 个等级:非常满意、满意、较满意、不满意;4 个等级对应的满意系数分别为 1.0、0.6、0.2、0.0。实验组织了一个 10 人的测试小组,根据上述规则对 50 个网页的满意系数进行评估,最后对传统 PageRank 算法和 BPR 算法得到的前 30 个网页的排序结果分别计算满意度值 S 。这里设定了 5 个搜索即大学、留学、英国、美国、澳大利亚,其分别对应的 PageRank 算法的 S 值为:524.7、547.6、516.4、486.9、524.3;对应的 BPR 算法的 S 值分别为:599.7、623.5、586.8、561.2、587.7。结果比较如图 5 所示。

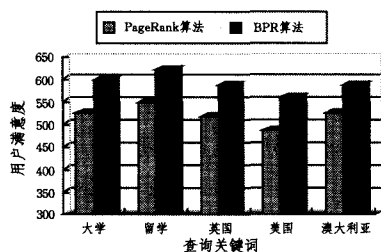


图 5 算法排序满意度对比

由图 5 可以看出,改进后的 BPR 算法能够更好地满足用户的需求,在一定程度上提高了用户满意度,达到预期目标。

因此,根据以上实验仿真结果,综合分析得出:改进后的 BPR 算法能够合理弥补 PageRank 算法均分网页权威值、偏向旧网页以及不考虑用户兴趣的不足;BPR 算法能够从用户和设计者角度出发,对网页权威值做出较合理的分配,使得高质量的新网页上浮,当前信息价值不高的旧网页下沉;改进算法的排序结果对用户信息需求的满意度有一定程度提升。

结束语 本文综合考虑了网页间链接结构以及用户反馈

信息两大要素,提出了一种改进 PageRank 的 BPR 算法。该算法能够从网页设计者和信息用户的角度,通过融合网页链接结构、点击量以及网页存在时间等信息,来对新旧网页的权威值进行较合理分配;通过使用热度因子对网页当前的信息价值做出修整,使得当前信息价值高的网页排序靠前,信息价值较低的网页下沉。实验仿真结果也表明,改进的算法能够改善网页结果集的排序效果,在一定程度上弥补了 PageRank 算法存在的一些不足,并在一定程度上提高了用户满意度。

参考文献

- [1] 中国互联网络信息中心(CNNIC). 第 31 次中国互联网络发展状况统计报告[OL]. [2013]. http://news.xinhuanet.com/tech/2013-01/15/c_124233840.htm
- [2] Brin S, Page L. The Anatomy of a Large-scale Hypertextual Web Search Engine[C]//Proceedings of the 7th ACM-WWW International Conference. Brisbane; ACM Press, 1998: 107-117
- [3] Buttcher S, Clarke C I, A, Cormack G V. 信息检索:实现和评价搜索引擎[M]. 陈健,黄晋,等译. 北京:机械工业出版社,2012
- [4] 田甜,倪林. 基于 PageRank 算法的权威值不均衡分配问题[J]. 计算机工程,2007,33(18):53-55
- [5] Xing Wen-pu, Ghorbani A. Weighted PageRank algorithm[C]//Proceedings of Second Annual Conference on Communication Networks and Services Research. IEEE, 2004:305-314
- [6] 段淮川,胡平. 基于主题特征和时间因子的改进 PageRank 算法[J]. 计算机工程与设计,2010,31(4):866-868
- [7] 彭聪,吴强,李仁发. 一种改进型的网页排序算法[J]. 微计算机信息,2010,26(11):72-74
- [8] 王德广,周志刚,梁旭. PageRank 算法的分析及其改进[J]. 计算机工程,2010,36(22):291-293
- [9] 冯海涛. 基于网页时间权值的 PageRank 算法改进[J]. 西安邮电大学学报,2013,18(2):121-124
- [10] 张砚明. 基于链接结构分析的 Web 页面排序算法[D]. 西安:西安电子科技大学,2010
- [11] 李稚楹. 基于网页内容和时间反馈的网页排 PageRank 算法研究[D]. 重庆:重庆理工大学,2012
- [12] Kumar G, Duhan N, Sharma A K. Page Ranking Based on Number of Visits of Links of Web Page[C]//International Conference on Computer & Communication Technology (ICCCT). 2011:11-14
- [13] Zhou Cai-lan, Chen Kai, Li Sha-sha. Improved PageRank Algorithm Based on Feedback of User Clicks[C]//2011 International Conference on Nanjing Computer Science and Service System (CSSS). China, IEEE, 2011:3949-3952
- [14] Kritikopoulos A, Varlamis M, et al. A Method for Ranking Web Pages Based on Content Similarity[C]//2007 Workshops of the 24th British National Conference on Databases (BNCOD). Glasgow, United Kingdom, IEEE, 2007:92-100
- [15] 温韬,朱敏,周科,等. 基于多权重因子的 PageRank 算法改进[J]. 软件时空,2012,28(9):422-424
- [16] 史铭茗. 加权 PageRank 算法研究综述[J]. 软件导刊,2013,12(2):30-32
- [17] 李稚楹,杨武,谢治军. PageRank 算法研究综述[J]. 计算机科学,2011,38(10A):185-188