|  |
| --- |
| **基于计量的网络用户行为研究分析** |
| 施夏玲 |
| MF1514046 |

**基于计量的网络用户行为研究分析**

**摘要**：随着网络的迅速发展，网络用户行为也逐渐成为学术界研究的热点，在实际应用中，网络用户行为分析逐渐成为网络信息检索系统进行架构分析、性能优化和系统维护的重要基石，商业网站等赢利性站点需要分析用户的行为及爱好来提供更满意的服务使得其利润最大化，政府、科研机构等非赢利性的网站也需要通过分析用户构成与其网络行为上的特点来构建科学的决策支持系统。本文旨在通过对web of science上关于“网络用户行为”研究文献的分析，利用相关工具进行相应处理，探究该主题目前研究的现状、热点以及基本分布。

**关键词**：网络用户行为，citespace，计量，可视化

**1研究背景**

**1.1基本概念**

网络用户行为是一个广义的概念，它是指用户在使用网络资源所呈现出的规律。[11]随着网络搜索用户的大规模增加，网络用户行为分析已成为网络信息检索系统进行架构分析、性能优化和系统维护的重要基石，是网络信息检索和知识挖掘的重要研究领域之一。[20]

用户行为可以分为信息查询行为、沟通交流行为、休闲娱乐行为、电子服务行为和电子商务服务行为等许多方面。互联网技术和大数据技术的发展，使用户对网络服务质量的要求逐步提升，网络服务提供者提供服务的方式也逐渐走向科学，通过对网络用户行为的分析和规律总结，可以将这些规律与产品或服务的经营策略相结合，发现目前经营服务中的问题，并为进一步提高服务质量和经营策略的制定都有非常重要的意义。[4]

用户行为分析方法可以有：以数据分析为导向；以产品设计反馈为导向；以用户调查为导向。

**1.2研究分类**

根据对相关文献的阅读可发现该主题下进行的研究多种多样，总结下来主要可以分为以下四类：

用户类型导向：包括不同类型网络用户行为研究，比如青少年网络行为、大学生网络行为等。

网络类型导向：不同类型网络中用户行为的研究，比如电子商务中用户行为或社交网站用户行为。

研究方法导向：比如网络日志挖掘方法或问卷调查方法或基于关键词分析方法。

行为影响导向：该部分主要研究网络用户行为所带来的社会或法律等的影响，比如很多论文论述到网络用户行为对网络安全的影响，还有针对社会问题比如青少年网络行为进行探究的研究。

**1.3文献综述**

在cnki中以“网络用户行为”为主题进行检索，检索结果按被引频次由高到低排列，选取被引频次较高具有代表性的若干篇文章进行分析，以了解该主题下研究的大体内容及重要分类和概念。相比于其他主题，该主题现在被研究的并不多，但随着网络的迅速发展，相信网络用户行为会得到越来越多的关注与研究。

论文[10]主要研究网络用户行为的可信度，对用户行为可信的研究不仅可以提高网络的安全性而且也可以提高网络的性能。文章[11]在系统研究Internet的用户行为特征的基础上，从监测和分析的角度给出了因特网用户行为的一种分类方式及表示方式，讨论了正常行为聚类研究的常用方法和异常行为的监测方法，设计了分布式的网络用户行为分析模型的体系结构，为网络用户行为的自动发现提出了一种有效的解决途径。论文[20]考察了用户搜索行为中的查询长度、查询修改率、相关搜索点击率、首次/最后一次点击位置分布以及查询内点击数分布等信息，基于不同类型的查询集合，考察用户在不同查询需求下的行为差异性。论文[8]基于位置的社交网络(MZ2F) 能够支持用户分享地理位置信息，提出基于划分层次，在不同的邻域半径下密度聚类的方法，探索基于位置的服务(MZ2) 平台上用户地理位置上相似性的度量，该方法在不同空间位置比例尺下观察用户访问各个聚类区域的次数，进而利用向量空间模型(d2c)计算用户在各个层级的相似性，最终以不同权重叠加各层级的用户相似性值，得出用户在地理空间行为上的相似性！论文[9]通过对不同领域间标签与关键词的相似度的分析入手，来分析领域间标签与关键词相似度的差异，进一步来探讨差异存在的原因和影响因素，从而揭示了网络用户的行为特点。论文[5]聚焦于社交网络，基于新浪微博平台，设计并实现了一个用户关系网络挖掘系统，对微博用户构建的关系网络进行分析，帮助理解网络用户行为和监控微博网络舆情传播，尤其对微博上某些敏感人物的行为进行追踪和调查。论文[16]提出一种基于Web 关联规则挖掘、页面内容和会话相似度相结合的研究方法，聚类出用户频繁访问的页面组，以发现网络用户行为的规律。论文[14]借鉴已有的研究工作，以复杂的网络环境为研究背景，通过对复杂网络环境中用户行为的研究，引入云模型理论，研究并提出了一个基于云模型理论的网络用户行为评判模型，实现了对网络用户行为定量描述及评判；同时研究结果还体现了网络用户行为的不确定性和评判标准的差异性，这样使网络行为评判结果更加符合客观事实。论文[15]通过分析目前典型的个性化推荐算法的特点，提出了基于贝叶斯网络的用户行为模型，该模型能准确反映用户的执行动向。为了解决同种服务类型内多个服务的偏好问题，提出了基于向量空间模型的用户兴趣模型。论文[6]同样基于社交网络，主要介绍了网络数据的抓取，实现数据的可视化以及舆情信息的分析三个方面。论文[1]同样基于社交网络，选取大约263万个微博用户的真实数据，对用户标签的分布进行了研究和分析。主要考察了用户标签的宏观分布特征，以及用户标签与关注对象的标签分布之间的联系，发现微博用户给自己添加标签时，在开始阶段倾向于使用反映个性的标签，之后会出于从众心理而选用大众化标签。论文[18]同样基于社交网站，社交网络的流行对用户的隐私保护提出了新的挑战。该文通过使用人类动力学和统计物理的方法，研究用户的网络行为与用户隐私量值的关系。论文[17]基于大规模真实网络用户的行为日志，对用户与网络搜索引擎系统的交互过程和用户决策过程展开研究。通过比较具有相关信息的用户点击和普通点击的分布，对用户点击的3 类上下文背景特征进行分析，从而实现对用户点击的可靠性评估。论文[2]在介绍建立网络用户行为模型相关工具的此基础上，建立起基于CBMG 的网络用户行为模型，最后通过实例证明该模型能有效地揭示网络用户行为，发现网络瓶颈，改善网络信息资源组织。论文[12]结合某高校主页的真实运行数据，通过Web 使用挖掘对于网站的运行日志文件进行全面的挖掘分析，分析用户对信息内容的兴趣度，并通过用户对网页的访问数据推算出各个页面受众的兴趣度高低，借此改良网站的内容和布局。论文[3]针对网络运营商所关心的用户行为分析问题，探讨如何对网络用户的行为进行分析，提出了一种用户行为分析分类的模型。论文[2]分别对中德两国18-35 岁青年开展调查，尝试从国别角度探讨中德两国青年网络社交行为特征差异及其文化成因。论文[4]阐述了大数据的概念和特征，分析了大数据对于网络用户行为的影响，并对基于大数据的网络用户行为分析模式进行了实证研究。

通过以上综述可以发现，针对网络用户行为的研究有很多，比如用户信任机制的建立、基于不同数据来源分析不同类型数据、对网络用户行为分析的方法探讨以及不同国家之间用户行为的不同等，但是其中研究最多的就是不同类型网站中用户的行为规律以及由此可以应用的可能性和网络用户行为研究的方法。

**2数据来源和研究方法**

以 SCI-EXPANDED SSCI A&HCI CPCI-S CPCI-SSHI 数据库为来源，以“online user behavior”为主题进行检索，结果有4851条，由于结果比较多，所以在基本分析的时候使用全部数据，而在共引分析时使用以该关键词为题名检索出来的156条结果，检索时间为2016年6月25日。使用软件为可视化软件citespace第三版，版本号为4.0.R5.SE.64-bit.12.29.2015。

研究主要分为基本分析和共引分析，基本分析包括研究文献的国家地区以及关键字等分布，共引分析主要包括作者共被引、期刊共被引、参考文献共被引。

**3可视化分析**

**3.1基本分析**

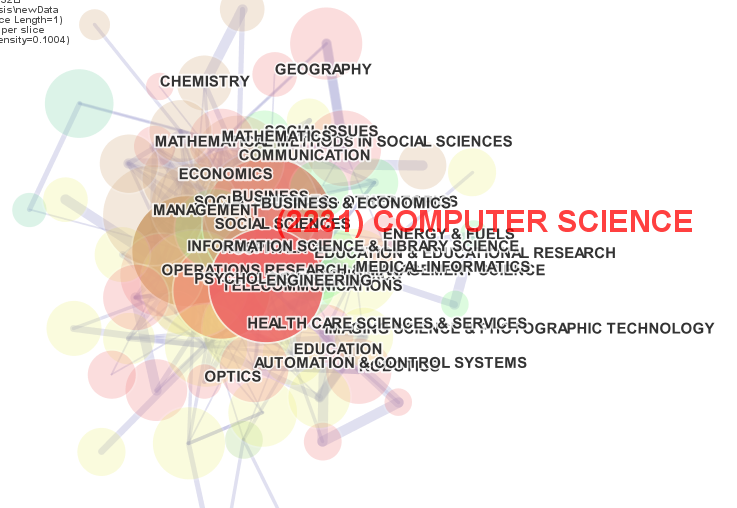
该部分包含该主题研究相关的基本分析，包括国家、机构、研究领域、关键字等分析，旨在了解该主题目前研究的主要热点及基本概况。

**3.1.1领域分析**

这些奠定知识基础的关键节点文献进行学科分类，节点类型选择Category。网络用户行为研究涉及多个学科，正在向更多领域渗透而且很多学科之间交叉合作，跨领域的情况越来越多见。由下表可以看到，研究最多的领域为计算机科学，频次达到2231，这也在预料之内。其次是经济领域，频次达到739，需要说明的是，通过对网络用户行为的研究从而探索用户在进行网上消费时的规律及偏好达到提高收益精准营销的目的，该主题是很多电子商务领域研究的热点，事实上，电子商务也是研究该主题的重要领域。还有一些常规的领域，包括心理学及管理学，在该主题上同样有很多研究，比如用户在网络环境下心理上不同于现实中的变化，类似的课题不仅是学术研究热点，也是社会关心的话题。

表一：网络用户行为研究涉及领域

|  |  |  |  |
| --- | --- | --- | --- |
| 排序 | 频次 | 中心度 | 来源领域 |
| 1 | 2231 | 0.28 | Computer science |
| 2 | 739 | 0.08 | Business&economics |
| 3 | 724 | 0.33 | Engineering |
| 4 | 512 | 0.08 | Information science&library science |
| 5 | 461 | 0.3 | Psychology |
| 6 | 376 | 0.04 | Business |
| 7 | 302 | 0.04 | Management |



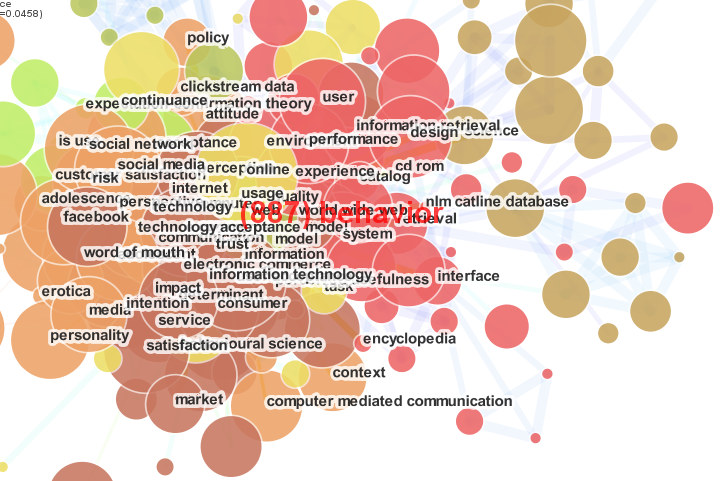
图一：computer science领域共现图

**3.1.2关键词分析**

通过对关键字进行分析可以了解该领域当前研究的主要内容，排名在前面的关键词多是高相关的直接语词，但其中也包含重要概念，比如模型，频次达到480，说明研究网络用户行为模型占很大比例，由前面综述也可以看出，利用各种算法构建模型或者探究构建方法是普遍研究的热点。信息技术排名第五，频次达到341，主要涉及研究该主题的技术手段。而接下来的两个关键词涉及用户信任度，这也是网络用户行为研究的一个重要方面，包括网站安全监控与用户信任机制的建立等。通过对频次最高的“行为”关键字进行共现分析也可以看到对应于研究分类，“信息检索”、“电子商务”、“社交网站”等出现频次都很高，特别的，甚至社交网站的代表facebook出现频次同样很高。

表二：网络用户行为研究关键词

|  |  |  |  |
| --- | --- | --- | --- |
| 排序 | 频次 | 中心度 | 关键字 |
| 1 | 887 | 0.35 | Behavior |
| 2 | 812 | 0.06 | Internet |
| 3 | 480 | 0.08 | Model |
| 4 | 449 | 0.14 | Online |
| 5 | 341 | 0.08 | Information technology |
| 6 | 337 | 0.04 | User acceptance |
| 7 | 327 | 0.03 | Trust |
| 8 | 268 | 0.07 | Information |
| 9 | 237 | 0.01 | System |



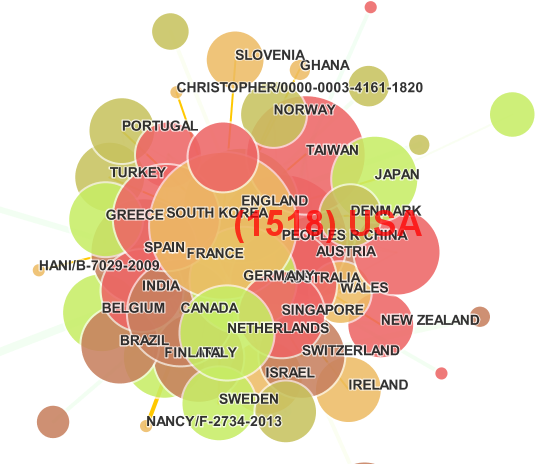
图二：行为相关关键词共现图

**3.1.3国家分析**

由表三可以看到，美国不出意外地在该主题研究中站绝对地位，频次第一，并且与排名第二的中国相差很大，值得注意的是台湾地区排到了第三名，说明其在该主题研究中占有重要比重，英国排名也比较高，同样还有英国、澳大利亚等。从共现图可以看到，国家之间的合作相当高，特别是排名在前列的国家之间合作覆盖面比较大，所以随着国际间的合作的日益加强，国家间合作也越来越多。

表三：网络用户行为研究国家分布

|  |  |  |  |
| --- | --- | --- | --- |
| 排序 | 频次 | 中心度 | 国家或地区 |
| 1 | 1518 | 0.51 | 美国 |
| 2 | 619 | 0.09 | 中国 |
| 3 | 337 | 0.05 | 台湾 |
| 4 | 322 | 0.3 | 英国 |
| 5 | 219 | 0.09 | 德国 |
| 6 | 216 | 0.09 | 澳大利亚 |
| 7 | 208 | 0.11 | 西班牙 |



图三：以美国为中心的国家共现分析

**3.2共现分析**

**3.2.1作者共被引**

由于wos在采取完整数据时每次下载的数据量有限制，所以由于时间有限，在分析共被引时选取了相关程度比较高的部分文章进行研究，这样在视图呈现以及清晰度方面都有保证。

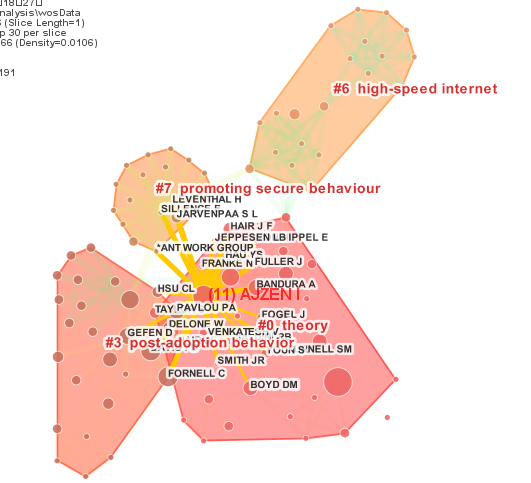
由表四可以看到匿名作者占据最多数，频次为26，排名第二的为Ajzen I，所属聚类号为0，即多为理论研究；排名第三的为Fornell C，所属聚类号为3，研究范畴多为采纳后行为。其他的研究类别有回溯性综述、安全性行为、网络用户等。通过共被引图也可以看到作者之间合作也比较多，特别是作品频次比较高的作者之间，所以合作在一定程度上促进了知识发展，而该主题领域的迅速发展也会引来更多的合作。

表四：作者频次及聚类分析

|  |  |  |
| --- | --- | --- |
| 频次 | 作者 | 聚类号 |
| 26 | [Anonymous], 2011, SO, V, P | 0 |
| 11 | Ajzen I, 2010, SO, V, P | 0 |
| 9 | Fornell C, 2011, SO, V, P | 3 |
| 9 | Davis FD, 2011, SO, V, P | 3 |
| 8 | Hoffman DL, 2005, SO, V, P | 0 |
| 8 | Venkatesh V, 2012, SO, V, P | 0 |
| 7 | Bhattacherjee A, 2005, SO, V, P | 3 |
| 7 | Newman MEJ, 2009, SO, V, P | 35 |

表五：作者研究内容类别

|  |  |  |
| --- | --- | --- |
| 聚类号 | 聚类大小 | 标签 |
| 0 | 38 | theory |
| 1 | 30 | web user |
| 2 | 28 | practitioner end user |
| 3 | 27 | post-adoption behavior |
| 4 | 22 | internet-based instrument |
| 5 | 22 | retrospective review |
| 6 | 20 | high-speed internet |
| 7 | 17 | promoting secure behavior |



图四：作者共被引及聚类分析

**3.2.2参考文献共引**

参考文献所属类别主要包括站点、网络行为、网络基础设施等，文献被引次数最多的为5，所属类别为“实证研究”；排名第二的文献为4，所属类别为“网站”。由于样本比较少，参考文献之间的共被引比较分散，可视化效果不是很好，但是在有限的样本里面依然可以看到参考文献之间有明显的共被引情况。

表六：被引文献聚类类别

|  |  |  |
| --- | --- | --- |
| 聚类号 | 大小 | 类别标签 |
| 0 | 31 | web site |
| 1 | 30 | online behavior |
| 2 | 21 | practitioner end user |
| 3 | 21 | internet-based instrument |
| 4 | 19 | empirical study |
| 5 | 14 | information behavior |
| 6 | 14 | high-speed internet |
| 7 | 13 | self-report |
| 8 | 13 | promoting secure behavior |
| 9 | 12 | ergotracer |

表七：参考文献被引频次及所属类别

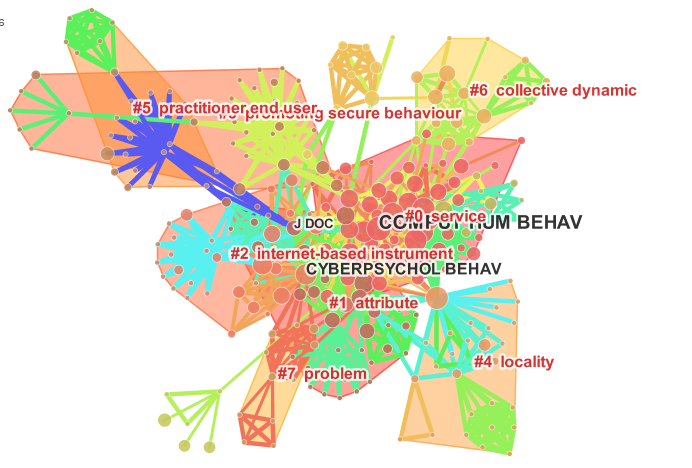
|  |  |  |
| --- | --- | --- |
| 被引频次 | 被引文献 | 聚类号 |
| 5 | Xiang Z, 2010, TOURISM MANAGE, V31, P179 | 4 |
| 4 | Buhalis D, 2008, TOURISM MANAGE, V29, P609 | 0 |
| 4 | Boyd DM, 2007, J COMPUT-MEDIAT COMM, V13, P210 | 4 |
| 3 | Spink A, 2004, WEB SEARCH PUBLIC SE, V, P | 5 |
| 3 | Bhattacherjee A, 2004, MIS QUART, V28, P229 | 12 |
| 2 | Borna C, 2000, TELECOMMUNICATIONS, V34, P83 | 11 |
| 2 | Albert R, 2002, REV MOD PHYS, V74, P47 | 15 |
| 2 | Ahn YY, 2007, P 16 INT C WORLD WID, V,, P | 46 |
| 2 | Giles DE, 2005, EWP0511, V, P | 64 |
| 2 | Newman MEJ, 2003, SIAM REV, V45, P167 | 65 |

**3.2.3期刊共引分析**

期刊共引分析遵循同样的研究路径，在citespace中先对期刊刊载的文献数进行排序找到刊文数比较多的期刊，随后对期刊所属研究领域进行聚类分析，结果如下所示。刊文数最多的期刊为“Comput HUM BEHAV”，刊文数为27，研究聚类属于服务类；排名第二的期刊为“J MARKETING RES”，同样研究聚类属于服务类，事实上由表格可以看到多数期刊研究类别都属于服务类，其余的还有“基础网络建设”、“可信任行为”等。

表八：期刊被引频次及所属类别

|  |  |  |
| --- | --- | --- |
| 刊文数 | 期刊名 | 聚类号 |
| 27 | Comput HUM BEHAV, 2003, COMPUT HUM BEHAV, V, P | 0 |
| 21 | J MARKETING RES, 2005, J MARKETING RES, V, P | 0 |
| 21 | Commun ACM, 2001, COMMUN ACM, V, P | 4 |
| 21 | Mis QUART, 2005, MIS QUART, V, P | 0 |
| 21 | Inform SYST RES, 2005, INFORM SYST RES, V, P | 0 |
| 18 | Manage SCI, 2007, MANAGE SCI, V, P | 0 |
| 17 | J MARKETING, 2007, J MARKETING, V, P | 0 |
| 16 | Cyberpsychol BEHAV, 2003, CYBERPSYCHOL BEHAV, V, P | 2 |
| 15 | J CONSUM RES, 2007, J CONSUM RES, V, P | 0 |
| 15 | J AM SOC INF SCI TEC, 2004, J AM SOC INF SCI TEC, V, P | 1 |



图五：文献共被引及聚类图

**4小结**

随着网络的日渐普及，网络用户行为研究逐渐获得越来越多的关注，研究网络用户行为也具有很多的实际意义，比如通过研究电子商务网站用户行为找到用户浏览或购买模式从而应用于实际提高收益，在实际研究中该类研究占了很大部分比例。除了理论上的模型构建，很多论文讨论到网络用户行为研究的实用意义，包括网站安全建设、用户信任机制建设、网站合理规划、社交网站用户行为分析等等，说明网络用户行为研究的实践指导意义是吸引学者研究的一个重要方面。以上皆为网络用户行为研究目前的主要内容，但是可以预见未来研究内容会随着网络的发展发生更为巨大的变化。

**参考文献**

[1]刘列,邢千里,刘奕群,张敏,马少平,. 社交网络用户标签预测研究[J]. 中文信息学报,2016,(2).

[2]张冰,钱燕云,罗小芳,. 中德青年网络社交行为比较及文化成因[J]. 安徽文学(下半月),2016,(2).

[3]赵丙秀,. 用户行为分析分类模型的研究[J]. 计算机时代,2016,(2).

[4]左军,. 基于大数据的网络用户行为分析[J]. 软件工程师,2014,(10).

[5]戴月卿,钟玲,林柏钢,陈明志,. 基于微博的人物关系网络挖掘系统[J]. 信息网络安全,2013,(2).

[6]许星,席鹏富,秦天,. 社会网络的舆情信息分析与可视化——以新浪微博为例[J]. 计算机光盘软件与应用,2013,(12).

[7]李常先,. 大学校园用户网络行为分析系统研究[J]. 统计与管理,2013,(4).

[8]袁书寒,陈维斌,傅顺开,. 位置服务社交网络用户行为相似性分析[J]. 计算机应用,2012,(2).

[9]潘婵,冯利飞,丁婉莹,贺芳,. 基于标签-关键词的用户行为分析[J]. 情报杂志,2010,(3).

[10]林闯,田立勤,王元卓,. 可信网络中用户行为可信的研究[J]. 计算机研究与发展,2008,(12).

[11]马力,焦李成,董富强. 一种Internet的网络用户行为分析方法的研究[J]. 微电子学与计算机,2005,(7).

[12]张亮,赵娜,. 网络用户浏览行为的分析[J]. 计算机系统应用,2016,(6).

[13]许星,席鹏富,秦天,. 社会网络的舆情信息分析与可视化——以新浪微博为例[J]. 计算机光盘软件与应用,2013,(12).

[14]张仕斌,肖姗,昌燕,闫丽丽,. 基于云模型理论的网络用户行为评判模型研究[J]. 小型微型计算机系统,2013,(10).

[15]许飒爽,曹健,. 面向服务环境的服务个性化推荐算法[J]. 计算机集成制造系统,2011,(11).

[16]阮光册,. 基于页面内容和会话相似的网络用户行为研究[J]. 图书情报工作,2011,(11).

[17]岑荣伟,刘奕群,张敏,茹立云,马少平,. 网络检索用户行为可靠性分析[J]. 软件学报,2010,(5).

[18]朱涵钰,吴联仁,吕廷杰,. 社交网络用户隐私量化研究:建模与实证分析[J]. 清华大学学报(自然科学版),2014,(3).

[19]余肖生,马费成,. 网络用户行为模型的构建方法研究[J]. 情报科学,2011,(4).

[20]岑荣伟,刘奕群,张敏,茹立云,马少平,. 基于日志挖掘的搜索引擎用户行为分析[J]. 中文信息学报,2010,(3).