

基于用户日志的查询扩展统计模型*

崔航¹, 文继荣², 李敏强¹⁺

¹(天津大学 系统工程研究所,天津 300072)

²(微软亚洲研究院,北京 100080)

A Statistical Query Expansion Model Based on Query Logs

CUI Hang¹, WEN Ji-Rong², LI Min-Qiang¹⁺

¹(Institute of Systems Engineering, Tianjin University, Tianjin 300072, China)

²(Microsoft Research Asia, Beijing 100080, China)

+Corresponding author: Fax: 86-22-27404796, E-mail: mqli@tju.edu.cn

<http://www.tju.edu.cn>

Received 2002-03-12; Accepted 2002-08-13

Cui H, Wen JR, Li MQ. A statistical query expansion model based on query Logs. *Journal of Software*, 2003, 14(9):1593~1599.

<http://www.jos.org.cn/1000-9825/14/1593.htm>

Abstract: Ambiguity of query terms has been a long-standing problem in information retrieval field, which becomes more serious in Web searching. A method for automatic query expansion based on query logs obtained from users' daily usage is suggested. This model establishes probabilistic relationship between terms in documents and in user queries through statistical learning from the log, and selects high-related expansion terms based on Bayesian theory. These expansion terms are added into the original query to formulate a new one in order to improve the effectiveness of retrieval. Experimental results show that this technique is more adaptive to Web searching, and can improve the precision of document retrieval markedly compared with conventional ones.

Key words: information retrieval; query expansion; user log; log mining

摘要: 信息检索长期存在着用词歧义性问题,在 Web 搜索上的表现更加突出.提出了一种基于用户查询日志的查询扩展统计模型,将用户查询中使用的词或短语与文档中出现的相应词或短语以条件概率的形式连接,利用贝叶斯公式挑选出文档中与该查询关联最紧密的词加入原查询,以达到扩展优化的目的.实验结果表明,该方法更适宜改进 Web 上的信息检索,相对传统的查询扩展算法可以大幅度提高查询精度.

关键词: 信息检索;查询扩展;用户日志;日志挖掘

中图法分类号: TP311 文献标识码: A

随着 Web 技术的巨大发展和日益普及,Internet 越来越成为人们搜寻各方面信息的主要来源,搜索引擎也在

* Supported by the National Natural Science Foundation of China under Grant Nos.69974026, 70171002 (国家自然科学基金)

第一作者简介: 崔航(1977-),男,北京人,博士生,主要研究领域为查询扩展与优化,结构化文档检索与索引,基于 Web 的信息检索.

人们的日常生活和学习中发挥着无法替代的重要作用.然而,由于大量同义词和多义词的存在,用户在提交查询时使用的词往往不尽规范,与文档索引使用的词或词组有很大差别,这就给现今基于关键词的查询系统带来巨大的困难,也成为长期困扰信息检索领域的基本问题.

Furnas 第一个发现了这个所谓的“词典问题”(dictionary problem)^[1].他们的实验表明,通常情况下,两个人使用同样的关键词描述同一物体的几率小于 20%.在当前的搜索引擎的使用过程中,这个问题变得更加尖锐.如果用户使用足够多的词描述查询内容,用词不一致的问题则会在一定程度上得到缓解.从对用户查询日志的分析来看,事实并非如此.通过对微软公司旗下的 MSN 中的 Encarta(<http://encarta.msn.com>)在线百科全书网站连续两个月的用户查询记录进行分析,我们发现 49%的用户查询仅有一个单词,33%的查询由两个单词组成,用户平均使用 1.4 个单词描述他们的查询^[2].这样,传统的基于关键词的向量空间模型无法发挥正常的作用.同时,在许多情况下,用户使用的词即使在文中出现,也未必在相关文章中具有足够的权重,即仅靠用户提交的短查询无法提供检索出相关文档的足够信息.

目前,解决“词典问题”的有效方法是自动查询扩展优化(query expansion,简称 QE),即在原来查询的基础上加入与用户用词相关联的词,组成新的更长、更准确的查询,这样就在一定程度上弥补了用户查询信息不足的缺陷.根据其计算查询用词与扩展用词相关度的方法的不同,可以大致将已有的查询扩展的方法分为全局分析和局部分析.实验结果表明,多数情形下,局部方法在计算效率和检索性能上均优于传统的全局分析^[3].Xu 和 Croft 在其论文中提出的局部上下文分析(local context analysis,简称 LCA)方法^[4]将全局分析的方法融入局部方法之中,取得了良好的检索结果.

以往的查询扩展算法主要基于对文档关键词的分析,而忽视了对用户用词的重要来源——查询记录的使用.本文提出一种崭新的查询扩展统计模型,充分使用用户的查询日志作为对词与词之间关系进行分析的主要依据,利用贝叶斯公式建立相应的计算模型,自动选择与原查询高度相关的词与词组来进行查询扩展.最后,在 Encarta 数据集上进行了应用实验,并系统地分析了实验结果.

1 相关研究方法

在信息检索领域,关于自动查询扩展优化方法基本上可以分为全局和局部两大类.

1.1 查询扩展的全局和局部方法

查询扩展早在 20 世纪 70 年代就被提出来,作为改善检索的一种方法.全局分析是较早出现的具有实际应用价值的查询扩展优化方法,其基本思想是对全部文档中的词或词组进行相关分析,计算每对词或词组间的关联程度.当一个新的查询到来时,则根据预先计算的词间相关关系,将与查询用词关联程度最高的词及词组加入原查询以生成新的查询.早期的典型全局分析的方法是词聚类方法,它是将文档中出现的词按共同发生的频率先行聚类,其后根据词的不同集合对查询进行扩展.目前常见的全局分析方法包括 LSI(Latent semantic indexing)^[5]、相似性词典^[6]等.全局分析的优势是可以最大限度地探求词间关系,并在词典建立之后以较高的效率进行查询扩展.但是,当文档集合非常大时,建立全局的词关系词典在时间和空间上往往是不可行的,并且在文档集合改变后的更新代价巨大.因此,近期的查询扩展研究主要集中在与之对应的局部分析上.

局部分析利用两次查询的方法解决扩展问题.局部分析利用初次检索得到的与原查询最相关的 N 篇文章作为扩展用词的来源,而非利用先前计算得到的全局词关系词典.局部分析的思想可以追溯到 Atter 和 Fraenkel 的 1977 年的论文^[7].目前流行的局部分析方法主要是局部反馈(local feedback,也称为 pseudo feedback),它是在相关反馈(relevance feedback)的基础上发展起来的^[8,9].相关反馈根据用户对初次检索的结果进行评判后,将用户认为相关的文章作为扩展用词的来源.而局部反馈解决了相关反馈必须与用户交互的问题,它将初次查询的前 N 篇文章认为是相关文章,并以此为依据对查询进行扩展.

局部分析的方法是应用最广泛的查询扩展方法,并在一些实际的信息检索系统中得以使用.但是,当初次查询后排在前面的文档与原查询相关度不大时,局部分析会把大量无关的词加入查询,从而严重降低查询精度,甚至低于不做扩展优化的情形.

1.2 局部上下文分析

局部上下文分析方法是由 Xu 和 Croft 提出来的^[4].它在整体上是一种局部分析方法,但利用全局分析的词共同出现频率的思想避免了向原查询加入不相关的词.局部上下文分析的方法被用于 INQUERY 系统中,并在 TREC 标准测试集上取得了良好的效果^[4].实验表明,该方法的检索结果明显优于传统的全局分析和局部分析方法.

局部上下文分析的工作流程如下:

- (1) 使用检索系统得到与原查询最相关的 n 篇文章,一般是初始检索得到的前 n 篇文章.
- (2) 从该 n 篇文章中选取与原查询最相关的词与词组,相关度计算为

$$bel(Q, c) = \prod_{t_i \in Q} [\delta + \log(af(c, t_i) + 1) \times idf_c / \log(n)]^{idf_i}, \quad (1)$$

$$af(c, t_i) = \sum_{k=1}^n f_{ik} f_{ck}, \quad (2)$$

其中, $af(c, t_i)$ 是文章中出现的词与查询用词共同出现的频率, bel 表示词与该查询之间的相关程度,用于决定与原查询相关的扩展用词.

- (3) 将 m 个最相关的(即 bel 值最高的)词或词组加入原查询构成新查询.

局部上下文分析的应用效果仍然高度依赖于初次检索的结果.如果初次检索返回的多数文档与原查询无关,该方法仍会将大量无关的词加入新查询,从而大大降低最终的检索精度.

2 基于用户查询日志的查询扩展统计模型

2.1 基本思想

以往的查询扩展方法仅考虑到将新的用户查询映射到文档集中,从文档中选择与其相近的词,而未将以往的用户查询记录考虑在内.用户查询日志是众多用户使用检索系统时多次“反馈”结果的积累,对它的分析相当于使用大量的用户相关反馈.相对于传统的即时相关反馈而言,对日志记录的分析更具普遍性和统计意义.

基于用户查询日志的查询扩展统计模型的基本思想是,在用户查询记录的基础上建立用户查询空间,在文档集上建立文档空间,根据用户日志将两个空间中的词按照用户提交某个查询所点击的文章以条件概率方式连接起来.当新查询到来时,系统选取当该查询出现时被选择成为扩展用词的条件概率最大的文档用词加入查询.

从 Encarta 网站获得的用户日志超过 1GB,经预处理后形成 88MB 的用户记录库,包含一个月的用户点击记录,共计 266 万条.其数据结构如下:记录号,用户查询,用户点击的文章.从该日志数据库,可以得到用户的查询用词(用户查询域)和当时用户点击查看的文档记录(点击文章域).因此,用户选取的认为相关的文章就成为连接两个空间用词的桥梁.具体来说,用户查询用词在日志记录中与相关文章相连接,这些文章本身包含众多关键词.这样,查询用词和文档用词通过用户以往选择的的文章被连接起来.

比如,用户提交的查询有“windows”和“window”等关键词,大多数用户认为“windows”是微软公司的操作系统,因此点击的文章中包含“Microsoft Corporation”和“Bill Gates”等关键词,这样用户的查询用词“windows”与这些描述该操作系统的文章建立了紧密的联系,也即与这些文章的关键词,如“Microsoft”,“Gates”或“operating system”等建立了关联.当包含“windows”的新查询出现时,这些与其高度关联的词自然就成为补充描述用户意图的最佳选择.

该算法介于全局分析和局部分析之间.当使用整个文档集作为分析对象时,该算法类似于全局分析,但同时引入了用户查询空间的关键词作为关联关系的一部分.也可以采用寻找与原查询高度相关的文章作为扩展用词来源,但其相关依据并不依赖于初始检索的结果,也不需要用户给出及时反馈来判断,而是建立在对大量用户长期行为的分析和学习的基础上,得出具有普遍意义的相关量度,称为“用户投票”的学习方法.显然,该方法从大量日志中得到的“先验知识”,远比个别用户的临时判断或系统在毫无人参与的情况下得到的结果更为准确,

并且将关于用户反馈的学习放在检索之前,省去了初始检索和用户参与的代价.

2.2 模型描述

基于用户查询日志的查询扩展的目的,是根据用户日志在新的查询来临时选取与该查询最相关的文章用词来形成新的查询.使用给出某个查询(由一个或多个词组成)的情况下,某文章关键词出现的条件概率表示该关键词与此查询的关联程度,则根据贝叶斯公式,该条件概率表示为

$$P(w_j^{(d)} | Q) = \frac{P(w_j^{(d)}, Q)}{p(Q)}. \quad (3)$$

用 Q 表示查询,由关键词 $w_i^{(q)}$ 组成.假设各查询关键词(含词组)的出现相互独立.使用 N 元(N -gram)的方法从用户查询中抽出所有可能的词组,将其中出现频率小于 5 次且没有在文档中出现的词组去掉,从而建立包含词组的词典,使模型符合元素独立的假设.这样,可以将该条件概率表示为

$$\begin{aligned} P(w_j^{(d)} | Q) &= \prod_{i=1}^n P(w_j^{(d)} | w_i^{(q)}) = \prod_{i=1}^n \frac{P(w_j^{(d)} | w_i^{(q)})}{P(w_i^{(q)})} \\ &= \prod_{i=1}^n \frac{\sum_{\forall D_k \in S} P(w_j^{(d)}, w_i^{(q)} | D_k) \times P(D_k)}{P(w_i^{(q)})} \\ &= \prod_{i=1}^n \frac{\sum_{\forall D_k \in S} P(w_j^{(d)}, w_i^{(q)}, D_k)}{P(w_i^{(q)})} \\ &= \prod_{i=1}^n \frac{\sum_{\forall D_k \in S} P(w_j^{(d)} | w_i^{(q)}, D_k) \times P(w_i^{(q)}, D_k)}{P(w_i^{(q)})} \\ &= \prod_{i=1}^n \frac{\sum_{\forall D_k \in S} P(w_j^{(d)} | D_k) \times P(D_k | w_i^{(q)}) \times P(w_i^{(q)})}{P(w_i^{(q)})} \\ &= \prod_{i=1}^n \left[\sum_{\forall D_k \in S} P(w_j^{(d)} | D_k) \times P(D_k | w_i^{(q)}) \right], \end{aligned} \quad (4)$$

其中, $w_i^{(d)}$ 代表文档关键词, S 为对于某个查询关键词用户选择的文档集合, D_k 即为其中的文档.因为文档 D_k 可以分隔查询用词 $w_i^{(q)}$ 和文档用词 $w_i^{(d)}$ 两个集合,故得到 $P(w_j^{(d)} | w_i^{(q)}, D_k) = P(w_j^{(d)} | D_k)$.

根据式(4),如果认为一个文档用词与该查询高度相关,则该查询中的每个词与包含此文档用词的文章在用户日志中以较高的频率共同出现,同时该文档用词在这些文档中具有较高的权重. $P(w_j^{(d)} | D_k)$ 可以用 $w_i^{(d)}$ 在 D_k 中出现的频率近似得到,而 $P(D_k | w_i^{(q)})$ 则可以使用用户日志的统计信息作为先验值.这样,我们得到以下的最终公式:

$$P(w_j^{(d)} | Q) = \prod_{i=1}^n \left[\sum_{D_k \in S} W_{jk}^{(d)} \times \frac{f_{ik}^{(q)}(w_i^{(q)}, D_k)}{f^{(q)}(w_i^{(q)})} \right], \quad (5)$$

其中, $W_{jk}^{(d)} = \frac{\ln(1 + tf_{jk}) \times idf_j}{\max(W_{jk})}$. tf_{jk} 是词 $w_i^{(d)}$ 在文档 D_k 中出现的次数, idf_j 表示词 $w_i^{(d)}$ 在整个文档集中出现频率

的倒数. $W_{jk}^{(d)}$ 被文档 D_k 中的词的最大权重除,以保证其值在 0 到 1 之间. $f_{ik}^{(q)}(w_i^{(q)}, D_k)$ 为用户日志中查询用词 $w_i^{(q)}$ 与用户点击的文章 D_k 同时出现的记录数目, $f^{(q)}(W_i^{(q)})$ 表示查询关键词出现的记录总数.

3 实验与结果分析

以 Encarta 数据集为基础,本节对该模型及对比模型进行应用实验,并对实验结果给予详细分析.

3.1 实验方法

采用 Encarta 网站提供的文档集作为实验数据,共计 41 942 篇各个领域的文章,文章语言为英文,长度从几十个词到上万个词不等.训练数据采用 2000 年 11 月的实际用户查询日志,处理后的记录总数为 266 万条.测试所用查询是从用户日志中随机抽取的 10 个问题,见表 1.

Table 1 Problems for test
表 1 测试问题列表

Number	Problems
1	Fossil fuels
2	What is Daoism
3	Cellular phone
4	Chinese music
5	Search engine
6	Apollo program
7	Desert storm
8	Six day war
9	Space shuttle
10	Table of elements

这些问题来源于用户的真实查询,平均长度为 2.2 个词,可以充分反映目前搜索引擎面临的实际情况.

将文档中出现频率极高的词,如 the,a,there 等词除去,建立了包含 190 000 个单词的词典,时态和复数没有加以区分.另外,使用第 2.2 节中描述的方法,从用户日志中抽取了 13 000 个词组.除非特别说明,实验中均把单词和词组作为索引的词典.

在比较查询性能时,采用标准的精度-查全率(precision-recall)指标,查询精度定义为

$$precision = \frac{\text{sum_of_relevent_documents_in_the_retrieval}}{\text{sum_of_the_retrieved_documents}} \times 100\%; \quad (6)$$

查全率定义为

$$recall = \frac{\text{sum_of_the_relevant_documents_in_the_retrieval}}{\text{sum_of_the_relevant_documents_in_the_collection}} \times 100\%. \quad (7)$$

为了进行对比实验,还引入了统计检验的方法来监测性能改进的显著性, t 检验用于信息检索性能分析在 Hull 1993 年的论文中有详细描述^[10].这里仅给出定义:

$$t = \frac{\bar{D}}{s(D_i) / \sqrt{n}}. \quad (8)$$

其中, $\bar{D} = \frac{1}{n} \sum_{i=1}^n D_i$, $s(D_i) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (D_i - \bar{D})^2}$, D_i 为两种对比实验数据的差值,该检验为自由度为 $n-1$ 的 t 分布.

3.2 对比算法

使用标准的向量空间模型(VSM)作为检索算法,将统计扩展模型与不加查询扩展及使用局部上下文分析方法的查询扩展效果进行比较.对比实验的基线由不加任何查询扩展的系统得到,查询与文档间的相似度定义为向量夹角的余弦函数值:

$$Sim = \frac{\sum_{i=1}^n W_i^{(q)} \times W_{ij}^{(d)}}{\sqrt{\sum_{i=1}^n W_i^{2(q)} \times \sum_{i=1}^n W_{ij}^{2(d)}}}, \quad (9)$$

其中, $W_{ij}^{(d)}$ 为关键词 $w_i^{(d)}$ 在文档 D_k 中的权重, $W_{ij}^{(d)} = \ln(1+tf_{ij}^{(d)}) \times \ln idf_i^{(d)}$, $W_i^{(q)}$ 是查询关键词的权重, $W_i^{(q)} = \ln idf_i^{(d)}$.

在实验中,关于局部上下文分析的算法,采用与 Croft 等实验相同的参数^[4],使用与原查询最相关的前 100 篇文章作为查询扩展的基础,并选择前 30 个与原查询关联度最高的词成为扩展用词.对于基于用户日志的统计扩展模型,同样采用前 30 个出现的条件概率最大的词加入原查询以产生新的查询.在所有实验中,检查前 100 篇返

回文章,并以精度-查全率作为主要的性能指标.

3.3 实验结果与讨论

对原查询结果、经局部上下文分析扩展优化后的结果以及使用我们的算法得到的结果,进行综合比较(见表 2 和图 1).

Table 2 Comparison of retrieval performance

表 2 查询性能比较

Precision Recall	Baseline	Query expansion on user log (improvement %)	Query expansion on local context analysis (improvement %)
10	46.36	62.73 (+35.29)	61.82 (+33.33)
20	27.27	43.18 (+58.33)	40.00 (+46.67)
30	20.00	34.24 (+71.21)	29.09 (+45.46)
40	16.14	28.86 (+78.87)	24.09 (+49.30)
50	14.73	24.91 (+69.14)	20.91 (+41.98)
60	12.88	22.12 (+71.77)	18.94 (+47.06)
70	11.04	20.78 (+88.24)	17.40 (+57.65)
80	9.77	19.43 (+98.84)	16.02 (+63.95)
90	8.69	17.78 (+104.65)	14.34 (+65.12)
100	8.09	16.36 (+102.25)	13.46 (+66.29)
Total averaged	17.50	29.04 (+77.86)	25.61 (+51.68)

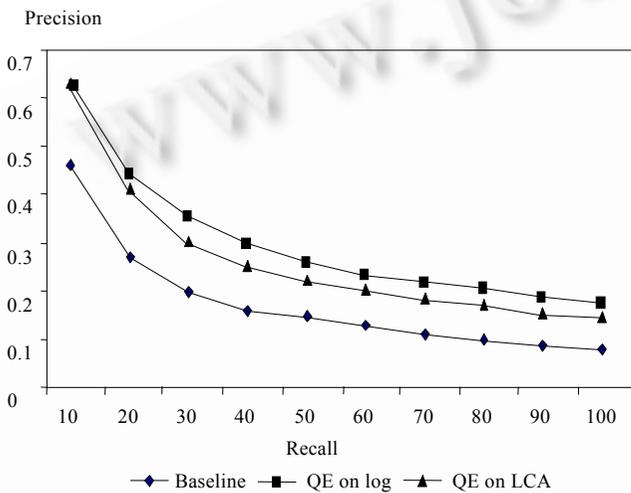


图 1 查询性能比较

图 1 查询性能比较

从表 2 可以看出,在测试数据集上,基于日志的统计模型确实获得了相当好的结果,它相对于基线的提高百分比最大可以达到 104.65%,平均为 77.86%.与基准算法比较的检验值 $t=11.38$,在自由度为 9 的情况下,其显著性水平为 0.000 000 35,证明该提高在统计上是有一定意义的.在与局部上下文分析的比较中,我们提出的算法同样获得 16.91%的提高,其显著性水平为 0.000 002 8,表明改善同样是显著的.另外,使用局部上下文分析方法进行查询扩展相对于未加查询扩展提高了 51.68%.

实验表明,查询扩展优化技术对于查询短小、文档集内容比较分散的情形尤为适用,可以极大地提高查询精度和查全率.同时,由于初始查询效果并不理想,使得受

此影响较大的局部上下文分析扩展算法在性能上低于基于日志的统计扩展算法,这主要是由于初始检索结果的偏差使得其扩展用词中加入大量无关词汇,从而未能达到最佳效果.本文所提出的算法建立在对大量用户长期查询行为的基础上,在选择扩展词汇时更具有针对性,有效防止了扩展用词的分散,并避免了初始查询结果对后续扩展查询的影响,更适应于当前 Internet 上信息检索的实际情况.

同时,还对该方法进行了词组影响对比和扩展用词数量影响的实验(见表 3),表明词组对查询性能起着重要作用.在加入词组的情况下,查询性能平均提高了 24.03%.其原因是,实际情况中查询长度短、少量的关键词,如果经常同时出现在查询中,则无法保证关键词相互独立出现的假设前提.如“apple”和“computer”是两个独立意义的词,但两者同时出现则代表另一个含义——苹果电脑公司.这种情况下只有通过将单词结合为词组,使得统计中各元素成为基本单位,才能更好地发挥该模型的统计意义.在系统中使用的词组也是从用户查询日志中抽取的.

查询扩展是对原查询的补充和优化,其扩展用词的数量也并非多多益善,实验表明,向原查询加入 30 个扩展词时查询性能达到最高,超过 30 个扩展词后查询性能下降较快,如图 2 所示.这是由于更多的相关度排在后面

的扩展词不但不会起到优化原查询的作用,反而会加入噪声,使得查询的歧义性增大。

Table 3 Comparison of query expansions with and without phrases

表 3 词组对查询性能的影响

Recall	Precision (%)		
	Without phrases	With phrases	Improvement
10	53.64	62.73	16.95
20	35.46	43.18	21.80
30	28.49	34.24	20.21
40	22.96	28.86	25.74
50	20.73	24.91	20.18
60	17.88	22.12	23.73
70	16.49	20.78	25.98
80	15.00	19.43	29.55
90	13.74	17.78	29.41
100	12.91	16.36	26.76
Total averaged	23.73	29.04	24.03

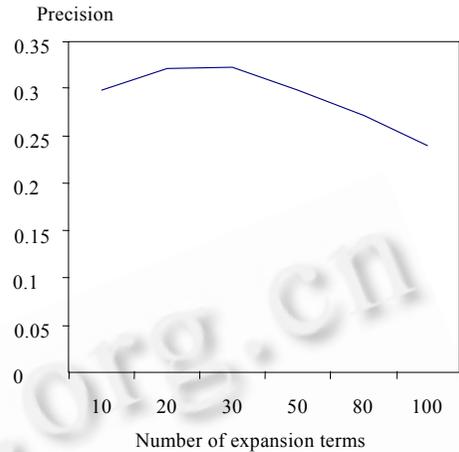


Fig.2 Retrieval performance with various number of expansion terms

图 2 扩展用词数量对查询性能的影响

4 结束语

本文提出了一种崭新的查询扩展优化算法——基于用户日志的统计模型。在典型的测试数据集上,将其与未加查询扩展及目前性能较好的局部上下文分析扩展算法进行了实验比较,结果表明,该方法在查询精度和查全率指标上均显著优于对比方法。同时,由于所使用的实验数据来源于真实的用户访问记录,因此,在消除用户查询用词歧义性方面具有更明显的优势。

在研究过程中我们发现,查询扩展优化仅是用户日志挖掘的应用之一,还可以将基于用户查询日志的信息挖掘应用到文档分类及用户行为模型的分析与建立中去,以最大限度地利用现存的大量用户日志。

References:

- [1] Furnas GW, Landauer TK, Gomez LM, Dumais ST. The vocabulary problem in human-system communication. *Communication of ACM*, 1987,30(11):964~971.
- [2] Wen JR, Nie JY, Zhang HJ. Clustering user queries of a search engine. In: *Proceedings of the 10th International World Wide Web Conference (WWW10)*. New York: ACM Press, 2001. 162~168.
- [3] Xu JX, Croft WB. Query expansion using local and global document analysis. In: Frei HP, Harman D, Schauble P, Wilkinson R, eds. *Proceedings of the 19th Annual International SIGIR Conference on Research and Development in Information Retrieval*. New York: ACM Press, 1996. 4~11.
- [4] Xu JX, Croft WB. Improving the effectiveness of information retrieval with local context analysis. *ACM Transactions on Information Systems*, 2000,18(1):79~112.
- [5] Deerwester S, Dumai ST, Furnas GW, Landauer TK, Harshman R. Indexing by latent semantic analysis. *Journal of ACM Transactions on Information Systems*, 2000,18(1):79~112.
- [6] Qiu Y, Frei H. Concept based query expansion. In: Korfhage R, Rasmussen EM, Willett P, eds. *Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York: ACM Press, 1993. 160~169.
- [7] Attar R, Fraenkel AS. Local feedback in full-text retrieval systems. *Journal of the ACM*, 1977,24(3):397~417.
- [8] Buckley C, Salton G, Allan J, Singhal A. Automatic query expansion using SMART. Technical Report, TREC-3, 1995. 69~80.
- [9] Ricardo B-Y, Berthier R-N. *Modern Information Retrieval*. England: Pearson Education Limited, 1999.
- [10] Hull D. Using statistical testing in the evaluation of retrieval experiments. In: Korfhage R, Rasmussen EM, Willett P, eds. *Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York: ACM Press, 1993. 329~338.