

大数据处理技术在交互设计中的应用研究

王晓慧, 覃京燕

(北京科技大学, 北京 100083)

摘要: **目的** 将分布式与并行计算、机器学习、假设检验、模式识别等大数据处理技术应用于交互设计的经典问题,包括用户调研、产品迭代和产品测试。**方法** 首先调研了交叉学科和交互设计的研究现状,然后提出了将大数据处理技术应用于交互设计的研究框架,进而分别研究了基于大数据处理技术的用户调研方法、基于A/B测试的版本更新方法以及基于模式识别的用户体验优化方法。**结果** 研究结果表明了将大数据处理技术应用于交互设计的可行性,大数据处理技术为设计学中的传统用户调研和测试方法提供了坚实的理论基础和数据支持。**结论** 将大数据处理技术应用于交互设计,大大降低了人工成本,并进一步为设计学与计算机科学的交叉进行了初步探索,提供了有效实用的理论与方法。

关键词: 交互设计; 大数据处理技术; 模式识别; 用户调研

中图分类号: TB472 **文献标识码:** A **文章编号:** 1001-3563(2015)22-0009-04

Application of Big Data Processing Technologies in Interactive Design

WANG Xiao-hui, QIN Jing-yan

(University of Science & Technology Beijing, Beijing 100083, China)

ABSTRACT: It applies big data processing technologies, such as distributed and parallel computing, machine learning, hypothesis testing and pattern recognition, to the classic problems in interactive design. First, it studies the related works about interdisciplinary research and interaction design. Then it proposes a framework on application of big data processing technologies in interactive design. The details include big data based user study, A/B test based product version update, and pattern recognition based user experience optimization. The results show the feasibility of application of big data processing technologies in interactive design. The big data processing technologies provide a solid theoretical basis and data support for research on user studies and user experiences. Application of big data processing technologies in interactive design greatly reduces labor costs. It provides effective and practical methods for interdisciplinary research of computer science and design science.

KEY WORDS: interactive design; big data processing technology; pattern recognition; user study

随着计算机技术的发展,互联网每天都在产生大量的数据,数据已经从量变走向了质变,成为了“大数据”^[1]。它已经进入生物信息学、天气预报等科学领域。在大数据时代,用户的喜好、行为等数据,如上网

的习惯、浏览的痕迹、参与的话题以及评论的内容都已经公开在互联网上,并且具有规模化。这些信息将改变传统的交互设计方法,例如不再需要通过用户访谈、问卷调查才能获得用户习惯,而且传统方法所调

收稿日期: 2015-06-22

基金项目: 中央高校基本科研业务费专项资金(FRF-TP-14-004C1)

作者简介: 王晓慧(1987—),女,山东人,博士,北京科技大学讲师,主要研究方向为计算机科学与设计学交叉、大数据处理和交互设计等。

通讯作者: 覃京燕(1976—),女,四川人,博士,北京科技大学教授,主要研究方向为交互设计、信息设计以及大数据的信息可视化等。

查的个别用户得到的结果远没有大数据统计分析的结果全面、有说服力。

1 研究现状

1.1 交叉学科的研究现状

随着许多科学前沿问题在交叉学科的联合攻关下取得可喜成果,交叉学科的发展得到了充分的认可^[2]。“交叉学科”是指超过一个学科范围的研究活动,随着计算机技术和互联网的发展,计算机技术与越来越多的自然科学学科产生交叉,如生物学^[3]、物理学^[4]等。计算机技术也被应用到艺术设计中,如对色彩的研究采用定量的方法^[5],将计算机技术中的模式识别应用到自动分类垃圾桶的概念设计中^[6]等。

近年来,艺术设计与计算机技术进行交叉出现了计算机艺术设计交叉学科,简称为CG,它是艺术语言和设计理念通过计算机这一媒介展现出来的新型应用模式。国外知名大学的实验室,如MIT媒体实验室,同时招收计算机科学与技术专业、艺术设计专业等多学科的学生共同参与科研工作。国内清华大学美术学院也设置有交叉学科,培养计算机科学与技术、电子科学与技术与设计学集合等交叉学科的学生。

这里关注设计学与计算机技术的交叉,它不同于计算机艺术设计交叉学科,不是设计学中的多媒体设计与计算机的图像处理和图形学技术交叉,而是将计算机科学中的大数据处理技术应用于设计学中的交互设计,为设计学中的传统用户调研和测试方法提供了坚实的理论基础和数据支持,并且大大降低了人工成本。

1.2 交互设计与大数据处理技术

交互设计是设计软件、移动设备、人造环境、服务、可佩戴装置等人造系统的行为的设计领域。在设计之前,首先要了解用户心理,通常需要进行设计调查,获得用户需求^[7]。传统的调查方法有观察法、访谈法和问卷法。观察法是在自然条件下,对心理现象的外部活动进行有系统、有计划观察,从中发现心理现象产生和发展的规律。访谈法是通过访谈者与受访者之间的交谈,了解受访者的动机、态度、个性和价值观念等的一种方法。问卷法是事先拟定出所要了解的问题列成问卷让消费者回答,通过对答案的分析和统计研究,得出相应结论的方法。这些方法都是基于观察再人为地总结用户需求。

随着大数据时代的到来,交互设计也迎来了变革与发展^[8-9]。用户的喜好、行为等数据,如上网的习惯、浏览的痕迹、参与的话题以及评论的内容都已经公开在互联网上,并且已具有规模。这些信息不再需要通过用户访谈、问卷调查才能获得,这一变化将改变传统的用户调研方法,而且传统方法所调查的个别用户得到的结果远没有大数据统计分析的结果全面、有说服力。在大数据时代下,如何使用互联网中的海量数据和大数据处理技术来进行用户调研,分析用户的喜好以及当下设计的趋势等;如何用机器学习、模式识别等大数据处理技术辅助产品设计;如何将设计学,特别是交互设计与用户体验的应用研究以及大数据处理技术结合起来已经成为迫切需要解决的问题。

2 研究框架

系统流程见图1,将大数据处理技术应用于交互设计的用户调研和测试两个阶段,具体包括:大数据处理技术用于用户调研;使用A/B测试确定版本更新的方案;基于模式识别优化用户体验,并以网站设计为例,进行实践。

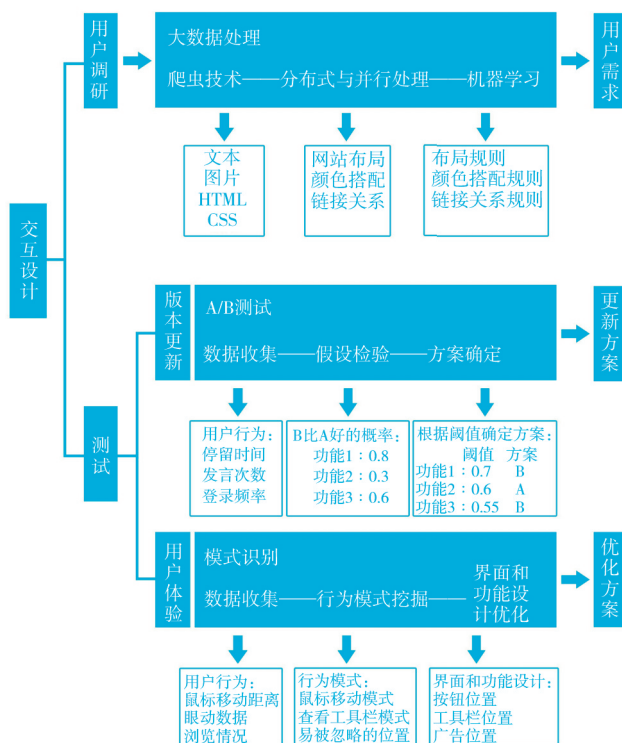


图1 系统流程

Fig.1 System flow

1) 基于大数据处理技术的用户调研。首先使用网络爬虫技术自动获取互联网上的有效数据;然后利

用分布式与并行处理技术对收集的海量数据进行处理;最后采用机器学习、数据挖掘的方法进行自动学习并得到用户喜好。

2) 基于A/B测试的版本更新。在更新软件产品时,将旧版本A和新版本B以随机的方式同时推送给所有用户,收集访问量、转化率等指标;然后采用假设检验的方法,计算新版本B中某个功能更新比旧版本A好的概率;最后根据设计师设置的阈值确定方案,即使用新版本B中的功能1方案,保留旧版本A中的功能2方案等。

3) 基于模式识别的用户体验优化。首先使用计算机技术,在软件程序的后台收集用户行为数据;然后基于大数据处理技术和信息可视化技术,对收集的原始数据进行分析,使用模式识别技术,挖掘用户行为模式;最后根据总结的用户习惯,优化界面和功能设计,使信息的呈现符合用户的心智模型。

3 基于大数据处理技术的用户调研

首先使用网络爬虫技术自动获取互联网上的有效数据;然后利用分布式与并行处理技术对收集的海量数据进行处理;最后采用机器学习、数据挖掘的方法自动学习并得到用户喜好。

以网站设计为例,首先使用网络爬虫技术自动获取不同访问流量的网站,如访问流量大的大型门户网站,类似搜狐、新浪等,以及访问流量小的专业性较强的网站,获取层次化的网站结构和内容,如文本、图片等。网络爬虫技术是一种按照一定的规则,自动抓取万维网信息的程序或者脚本。

然后利用分布式与并行处理技术对收集的海量数据进行处理,整理出网站设计中不同关注点的数据,如网站布局、颜色搭配、链接关系等。分布式处理和并行处理都是为了提高计算机处理速度采用的两种不同的体系架构。并行处理是利用多个功能部件或多个处理机同时工作来提高系统性能或可靠性;分布式处理则是将不同地点的,或具有不同功能的,或拥有不同数据的多台计算机通过通信网络连接起来,在控制系统的统一管理下,协调完成大规模信息处理任务。

最后采用机器学习、数据挖掘的方法自动学习并得到用户喜好,得到网站布局、颜色搭配、链接关系等。大数据的主要研究方法是计算机科学中的数据挖掘,其基本目标有两个,即描述与预测。通过描述以刻画海量数据中潜在的模式,并根据数据中的潜在

模式进行预测,从而发现数据中有价值的模型和规律。基于大数据处理技术,整理成人类所能解读的信息形式。通过大数据和信息可视化的方法对原始数据信息进行分析,使信息的呈现符合用户的心智模型。

4 基于A/B测试的版本更新

A/B测试是一种新兴的网页优化方法,可以用于增加转化率、注册率等网页指标。这里将A/B测试的思想推广到一般软件的版本更新问题,设软件旧版本是A,新版本是B。新版本在某些功能上作出了修改,如网站设计中,网页可能在标题字体、背景颜色、措辞等方面与原有页面有所不同,分别设为功能1、功能2和功能3。具体做法如下。

首先,将旧版本A和新版本B以随机的方式同时推送给所有用户,收集使用两个版本的用户行为,如停留时间、发言次数、登陆频率等,进而计算访问量、转化率等指标,即可清晰地了解到两种设计的优劣。网站转化是进行了相应修改的访问量与总访问量的比值,是衡量网站内容对访问者的吸引程度以及网站的宣传效果的重要指标。

然后,采用假设检验的方法,计算新版本B中某个更新比旧版本A好的概率。假设检验是数理统计学中根据一定假设条件由样本推断总体的一种方法。具体作法是:根据问题的需要对所研究的总体进行某种假设,记作 H_0 ;选取合适的统计量,这个统计量的选取要使得在假设 H_0 成立时,其分布为已知;由实测的样本,计算出统计量的值,并根据预先给定的显著性水平进行检验,作出拒绝或接受假设 H_0 的判断。计算新版本B中每个更新功能比旧版本A好的概率。

最后,根据设计师设定的阈值,确定更新方案。如果假设检验结果大于所设阈值,则接收新版本的更新,如果小于所设阈值,则保留旧版本的设计。

5 基于模式识别的用户体验优化

首先在软件程序的后台收集用户行为数据;然后基于大数据处理技术和信息可视化技术,对收集的原始数据进行分析,使用模式识别技术,挖掘用户行为模式;最后根据总结的用户习惯,优化界面和功能设计,使信息的呈现符合用户的心智模型。

以网站设计为例,首先在软件程序的后台收集用户行为数据,如鼠标移动距离、用户浏览情况。此外,

还可依靠眼动仪快速捕捉变化的眼动数据,包括聚焦点、注视时间、运动路线等。

然后,基于大数据处理技术和信息可视化技术,对收集的原始数据进行分析,使用模式识别技术,挖掘用户行为模式。模式识别是通过计算机用数学技术方法来研究模式的自动处理和判读,用户的行为、使用方式统称为“模式”。挖掘的用户行为模式包括鼠标移动模式、查看工具栏模式、易被忽略的位置、浏览顺序和运动趋势等。

最后,根据挖掘的用户模式,优化界面和功能设计,例如根据鼠标移动模式设置按钮的位置,根据用户查看工具栏模式设置工具栏的位置,根据易被忽略的位置设置广告的位置等。使软件信息的呈现和功能的设计更符合用户的心智模型。

6 结语

这里首先调研了交叉学科和交互设计的研究现状。然后基于交互设计的研究现状和特点,提出了将大数据处理技术应用于交互设计的研究框架,将分布式与并行计算、机器学习、假设检验、模式识别等应用于交互设计的传统经典问题,包括用户调研、产品迭代和产品测试。进而分别研究了基于大数据处理技术的用户调研方法、基于A/B测试的版本更新方法和基于模式识别的用户体验优化方法。研究结果表明了将大数据处理技术应用于交互设计的可行性,大数据处理技术为设计学中的传统用户调研和测试方法提供了坚实的理论基础和数据支持,大大降低了人工成本,并进一步为设计学与计算机科学的交叉作出了初步探索,提供了有效实用的理论与方法。

参考文献:

- [1] 库克耶·肯尼思.大数据时代:生活、工作与思维的大变革[M].杭州:浙江人民出版社,2013.
CUKIER K.Big Data: A Revolution That Will Transform How We Live, Work and Think[M].Hangzhou: Zhejiang People's Publishing House, 2013.
- [2] 王续琨.交叉学科、交叉科学及其在科学体系中的地位[J].自然辩证法研究,2000,16(1): 43—47.
WANG Xu-kun.Interdisciplinarity, Interdisciplinary Science and Their Position in the System of Sciences[J].Studies in Dialectics of Nature, 2000, 16(1): 43—47.
- [3] 卢欣华,孙吉贵,韩霄松.人工生命:计算机与生物学交叉的前沿学科[J].计算机科学, 2008,35(5):9—13.
LU Xin-hua, SUN Ji-gui, HAN Xiao-song.Artificial Life: Front Porch Subject Crossed Computer and Biology[J].Computer Science, 2008,35(5):9—13.
- [4] 倪军,刘华.计算机物理前沿及其与计算机技术的交叉[J].物理,2002,31(7): 461—465.
NI Jun, LIU Hua.Forefront of Computational Physics and Its Intercross with Computational Technology[J].Physics, 2002, 31(7):461—465.
- [5] 李玉凤,黄义萍.色彩规划的定量化和理性化方法研究[J].包装工程,2004,25(6): 92—94.
LI Yu-feng, HUANG Yi-ping.The Study of the Quantitative and Rational Method for Color Layout[J].Packaging Engineering, 2004,25(6): 92—94.
- [6] 胡中艳,曹阳,孙建华.模式识别技术在自动分类垃圾桶概念设计中的应用[J].包装工程,2008,29(12):214—216.
HU Zhong-yan, CAO Yang, SUN Jian-hua.Application of Pattern Recognition Technique in Concept Design of Automatic Classification Garbage Can[J].Packaging Engineering, 2008,29(12): 214—216.
- [7] 诺曼.设计心理学[M].北京:中信出版社,2010年.
NORMAN D.The Design of Everyday Things[M].Beijing: China Citic Press, 2010.
- [8] 覃京燕.大数据时代的大交互设计[J].包装工程,2015,36(8):1—5.
QIN Jing-yan.Grand Interaction Design in Big Data Information Era[J].Packaging Engineering, 2015,36(8):1—5.
- [9] 陈志刚,鲁晓波.大数据背景下信息与交互设计的变革和发展[J].包装工程,2015,36(8):6—9.
CHEN Zhi-gang, LU Xiao-bo.Reformation and Development of Information and Interaction Design Based on the Big Data [J].Packaging Engineering, 2015,36(8):6—9.
- [10] 郑柳和,刘鹏.艺术博物馆生态系统:一种新的替换模式[J].文化艺术研究,2013,6(2):152—164.
ZHENG Liu-he, LIU Peng.Eco-system of Art Museum: A New Replacement Model[J].Studies on Culture & Art, 2013, 6(2):152—164.

(上接第4页)

渴望的是什么[M].北京:中信出版社,2010.

GILMORE H J, PINE J.Authenticity: What Consumers Really Want[M].Beijing: China Citic Press, 2010.

[9] 高嘉蔚.服务设计视角下的虚拟展示设计研究[J].包装工程,2014,35(14):89—93.

GAO Jia-wei.The Virtual Display Design Based on the Ser-