用户行为分析在网络信息检索中的应用概述*

刘奕群　张敏　马少平

清华大学计算机科学与技术系 北京 100084

E-mail: liuyiqun03@mails.tsinghua.edu.cn

摘　要：在网络信息资源持续膨胀的情况下，用户行为分析已经成为网络信息检索研究的重要热点。用户行为分析不仅对改进信息检索算法指出有益的方向，而且事实上已经成为任何一种成熟的网络信息检索评测方案不可缺少的一部分。本文试图从改进检索算法以及评测检索效果两方面对用户行为分析的研究情况进行概述，并对用户行为分析在网络信息检索中的应用做出展望。

关键词：用户日志分析，网络信息检索，效果评测

Review of User Behavior Analysis in Web IR Research

Yiqun Liu, Min Zhang, Shaoping Ma

Department of Computer Science and Technology, Tsinghua University, Beijing 100084

E-mail: liuyiqun03@mails.tsinghua.edu.cn

Abstract: With the page explosion of Web environment, user log analysis has become more and more important and interesting for Web information retrieval. User log analysis proposes possible methods to improve retrieval algorithms. Almost all kinds of retrieval performance evaluation methods also make use of user log information. This paper tries to give a brief review of user log analysis in Web IR research in two aspects: how it improves retrieval algorithm and how it is used for performance evaluation.

Keywords: User log analysis, Web Information Retrieval, Performance evaluation.

1　引言

用户行为分析是网络信息检索技术得以前进的重要基石，也是能够在商用搜索引擎中发挥重要作用的各种算法的基本出发点之一。检索用户的行为对于网络信息检索系统研究而言是最可宝贵的反馈信息。对于网络通用搜索引擎这种反馈显得格外重要，庞大的用户群体为搜索引擎带来的不仅是巨大的挑战，其群体行为也带来了调整检索系统算法的必要依据，进而成为推动网络信息检索技术发展的重要动力。面对几乎是不可穷尽的数据对象，用户在使

*本文工作得到国家重点基础研究与发展 973 计划（2004CB318108），国家自然科学基金（60223004, 60321002, 60303005）和教育部科学技术研究重大项目资助（104236）支持。

用搜索引擎时的行为不仅直接的对查询结果进行了评价，也间接的对 Web 页面的质量，进而甚至对搜索引擎相关的商业产品（如竞价广告等）的优劣程度给出一个相对客观的评价。

国内外研究者一直对网络信息检索工具（主要是搜索引擎）的用户行为分析给予了足够的重视，早在网络信息检索研究开始之前，用户相关反馈在文本检索中的研究就被认为是提高检索精度的主要途径之一。面对复杂的用户需求与更加复杂浩繁的文档集合，通常只有几个字词组成的用户查询成为了影响用户与检索系统之间信息传递的瓶颈，而相关反馈则成为了更明确表达用户意图的有效方案。尽管由于增加了使用成本，用户相关反馈并没有能够在文本检索中得到大规模的应用，但这方面的工作后来却成为了伪相关反馈（pseudo relevance feedback）、查询扩展（Query expansion, QE）等信息检索核心技术的基础。与此同时，对网络用户的行为分析自从 WWW 得到大规模应用以来，也得到了足够的关注。Cockburn[1]，Catledge[2]和 Tauscher[3]等人就分别在 90 年代中期左右对 Web 用户的浏览行为进行了调研和分析，Byrne[4]等人还在 1999 年提出了 Web 用户目的分类的概念，以此对用户浏览 WWW 的整个行为进行建模，但上述工作都没有将注意力集中到网络信息检索的范畴中来。

网络信息检索工具得到普及之后，面向网络信息检索的用户行为分析得到了更多的关注，主要的搜索引擎技术供应商如 AltaVista，Yahoo!，Excite 等都花费了巨大的人力物力进行基于用户行为日志的相关研究。总体而言，这方面的研究成果可以按照目的不同划分为两类，即针对检索系统算法设计改进的用户行为分析，以及针对检索系统性能评测的用户行为分析。

在下文中，我们将首先介绍网络信息检索用户群体的大致特性，以及用户日志的主要组成；然后分两部分针对算法改进和性能评测介绍当前用户行为分析的主要研究成果；最后根据前人的研究成果，对利用用户日志分析改进检索性能做出展望。

2. 网络信息检索的用户群体以及用户日志构成

根据 Sullivan 的统计[5]，Google 是世界上访问频率最高的搜索引擎，2004 年底，Google 每天处理的用户查询超过 2.5 亿个。而最近两期公布的中国互联网络发展状况统计报告[6][7]则指出，2004 年中国搜索引擎用户已占互联网用户的 95.2%，绝对用户数超过 8000 万人，每天的搜索请求量达到近 1.9 亿次。包括搜索引擎在内的网络信息检索工具已经成为网络用户获取信息的主要手段，86.6%的用户指出，搜索引擎已经是他们得知新网站的主要途径，而 65.0%的用户指出搜索引擎是他们经常使用的网络服务功能。

面对如此巨大的用户群体，搜索引擎用户日志的组织形式变得格外重要，如何将用户使用搜索引擎过程中最重要的行为加以记录又适当忽略掉次要的行为，是十分值得探讨的问题。在这方面，AltaVista 的用户日志组成形式[8]得到了广泛的承认和应用，也成为了许多用户日志分析[8][9][10]研究普遍使用的对象。该日志结构主要由一系列的查询需求组成，而每一个查询需求都包括时间戳（timestamp），浏览器标识符（cookie），查询词（query terms），每页返回的结果数（bf result screen），其他用户需求（other user-specified modifiers），用户信息（submitter information）等方面的内容。除了记录用户查询需求之外，AltaVista 最近制定的日志标准还记录了用户点击结果页面的信息，以及用户使用查询辅助工具的信息等。在增加这些信息之后，从日志中就可以提取出某个用户一次完整的查询行为。

针对用户日志记录的大量信息所进行的用户行为分析可以分为两个层次，其一是建立在对大量的用户需求进行统计的基础之上的宏观分析，分析的目的主要是寻找用户需求中的热点、词频分布规律、查询行为特点等，进而对检索系统的系统结构和算法设计做出改进；其二则是针对具体用户查询需求的微观分析，分析的目的则是找出某个用户查询对应的真实用户需求以及这个需求是否得到满足，进而确定检索系统的检索效果优劣。AltaVista 用户日志的记录格式无疑满足了这两个层次的分析需求，因此在研究界和产业界都得到了广泛的应用。

3. 用户行为分析在检索系统与算法改进中的应用

用户行为分析的结果对于检索系统结构与算法改进有着重要的指导作用，这是由于只有借助于这种分析，才可能了解纷繁复杂的用户需求背后的统计性规律，并找出隐藏于用户查询背后的真实需求，从而提高检索的效率和效果。

3.1 用户查询行为的宏观统计分析

对用户查询行为的宏观分析是针对检索系统用户行为分析中较早开展的一部分工作。在 1998 年左右，部分研究者如[8][11]等就开始对商业搜索引擎的用户日志进行大规模的分析，他们的研究着眼于描述网络检索用户的一般性行为，其主要分析对象包括查询需求的统计特征分布、用户使用检索工具的习惯、检索词同现情况等。Silverstein 等在[8]中指出了一系列对检索系统设计有深远影响的用户行为分析结果，这些结果包括：

- 用户需求中有大量的重复项，在被分析的近 5.8 亿个用户查询中，有 4.2 亿个用户查询是重复冗余的。这充分证明了在搜索引擎设计中引入缓存（cache）机制的必要性。
- 77.6%的用户在进行检索的过程中不对其查询进行修改。这突出了如查询扩展、伪相关反馈等协助系统理解用户查询的工作的重要性。
- 超过 85%的用户只翻看搜索引擎返回结果页面的第一页。针对这个现象而设计的各种评测指标应运而生，关于它们的信息将在第 4 部分详述。

上述不少分析方法对于中文搜索引擎的日志分析同样适用，但也同样有改进的余地，例如 Silverstein 和 Jansen 等人的分析中，并没有涉及对查询词频分布规律的分析，而这种分析对于改进检索系统的索引结构是至关重要的。

3.2 利用日志挖掘的用户查询目的分析

对用户检索目的的分析是近年来用户行为分析研究的热点之一，IBM 研究院的 Broder 首先提出了检索是由“任务驱动”（task oriented）的概念，在他构想的用户检索流程模型中，查询任务决定了用户的查询需求，进而反映在用户的查询词上。Broder 在[9]中指出，用户的查询任务包括以下三类：

1. 导航类（Navigational）：目标是查找某个特定的站点或者网页。
2. 信息类（Informational）：目标是获取可能位于一个或某几个网页上的信息。
3. 事务类（Transactional）：目标是查找能够处理某些以 Web 为媒介的事务的网页。

Broder 同时利用对 AltaVista 用户日志的分析和对用户调研的结果给出了三类检索分别所占的比例，即 20% : 50% : 30%。根据这个分析，信息类检索仍然是用户使用搜索引擎的主要任务，但事务类检索也占有了相当大的比重，而导航类检索所占的比重就较低。随后，雅虎公司的 Daniel 等人也提出了类似的任务分类概念，他们对 Broder 提出的三类体系进行了更详细的划分，并制定了利用日志对用户查询进行手工分类的详细步骤。他们分析的结果说明，三类检索所占的用户查询比例大致为 15% : 60% : 25%，这一结果与 Broder 的分析基本一致。

对查询任务进行划分的出发点在于，针对三类检索可以使用不同的检索模型、参数，甚至评价方法也随着检索类别的变化而有区别。因此实现检索类别的自动划分对于提高检索性能和增加检索评价的可信度都有非常重要的意义。TREC 在 2004 年引入了一个检索任务自动分类的子任务，最后的实验结果[12]说明自动分类的正确率约为 60%，能够与手工分类的效果比较接近。Lee 等人在加州大学伯克利分校内部使用的一个搜索引擎的用户日志上，得到了 90%左右的自动分类正确率[7]，但他同时也指出，这个正确率是在充分使用了用户点击信息的基础上得到的，如果不使用此类信息（在预测查询分类时，是无法得知用户点击信息的），则正确率与 TREC 的结论类似，即维持在 60%左右的水平。

3.3 针对特定检索用户群体的行为分析

针对特定用户的检索系统是近年来备受关注的一个发展方向，通常认为对特定用户行为的学习有助于提高检索系统的检索效果，因此对特定用户的行为分析研究也一直在进行当中。Navarro-Prieto 等人在 1999 年[14]对不同 Web 使用经验的检索用户的不同行为进行了比较，他们得出的结论是：有 Web 使用经验的检索用户在使用检索工具时比较不容易受检索结果表现形式的干扰。Zukerman 等人[15]则根据用户以往查询需求的信息，利用马尔科夫模型对接下来的用户需求进行预测。这方面的工作被认为对于提高检索工具界面与功能设计有一定的指导作用。在个人信息管理工具日渐强大的今天，这类行为分析也得到了越来越多的重视。

4. 用户行为分析在网络信息检索系统评价中的应用

网络信息检索系统的效果评价是研究领域面临的重大挑战，对于一般的信息检索系统而言，查全率（recall）和查准率（precision）是最基本的效果评价指标。然而一个不漏的在文档集合中找出所有相关文档，对于检索文档动辄以亿计数的网络信息检索而言几乎是不可完成的任务。这造成了传统的查全率和查准率模型对网络信息检索任务并不适用。无论是定位相关文档集合，还是建立新的评价模型，都需要涉及对用户行为的分析以及用户日志信息的挖掘，下面即分两部分介绍用户行为分析在网络信息检索系统评价中的应用。

4.1 定位相关文档集合

网络信息检索的数据对象即 Web 数据为定位相关文档集合带来了几乎不可逾越的障碍，根据美国加州伯克利大学的统计报告[13]，2002 年时 Web 数据就包括了超过 100 亿静态页面和超过 1500 亿个动态生成的页面。这使得对于任何查询主题而言，定位完整的相关文档集合

都变得不现实，因此这个问题的解决方案注定不可能完美的解决[17]。

文本检索会议（Text retrieval conference, TREC）[18]从 1992 年创立之初就将促进大规模文本信息检索的研究作为其首要目的，通过每年组织各种形式的检索评测，TREC 积累了丰富的在大规模文档集合中定位相关文档的经验，其核心技术被称为结果池过滤技术（pooling）。结果池过滤技术是基于对检索用户的行为分析而得出的。他假定用户在面临大规模文档集合时只能借助具有一定检索精度的检索工具进行信息获取，因此候选答案集合只可能通过检索工具进行定位。这一定程度上是出于无法手工筛选整个文档集合的权宜之举，但在处理类似网络数据集合这样规模的文档集时却又是必要与合理的。TREC 在应用结果池过滤技术构建相关文档集合方面积累了相当丰富的经验[19]。国内从 2003 年开始，也逐渐开始进行针对大规模网络文档集合检索的相关评测（如[20][22]等），而所采用的构建相关文档集合的方法，也基本沿用了结果池过滤方法。

除结果池过滤技术之外，对于某些特定种类的用户需求（主要是导航类查询需求）而言，还可以利用已有的网络信息资源自动寻找相关文档集合。美国在线公司的 Chowdhury 在[23]中就研究了利用开放目录计划[24]（ODP, Open Directory Project，一个利用志愿者标注网络资源的项目）自动查找导航类查询目标页面的可能性。由于导航类查询目标页面的唯一性，这种尝试得到了成功，但由于缺乏相应的网络资源，对于目标页面集合较大的查询需求而言，这种自动定位不可能得到广泛的应用。

4.2 确定评价指标

网络信息检索的特殊性质决定了基于对整个结果集合考察查准率和查全率的评价方式不再适用于网络信息检索的评价。Hawking 等在[19]中指出，网络信息检索的主要评价指标包括：

- 前 n 选精度（Precision at n, P@n）：前 n 篇中满足用户需求文档的比例。这个指标通常用来评测信息类或者事务类查询的性能。
- 前 n 选成功率（Success at n , S@n）：在前 n 篇结果文档中能否有满足用户需求的文档。这个指标通常被用于评测网络信息检索系统对各种类型查询的综合性能。
- 首现正确结果排序倒数（Reciprocal rank of first correct, RR）：检索系统返回的结果序列中第一个满足用户需求文档出现序号的倒数。这个指标通常用来评价导航类检索的性能。
- 平均精度（Average precision, AP）：检索系统返回每一个标准答案文档时精确度的平均值。这个指标可以用来衡量检索系统对各种类型查询的综合性能。

与前两个评价指标密切相关的典型用户行为是：根据[17]的统计分析，超过 85%的用户只翻看搜索引擎返回结果的第 1 页（一般来说，也就是前十个结果）。这个用户行为决定了尽管搜索引擎返回的结果数目十分庞大，但真正可能被绝大部分用户所浏览的，只有排在最前面的很小一部分而已。后两个评价指标 RR 和 AP 都是着重强调结果序列中最靠前文档相关程度的评价指标。这种偏向性是与网络信息检索用户期望尽快找到满足需求的结果页面的想法一致的，因此也是一个合理的设置。

由此可见，以上的主要评价指标都试图从某个侧面反映用户需求，这是他们得以在网络信息检索评价中得到广泛的应用的原因。TREC 在近年组织的网络信息检索评测[25][26][27][28]中，无一例外的采用了某个或某几个上述评价指标作为衡量检索系统性能的标

准。而针对中文网络信息检索的评测如[20][21][22]也都采用了类似的评价方式。

除此之外，部分研究者还设计了与用户行为直接相关的评价指标对检索系统的性能给出评价，如 Tang 等人在中提出的查找长度（用户在找到相邻两篇相关文档时需要阅读的其他文档个数）以及排序相关度（用户认为的相关程度排序与系统实际排序的相关程度）等。这类评价指标更精确的描述了用户检索行为，但由于数据的采集需要给用户的使用带来过多的不便，因此并不适用于大规模的检索系统尝试。

5. 结论与展望

在本文中，我们力图详尽地回顾近些年来用户行为分析在网络信息检索领域中各个方面的研究和应用。用户行为分析作为一种研究检索系统算法设计及性能评价的极为重要的工具，在网络信息检索研究领域具有极其重要的研究价值和广泛的应用背景。同时，网络信息检索毕竟是一个崭新的研究方向，用户行为分析的研究从总体上说尚处于一个起步的阶段，已有的研究工作正为这个领域提出越来越多需要解决的问题，下面是一些可能的研究方向：

1. 基于用户查询本身内容以及用户点击信息的查询自动分类系统已经能获得相当高的分类准确率，但在网络信息检索系统的实际应用中，如何在用户进行点击之前根据先验知识学习出查询所属的任务类别更为重要，这是一个十分值得关注的研究课题。
2. 正如前文（第 3.1 节）所指出的，当前大多数用户查询统计方面的研究并没有涉及对查询词频分布规律的分析，在中文检索系统中进行此类研究的更是凤毛麟角，但究其应用价值而言，这种分析对于改进检索系统的索引结构，确定哪些索引条目应当放入更容易访问到的存储区域是至关重要的。

现有的检索系统评测方法与指标已经能够比较完善的评价网络信息检索系统的性能，然而还缺乏在此基础上建立一个比较完善的针对中文搜索引擎性能的评价系统的努力，这也是我们未来可能关注的研究方向之一。

参 考 文 献

[1] Cockburn, A., & Jones, S. (1996). Which way now? Analysing and easing inadequacies in WWW navigation. International Journal of Human-Computer Studies, 45, 105-129.

[2] Catledge, L. D., & Pitkow, J. E. (1995). Characterizing Browsing Strategies in the World-Wide Web. Computer Networks and ISDN Systems, 27, 1065-1073.

[3] Tauscher, L., & Greenberg, S. (1997). How people revisit web pages: Empirical findings and implications for the design of history systems. International Journal of Human-Computer Studies, 47, 97-137.

[4] Byrne, M. D., John, B. E., Wehrle, N. S., & Crow, D. C. (1999). The tangled web we wove: A taskonomy of WWW use. In Human Factors in Computing Systems: Proceedings of CHI 99 (pp. 544-551).

[5] Danny Sullivan, Search Engine Sizes. In search engine watch website , http://searchenginewatch.com/reports/article.php/2156481

[6] 第 14 次中国互联网络发展状况统计报告，中国互联网络信息中心（CNNIC），2004 年 7 月。

[7] 第 15 次中国互联网络发展状况统计报告，中国互联网络信息中心（CNNIC），2005 年 1 月。

[8] Craig Silverstein, Monika Henzinger, Hannes Marais and Michael Moricz. Analysis of a very large Web search engine query log. In SIGIR Forum , fall 1998, Volumn 33 Number 1, 6-12.

[9] Andrei Broder, A taxonomy of web search. In SIGIR Forum, fall 2002, Volume 36 Number 2.

[10] Daniel E. Rose and Danny Levinson, Understanding User Goals in Web Search. In proceedings of the 13th World-Wide Web Conference (WWW13), 2004.

[11] Jansen, B. J., Spink, A., Bateman, J., & Saracevic, T. (1998). Real life information retrieval: A study of user queries on the Web. SIGIR Forum, 32(1), 5-17.

[12] N. Craswell and D. Hawking. Overview of the TREC-2004 Web track. In NIST Special Publication 500-261: The Thirteenth Text REtrieval Conference Proceedings (TREC 2004)

[13] Uichin Lee, Zhenyu Liu and Junghoo Cho, Automatic Identification of User Goals in Web Search, to be appeared in proceedings of the 14th World-Wide Web Conference (WWW14), 2005.

[14] Navarro-Prieto, R., Scaife, M., & Rogers, Y. (1999). Cognitive Strategies in Web Searching. Proceedings of the 5th Conference on Human Factors & the Web, June 1999.

[15] Zukerman, I., Albrecht, D. W., & Nicholson, A. E. (1999). Predicting Users' Requests on the WWW. User Modeling: Proceedings of the Seventh International Conference, UM99, 275-284.

[16] Lyman, Peter and Hal R. Varian, "How Much Information", 2003. Retrieved from http://www.sims.berkeley.edu/how-much-info-2003 on April 2th, 2004.

[17] D. Hawking, N. Craswell, P. Thistlewaite, and D. Harman. "Results and challenges in web search evaluation. ", in Proc. Eighth World Wide Web Conf., pages 243-252, May 1999.

[18] Text Retrieval Conference web site: http://trec.nist.gov/

[19] David Hawking and Nick Craswell, Very Large Scale Retrieval and Web Search, in Ellen Voorhees and Donna Harman, TREC: Experiment and Evaluation in Information Retrieval, MIT press, 2005.

[20] 国家 863 计划基础资源与评测，2003 年度信息检索评测大纲，http://www.863data.org.cn/src/863history/2003/2003fulltextretrieval_s.zip

[21] 国家 863 计划基础资源与评测，2004 年度信息检索评测大纲，http://www.863data.org.cn/src/2004eval/

[22] 北大网络实验室，SEWM-2004 中文 Web 检索测试指南，2004 年 10 月。

[23] Chowdhury, A., and Soboroff, I. Automatic Evaluation of World Wide Web Search Services. SIGIR'02, 421-422

[24] Open Directory Project, http://www.dmoz.org

[25] E. M. Voorhees and D. K. Harman, editors. The Tenth Text Retrieval Conference (TREC-2001), volume 10. National Institute of Standards and Technology, NIST, 2002.

[26] E. M. Voorhees and Lori P. Buckland, editors. The eleventh Text Retrieval Conference (TREC-2002), volume 11. National Institute of Standards and Technology, NIST, 2003

[27] D. Hawking and N. Craswell. Overview of the TREC-2002 web track. In Voorhees and Buckland [26].

[28] D. Hawking and N. Craswell. Overview of the TREC-2003 Web track. In NIST Special Publication: SP 500-255, The Twelfth Text Retrieval Conference (TREC 2003).

[29] Tang, M.-C. Sun, Y. (2003). Evaluation of Web-Based Search Engines Using User Effort Measures. Library and Information Science Research Electronic Journal 13(2).