

doi:10.3772/j.issn.2095-915x.2015.05.008

国内信息检索用户相关性研究现状分析

李洋

(西南大学计算机与信息科学学院 重庆 400715)

摘要: 本文采用文献计量和社会网络分析方法,从文献年代、学科、期刊来源、作者的分布以及研究主题等方面,对国内信息检索用户相关性的研究现状进行了分析。指出了我国现有研究中的不足,并给出了相应的改进建议。

关键词: 信息检索, 用户相关性, 相关性, 文献计量

中图分类号: G354, G353.1

Analysis of User-oriented Relevance in Information Retrieval Study of China

LI Yang

(College of Computer and Information Science, Southwest University, Chongqing 400715, China)

Abstract: Based on the methods of bibliometrics and social network analysis, the author analyzes the user-oriented relevance in information retrieval research of China by document's annual distribution, discipline, source of journal, authors distribution and research themes. It points out the deficiencies of existing researches in China, and puts forward suggestions for improvement.

Keywords: information retrieval, user-oriented relevance, relevance, bibliometrics

作者简介: 李洋,男(1991-),西南大学计算机与信息科学学院2014级情报学专业硕士研究生;研究方向:信息用户与服务,E-mail: youngli0328@163.com。

1 引言

自19世纪50年代人们对信息检索相关问题开始探讨以来^[1],相关性一直是信息检索领域的核心研究内容之一^[2],也是该领域一个非常重要的概念,得到了国内外研究人员的青睐,也取得了颇为丰硕的研究成果。从目前的研究成果来看,该领域的研究主题主要包括以下几个方面:相关性基本概念研究^[1]、相关性理论模型(框架)研究^[3]、相关性判断及其标准研究^[4,5]、相关性类型研究^[6]、相关性影响因素研究^[7]、相关性评价研究^[8]以及相关反馈研究^[9]等。可见,相关性研究的内容非常丰富。

目前学界已经普遍认同将信息检索相关性划分为两个研究范式,即面向系统的相关性和面向用户的相关性。用户相关性研究,对于提高检索系统的检索效率和效果有很大的现实意义和理论意义,同时对于我们更好地理解用户在不同情境下的检索需求有很大的帮助。虽然学界非常明确用户相关性研究的重要性,而且该方面的研究也在如火如荼的进行着,但是检索系统研究只能对此作无奈的观望,系统判断相关性仍然要通过词频计算等机械手段而无法利用用户相关性研究成果^[1]。就目前用户相关性的应用来看,相关反馈机制是最成熟的,但相关研究表明该机制虽然能够提高用户信息检索的效率^[10],但是在实践中用户检索信息时却很少运用^[11]。总体而言,用户相关性的研究取得了一定的成果,但就应用而言,情况还很不尽如人意。

今天,社会科学越来越重视用户研究,信息检索领域也不例外,虽然技术是信息检索相关性实现的基础,但是用户因素在相关性判断中也起着举足轻重的作用,随着信息检索方式从传统发展到现代,信息检索用户从专业人员普及到普通大众,当前大部分相关性研究也是面向用户展开

的。基于用户相关性研究的重要性,同时为了明确国内相关研究的进展,笔者认为有必要对该领域的研究现状进行分析。

2 数据来源与处理方法

2.1 数据来源

笔者以中国期刊全文数据库(CNKI)为数据统计来源,采取的检索策略为:“SU=(‘用户’+‘认知’+‘心理’+‘行为’+‘情境’+‘上下文’+‘专家’)*信息检索*相关性 OR TI=(‘用户’+‘认知’+‘心理’+‘行为’+‘情境’+‘上下文’+‘专家’)*信息检索*相关性 OR KY=(‘用户’+‘认知’+‘心理’+‘行为’+‘情境’+‘上下文’+‘专家’)*信息检索*相关性”,初步检索到相关文献509篇,检索日期为2015年6月。

2.2 数据处理方式

首先对检索到的文献进行处理,剔除了不相关及重复文献,接着利用Excel对检索结果进行分类汇总,将数据转化为规范化的易处理格式,然后结合文献计量学和社会网络统计分析方法,从文献的年代分布、期刊来源、作者和研究主题等方面进行分析,最后利用Sati、Ucinet软件对研究团体和关键词共现进行可视化展示并得出相关结论。

3 国内用户相关性研究现状分析

3.1 文献年代分布

对某一学科或领域的论文年度发文量进行统计分析,可以了解该学科或领域研究的发展历程和研究热度,揭示研究的纵向规律^[12]。近年来国

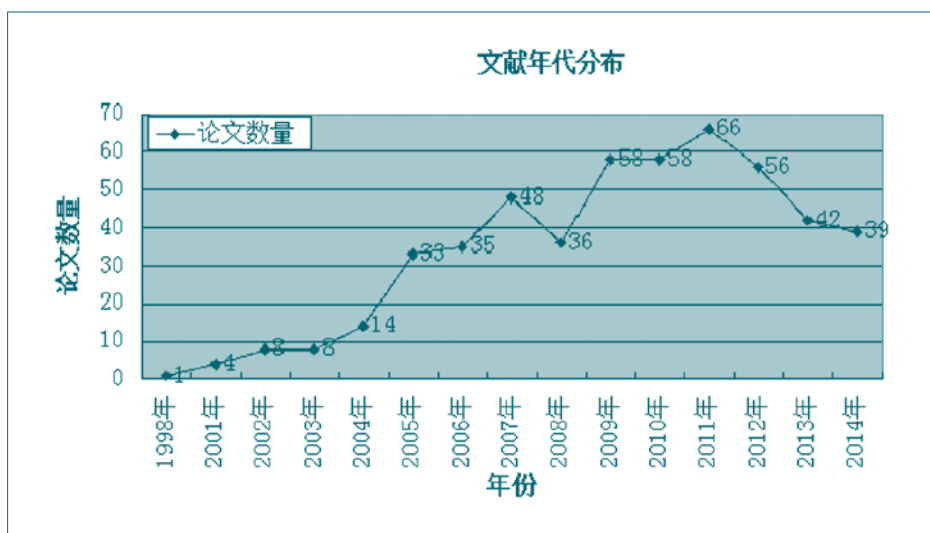


图1 国内信息检索用户相关性文献年代分布

内用户相关性研究的文献年代分布情况如图1所示（2015年的数据暂不统计）。

从图1可以看出，国内关于用户相关性的研究始于1998年，之后到2011年该领域的研究论文数量呈总体上升趋势，2011年之后呈下降之势。从上图中反映的结果来看，1998—2004年论文数量相对较少，只有2004年的论文数量达到两位数的14篇，说明这个阶段国内用户相关性的研究还处于起步、探索阶段；2005—2011年每年的发文量均在30篇以上，2011年达到最多的66篇，说明在此阶段伴随着各学科用户研究的不断发展，信息检索用户相关性的研究也日益受到关注；2011年之后，每年的发文量仍然在30篇以上，但是却呈逐年下降趋势，说明此领域的研究热潮已经过去，或者由于该领域的研究太过困难，许多学者都不再涉足。

3.2 论文学科分布情况

信息检索的实现需要强大的技术支持，在当前跨学科研究的背景下，某一领域的研究需要多学科专家共同努力，用户相关性研究也不例外。

表1统计了用户相关性研究相关的学科，由

于有些文献分属多个研究领域，所以论文数的统计结果（532篇）会多于最初检索到的数量。在CNKI数据库中选择“学科”为分组浏览方式统计分析论文的学科分布，可以看出，该领域研究最多的是计算机学科，占到所有论文的77.26%，其次是图书情报学科，所占比例为15.23%，还有自动化与电信类、语言学类、数学类、新闻类等，不过所占比例均相对较少。虽然用户相关性研究呈多学科态势，但是计算机和图书情报类才是该领域研究的核心学科。

表1 用户相关性研究论文学科分布

学科	论文数	学科	论文数
计算机类	411	图书情报类	81
自动化、电信技术类	15	经济类	6
医学类	6	数学类	4
地理类	3	语言类	3
新闻类	2	教育类	1

3.3 图书情报类论文来源核心期刊分布情况

论文的来源期刊可以在一定程度上反映该研究领域的空间分布特点，通常来说，文献分布

有其自身规律,大多数影响力较高的科研论文集中在核心期刊上,掌握了其核心期刊群,有利于了解该研究领域的研究视角和研究群体^[3]。在CNKI数据库中选择“学科”为分组浏览方式,数据显示与用户相关性研究相关的图书情报与数字图书馆类的论文共有81篇,占到全部论文的15.91%,相对而言论文数量偏少。其中学位论文共计21篇,占到25.93%,所占比例相对较高,

说明用户相关性在研究生学位论文选题中颇受垂青,其中有2篇以上的单位有3个,分别是南京大学5篇、武汉大学2篇、西南大学2篇,可见南京大学对该领域的研究比较成熟。其余60篇均为期刊论文,其中核心41篇,占到期刊论文的68.33%,表明本领域的研究成果质量相对较好。下面就这81篇论文中的专业核心来源期刊进行的分析,具体分布见表2。

表2 图书情报类论文来源核心期刊分布

序号	期刊名称	论文数量	所占比例
1	图书情报工作(含2篇网刊)	11	26.83%
2	中国图书馆学报	5	12.20%
3	情报杂志	5	12.20%
4	情报理论与实践	5	12.20%
5	图书情报知识	4	9.76%
6	情报科学	4	9.76%
7	现代图书情报技术	3	7.32%
8	图书馆学研究	2	4.88%
9	图书馆杂志	1	2.44%
10	图书与情报	1	2.44%

注:以《CSSCI(2014-2015)来源期刊目录》的“图书馆、情报与文献学来源期刊目录”现刊为图书情报类核心期刊来源。

从表2可以看出,41篇论文分布在10种期刊上,其中《图书情报工作》收录最多,共计11篇(含2篇网刊),占到核心期刊论文的26.83%,其余9种期刊载文均在5篇及以下。

3.4 论文作者分析

3.4.1 研究团体分析

大部分研究课题是基于科研团队之下的,研究成果也是团队成员共同取得的,团队合作对于科研的重要性是不言而喻的。为明确用户相关性研究的团队合作情况,本文借助Sati文献题录信息统计分析工具生成作者共现矩阵,接着将该矩阵导入社会网络分析工具Ucinet中生成了本领域

研究团队的整体网络示意图,如图2所示。

从图2可以看出,本领域研究团队共计15个,其中以二人居多,共9个,三人及其以上的团队共有5个,其中三人团队2个、四人团队2个,其余一个是7人团队。总体来讲,科研团队数量偏少,且以二人团队为主,说明本领域的科研合作情况不太理想。图中线条的粗细代表合作的强度的大小,只有孙建军和成颖的合作强度为3,其余为2或1,表明本领域科研团队的成员之间的合作强度还偏低。总体来讲,关于用户相关性的研究团体,不仅在数量上偏低而且就其合作强度而言也较低,所以这方面急需加强和改善,尤其需要不同学科、不同高校的研究人员进行合作。

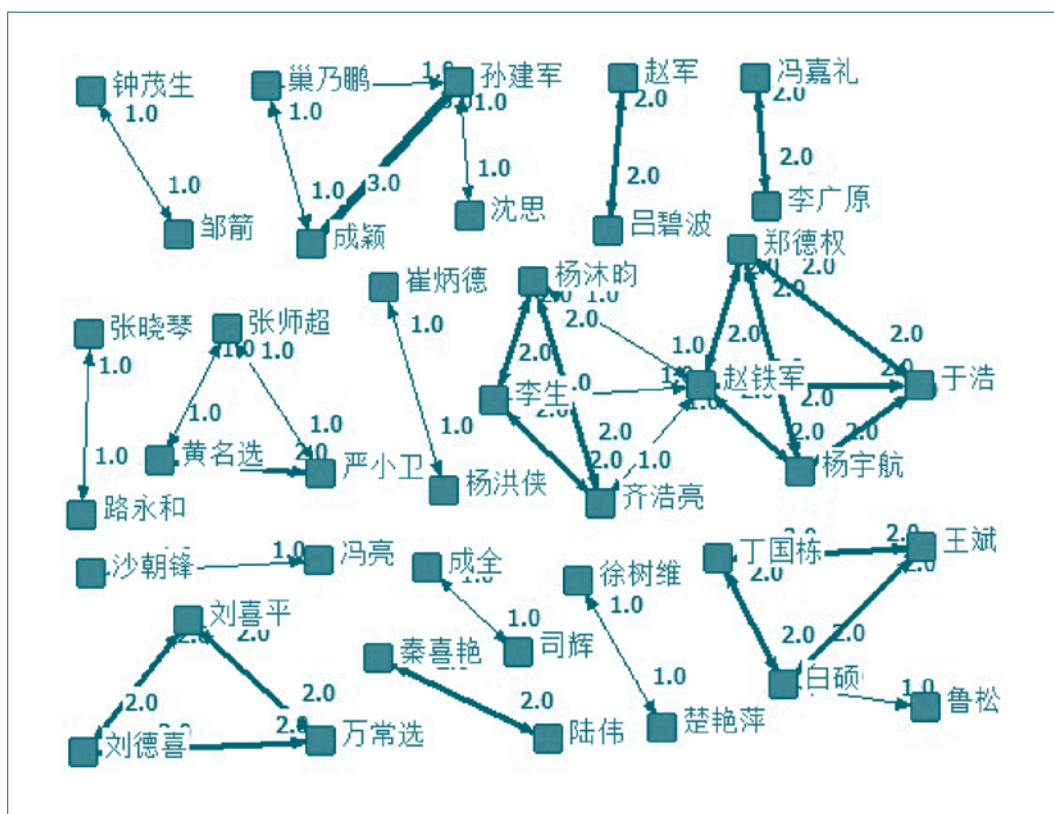


图2 用户相关性研究团体整体网络示意图

3.4.2 核心作者分析

核心作者是在某一学科领域研究中造诣较深、获得科研成果较多的学者，他们的研究水平和思想观点对该领域的发展具有重要影响^[14]。用户相关性的这些论文是由许多作者共同完成的，其中发文量最多的仅为4篇，共有3位学者，分别是南京大学的孙建军、成颖和武汉大学的陆伟；发表3篇论文的同有3人，分别是南开大学的王知津、哈尔滨工业大学的赵铁军和河南大学的徐树维；发表2篇论文的有武汉大学的吴丹、南京大学的巢乃鹏等共21人，其余均仅发表1篇论文。可见，关于用户相关性的研究学者虽然较多，但是比较分散，目前尚未形成具有代表性的核心作者群体，从机构方面来看，南京大学和武汉大学在用户相关性方面的研究实力相对较为强劲。

3.5 基于关键词的研究主题分析

关键词是从论文中精选出来的能够反映论文主题内容和关注点的词或词组，可以帮助读者快速了解文章的主题内容^[14]。而对某一研究主题的相关论文的关键词进行统计分析，能够很好地反映出该领域所关注的焦点以及今后关注点的发展趋势^[15]。

3.5.1 总体论文关键词分析

利用Sati文献题录信息统计分析工具对检索到的文献进行关键词词频统计，共得到关键词1438个，其中出现2次以上的关键词共计293个，全部关键词共计出现2519次，平均每篇约5个关键词。为了对研究主题进行分析，首先要对关键词进行合并（如：本体、Ontology等）和清洗（如：知网、综述等）等处理，然后将处理后的关键词按其出现的频次由高到底进行排序，从中选择词

连接分析等关键词, 该方面的研究与搜索引擎的相关研究较为密切, 因为搜索引擎返回给用户的结果一般是网页, 网络主题的研究可以为优化搜索引擎提供一定的参考, 因此也是相关性研究的主题之一。

图 3 中显示的研究主题不至以上提到的五个方面, 还有用户相关性的影响因素、评价、检索模型等方面的研究等, 总体来讲系统方面的研究主题还是偏多一些。

3.5.2 图书情报类论文关键词分析

相关性是情报学的基本原理之一^[17], 同时也是本学科非常重要的一个研究热点, 尤其是信息检索研究中最基本的概念^[18], 因此有必要对本学科关于用户相关性的研究主题进行梳理。

采用和上文同样的方法, 对图书情报类用户相关性的论文进行关键词(这里是合并和清洗之后的全部关键词)分析, 生成如图 4 所示的关键词共现可视化图形。

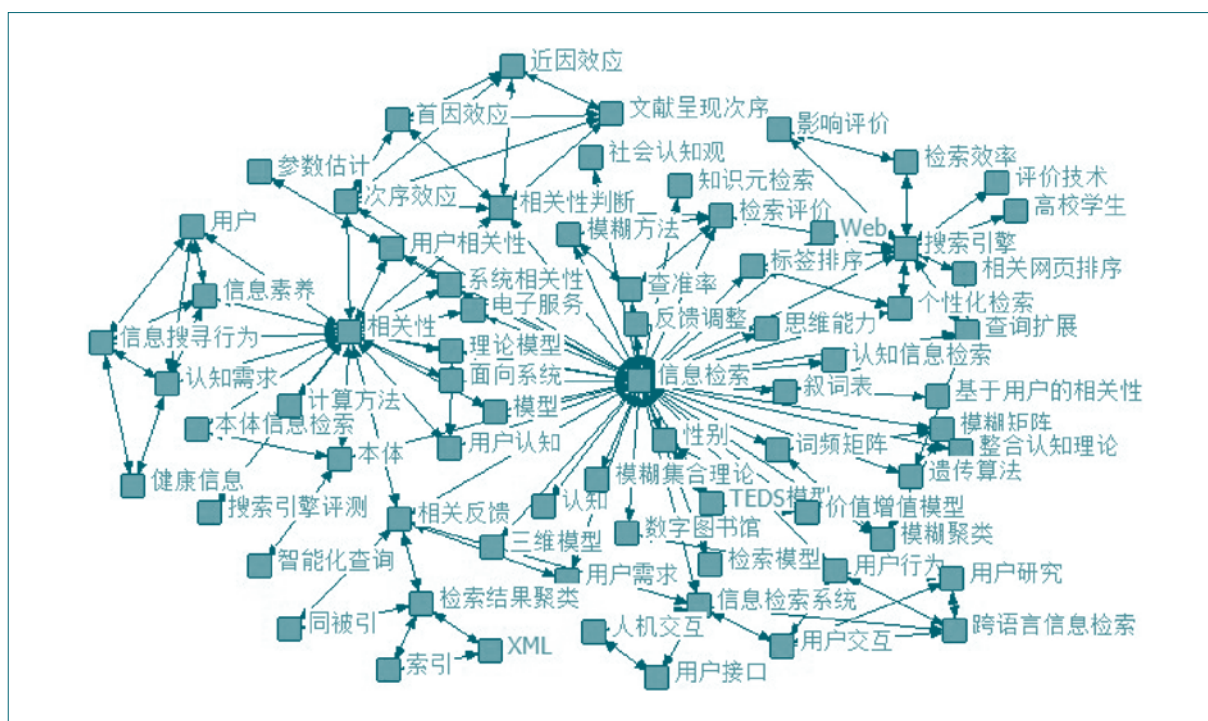


图 4 图情类论文用户相关性高频关键词整体网络示意图

与图 3 相比, 图 4 显示出了更多非技术层面的关键词, 说明本学科该领域的研究主题确实是以用户为中心来展开的, 出现了如认知信息检索、用户认知、思维能力、用户交互、用户需求等关键词。

总体来讲, 关于该领域的研究, 计算机学科偏向于技术, 图书情报学科偏向于理论, 这也符合两个学科研究方向的客观现实。计算机学科与图书情报学科是信息检索相关性研究最为贴近的两个学科, 因此, 在当前跨学科合作研究的大环

境下, 两个学科要各自发挥出自己的优势, 计算机学科要以图书情报学科的理论研究为导向, 图书情报学科要以计算机学科的技术为支撑, 努力做到信息检索用户相关性研究领域的理论与实践的紧密结合。但是从目前的研究现状以及用户的体验来看, 相关研究成果还没有具体应用到检索系统中, 用户还无法利用到这些研究成果去提高自己信息检索效率, 因此, 用户相关性方面的研究不能只停留在理论层面。

4 结语

理论具有解惑、预测的能力,好的理论能在一段时期解释和预测许多现象,所以理论研究属于研究工作的一部分,决定着某一领域的发展方向^[19],构建主义者甚至认为理论是实践研究的前期准备,在这个意义上,理论研究是一种隐性的实践^[20],由此可见,理论研究对于科学研究的重要意义及其价值所在,然而,科学研究的最终落脚点是实践,只有以理论为基础的实践应用研究才能推进科学研究的进一步发展。但是目前国内情报学理论与实践之间却存在着严重的错位现象,而且长期以来我国情报学理论与实践与服务实践严重脱节,两方的矛盾冲突始终成为社

会各界质疑情报学科独立地位的焦点话题^[21],信息检索作为情报学最重要的实践研究方向之一,此类现象更为突出,理论研究成果过于理想化。

从本文的分析来看,国内信息检索相关性研究起步较晚,虽然已经取得了比较丰硕的研究成果,但大多还只是处在理论研究阶段,实践应用方面的研究进展还不太理想,用户相关性方面的研究更是如此。因此,在未来的研究中,要以理论为指导,进行实践性研究,实现理论与实践应用的良好沟通,借助跨学科研究平台做到理论与实践的相结合,使信息检索用户相关性方面的研究成果为改善用户信息需求与查询结果之间的矛盾做出贡献。

参考文献

- [1] 王家铨. 信息检索中“相关性”概念的研究[J]. 现代外语, 2004, 24(2): 181-191.
- [2] 王雅坤, 成全. 信息检索相关性研究综述及发展趋势[J]. 图书与情报, 2012 (1): 88-94.
- [3] 成颖, 孙建军, 巢乃鹏. 信息检索中的相关性模型[J]. 图书情报工作, 2005, 48(12): 46-50.
- [4] 成颖. 相关性判断研究综述(2000-2010)[J]. 情报杂志, 2011, 30(9): 79-84.
- [5] 成颖, 孙建军, 宋玲丽. 相关性标准研究[J]. 中国图书馆学报, 2005, 31(3): 55-60.
- [6] 张新民, 化柏林, 罗卫东. 相关性研究探析[J]. 情报学报, 2008, 27(5): 691-697.
- [7] 于春, 彭爱东, 王波, 等. 信息用户对信息检索相关性判断的因素分析[J]. 图书情报工作, 2009, 53(3): 103-107.
- [8] 秦喜艳, 陆伟, 姜捷璞. 信息检索中的相关性判断和系统评价述评[J]. 图书情报知识, 2009(4): 89-94.
- [9] 石艳霞. 信息检索中“相关性”与“相关反馈”研究概述[J]. 晋图学刊, 2002 (2): 13-15.
- [10] Ruthven I, Lalmas M, & Van Rijsbergen K. Incorporating user search behavior into relevance feedback[J]. Journal of the American Society for Information Science and echnology, 2003, 54(6):529-549.
- [11] Saracevic T. Relevance: A review of the literature and a framework for thinking on the notion in information science. Part III: Behavior and effects of relevance[J]. Journal of the American Society for Information Science and Technology, 2007, 58(13): 2126-2144.
- [12] 巩海霞, 王明芝, 谷丽娜. 基于文献计量的个性化信息服务研究现状分析[J]. 情报科学, 2011, 29(3): 391-395.
- [13] 刘圣君, 屈宝强. 我国图书馆联盟研究的文献计量分析[J]. 情报科学, 2011, 29(3): 396-400.
- [14] 任红娟, 张志强. 基于文献计量视角的观点挖掘发展研究——情报学未来潜在研究领域解析[J]. 图书情报知识, 2010 (2): 55-63.
- [15] 王莹莉, 张敏. 国内微博研究现状综述[J]. 图书馆学研究, 2012(12): 2-8.

- [16] 黄敏, 赖茂生. 语义检索研究综述 [J]. 图书情报工作, 2008, 52(6): 63-66.
- [17] 马费成. 论情报学的基本原理及理论体系构建 [J]. 情报学报, 2007, 26(1): 3-13.
- [18] 张新民, 罗卫东. 相关性与情报学 [J]. 情报理论与实践, 2008, 31(1): 12-14.
- [19] 邹永利, 樊国晓. 信息检索的大理论及其实践 [J]. 情报科学, 2012, 30(5): 787-790.
- [20] Cornelius I. Theorizing information for information science [J]. Annual review of information science and technology, 2002, 36(1): 392-425.
- [21] 彭靖里, 李建平, 王澜瑾. 对中国情报学理论与实践研究错位及其原因的探究 [J]. 情报理论与实践, 2014, 37(11): 1-6.