Inline Text Wrapping Picture

北京邮电大学

硕士研究生学位论文开题报告

学 号: 2013140249

姓 名: 尹彦龙

学 院: 信息与通信工程学院

专业(领域): 电子与通信工程

研究方向:

导师姓名: 张琳

攻 读 学 位: 工程硕士

2014年12月12日

|  |  |  |  |
| --- | --- | --- | --- |
| 论文题目 | 面向用户体验的智能应用使用模式分析与优化研究 | | |
| 选题来源 | 973、863项目 | 论文类型 | 应用研究 |
| 开题日期 | 2014-12-12 | 开题地点 | 主楼1006 |
| **一、立题依据（包括研究目的、意义、国内外研究现状和发展趋势，需结合科学研究发展趋势来论述科学意义；或结合国民经济和社会发展中迫切需要解决的关键科技问题来论述其应用前景。附主要参考文献目录）（不少于800字）**  随着Android,ios,Windows Phone等智能操作系统的崛起，移动互联网时代正处于快速成长期，在北京召开的移动互联发展大会上发布的由人民网研究院主编的《中国移动互联网发展报告（2014）》蓝皮书显示，截至2014年1月，我国移动互联网用户总数达8.38亿户，在移动电话用户中的渗透率达67.8%；手机网民规模达5亿，占总网民数的八成多，手机保持第一大上网终端地位。我国移动互联网发展进入全民时代。通过智能终端设备接入互联网获取信息和服务成为当今信息时代的一种新的主流交互方式。紧随着移动互联网发展的大数据时代也接踵而来,“移动互联网”和“大数据”成为当前互联网领域内最热火的两大话题。因此各大互联网公司提供移动互联网对用户行为分析认识程度直接关系到自身能否在新的浪潮得到新的发展,所以对移动互联网的用户行为相关领域的学术研究有着深远的价值,无论是对移动互联网相关企业的市场发展的推动还是移动互联网的格局都有着重要的意义。  在新的市场环境下,运营商必须在原有优势的基础上,做出顺应移动互联网发展的调整。在移动互联网络中,用户的需求趋于娱乐化、媒体化、实用化。目前用户付费使用最多的移动互联网应用莫过于手机游戏,这足以体现娱乐化应用的发展已趋于成熟。现在正处于微博式媒体时代,用户参与内容创造的个性化服务平台的兴盛,表示了移动互联网在某种程度上代替了传统媒体,并赋予了新的媒体特征。在机遇与挑战的面前,运营商必须第一时间了解潜在市场的用户需求,就能适应市场,为用户提供更完善更个性化的服务,就能在竞争中占有一席之地。如何获取用户的需求信息,最根本的途径是针对用户的上网数据进行分析,即实现用户行为分析。用户行为分析对精准营销的巨大作用是不言而喻的。在移动互联网潜在客户类型识别过程中,移动互联网用户行为分析可以有效、快捷地获取用户潜在需求,能够持续不断地吸引客户。对于成熟稳定的客户群体可以通过用户行为分析不断增强和改善用户服务感受,提升用户体验。  随着移动互联网的爆炸式发展,用户行为的研究越来越受到关注。2012年的移动互联网用户规模仍然在相对稳定中持续增长。用户在网络中的地位不容小觑,以用户为中心的通信模型正在逐步成型。很多国内外研究机构利用分析伩器获取网络日志等相关数据对用户行为进行分析研究,这类研究大部分是针对构造用户行为模型。还有一部分研究主要集中在用户行为习惯及其对整个网络性能的影响分析。文献【1】研究学习用户行为数据分析的系列理论，结合移动互联业务的特点，采集有效数据并形成一套移动数据业务的个性化推荐系统。文献【2】在分析传统用户行为方法模型的基础上,研究并提出了针对移动互联网用户行为的研究方法,并借助于移动终端设备釆集到的实测数据，对数据进行了分析得出用户行为规律进行说明。  参考文献：  [1]常凯. 基于神经网络的数据挖掘分类算法比较和分析研究[D]. 安徽大学 2014  [2]郑桂凤. 移动互联网的用户行为分析系统的设计与实现[D].北京邮电大学,2010.  [3]刘路. 基于云计算的移动互联网大数据用户行为分析引擎设计[J]. 电子制作,2014,04:174-175.  [4]赖志伟,梁勇华. 移动互联网用户行为分析[J]. 科技风,2014,04:249.  [5]冯铭,王保进,蔡建宇. 基于云计算的可重构移动互联网用户行为分析系统的设计[A]. 中国计算机学会(CCF).CCF NCSC 2011——第二届中国计算机学会服务计算学术会议论文集[C].中国计算机学会(CCF):,2011:4.  [6]赵勇. 移动互联网用户行为分析系统技术架构浅析[A]. 中国通信学会无线及移动通信委员会.2012全国无线及移动通信学术大会论文集（下）[C].中国通信学会无线及移动通信委员会:,2012:3.  [7]唐家琳. 移动互联网用户行为比较分析[J]. 西安邮电大学学报,2013,05:90-94+99.  [8]王璐. 移动互联网用户行为分析[D].重庆邮电大学,2012.  [9]Lei X,Wu S,Ge L,et al. Clustering and overlapping modules detection in PPI network based on IBFO[J]. Proteomics, 2013,13(2): 278-290.  [10]Ostermann S, Iosup A, Yigitbasi N, et al. A performance analysis of EC2 cloudcomputing services for scientific computing[M]//Cloud Computing. SpringerBerlin Heidelberg, 2010: 115-131.  [11]Kuo-Wei Su,Chao-Hung Wang. Usability Testing on the Interface of the Location-Based M-tourism Application[A]. IEEE Beijing Section,China、Guangzhou University,China.Proceedings of 2013 IEEE International Conference on Computer Science and Automation Engineering VOL01[C].IEEE Beijing Section,China、Guangzhou University,China:,2013:4.  [12]Xinye Lin,Xiao Xia,Shaohe Lv,Xiaodong Wang School of Computer Science,National University of Defense Technology,Changsha 410072,China. Research on the Predictability of Mobile App Usage[A]. 中国计算机学会多媒体技术专业委员会、中国图象图形学学会多媒体专业委员会、中国计算机学会普适计算专业委员会、ACM SIGCHI中国分会、中国自动化学会计算机图形学和人机交互专委会.第七届和谐人机环境联合学术会议（HHME2011)论文集【oral】[C].中国计算机学会多媒体技术专业委员会、中国图象图形学学会多媒体专业委员会、中国计算机学会普适计算专业委员会、ACM SIGCHI中国分会、中国自动化学会计算机图形学和人机交互专委会:,2011:13.  [13]Jianlin Xu,Yifan Yu,Zhen Chen,Bin Cao,Wenyu Dong,Yu Guo,Junwei Cao. MobSafe:Cloud Computing Based Forensic Analysis for Massive Mobile Applications Using Data Mining[J]. Tsinghua Science and Technology,2013,04:418-427. | | | |

|  |
| --- |
| **二、研究内容和目标（说明课题的具体研究内容，研究目标和效果，以及拟解决的关键科学问题。此部分为重点阐述内容）（不少于2500字）**  移动互联网用户行为特征分析主要研究用户使用网络的行为模式和兴趣习惯。人们日常生活中通过智能终端的各种应用软件获取导航、健康、学习、天气、新闻、社交等服务成为人们生活的一部分进而成为一种生活习惯。智能终端设备的发展也呈现出突飞猛进的态势,智能设备软硬件的发展升级速度巳经远远超过PC。首先数据统计,分析用户确定时间段内的行为,从用户使用时间模式、使用流量及兴趣习惯等角度研究移动互联网在线用户的行为特征,进而为用户提供更友好的使用体验。  本文从移动终端获取实测数据出发,提出一种通过移动终端应用采集用户行为数据、分析数据得到用户行为的方法。利用云计算数据分析平台实现终端应用数据的存储和分析,构建了基于Hadoop的云计算海量终端数据的分析平台,并且通过观察实测数据研究了移动互联网用户行为规律,证实了该方法的可行性。该方法有三部分组成:移动终端数据釆集应用设计、数据分析平台的搭建、用户行为分析。在终端数据采集应用设计部分,本文采用当前移动终端主流操作系统之一的Android系统上进行设计。用户在使用智能应用享受网络服务的时候，数据采集系统会手机用户使用应用所产生的一些信息，然后系统会将这些信息定时的收集起来，然后再将信息上传到远端数据库存储。由于远端数据库空间有限，而数据源源不断的在上传，所以需要将远端数据库里的数据定时删除，这也就需要我们在删除远端数据之前将数据备份到本地。通过上述终端数据釆集到数据,只能得到某一时刻某一网络某一项应用单点的数据是最原始未处理的数据,通过这些原始数据还无法分析出用户的行为特征。通过挖掘数据之间的深层次的联系,从而获得用户使用智能应用的一些情况。在得到移动互联网用户使用智能应用行为数据基础上,进一步采用云计算数据分析平台研究移动互联网用户的行为的一些特征。  本文提出面向用户体验的智能应用使用模式与优化的研究,主要包括以下三大部分:  (1)根据智能终端在移动互联网发展的地位,并且结合智能终端设备在釆集移动互联网数据的强大优势——真实可靠用户实际的效果数据上。采用了智能终端应用的方式进行用户接入服务及用户行为的数据采集,完成了以往单纯依靠流量统计及企业单纯一方的数据采集上缺陷,不能够釆集到完整真实的用户使用数据。这里通过智能终端设备的数据采集就可以解决这些问题,并且终端应用釆用数据上传机制等完成智能终端设备数据釆集的设计。通过在Android平台上开发一款软件，去实时的记录用户使用App的情况。具体信息包括，用户使用的App名称，使用App的时间以及使用这款App消耗的流量。在此基础上，每天产生的数据多达几十万条。基于这些数据信息，我们对用户的一些行为进行分析。  (2)搭建Hadoop平台进行数据分析。由于移动终端的设备每日产生的数据量庞大,而且数据的内容复杂包含了用户行为的数据信息,面对如此庞大的数据量,存储分析是一个很重要的问题,这里将云计算技术引入到我们数据的存储和分析处理中来。构建了基于Hadoop云计算海量数据的分析平台,该平台可以将上传到云端的数据分布式存储,实现了分布式计算,极大的便利了海量数据的存储和处理。在大数据时代，Hadoop几乎快成了大数据的代名词，在Hadoop平台上能够轻松的处理GB级甚至TB级数据。Hadoop原本仅仅是一个基于分布式计算环境的开源软件库，由于其符合大数据计算环境的分布式要求，“时势造英雄”一炮走红，犹如当年的Linux开源软件系统一样，成为了研究和设计大数据解决方案的主流平台。本文的Hadoop平台是基于Ambari搭建起来的。Ambari旨在帮助系统管理员部署和配置Hadoop，升级集群以及监控服务。通过Ambari能够很好的搭建并且管理Hadoop平台的其他组件，比如说数据仓库Hive，高级过程语言Pig等。以及基于Hadoop的机器学习工具Mahout。  (3)基于上述两部分的数据分析和用户行为分析，并根据分析结果对用户的智能应用使用行为提出建议，进而改善用户的使用体验。这一部分是本文的重点内容。首先提取用户行为和App使用模式的参数。具体来讲，用户行为的一些参数包括：使用app的时间，使用的app的名字以及使用的app的地点以及区分用户唯一标识的IMEI。这些参数能够表明用户的一些使用app的行为模式，比如用户可能会经常在特定的时间内使用某一类特定的app。有关研究表明，用户通过移动终端登录互联网的时间集中在晚上20:00-22:00,说明此时段是移动终端用户登录互联网最活跃的时段，而在上下班时间，通过智能手机上网的人数比例超过50%，说明在上下班时间用户更习惯用智能手机上网。但是，这些时间用户都用了什么软件上网是一个需要研究的问题。而这个问题通过我们采集的数据，就能得出一个大题的结论。通过得出的结论我们就能够很据用户的习惯向用户提出相关建议。其次app使用模式的问题，在本文收集的数据中，并没有app类别这个概念，但是通过使用相关的聚类分类算法，可以对我们采集到的数据里的app的类别进行分类。这样一来，就可以对用户在什么样的时间使用什么类别的app进行统计分析。具备这些准备条件之后，就可以建立模型对采集到的数据进行分析。我们要分析用户行为，从我们的目的出发，最强有力的工具就是数据挖掘，即获取用户行为产生的海量数据，并将这些数据转换成有用的信息和知识。数据挖掘里有很多优秀的算法可以为我们分析用户行为模式提供基础，选取哪一种算法进行数据的分类以及其他分析，还需要多做调研。  综上所述，移动互联网用户行为模式是我们要分析的目标,移动终端数据采集是我们数据来源,云计算数据分析平台是我们数据处理分析的方式。这三个方面构成了我们面向用户体验智能应用使用模式与优化分析研究的整体框架。本文提出的面向用户体验智能应用使用模式与优化分析方法最大优势在于能够兼容移动终端发展迅速,移动终端的操作系统更新换代快、移动终端涵盖面很广。本文介绍的基于移动终端进行数据采集,云计算数据分析平台进行数据的存储分析得出移动互联网户行为分析的方法,最大程度上得到了用户使用时真实数据,而且通过基于Android操作系统应用的方式,将我们数据采集模块嵌入到我们的服务中。这样就可以记录用户每次使用服务整个过程流量以及用户行为的数据。通过建立数据指标借助于云计算平台对数据进行过滤存储。在通过云计算的计算模块对数据进行深度挖掘计算得到移动互联网用户智能应用使用模式。 |

|  |
| --- |
| **三、研究方案设计及可行性分析（包括：研究方法，技术路线，理论分析、计算、实验方法和步骤及其可行性等）（不少于800字）**  首先是数据采集软件的开发，数据采集是以监控的形式在后台时刻运行的。程序每隔20s去检查一下网络流量变化情况，如果用户的总的网络流量或者某个智能应用的流量发生了大于1KB的变化，则程序会去存储这次变化的详细信息，存储的信息包括发生流量变化的app的包名，App名，软件的uid。，然后是用户此时的网络类型，包括具体的网络制式。软件还会采集用户所在的蜂窝小区的id，如果用户使用的是WiFi网络，软件还会采集WIFi相关的信息。除此之外，软件还会存储发生网络变化的开始时间，结束时间，本次变化上行和下行产生的流量以及用户当前的经纬度等信息。就这样，每隔20秒去检测一下。信息存储到手机上之后，再设置一个定时器和一个获取存储信息的方法，定时去调用这个方法，取得存储的信息上传到远端服务器数据库。在这个过程中，传数据到远端服务器成功后，终端就会删除已经上传的信息。  数据上传到远端数据库之后，在本地服务器上搭建数据库环境，其中数据表的内容等同于远端数据库中数据表的内容，然后写程序将远端数据库里的数据同步备份到本地服务器的数据库中。  搭建Hadoop平台。平台的搭建是基于Ambari，使用Ambari的好处就是，搭建平台的时候可以选择安装你需要的组件，例如Hive,pig,Hbase等，以及一些集群监控组件nigios,ganglia,这些监控组件能够帮助你更好的管理集群，能够实时的检测集群的健康状况、空间信息、配置信息等。平台搭建完成之后，利用sqoop将本地数据库里的数据导入到Hadoop的文件系统HDFS里，以便于利用Hadoop的mapreduce分析数据。  数据收集和平台的搭建完成之后，接下来就是利用平台分析这些数据，对用户行为作出一些分析。初步使用的数据项有，软件的包名，软件的时候时间以及一些使用的流量信息。通过使用相关算法对用户使用软件的行为进行分析，进而对用户体验提出建议。首先提取用户行为模式和app使用模式的相关参数，再建立相关的分析模型并选取相关的分析算法进行分析。最终基于收集的用户软件使用的数据，通过相关的数据挖掘算法分析得出相关结论，并达到提升用户体验的目的。 |

|  |
| --- |
| **四、本研究课题可能的创新之处（不少于500字）**  （1）本文采用的智能应用使用模式的分析方法是基于用户使用智能应用时产生的一些流量信息、时间信息、地理位置信息、网络信息等来对用户使用应用的行为模式进行分析预测。现有的一些用户行为分析所采用的数据大多是一些日志文件，浏览记录等，并没有如此详细的数据，是纯粹的为了分析用户的上网特征而分析。本文使用的数据源每天都在更新，在终端应用启动后以后台服务的形式在移动终端定时执行常规数据釆集任务并上传测试信息日志。该数据采集模块具有自动从服务端获取最新版本的数据采集脚本,自动采集反馈在真实环境下的服务数据  （2）通过收集用户在使用智能应用时产生的一些用户行为信息，对移动互联网用户使用智能应用的情况进行分析处理，得出移动互联网用户行为的一些特征，根据这些特征反过来去对用户使用智能应用的情况提出一些建议，达到提升用户使用智能应用体验的目的。大多数相关的分析用户行为的文章，出发点都是提出一套系统或者方法，然后基于这些去分析用户的行为，得出一些结论，然后利用这些结论去让运营商或者应用开发人员来作出相应的改变去适应用户，为用户提供更好的服务。本文是基于这些结论去对用户的一些行为特征提出相关的建议，旨在帮助用户有一个更好的智能应用的使用体验。 |
| **五、研究基础与工作条件（1.与本项目相关的研究工作积累基础 2.包括已具备的实验条件，尚缺少的实验条件和拟解决途径）（不少于500字）**  截止目前为止，基于Android操作系统的数据采集的APP已经开发完成，数据采集工作每天都在进行，截止目前采集的数据也有一定的规模，累计1亿多条。另外，进行数据分析的Hadoop平台已经搭建完成。该平台不仅包括Hadoop的核心组件HDFS和MapReduce，而且还安装了Hadoop数据仓库Hive，通过Hive可以对HDFS上的数据利用类Sql语言进行更好的操作和管理，这样一来可以随时进行数据分析和处理。其次数据备份工作还在一直进行，由于app采集的信息上传到了远端服务器上，如果我们想利用Hadoop进行数据分析，必须将数据库数据备份到本地，这项工作时刻都在进行。最后将数据备份到本地后，为在Hadoop平台上进行数据分析，还需要将本地数据库的数据导入到Hadoop平台上的HDFS上。这一部分工作也是每天都在进行着。  基于现在的条件，已经进行过一些基于Hadoop的数据分析工作，能熟练的运行Hadoop对数据进行分析。通过这些数据对用户的的流量使用情况以及用户的智能应用使用情况，进行过简单的分析处理。目前正在研究基于这些数据的分析用户相关行为模式算法问题。目前还缺乏分析用户行为模式的合适的算法，以及分析用户行为模式的系统，如何将采集到的信息进行合理的分析是本文的重点内容。初步打算查看相关的论文资料，并寻找合适的数据挖掘算法，尽快找到能合理的分析用户行为模式的相关算法。 |

**学位论文工作计划**

|  |  |  |
| --- | --- | --- |
| 时间 | 研究内容 | 预期效果 |
| 2014年9月-2015年1月 | 查询文献，分析移动互联网用户相关行为模式，并开发数据收集系统和搭建数据分析平台 | 完成数据收集系统的开发和数据分析平台的搭建 |
| 2015年2月-2015年7月 | 测试数据手机系统和分析平台，开始撰写论文 | 实现目标系统，并完成论文初稿 |
| 2015年9月-2015年12月 | 修改论文细节 | 完成论文终稿 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 评  定  小  组  成  员 | 姓 名 | 职 称 | 单位名称 | 职务 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
| 导师意见： | | | | |
|  | | | | |
| 导师（签名）：  日期： 年 月 日 | | | | |
| 开题报告小组意见： | | | | |
| 组长（签名）：  日期： 年 月 日 | | | | |
| 学院意见（签章）： | | | | |
| 负责人：  日期： 年 月 日 | | | | |