

远程用户体验评估综述:工具、方法和挑战

韩立 刘正捷 张军 陈媛嫒

(大连海事大学信息科学技术学院 大连 116026)

摘要 远程用户体验评估允许评估人员和用户在时间和(或)空间上分离,它具有经济、便捷、对用户打扰少,能在真实环境中对大量多样的用户进行研究等特点,被认为是解决传统用户体验评估问题的一个理想方案,并逐渐成为人机交互领域中一个热门的研究课题。首先介绍了远程用户体验评估的概念和分类;然后描述了远程评估的研究概况;接着提出了远程评估工具的系统架构,并依据系统架构,详细描述了数据采集、分析和呈现的各种方法;最后,总结了远程评估所面临的挑战以及未来的研究方向。

关键词 远程用户体验,可用性,数据采集,自动分析,可视化

中图分类号 TP311 文献标识码 A DOI 10.11896/j.issn.1002-137X.2014.10.043

Remote User Experience Evaluation Review: Tools, Methods and Challenges

HAN Li LIU Zheng-jie ZHANG Jun CHEN Yuan-yuan

(College of Information Science and Technology, Dalian Maritime University, Dalian 116026, China)

Abstract Having many advantages, such as economy, convenient and fast, non-intrusive, allowing to study the plenty of diversity users in field, the remote user experience evaluation is considered as the potential solution for the problems of traditional user experience evaluation, and it becomes a research hotspot in the field of human-computer interaction. This paper reviews the state of remote user experience evaluation, where the architecture is given. According this architecture, the issues are classified into three parts which are data capturing, automatic analysis and visualization to discuss. At the end of this paper, the challenges and future research topics are listed.

Keywords Remote user experience, Usability, Data capturing, Automatic analysis, Visualization

1 引言

对移动设备、软件、网站等交互式信息产品来说,用户体验(User Experience)是决定其成败的一个关键因素,这也是近年来人机交互领域的一个研究热点。随着移动互联网的发展,信息通信产品的使用日益深入生活和工作的各个方面,人们在家庭、办公室、公共场所、旅途中以及其它许多生活和在工作场所使用这些产品,这使得研究人员越来越难以获取关于产品实际使用的体验数据,基于人工和面对面的用户研究方法和手段也无法覆盖足够的用户样本,无法采集足够的用户体验数据。远程用户体验评估的出现和发展,在一定程度上缓解了这些问题。

1.1 用户体验

ISO 9241-110^[1]给出了用户体验的标准化定义:一个人对一个产品、系统或者服务的使用和(或)期望所产生的观念和反应。用户体验涵盖了人与系统发生交互时的所有情形^[2,3],在用户体验概念没有出现之前,可用性(Usability)一直被作为衡量人和系统之间交互质量的标准。用户体验与可用性是一种相互缠绕的关系,在概念和实际操作上很难分清

它们的界限,文献[4]把可用性归入到用户体验中,认为可用性是用户体验非常重要的一个部分。从实际情况来看,用户体验评估也继承和发展了可用性测试的技术方法,可用性也通常被认为是用户体验评估的关键指标^[5]。

传统的用户体验研究,分为现场和实验室两种方式。在实验室中,实验环境虽然是可控的,但是却无法通过模拟来还原现场环境^[6],而真实的使用环境对用户体验的影响很大,尤其是对移动应用程序来说,使用场景和移动性是具有决定性的影响因素^[7-10],用户体验也非常依赖于用户的内在状态,这种内在状态会受到真实环境的影响^[3]。

与实验室评估相比,现场评估更关注系统在真实环境中的使用状况,而不是简单的程序可用性^[11]。现场评估能为研究提供更真实可靠的情境数据^[4,12],尤其在研究用户对移动设备的总体接受度时更有效^[13]。但现场评估也存在很多缺点,例如:需要更多的经费、花费更多的时间、较难保证用户的参与度、数据收集困难、实验可控性不强等^[4,14-16]。尤其是开发者、评估者和用户处在不同的组织和国家时,现场测试几乎是不可能进行的^[17]。

信息通信产品日益无处不在地使用,使得产品设计和开

到稿日期:2013-12-12 返修日期:2014-02-14 本文受中央高校基本科研业务费专项资金(3132013041)资助。

韩立(1980—),男,博士生,CCF会员,主要研究方向为用户体验、人机交互等,E-mail:hherosoul@163.com;刘正捷(1958—),男,教授,博士生导师,CCF高级会员,主要研究方向为用户体验、人机交互等;张军(1977—),男,博士,副教授,CCF会员,主要研究方向为用户体验、人机交互等;陈媛嫒(1981—),女,博士,CCF会员,主要研究方向为用户体验、人机交互等。

发人员很难准确获得真实可靠的用户体验数据,难以对用户行为进行准确的评估。实验室和现场评估方法虽然应用广泛,但是解决这一问题的能力比较有限,远程用户体验评估是一个比较有潜力的解决办法。

1.2 远程用户体验评估

1.2.1 概念

远程评估的概念大概出现在 20 世纪 90 年代中期,是指用户和研究人员在时间和(或)空间上是分离的^[18],这是与传统方法最大的区别。远程用户体验评估的优点有省时、省力、对用户的干扰少^[19],尤其是移动互联网技术的发展和普及,使远程评估能以更经济、便捷的方式收集和分析用户的体验数据^[20]。远程评估的另一个关键优点是可以针对大量多样的用户进行研究^[22],如果与一个项目有关联的人分布在广泛的区域,就很难利用传统方法对这些用户进行研究,远程方法可以有效地解决这种问题。

目前远程评估方法已经在信息共享产品、协同工作工具、网站等的开发及评估上得到较为广泛的应用^[23-26]。

1.2.2 分类

按照时间特征,远程用户体验评估可以分为两种形式:同步方式和异步方式。

同步方式:用户体验评估在实时状态下进行,评估人员和用户在空间上是分离的。大多数远程同步评估都模拟传统的基于实验室的出声思考(Think Aloud)方法,可以利用远程共享桌面、视频和音频工具来实现^[19],常用的工具有 WebEx、NetMeeting 等。

同步方式的优点是可以在真实的环境中研究多样的用户,能发现和传统可用性测试相同数量的可用性问题^[22]。但是远程同步方式需要评估人员实时控制和监视评估过程,这个方法所需要的资源几乎和传统的实验室评估一样多^[17]。

异步方式:用户和评估人员在时间和空间上都是分离的,体验数据是在特定的时间以异步方式获取的^[22]。异步方式解决了同步方式需要过多资源的缺点,实验以异步的方式进行,还会减轻评估人员时间上的压力^[27]。

异步方式最主要的缺点是评估人员无法对实验现场进行控制,这导致无法及时解决现场可能出现的问题,由于无法对用户的行为进行实时观察,后期的数据解读会存在一定的困难^[19]。

本文第 2 节回顾远程用户体验评估的研究概况;第 3 节从数据处理的角度描述评估工具的系统架构,并依据此架构,重点介绍远程用户体验评估的数据采集、分析和呈现方法;第 4 节讨论远程评估所面对的困难和挑战;第 5 节探讨未来的研究方向;最后总结全文。

2 研究概况

远程用户体验评估的很多工作是对远程方式和传统方式的有效性进行对比研究,例如,文献^[28]认为远程评估与传统的实验室评估相比可以得到相似的结果。Brush 等人^[29]从用户和评估者两个角度出发来研究远程研究方法和传统方法的区别,研究者使用远程同步方式与传统方式进行对比,结果表明远程同步与实验室在发现可用性问题的数量上没有显著的差异,在两种方法中用户觉得他们对设计的贡献也是相似的;

评估人员则认为准备远程研究很耗费精力,但招募参与者相对比较容易。

一些早期研究所使用的工具只能收集数据,不能对数据进行分析,例如 TEA^[33]、UsaProxy^[34]等。单从数据收集方式来看,自动的日志记录是最常用的一种方法,评估工具通常使用日志记录来自动记录应用程序和相关设备的使用情况^[35,36]。

另有一些工具除了数据收集功能之外,还能进行一定程度的自动分析,例如 WebRemUsine^[37]、AWUSA^[38]、WELFIT^[40]、WebQuilt^[41]、W3Touch^[42]、MATE^[43]、RMTS^[44]等工具。W3Touch 是最新的远程评估工具,它记录触摸智能手机上产生的应用交互行为,这个研究主要考虑两个独立评估维度,即缩放事件的数量和触摸交互中丢失的链接数。MATE 能够帮助用户在特定的应用程序和用户界面上发现用户体验问题,评估人员事先定义一个剧情,这个剧情包含一系列需要用户执行的任务,并将最优任务序列与用户实际操作序列进行对比来发现用户体验问题。

自动分析在远程评估工具中的应用越来越普遍,为了提高分析效率,减轻评估人员的工作负荷,研究人员提出了一些数据分析方法(例如任务模型分析法^[45]、任务序列比较法^[46-50]等)和数据可视化方法(例如时间轴^[25]、交互有向图^[41]、热点图^[51]等)。

从目前的研究来看,远程评估工具最常见的系统结构是中介代理和客户端/服务器这两种形式。

1)中介代理结构:这种结构主要是基于一个中介代理服务,这个代理是服务器和客户端之间的一个中间件^[24,25](如图 1 所示),在评估期间,用户的所有行为都被保存成日志记录,通过实时或者非实时的方式传到服务器上。这种结构可以简化评估人员的操作^[34],避免可能出现的安全冲突^[24]。

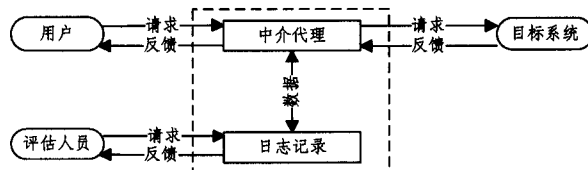


图 1 中介代理结构

2)客户端/服务器结构:评估人员可以把任何连接了互联网的桌面电脑作为服务器,连接了互联网的桌面电脑或者移动设备作为客户端(如图 2 所示),客户端需要安装测试程序、记录程序、分析和问卷系统^[40,55]。工具以客户端的用户行为为数据源,可以在服务器的后台以可视化的方式将数据呈现给评估人员。这种结构的好处是整个实验设置起来比较简单,在客户端安装好需要的工具后,就可以开始工作,不需要进行额外的配置。

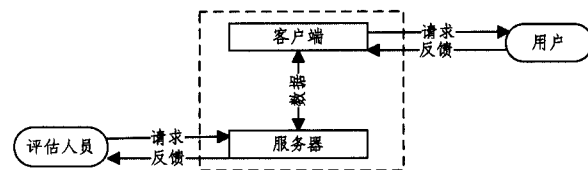


图 2 客户端/服务器结构

远程评估工具从本质上来说仍然是一种软件工具,除了其自身的特点外,它还具有一般软件工具的所有要求和特征,

所以远程评估工具的目标与文献[57]提出的自动软件工具在很多方面是相似的。以下是远程用户体验评估工具应该尽量满足的要求:

评估有效性:工具至少能与传统的评估方法发现同样多的用户体验问题。

评估效率:工具应该比人工的方式快。

评估正确性:产生的结果应该尽量正确,由评估人员不当的输入和指导而引发的错误应该尽可能少。

适应性:工具应该能用在—个范围内的用户体验评估上,并允许环境的可定制化。

一致性:这个工具应该提供一种贯穿整个开发过程的评估标准。

合理的分工:将评估人员从不必要的工作中解放出来,将精力集中在数据解读上。

减少资源消耗:应该能减少时间、资金、设备和人力的需求。

工具本身具有好的用户体验:工具本身的设计和开发也应该使用有效的规范来提升用户体验。

3 数据采集、分析和呈现

远程用户体验评估的核心是数据处理,从数据处理的角度来看,绝大多数工具都包含数据采集、数据分析和数据呈现3个模块。图3描述了远程用户体验评估工具的一般性系统架构。

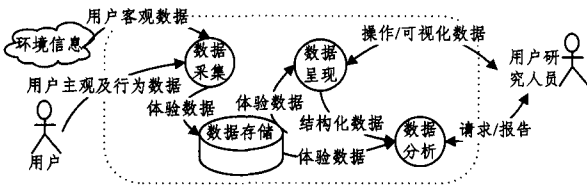


图3 远程评估工具系统架构

依据这一系统架构,本节将主要从数据采集、数据分析和数据呈现这3个方面来介绍数据处理的方法。

3.1 数据采集

通常来说,数据采集的步骤分为两部分:首先,根据研究目的决定需要收集什么数据;其次,选择合适的采集方法。

3.1.1 确定要采集的数据

远程用户体验评估所要收集的数据有以下4类:交互行为[58]、系统状态[58]、用户情境[61]和用户状态[51,60],对于不同的评估目标和标准,获取数据的类型、数量和侧重会有不同。

首先确定目标任务,通常来说,评估分为[58]界面设计评估、交互流程评估、系统功能评估和总体接受度评估4类。不同的目标需要获取不同类型和不同粒度的数据。

其次确定评估标准,对由目标任务所确定的数据类型进行细分。在远程方式中常用的评估标准有:Nielsen[68]提出的可用性5个维度(可学习性、效率、可记忆性、出错和满意度)和审美、享乐等情感体验维度。在实际的研究中,很难对这些维度直接进行评估,一般都会根据实际的目标对这些维度进行分解,使用间接的方式去评估,例如:文献[43]将评估标准分为3类:有效性、效率和频率。文献[69]利用复杂度来对应用程序界面的可用性进行评估。文献[70]认为用户操作出现

错误的地方代表了存在用户体验问题,它利用以下3个标准来评估系统中存在的体验问题:导航错误、显示错误和控制错误。W3Touch[42]主要考虑了两个独立的评估标准,即缩放事件的数量和交互过程中丢失的链接数。

3.1.2 采集数据

传统的实验室和现场研究已经对如何收集数据进行了大量卓有成效的研究,并提出很多被证明行之有效的数据收集方法。远程评估中常用的方法有:出声思考[7,72]、自动记录[76,78]、观察法[82,84]、用户自报告[22,85,88]、测试后访谈[85]、问卷[18,55,77]、视频[18,83]和生理测量[90,91]等等。

3.2 数据分析

自动分析在远程评估中的使用已经越来越普遍,通常可以分为定性和定量两种方式。

3.2.1 定性数据分析

传统评估中使用的定性数据分析方法有很多,可用性工程、人机交互、数据挖掘、统计学等相关文献和书籍已经对这些方法进行过大量的研究,例如音频分析[56]、图像分析[62]、文本分析[64]、即时数据分析[67]、专家评估[75]等等。从目前的研究情况来看,受网络链接、网络带宽、设备性能和评估人员技术能力的限制,远程评估中最常用的是由传统方法发展而来的任务模型分析法和任务序列比较法。

任务模型分析法:任务模型分析已经被用来对智能手机[48]、交互系统原型[63]、用户界面[45]和网站[80]等进行评估,这种方法的分析流程如图4所示。

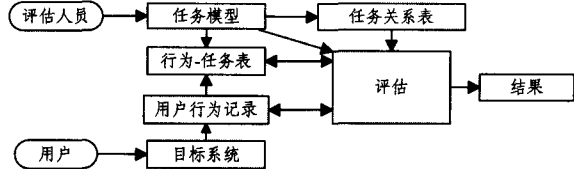


图4 任务模型分析法

1) 建立任务模型:任务模型对于理解用户当前的意图是非常有用的。有很多工具可以用来建立任务模型,例如 ConcurTaskTrees[81],它利用层次结构来表示任务,高层任务是抽象的,底层的任务更细化和具体。

2) 将用户与系统的交互行为记录在日志文件中。

3) 根据任务模型定义任务之间的关系,任务关系表中存有每个任务的名称,以及在每个任务执行之前必须完成的任务名称,任务之间的关系是动态的,随着评估过程的进行,表中的关系是不断更新的。

4) 建立行为-任务表,有很多工具可以用来建立行为和任务之间的关联关系,例如 USINE[45],它可以很方便地在用户的具体行为和任务之间建立如表1所列的关联关系。

表1 行为-任务表

行为	任务
点击{/邮件}	发送邮件
点击{/评论}	写评论
拖动{/文件}	复制文件

5) 评估:首先从用户行为记录中提取一个行为,如果这个行为在行为-任务表中存在,就把相关的任务提取出来,否则就算是发生了一个错误。如果当前任务有一个前提任务,就要查看这个前提任务是否已经完成,若没有完成就认为存在

一个错误。接着重新开始,在用户行为记录中提取下一个用户行为。如此循环执行,直到评估结束。

任务模型分析法能够较为准确和详细地对用户的每一个行为进行评估,但任务模型的建立却比较困难,有时可能要对应用程序的所有交互流程进行建模,需要花费评估人员很多时间和精力。

任务序列比较法:任务序列比较是对任务模型分析的一种简化,不需要建立任务模型,它利用文本分析技术^[86,87]将日志记录中的各种任务序列进行分析、组织和提取,再利用比较法^[51]来寻找用户体验问题。但是这种方法与任务模型分析法一样,无法解读用户的认知任务^[80],一个主要的解决办法就是利用用户自报告来进行补充。任务序列比较法的一般流程如下:

1)采集任务:评估人员利用文本分析技术从原始的日志记录中提取出用户执行过的任务。这些任务以行为或者事件序列的形式存在日志记录中^[58]。

2)分裂日志:把包含所有用户交互行为的日志分裂成多个日志,每一个日志文件只含有一个用户的交互任务记录。

3)确定序列:根据评估目标确定哪些任务序列是需要进行比较的,并将这些序列提取出来。可以利用最大共同子序列算法^[89]去测量各个序列与期望的任务执行序列之间的相关性,当相关性超过一定的阈值时,就把这个任务序列提取出来。

4)序列比较:就是将实际执行序列与标准执行序列进行比较,如果存在不一致的地方,就说明存在用户体验问题。通常有3种比较方法:最一般任务序列比较法^[50]、期望序列比较法^[36,51]和启发式序列比较法^[58]。

3.2.2 定量数据分析

对于定量数据来说,自动分析方法相对比较简单。定量分析主要是以任务为基础,评估人员根据目标任务设计用户需要执行的任务序列。以任务执行过程中所产生的例如任务完成率、任务完成时间、数据流量、点击次数、出错数等定量数据为分析对象,利用传统的统计学方法进行统计分析。

很多的统计分析方法和工具可以很方便地嵌入到远程评估工具中^[43,69]。也有一些商业化的工具能够对用户行为数据提供比较准确的定量统计分析,例如谷歌分析^[53]和百度移动统计^[54]等。

3.3 数据呈现

数据以可视化的方式呈现给评估人员,有助于他们对体验数据有更直观的认识和深入的理解。以下是目前比较常用的可视化方法。

视频图形展示:视频展示主要是通过实时播放或回放来进行^[73,74],图形展示通常是将用户的操作流程以截图的方式保持起来,并根据特定的任务,将截图以任务流的形式组织起来进行呈现^[55]。

任务序列展示:是目前使用比较多的一种可视化技术,常用的展示方式有以下两种:

1)时间轴^[24,25]:每个用户执行的每个任务都有一个时间轴(如图5所示)。第一个时间轴表示由评估人员或者设计人员给出的期望执行序列,其它是用户的实际执行序列,通过时间轴评估人员可以很容易地发现用户的异常行为。

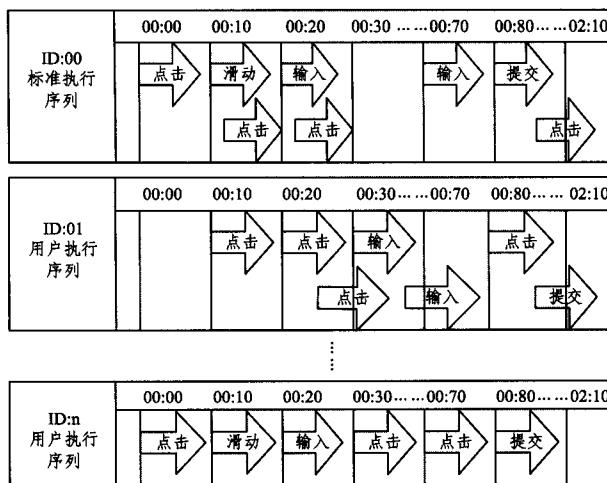


图5 时间轴

2)交互有向图^[36,41]:有向图的节点代表活动,边代表事件,通常的做法是先建立一个基准有向图(如图6所示),基准图表示设计者或者研究人员所期望的交互行为。再将用户的实际执行序列有向图(如图7所示)与基准有向图进行对比来发现用户体验问题。

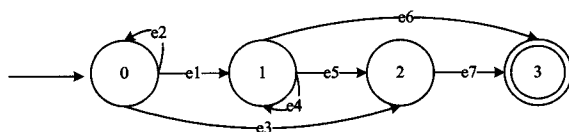


图6 基准有向图

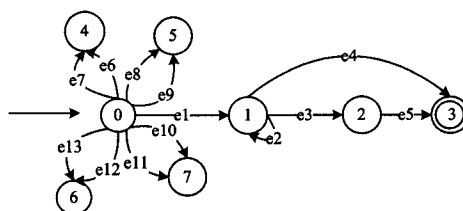


图7 实际执行有向图

统计展示:统计学上有很多定量数据的可视化方法,例如直方图、雷达图、柱状图、热点图^[51]等等,现在的工具基本都能给出定量数据统计图。

4 困难与挑战

与传统的评估方法相比,远程用户体验评估有很多的优势,但同时也存在很多的问题,本节将从以下3个方面介绍远程用户体验评估所面临的困难和挑战。

4.1 基础设施受限

移动环境中的远程用户体验评估所面对的问题最多,其中最主要的问题是网络链接和设备性能有限,例如对手机进行的远程用户体验评估就存在诸多限制^[55]:屏幕小、运算能力差、电池续航能力差、数据流量受限、网络链接不稳定等等。尤其是移动设备的计算能力有限和网络连接问题,对于数据种类多、数据量大的研究来说,利用移动设备进行远程评估基本是行不通的。

4.2 实验设置和维护困难

远程评估是在用户的生活环境里开展,中断、移动、噪音、多重任务等^[65]都会影响用户的操作,这些问题在实验室环境下是不会出现的,而且由于评估人员不在现场,发生软硬件故

障时重新恢复实验是非常困难的^[22],因此需要做好计划无法进行的准备。

管理实验过程中的人际关系也是远程评估的一个主要挑战,由于招募和测试都是以远程方式进行,评估人员无法和用户面对面交流,在这种情况下,评估人员很难与用户建立起信任关系^[19],而信任是影响评估结果的一个关键因素,它会影响用户的参与度、积极性、反馈的准确性和绩效表现等。

4.3 评估方法上的挑战

远程评估决定了评估人员和用户是分开的,评估人员对用户行为的观察是很有限的,有限的观察使评估人员很难捕捉用户的非语言线索,这会影响评估人员理解用户的瞬时意图,也会影响对用户情感的解读^[19]。远程方式对定性数据的收集能力有限,例如自动日志记录方法无法获取用户的心理状态^[3],缺乏准确的定性数据会给后期数据解读带来困难^[22]。虽然远程评估工具对用户的干扰比较少,但是由于缺乏现场观察,评估人员通常很难在用户的自报告和用户体验问题之间建立有效的关联^[19]。

虽然远程用户体验评估存在很多的困难和挑战,但是它的价值不容忽视,随着科学技术的发展,尤其是移动互联网技术的快速发展和普及,以上诸多问题有望得到进一步的解决。

5 未来的研究工作

由于具有经济、便捷、可以进行大量多样的用户研究等特点,远程用户体验评估越来越受到学术界和工业界的重视。受到移动互联网、可穿戴技术、普适计算,以及大数据等技术方法的影响,在未来一段时间内,以下一些方面将可能成为研究的热点。

5.1 智能分析

目前所使用的自动分析方法,对定量数据的分析已经相对比较完善,但是对定性数据的分析能力还很有限^[77],主要的原因是定性数据很难用量化指标进行度量,定性数据的分析在很大程度上依赖于评估人员的经验,只能依赖于模型进行演绎和归纳。

近期的一些研究^[52,79,92]尝试利用专家系统来建立用户评估人员的知识和经验模型,以更智能的方式处理用户体验数据,但是目前这种方法还很不完善,要想利用专家系统对用户体验数据进行准确的分析,还需要较长的时间去进行迭代和改进。虽然构造专家系统的难度高、代价大,但智能分析将是今后用户体验数据处理的重要发展方向。

5.2 眼动跟踪

越来越多的证据显示,信息可视化对人的任务处理和认知效率有很大的影响,为了研究这种影响,越来越多的研究利用眼动跟踪技术分析人的视觉交互过程。目前比较先进的技术产品是眼动跟踪眼镜^[66],这是一种可穿戴的眼动跟踪仪器,能够在真实和虚拟的环境中收集眼动数据^[52,71]。随着眼动跟踪技术的发展,眼动跟踪仪器的应用会越来越广泛,如何结合眼动数据对用户体验进行更准确和更智能的远程评估将是今后一个主要的研究问题。

5.3 数据挖掘

目前的远程评估通常都利用比较简单的方式进行自动分析,例如任务序列比较、数据统计等,这些方法对于理解用户体验数据来说还远远不够,尤其是随着体验数据日益剧增,这

种简单的分析方法很难深刻地认识数据间的联系。数据挖掘技术是从大数据中获取知识,将数据挖掘技术同用户体验相结合也是人机交互领域新的发展方向^[39,59],利用数据挖掘技术将会极大地提高远程用户体验评估的有效性和效率。

5.4 可穿戴技术

随着移动技术的不断发展,可穿戴的仪器设备也不断涌现,例如智能手机、智能眼镜、智能手表等等,这些设备的用户数量在快速增加,功能也越来越完善,可以利用这些可穿戴设备获取一些传统方法无法获取的体验数据,例如可以利用谷歌眼镜来收集用户在日常生活中的行为信息^[117],这种方式很自然,最大限度地避免了对用户的侵扰。目前的数据收集技术很难获取用户的心理状态和认知数据,随着可穿戴技术的不断发展,可以期望利用设备上的传感器获取用户的一些心理生理指标^[91],通过这些指标来解读用户的心理状态和认知过程。

5.5 情境感知

情境因素是用户体验数据非常重要的组成部分,对于设计者来说,对情境信息的深入理解可以帮助他们明确应用程序要在什么样的情境中如何去使用,对评估人员来说,对情境的理解能够更好地明确用户是如何与复杂系统进行交互的。但是从目前的远程用户体验评估来看,情境因素主要是用来帮助评估人员对体验数据进行解读,并没有利用情境信息去解决远程评估中存在的一些问题,例如评估人员很难有效地指导用户、对现场无法控制、任务设置固定、无法将数据和用户意图有效地联系起来等等。

情境感知技术可以在一定程度上解决这些问题^[32],利用情境感知技术模拟评估人员的情境感知机理,在评估的过程中,根据所感知到的情境因素对用户进行实时的指导,并与用户实时交流,这是一个非常具有前景的研究方向。

5.6 虚拟现实

远程用户体验评估最突出的问题是现场不可控,研究人员对用户行为的实时观察很有限。利用虚拟现实技术模拟评估实验,可以在一定程度上解决这种问题,这种虚拟技术能够让用户沉浸在使用情境中,可以获取丰富可靠的体验数据^[31],同时也可以让评估人员在这种虚拟的环境中与用户进行互动,并观察和指导用户执行任务,就像在传统的实验室环境中一样。

5.7 评估方法的改进和创新

传统用户体验评估方法的有效性已经在大量的研究中得到了验证,但是这些方法并不完全适用于远程用户体验评估,使用传统的研究方法去分析新形式、新技术下所产生的体验数据必然会存在一些问题,例如,出声思考是目前在远程评估中使用比较普遍的一种方法,但是有研究认为这种方法获取的定性数据比较有限,而且由于缺乏现场观察,评估人员难以建立定性数据与用户行为之间的关联^[7,19]。为了更准确有效地采集和分析数据,需要一些新方法和新框架的支持,例如目前流行的“去现场”方法^[21]和基于实践的研究方法^[30]指出了研究需要更深入到用户的真实环境中,并需要将用户行为和情境信息进行紧密的结合。对传统方法的改进和创新将是今后远程用户体验研究的一个重要工作。

结束语 随着越来越多的研究人员认识到远程用户体验评估的优势,远程评估也迎来了新的发展机遇,逐渐地成为人

机交互领域一个热门的研究课题。虽然远程评估方法具有很多的优点,但是它也不可避免地存在一些问题,例如缺乏对实验环境的有效控制、视觉观察受限、不同测试部分的保密性差、用户绩效容易受到影响、用户认知和心理状态难以处理、定性数据获取和分析能力有限等等,这些都是需要进一步解决的问题。

新技术的不断涌现为解决上述问题带来了可能,情境感知技术能对实验环境进行有效的控制;随着移动网络技术的发展,通过移动设备进行实时互动也逐渐成为可能;可穿戴技术的发展也为实时获取用户心理状态和认知过程提供了一种可行的办法;用户体验评估与专家系统和知识库的结合,也会进一步提升定性数据的处理能力。在不远的将来,远程用户体验评估会发展得更智能、更无处不在,能将研究人员从不必要的工作中解放出来,让他们把精力集中在模型构建、标准制定和数据解读这些更有价值的工作上。

参 考 文 献

- [1] ISO DIS. Ergonomics of human system interaction-Part 210: Human-centred design for interactive systems[S]. Switzerland: International Standardization Organization (ISO), 2010: 9241-210
- [2] Law E, Roto V, et al. Understanding, scoping and defining user experience; a survey approach[C]// Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2009. Boston, USA, 2009: 719-728
- [3] Hassenzahl M, Tractinsky N. User experience-a research agenda [J]. Behaviour and Information Technology, 2006, 25: 91-97
- [4] Law E, Roto V, et al. User experience evaluation methods: current state and development needs[C]// Proceedings of Nordic Conference on Human-Computer Interaction. Reykjavik, Iceland, 2010: 521-530
- [5] Agarwal A, Meyer A. Beyond usability: evaluating emotional response as an integral part of the user experience[C]// Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Boston, USA, 2009: 2919-2930
- [6] Johnson P. Usability and mobility: interactions on the move [C]// Proceedings of Human-Computer Interaction with Mobile Devices. Glasgow, Scotland, 1998(2): 1-4
- [7] Kallio T, Kaikkonen A, et al. Usability testing of mobile applications: a comparison between laboratory and field testing [J]. Usability studies, 2005, 1: 4-16
- [8] Elzakker V, Delikostidis I, et al. Field-based usability evaluation methodology for mobile geo-applications [J]. The Cartographic Journal, 2008, 45: 139-149
- [9] Christensen P, Romero M, et al. Children, mobility, and space: using GPS and mobile phone technologies in ethnographic research [J]. Mixed Methods Research, 2011, 5: 227-246
- [10] Coleman E. Ethnographic approaches to digital media [J]. Annual Review of Anthropology, 2010, 39: 487-505
- [11] Duh H, Tan G. Usability evaluation for mobile device: a comparison of laboratory and field tests[C]// Proceedings of International Conference on Human-Computer Interaction with Mobile Devices and Services. Helsinki, Finland, 2006: 181-186
- [12] Monahan K, Lahteenmaki M, et al. An investigation into the use of field methods in the design and evaluation of interactive systems[C]// Proceedings of British HCI Group Annual Conference. Swinton, UK, 2008: 99-108
- [13] Xu Sun, Andrew M. A comparison of field-based and lab-based experiments to evaluate user experience of personalised mobile devices [J]. Advances in Human-Computer Interaction, 2013, 2013: 64-72
- [14] Raptis D, Tselios N, et al. Context-based design of mobile applications for museums: a survey of existing practices[C]// Proceedings of International Conference on Human Computer Interaction with Mobile Devices and Services. Salzburg, Austria, 2005: 153-160
- [15] Kjeldskov J, Skov M, et al. Is it worth the hassle? Exploring the added value of evaluating the usability of context-aware mobile systems in the field[C]// Proceedings of International Mobile Conference. Glasgow, UK, 2004: 61-73
- [16] Kjeldskov J, Stage J. New techniques for usability evaluation of mobile systems [J]. Human Computer Studies, 2004, 60: 599-620
- [17] Bruun A, Gull P, et al. Let your users do the testing: a comparison of three remote asynchronous usability testing methods, 2009[C]// Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Boston, USA, 2009: 1619-1628
- [18] Hartson H, Castillo J, et al. Remote evaluation: the network as an extension of the usability laboratory[C]// Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Vancouver, Canada, 1996: 228-235
- [19] Susan D, David S. Remote possibilities? International usability testing at a distance [J]. Interactions, 2004, 11: 10-17
- [20] Kaasila J, Ferreira D, et al. Testdroid: automated remote UI testing on Android[C]// Proceedings of International Conference on Mobile and Ubiquitous Multimedia. Ulm, Germany, 2012: 1-4
- [21] Crabtree A, Chamberlain A, et al. Introduction to the Special Issue of "The Turn to The Wild" [J]. ACM Transactions on Computer-Human Interaction, 2013, 3(3): 1-4
- [22] Andreasen M, Nielsen H, et al. What happened to remote usability testing? An empirical study of three methods[C]// Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. San Jose, USA, 2007: 1405-1414
- [23] Thompson K, Rozanski E, et al. Here, there, anywhere: remote usability testing that works[C]// Proceedings of Conference on Information Technology Education. Salt Lake, USA, 2004: 132-137
- [24] Carta T, Paternò F, et al. Web usability probe: a tool for supporting remote usability evaluation of web sites[C]// Proceedings of IFIP TC International Conference Human-Computer Interaction. Lisbon, Portugal, 2011: 349-357
- [25] Carta T, Paternò F, et al. Support for remote usability evaluation of web mobile applications[C]// Proceeding of ACM International Conference on Design of Communications. Pisa, Italy, 2011: 129-136
- [26] DeShazo J, Harris L, et al. Designing and remotely testing mobile diabetes video games [J]. Telemedicine and Telecare, 2010, 16: 378-382
- [27] Scholtz J. Adaptation of traditional usability testing methods for remote testing[C]// Proceedings of Annual Hawaii International Conference on System Sciences. Maui, HI, 2001: 8-15

- [28] Kasik D, George H. Towards automatic generation of novice user test scripts[C]//Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Vancouver, Canada, 1996; 224-251
- [29] Brush A, Ames M, et al. A comparison of synchronous remote and local usability studies for an expert interface[C]//Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Vienna, Austria, 2004; 1179-1182
- [30] Pierce J, Strengers Y. Introduction to the Special Issue on Practice-Oriented Approaches to Sustainable HCI [J]. ACM Transactions on Computer-Human Interaction, 2013, 4(20): 1-8
- [31] Carlos A, Francisco A. Hands-on experiences of undergraduate students in Automatics and Robotics using a virtual and remote laboratory [J]. Computers & Education, 2011, 57: 2451-2461
- [32] 贡宏印. 情境感知的远程用户研究工具研究与设计 [D]. 大连: 大连海事大学信息科学技术学院, 2013
- [33] Obendorf H, Weinreich H, et al. Automatic support for web user studies with SCONE and TEA, 2004[C]//Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Vienna, Austria, 2004; 1135-1138
- [34] Atterer R. Logging usage of AJAX applications with the "UsaProxy" HTTP proxy[C]//Proceedings of International World Wide Web Conference. Edinburgh, Scotland, 2006; 1-4
- [35] Ferreira D, Dey A, et al. Understanding human-smartphone concerns; a study of battery life[C]//Proceedings of International Conference on Pervasive Computing. San Francisco, USA, 2011; 19-33
- [36] Ma X, Yan B, et al. A toolkit for usability testing of mobile applications[C]//Proceedings of International Conference on Mobile Computing, Applications and Services computer Sciences. Los Angeles, USA, 2011; 226-245
- [37] Paganelli L, Paterno F. Intelligent analysis of user interactions with web applications[C]//Proceedings of International Conference on Intelligent User Interfaces. San Francisco, USA, 2002; 111-118
- [38] Tiedtke T, Martin C, et al. AWUSA - A tool for automated website analysis, 2002[C]//Proceedings of Design, Verification and Specification of Interactive Systems conference. Rostock, Germany, 2002; 251- 266
- [39] Drachen A, Thureau C, et al. Game Analytics [M]. London: Springer, 2013; 205-253
- [40] Santana V, Baranauskas M. A Prospect of Websites Evaluation Tools Based on Event Logs[C]//Proceedings of TC 13 Human-Computer Interaction Symposium. Milano, Italy, 2008; 99-104
- [41] Hong J, Heer J, et al. WebQuilt: A proxy-based approach to remote web usability testing [J]. ACM Transactions on Information Systems, 2001, 19: 263-285
- [42] Nebeling M, Speicher M, et al. W3Touch: metrics-based Web content adaptation for touch[C]//Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Paris, France, 2013; 2311-2320
- [43] Porat T, Schlar A. MATE: a mobile analysis tool for usability experts[C]//Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Paris, France, 2013; 265-270
- [44] Huang Jin-feng. Remote mobile test system: a mobile phone cloud for application testing[C]//Proceedings of IEEE International Conference Cloud Computing Technology and Science. Taipei, Taiwan, 2012; 1-4
- [45] Lecerof A, Paternò F. Automatic Support for Usability Evaluation [J]. IEEE Transactions on Software Engineering, 1998, 24: 863-888
- [46] Maly I, Slavik P. Towards Visual Analysis of Usability Test Logs Using Task Models[C]//Proceedings of International Workshop on Task Models and Diagrams for User Interface Design. Hasselt, Belgium, 2006; 24-38
- [47] Propp S, Forbrig P. ViSE: A Virtual Smart Environment for Usability Evaluation[C]//Proceedings of Human-Centred Software Engineering conference. Reykjavik, Iceland, 2010; 38-45
- [48] Levy M, Dayan A, et al. Task modeling infrastructure for analyzing smart phone usage[C]//Proceedings of Mobile Business and Global Mobility Roundtable Conference. Athens, Greece, 2010; 264-271
- [49] Hay B, Wets G, et al. Mining navigation patterns using a sequence alignment method [J]. Knowledge and Information Systems, 2004, 6: 150-163
- [50] Charfi S, Ezzedine H. Towards an automatic analysis of interaction data for HCI evaluation application to a transport network supervision system[C]//Proceedings of International Conference on Human-Computer Interaction. Orlando, USA, 2011, 6761; 175-184
- [51] Au R W F, Baker S, et al. Automated Usability Testing Framework[C]//Proceedings of conference on Australasian user interface. Wollongong, NSW, Victoria, Australia, 2008; 55-64
- [52] Chyna P, Sobecki J. Remote Usability Evaluation Using Eye Tracking Enhanced with Intelligent Data Analysis[C]//Proceedings of Design, User Experience, and Usability Conference. Las Vegas, USA, 2013; 212-221
- [53] Google. Google Analytics[OL]. <http://www.google.com/analytics/index.html>
- [54] Baidu. Baidu mobile statistics[OL]. <http://mtj.baidu.com/web/welcome>
- [55] Liang H, Song H. A remote usability testing platform for mobile phones[C]//Proceeding of IEEE International Conference Computer Science and Automation Engineering. Shanghai, China, 2011; 312-316
- [56] Pfeiffer S, Fischer S. Automatic audio content analysis[C]//Proceedings of the fourth ACM international conference on Multimedia. Rostock, Germany, 1996; 21-30
- [57] Patton R. Software Testing(2nd edition)[M]. USA: Sams Publishing, 2006; 387-419
- [58] Vargas A, Weffers H. A method for remote and semi-automatic usability Evaluation of web-based applications through user behavior analysis[C]//Proceedings of International Conference on Methods and Techniques in Behavioral Research. Utrecht, Netherlands, 2010(19): 1-4
- [59] Horng D. Data Mining Meets HCI: Making Sense of Large Graphs [D]. Pittsburgh: Machine Learning Department Carnegie Mellon University, 2012
- [60] Hammontree M, Weiler P, et al. Remote usability testing [J]. Interactions, 1994, 1: 21-25
- [61] Arhippainen L, Ahti M. Empirical evaluation of user experience in two adaptive mobile application prototype[C]//Proceedings

- of International Conference on Mobile and Ubiquitous Multimedia, Lulea, Sweden, 2003; 27-34
- [62] Pantic M, Rothkrantz L. Automatic analysis of facial expressions; The state of the art [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22; 1424-1445
 - [63] Johnson P, Wilson S. ADEPT; advanced design environment for prototyping with task models[C]// *Proceedings of the INTERACT and CHI Conference on Human Factors in Computing Systems*. Amsterdam, Netherlands, 1993; 56
 - [64] Salton G, Buckley C. Term-weighting approaches in automatic text retrieval [J]. *Information Processing & Management*, 1988, 24; 513-523
 - [65] Tamminen S, Oulasvirta A, et al. Understanding mobile contexts [J]. *Personal and Ubiquitous Computing*, 2004, 8; 135-143
 - [66] Thad S. Wearable computing; through the looking glass[C]// *Proceedings of the 2013 International Symposium on Wearable Computers*. Zurich Switzerland, 2013; 125-126
 - [67] Kjeldskov J, Skov M. Instant data analysis; conducting usability evaluations in a day[C]// *Proceedings of Nordic conference on Human-computer interaction*. Tampere, Finland, 2004; 233-240
 - [68] Nielsen J. *Usability engineering* [M]. Boston: Academic Press, 1993; 23-43
 - [69] Souza C, Bentolila D. Automatic evaluation of API usability using complexity metrics and visualizations[C]// *Proceedings of International Software Engineering Conference*. Vancouver, Canada, 2009; 299-302
 - [70] Liu D, Bias R. Crowdsourcing for usability testing [C]// *Proceedings of the American Society for Information Science and Technology*. Baltimore, USA, 2012; 1-10
 - [71] Mele M, Federici S. Gaze and eye-tracking solutions for psychological research [J]. *Cognitive Processing*, 2012, 13; 261-265
 - [72] Virzi R, Sorce J. A comparison of three usability evaluation methods; heuristic, think-aloud, and performance testing[C]// *Proceedings of Human Factors and Ergonomics Society Annual Meeting*. Washington, USA, 1993; 309-313
 - [73] Schnepf J, Shiver B. Improving deaf accessibility in remote usability testing[C]// *Proceedings of international ACM SIGACCESS conference on Computers and accessibility*. Dundee, Scotland, 2011; 255-256
 - [74] Alcantud F, Coret J. Usability remote evaluation; METBA system[C]// *Proceedings of International Conference on Interactive Collaborative Learning*. Villach, Austria, 2012; 1-8
 - [75] Nielsen J, Molich R. Heuristic evaluation of user interfaces [C]// *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Seattle, USA, 1990; 249-256
 - [76] Sohn T, Li K. A diary study of mobile information needs[C]// *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Florence, Italy, 2008; 433-442
 - [77] Steves M, et al. A comparison of usage evaluation and inspection methods for assessing groupware usability[C]// *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Seattle, Washington, USA, 2001; 125-134
 - [78] Satoh I. A testing framework for mobile computing software [J]. *IEEE Transactions on Software Engineering*, 2003, 29; 1112-1121
 - [79] Engelbrecht K, Quade M, et al. Analysis of a new simulation approach to dialog system evaluation [J]. *Speech Communication*, 2009, 51; 1234-1252
 - [80] Paganelli L, Paternò F. Tools for remote usability evaluation of Web applications through browser logs and task models [J]. *Behavior Research Methods, Instruments, & Computers*, 2003, 35; 369-378
 - [81] Paterno F, Mancini C. ConcurTaskTrees; a diagrammatic notation for specifying task models[C]// *Proceedings of International Conference on Human-Computer Interaction*. Sydney, Australia, 1997; 362-369
 - [82] Chalil K, Greenstein JS. Synchronous remote usability testing - a new approach facilitated by virtual worlds[C]// *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Vancouver, Canada, 2011; 2225-2234
 - [83] Vreede G, Fruhling A, et al. A repeatable collaboration process for usability testing[C]// *Proceedings of Hawaii International Conference on System Sciences*. Big Island, USA, 2005; 6
 - [84] Schusteritsch R, Wei Towards C, et al. Towards the perfect infrastructure for usability testing on mobile devices[C]// *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. San Jose, USA, 2007; 1839-1844
 - [85] Petrie H, Hamilton F, et al. Remote usability evaluation with disabled people[C]// *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Montréal, Canada, 2006; 1133-1141
 - [86] Fournay A, Mann R. Characterizing the usability of interactive applications through query log analysis[C]// *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Vancouver, Canada, 2011; 1817-1826
 - [87] Song C, Ranganath S. Automatic sign language analysis; a survey and the future beyond lexical meaning [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27; 873-891
 - [88] Castillo J, Hartson H, et al. Remote usability evaluation; Can users report their own critical incidents? [C]// *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Los Angeles, USA, 1998; 253-254
 - [89] Paterson M, Dancik V. Longest Common Subsequences [C]// *Proceedings of Mathematical Foundations of Computer Science*. Kosice, Slovakia, 1994; 127-142
 - [90] Mandryk R, Inkpen K. Using psychophysiological techniques to measure user experience with entertainment technologies [J]. *Behaviour & Information Technology*, 2006, 25; 141-158
 - [91] Dirican A, Göktürk M. Psychophysiological measures of human cognitive states applied in human computer interaction[C]// *Proceedings of World Conference on Information Technology*. Istanbul, Turkey, 2011; 1361-1367
 - [92] Feuerstack S, Blumendorf M, et al. Automated usability evaluation during model-based interactive system development[C]// *Proceedings of Second Conference on Human-Centered Software Engineering and International Workshop on Task Models and Diagrams*. Pisa, Italy, 2008; 134-141