

互联网数据中的文化差异与文化变迁*

李永帅¹ 黄桢炜² 谭旭运³ 刘力¹

(¹北京师范大学心理学院, 应用实验心理北京市重点实验室, 北京 100875)

(²北京行政学院, 北京 100000) (³中国社会科学院社会学研究所, 北京 100732)

摘要 互联网时代, 人们在网络上留下了各种反映其心理过程与文化特征的信息。这些庞大多样的互联网数据为文化心理学研究提供了新视角。首先, 当前文化心理学存在文化差异和文化变迁两种研究取向, 互联网数据相较传统来源数据在这两种研究上均存在若干优势; 其次, 针对4种类型的互联网数据, 文化心理学家利用文本分析、多媒体分析、社会网络分析和互联网使用行为分析的方法进行研究; 再次, 基于互联网数据及其分析方法, 文化差异与变迁研究取得了丰硕的成果; 最后, 当前基于互联网数据的文化心理学研究存在效度、技术与理论局限, 未来研究需通过合理抽样、检验新指标有效性、准因果分析、充分运用新技术、数据驱动等方法来提升方法效度、结果深度及理论多样性。

关键词 互联网数据; 大数据; 文化心理学; 文化变迁; 文化差异

分类号 B849:C91

近年来, 互联网越来越成为人类生活重要的组成部分。截止 2015 年底, 全球已有超过 30 亿互联网用户, 中国网民数量也达到了 6.88 亿(中国互联网络信息中心, 2016)。互联网数据一方面反映着不同文化下人们的社会生活状况的差异, 另一方面也记录和塑造着社会生活形态的变迁。越来越多文化心理学研究者开始使用互联网数据进行研究。本研究对基于互联网的文化心理学研究方法进行总结, 并从文化差异与文化变迁两个角度对应用互联网数据的文化心理学研究进行整理, 以展示互联网数据在文化心理学研究中的应用概貌, 揭示互联网数据在研究人类文化与心理、行为关系上的独特优势, 为当下及未来中国文化心理学研究提供参考。

1 文化心理学研究的差异取向与变迁取向

文化是一套共享知识的网络, 这些共享知识

在一个相互关联的个体集合中被生产、散布和再生产(赵志裕, 康莹仪, 2011), 包括无形的价值观、信仰和社会关系, 以及有形的文化产品、民俗、语言、规范和制度等两种形式(Kroeber & Kluckhohn, 1952)。文化心理学探讨文化与心理的相互建构关系, 横向文化差异比较与纵向文化变迁分析是两种主要的研究取向(Shweder, 1995)。

这两种研究取向在对文化与心理关系的理解、研究方法和研究现状这三方面均有不同。文化差异研究本质上是将文化看作相对静态的前因变量, 将不同地域、文化间的人类心理行为差异作为不同文化影响的结果。其研究方法主要是问卷测量法与实验启动法。Hofstede (1983)提出的文化价值观维度(权力距离、不确定性规避、个人-集体主义、男子气-女子气、长期-短期取向)是研究文化差异的经典指标。后来, Heine (2010)将文化带来的心理行为差异作了较为系统的归类, 分为自我概念、人格、人际关系、认知与感知、语言、情绪、道德等方面, 形成了较为完整的文化差异研究图景。

与文化差异研究不同, 文化变迁研究则本质上将文化看作相对动态的变量, 心理行为与文化相互影响、共同变化。其研究方法主要包括: 对

收稿日期: 2016-09-12

* 国家自然科学基金项目(31571145)、北京市哲学社会科学规划项目(13ZHB027)资助。

通讯作者: 刘力, E-mail: l.liu@bnu.edu.cn

一个文化群体中的几代人进行横断研究,以及对文化意义的外部载体(如文化产品、民俗、语言的使用)进行纵向历史分析。文化变迁研究主要从内生(新工具产生,如技术发明、意识形态创新)与外生(跨文化接触)两类文化变迁起因入手(赵志裕,康莹仪,2011)。一方面,有学者致力于刻画与解释漫长历史中的文化变迁历程(Inglehart, 1997; Lesthaeghe, 1983)。另一方面,也有研究者关注伴随信息技术发展和全球化进程深入而不断变化着的当代文化(Deuze, 2006; Kraidy, 2002)。

传统的研究方法一定程度局限了文化差异与文化变迁研究的深入。文化差异研究常用的问卷法中被试的自我报告是否代表了真实的文化表征受到群体参照效应(Heine, Lehman, Peng, & Greenholtz, 2002)、被试作答风格(Schimmack, Oishi, & Diener, 2005)、测量对等性(Chen, 2007)等因素的影响;而实验法虽然规避了以上自我报告法的问题,但实验室环境中被试表现出的行为可能缺乏生态效度(Lee, Hallahan, & Herzog, 1996)。文化变迁研究常用的横断研究法难以排除年龄效应对于不同代人文化差异的干扰;而对一定时期内文化产品的纵向分析虽然可以避免横断法的问题,但文化产品完整性与代表性的局限仍制约着文化变迁研究。

互联网作为一种新的文化载体,持续记录着人们自然的、非反应性的行为与思想。通过对网络上已有的非反应性数据进行分析,继而探究人们心理与行为规律的“非反应性(nonreactive)研究”,为文化差异和文化变迁研究带来了新契机(王娱琦,余震坤,罗宇,陈杰明,蔡华俭,2015)。一方面,互联网数据具有数据量大(如 Twitter 每天产生超过5亿条推文)、共享程度高(如陌生人可方便地查看他人 Facebook 主页)、生态效度高(如利用用户在使用 Google 等搜索引擎的实际行为了解其关注的话题)的特点。这些特点扩大了文化产品的种类,增加了文化研究对象的可及性,并提高了文化研究的信效度。另一方面,互联网数据巨大的时间跨度(如 Google Books 包含图书的出版时间跨度超过400年)拓宽了文化变迁纵向对比研究的时间区间,其巨大的空间跨度(如 Twitter 覆盖全球超过130个国家)扩展了文化差异横向对比研究的空间范围。

2 基于互联网数据的文化心理学研究方法

互联网非结构化数据按照类型不同可以分为文本数据、多媒体数据、社会网络关系数据和互联网使用行为数据(张引,陈敏,廖小飞,2013)。不同类型的数据需要应用不同的分析方法。本部分将展示当前文化心理学研究者针对不同类型互联网数据进行分析的方法概况,并举例说明这些方法的应用过程。

2.1 文本分析

网络文本从广义上讲是网络上的文本信息,是一种文化产品(丰怡,蔡华俭,施媛媛,2013)。网络文本数据具有来源平台广、涵义丰富和较易处理等特点。网络文本分析是指利用自然语言处理技术(如信息提取、主题模型、文本摘要、分类、聚类、答疑和意见挖掘等)和信息检索等方法,在网络文本中发现隐含模式(Quek & Mitchell, 1997; 袁军鹏,朱东华,李毅,李连宏,黄进,2006)。目前研究者多使用人工方式进行信息提取(Church, 2013)、主题分类(李彪,2013)、意见挖掘(Goode, McCullough, & O'Hare, 2011)等。最近,基于词语分析的语言探索与字词计算软件(Linguistic Inquiry and Word Count, LIWC, <http://liwc.wpengine.com>)和谷歌全球书籍词语统计工具(Google Ngram Viewer, <https://books.google.com/ngrams/>)是两种文化心理研究者广泛使用的网络文本分析工具。

LIWC 是一款基于心理学词库的文本分析软件。其词库所采用的词汇来源于各种词典、问卷、词语分类包括不同词性,涉及情绪、社会及环境等不同方面(Pennebaker, Booth, & Francis, 2007)。研究发现 LIWC 具有良好的结构效度,能够较好地测量情绪表达(Kahn, Tobin, Massey, & Anderson, 2007)。中科院参照 LIWC2007 与 LIWC 中文词库开发了文心中文心理分析系统(<http://ccpl.psych.ac.cn/textmind/>),提供简体中文分词及语言心理分析服务。使用 LIWC 省去了人工进行文本分析的繁琐步骤,将文本输入软件即可得到不同种类词语的词频统计结果,并且可以用词频来做更深入的关联分析。如有研究者首先将不同文化背景 Facebook 用户的“关于我(About Me)”文本输入 LIWC,接着通过 LIWC 内置词库和计数程序得到心理特质、家庭等词语的频率,继而对来自不同

文化用户的词频进行差异比较,最后通过相关法分析词频与自我构念的关系(DeAndrea, Shaw, & Levine, 2010)。

Michel 等(2011)提出了文化组学(Culturomics)的概念,即用自然语言处理方法在多语种的谷歌图书数据库(收录了占全球所有出版图书 4%的数字图书)中研究文化变迁。Google Ngram Viewer 是基于谷歌图书的词语统计工具,它以图示形式显示查询单词/词组从 1600 年至今每一年的图书中出现的词频(该年包含查询词的图书数量除以该年图书总数)。除了基本查询方法,Google Ngram Viewer 还提供通配符搜索、拐点搜索、不区分大小写搜索、词性标签、以及 ngram 运算等几种高级搜索方法。另外,研究者可以通过网页提供的入口下载原始数据(包括查询单词/词组每年出现次数、所在不同书籍的册数),或通过网络爬虫抓取图形中的词频数据。在获取词频数据后,研究者还可进行更为深入的数据分析,如计算词频与年份的相关以发现词语相关的文化概念随时间变化的历程(P. Kesebir & S. Kesebir, 2012),还可以将词频与其他变量一起建立回归模型(Konrath & Anderson, 2011)等。

2.2 多媒体分析

多媒体数据包括社交网络、资源资讯、实用工具等平台中的图片、音频和视频。多媒体数据具有非结构化、语义丰富的特点,因此对它的意义进行分析更具难度。计算机辅助的多媒体分析常常应用多媒体摘要、多媒体注解、多媒体索引和检索、多媒体事件监测等技术进行自动化分析(张引等, 2013)。目前文化心理研究者仍多使用人工编码统计或诠释学等传统方法进行多媒体分析,计算机辅助多媒体分析则更多由计算机学者实施。

在人工分析方面,研究者通常首先选择平台及所需分析的多媒体类型,如 Flickr 上的图片, Twitter 上的视频。接着根据研究问题对筛选后的视频进行编码、标记、统计(高崇, 杨伯淑, 2016)。接下来,在文化变迁研究方面,有些研究者结合访谈结果对多媒体编码统计结果进行诠释,如某文化中热门视频男主角多具有温和(而非阳刚)特质,这体现了该文化中对同性恋恐惧和排斥的价值观有所减弱,因此男性不再需要通过表现得阳刚来避免同性恋污名(Morris & Anderson, 2015)。在文化差异方面,有些研究者会将多媒体分析统

计结果进行跨文化比较与解释,如权力距离更高的国家在网页界面更突出重要人物(Marcus & Gould, 2000)。

在计算机辅助多媒体分析方面,计算机技术通常应用于多媒体分析的某些步骤。如在最初的多媒体标注环节,有研究者将物体识别技术用于识别不同文化图片中与某些特定概念(如婚礼)相关的元素,进而揭示概念表征的文化差异(Yanai, Yaegashi, & Qiu, 2009)。又如在多媒体分析环节,有研究者使用图片可视化分析软件 ImageJ 和 ImagePlot 分析 Instagram 上不同文化的图片颜色分布(Hochman & Schwartz, 2012),进而揭示颜色偏好的文化差异。

2.3 社会网络分析

社会网络是指由某些特定社会群体间的社会关系构成的、相对稳定的关系网络 (Wellman & Berkowitz, 1988)。社会网络信息包括各种社交网络平台中的人物节点及其关系的信息。互联网上社会网络分析具体包括对人物节点的分析、社群挖掘和社会信息的网络化分析(陈浩, 乐国安, 李萌, 董颖红, 2013)。此领域研究多为计算机学者使用量化分析和可视化分析相结合的方法对社交网站上的社会网络进行分析,有待文化心理研究者未来更多涉足。

研究者通常首先选定具有丰富社会网络信息的互联网平台,如 Facebook、Twitter 等。接着,通过社会网络分析方法,如 UCINET 软件,对感兴趣的社会网络特性进行刻画:(1)社会网络的结构特性,如网络规模(network size)、网络密度(network density)、中介性(betweenness centrality)、网络异质性(network heterogeneity)等(Lewis, Kaufman, Gonzalez, Wimmer, & Christakis, 2008);(2)关键人物节点特性,如意见领袖(opinion leader)的推文特点及其社交网络特点(Quercia, Capra, & Crowcroft, 2012; 王嘉, 戴艳军, 王智宇, 2012);(3)社群挖掘,如关注不同性别、种族、社会经济状况的用户子群体间的差异(Lewis et al., 2008),又如从用户行为模式差异发现子群体(Conover et al., 2011)。同时,研究者可以对获得的社会网络信息进行可视化分析,如社群的可视化(Conover et al., 2011)。最后,研究者可发掘社会网络与文化之间的关联。此时文化差异研究主要关注不同文化下社会网络特征的差异(Poblete, Garcia, Mendoz,

& Jaimes, 2011); 文化变迁研究主要关注线上社会网络的状况描绘、相较线下社会网络有何异同、对文化变迁产生了哪些影响(Lewis et al., 2008)。

2.4 互联网使用行为分析

互联网使用行为数据是指用户在使用互联网过程中产生的大量点击流、事务数据和用户模型数据,包括服务器日志文件(用户请求的时间、客户端 IP 地址、所请求的资源等)和用户统计数据(浏览记录、购买历史、评分等)(张引等, 2013)。这类数据具有实时性较高、来源平台广、种类丰富的特点。Web 使用挖掘是从这些互联网使用行为中捕捉、建模并分析用户与网站交互的行为模式和整体状况。文化心理研究者当前更多对社交网站、工具网站等平台的前台可见用户行为数据进行量化分析。服务器日志文件较少被使用的原因可能是,目前互联网平台提供的后台数据接口较为有限。

研究者进行互联网使用行为分析时,通常首先选择平台及关心的互联网行为指标,如维基百科上的词条修改行为(Ferron & Massa, 2011; Yasseri, Spoorri, Graham, & Kertész, 2014), Springer 出版公司网站上的文献下载行为(Wang et al., 2012),日程安排网页工具 Doodle 上的日程安排行为(Reinecke, Nguyen, Bernstein, Näf, & Gajos, 2013)。接下来,文化变迁研究通过关注行为发生的过程来研究文化过程与趋势,如利用百度搜索指数对中国城市国知名度进行历史变迁分析(柳建坤, 吴愈晓, 刘伟峰, 2016)。文化差异研究则通过统计行为频率等方式对比不同文化背景下互联网使用行为的差异;在此之后,还可通过建立行为与文化价值观的联系来解释不同文化用户的行为差异。如有研究发现权力距离高的文化下, Flickr 图片标签中更多出现首府或国家名称(Dotan & Zaphiris, 2010)。

2.5 数据分析方法与互联网平台的关系

值得注意的是,互联网数据类型(及其对应分析方法)与互联网数据来源平台并非一一对应关系。互联网数据来源平台,包括社交网络服务(Social Networking Services, SNS)类、资源资讯类、搜索引擎(Search Engine)类、实用工具类等(如表 1)。一个互联网平台可能包含多种类型的互联网数据,如微博既反映用户之间的关注与被关注关系(社会网络关系数据),又包含用户发表的微

博(文本数据)、照片和视频(多媒体数据),还记录了用户的评论、转发和搜索等行为(互联网使用行为数据)。因此需要针对互联网平台包含的数据类型,综合应用各种互联网数据分析方法(见表 2)。

3 基于互联网数据的文化差异与变迁实证研究

3.1 文化差异

互联网记录了全球各个文化背景下人们的所思所想、所作所为,因此互联网数据被广泛应用于文化差异研究。下面分别介绍应用不同类型的互联网数据进行文化差异研究的进展。

3.1.1 文本数据

文本资源库(如谷歌图书)和社交网络平台(如 Facebook, 维基百科)中都包含着丰富的文本信息。这些文本信息体现了文化中价值观和自我概念等方面的差异。

社交网络中的文本可以体现不同文化下个体自我概念上的差异。DeAndrea, Shaw 和 Levine (2010)在美国某大学社交网络中抽取美国白人学生、美国非裔学生与亚裔学生(包含非美国国籍者) Facebook 用户各 40 名,借助用于分析“20 句自我陈述测验(Twenty Statements Test)”的编码系统及 LIWC 对用户主页上的“关于我(About Me)”文本进行分析。结果发现自我概念(个体-集体主义)对社交网络中的心理特质、个人细节、群体和关系归属、家庭宗教观念等的表达都有影响。

不同国家的语言使用可以体现文化价值观差异, Uz (2014)使用 10 种不同语言的 Google Book Ngram 人称代词数据及语言层面的个体主义得分来研究个体主义与人称代词使用的关系。结果发现,语言中第一人称单数代词的使用反映了该语言文化中的个体主义水平。

社交网络中的文本也可以体现文化价值观差异。Tsai 和 Men (2012)使用内容分析法研究中国和美国流行社交网站(人人网与 Facebook)上各 43 家公司的主页所反映的文化价值观,包括高-低语境、权力距离与个体-集体主义。结果发现,中国人人网上的公司主页内容更多反映互赖、受欢迎、高社会地位等价值观;美国 Facebook 上的公司主页则更多强调个体性与享乐主义。Chatzithomas, Boutsouki, Hatzithomas 和 Zotos (2014)使用文本分析法比较希腊和美国大型公司的社交网站广告使

表1 互联网平台分类及最受欢迎平台*一览表

平台类型	平台子类型	国外平台	国内平台
1. 社交媒体/ 社交网络	1.1 博客	Blogspot; Blogger; WordPress.com; Tumblr; Diply; Google+	新浪博客; Pixnet (痞客邦, 台湾)
	1.2 微博	Twitter; t.co	新浪微博
	1.3 论坛/社交新闻	Reddit	天涯论坛、百度贴吧
	1.4 图片分享	Imgur; Instagram; Pinterest; Flickr	Nice; In; LOFTER
	1.5 视频分享	YouTube; Nicovideo	优酷网; 土豆网; 爱奇艺
	1.6 百科与问答	Wikipedia; Stack Overflow	百度百科、百度知道、互动百科、知乎
	1.7 社交主页	Facebook; LinkedIn; VK; Odnoklassniki; MySpace	人人网; QQ 空间
2. 资源资讯	2.1 新闻网站	CNN; Guangming Daily; New York Times	新华网(Xinhua News Agency); 中国日报网(China Daily);
	2.2 门户网站	MSN; Yahoo!; Naver; FC2 Portal; Go.com	腾讯网; 新浪网; 搜狐网; Hao123;
	2.3 分类广告	Craigslist; Onlickads	58 同城; 智联招聘
	2.4 多媒体资源	Internet Movie Database; Netflix; Kickass Torrents	中国网络电视台; QQ 音乐
	2.5 文档、代码资源	GitHub; Google book	百度文库; CSDN
3. 搜索引擎	搜索引擎	Google; Bing; Yandex; Ask.com; Cốc Cốc	百度; 搜狐; 360 搜索
4. 实用工具 及其他	4.1 即时通讯	WhatsApp	微信; QQ;
	4.2 电子邮件	Mail.ru	网易邮箱; QQ 邮箱
	4.3 电子商务	Amazon; Rakuten; Alibaba Group; eBay; AliExpress; Flipkart	阿里巴巴; 淘宝; 天猫; 大众点评网
	4.4 支付系统	PayPal; Chase	支付宝; 中国工商银行
	4.5 文件存储	Google file storage; Dropbox	百度云; 微云
	4.6 系统支持	Apple Inc.; Microsoft; Adobe; Outbrain.com	奇虎 360(Qihoo 360); 360 安全卫士
	4.7 办公	Microsoft Office; Microsoft Online	WPS

注: *最受欢迎网站举例来源 1.Alexa Internet 发布的最受欢迎的 100 个网站排名 <http://www.alexa.com/topsites> (2016-03-23 提取); 2. 中国互联网协会中国网站排名 <http://www.chinarank.org.cn> (2016-06-02 提取); 3. China Webmaster 中国网站排行榜 <http://top.chinaz.com/all/> (2016-05-29 提取)。

表2 互联网平台类型对应的数据分析方法一览表

互联网平台类型	主要应用的数据分析方法
1. 社交媒体/社交网络	社会网络分析、互联网使用行为分析、文本分析、多媒体分析
2. 资源资讯	文本分析、多媒体分析、互联网使用行为分析
3. 搜索引擎	互联网使用行为分析、文本分析
4. 实用工具及其他	互联网使用行为分析、文本分析、社会网络分析、多媒体分析

用情况, 结果发现美国公司比希腊公司更多使用社交网站来宣传其产品或服务, 这可能是因为高个体主义、低不确定性规避、低权力距离和高男子气文化下的公司更敢于创新、更注重财务增长。Waters 和 Lo (2012)对不同文化下的 225 家非营利组织的 Facebook 主页进行内容分析, 探讨文化对信息展露、信息传播和参与度的影响。结果发现, 相较于高语境(high context)的中国和土耳其非营利

组织, 低语境(low context)的美国非营利组织主页有更多自发信息展露; 而中国的非营利组织网页上有最多与利益相关者的互动, 这可能是中国传统关系文化的体现。Pfeil, Zaphiris 和 Ang (2006)使用内容分析法研究了法国、德国、日本和荷兰版本的维基百科网页改动记录。结果发现, 权力距离越高的国家对已经编辑好的维基词条信息和链接删除和改动越少; 男子气越高的国家网页增

加和澄清信息越多;不确定性规避越高的国家网页内容增加越多。

3.1.2 多媒体数据

利用多媒体数据进行的文化差异研究,主要集中在不同文化的认知差异方面,如认知偏好及概念表征的文化差异。主要使用的平台有门户网站、社交网络等。

不同文化下网页界面与社交网络流行图片可以体现出认知偏好的文化差异。Marcus 和 Gould (2000)从 Hofstede 文化模型出发,发现了网页用户界面设计中的文化差异。如,由于权力距离更高的国家在网页设计中可能会更突出重要人物,因此在人机交互的心智模型(mental model)中会体现出更复杂的层级关系。类似地,Robbins 和 Stylianou (2002)发现商业网站的中设计元素反映了文化的差异,如组织架构图出现的频率高反映了高水平的权力距离。Barber 和 Badre (1998)发现不同文化群体成员喜欢不同的图标、颜色与网站结构,也就是说网页的一些设计元素可以看作一种文化标记(cultural marker)。Schmid-Isler (2000)发现了西方和中国的新闻网站的布局差异,表明文化影响人们对信息存储和展示的感知。Hochman 和 Schwartz (2012)使用图片可视化分析软件 ImageJ 和 ImagePlot 分析 Instagram 上不同文化的图片颜色分布,发现东京图片比纽约图片在一天内的亮度变化更小、总体色调更偏红黄。

社交网络上不同文化对相同概念具有不同表征。Yanai, Yaegashi 和 Qiu (2009)使用最新的物体识别技术对 Flickr 上标记了地理信息的图片进行跨文化比较,以发现与某些概念相关图片的文化差异,如美国图片中的婚礼蛋糕层数比欧洲图片中的层数更多,亚洲图片中的瀑布相较北美图片中的瀑布更为柔美,这表明同样的概念在不同文化中的表征具有差异。

3.1.3 互联网社会网络关系数据

社交平台毋庸置疑蕴含了全球丰富的社会网络数据,因此被不少研究者用来研究社会网络特点的文化差异。Lewis 等(2008)使用 Facebook 研究了不同性别、种族、社会经济地位和线上活跃度学生的社会网络特点,结果发现,非裔用户的社会网络拥有最大的网络规模(network size)和中介性(betweenness centrality)、亚裔用户的社会网络拥有最小的网络密度(network density)、亚裔

和非裔社会网络相较白人社会网络都具有更高的网络异质性(network heterogeneity)。Poblete 等(2011)研究 10 个最活跃国家超过的 450 万 Twitter 用户的连结关系,发现 Twitter 上用户的社会网络特点存在文化差异,如印尼 Twitter 社会网络的模块化(modularity)程度远高于澳大利亚,说明印尼用户更多分为内部联系紧密、外部联系松散的小群体。日本等国的 Twitter 社会网络规模较小,节点互通性(reciprocity)较高,体现了这些国家用户间较高的凝聚力与信任。

3.1.4 互联网使用行为数据

互联网用户在社交网络、门户网站、工具网站等各种平台上的行为被记录下来成为互联网使用行为数据。这些数据体现了不同文化中个体的价值观及行为模式的差异。

Morio 和 Buchholz (2009)通过观察科技资讯社区 Slashdot 的美国版与日本版上用户对热门文章评论的匿名情况,结果发现,Slashdot 美国版上访问最多的 10 个话题和最活跃的 10 个话题的匿名评论比例都比日本版上的更低。他们认为这种差异的原因是个体主义文化下的个体社会互动的动机更多是追求自主性,集体主义文化下个体社会互动的动机则更多是从属、融入群体。Dotan 和 Zaphiris (2010)比较了 5 个不同文化下的用户在 Flickr 上行为的特点。结果发现,男子气高的文化下,更多用户选择使用功能强大的专业账户(而非普通账户);不确定性回避高文化下的用户倾向于有更少的公开联系人和更多的图片标签;权力距离高的文化下,图片标签中更多出现首府或国家名称。

Wang 等人(2012)对 Springer 出版公司的网站 5 个工作日和 4 个休息日的下载情况的实时数据进行分析,发现周末下载研究文献在亚洲国家最常见,在德国最不常见。Reinecke 等(2013)对日程安排网页工具 Doodle 在 211 个国家的使用状况进行研究,结果发现德国用户倾向于提前大约 28 天安排日程,而哥伦比亚用户则通常最多提前 12 天提前安排日程。这是因为文化因素(如个体主义-集体主义)会影响活动的特征(如活动参与者对活动时间达成一致的难易程度)。

3.2 文化变迁

互联网数据不仅反映了不同文化之间的差异,还反映了文化的变迁历程。其中具有巨大时间跨

度和丰富语义信息的数字化文本数据在文化历史变迁研究中最受瞩目。而多媒体数据、社会网络数据、互联网使用行为数据与文本数据一起,记录了文化在当代的变迁。国内学者基于互联网数据的文化心理研究也多集中在文化变迁领域。

3.2.1 文本数据

自 Michel 等人(2011)提出利用谷歌数字化图书进行文化研究的文化组学概念后,越来越多文化历史变迁研究开始使用谷歌图书数据集及 Google Ngram 工具。使用如此大规模、长时间跨度的文本数据为研究文化价值观与文化现象的变迁研究带来了新的突破。

使用谷歌图书数据可以研究个体主义-集体主义的变迁。研究者对美国 1960 年到 2008 年出版图书中的代词作分析发现,第一人称复数代词(如 we, us)出现频率下降 10%, 第一人称单数代词(I, me)出现频率翻倍,第二人称代词(you, your)出现频率变为开始的 4 倍。这些结果补充了个体主义随时间变化的相关研究(Twenge, Campbell, & Gentile, 2013)。Twenge, Campbell 和 Gentile (2012a)还对同一时期的个体主义词和短语做了类似分析,发现个体主义上升的结果是稳定的。Greenfield (2013)将时间跨度扩大到 1800 年至 2000 年,将集体主义及其相关词也加入分析,发现集体主义相关词频率下降,个体主义相关词频率上升。他将上述结果与美国人口普查中的居住地数据联系起来,认为这个结果支持了人口城市化进程对文化的影响,符合以往理论对人类发展的预测。Grossmann 和 Varnum (2015)使用 8 个国家层面的文化主义指标(图书中的个体主义主题频率、婴儿名字的独特性、独生子女家庭比例、单代家庭的比例、独居成年人与老人比例、家庭规模、离婚率)研究美国 20 世纪的文化变迁,得到了一致的个体主义上升结果。使用交叉滞后分析发现只有社会经济结构变化可能是美国 20 世纪个体主义文化增长的原因。Zeng 和 Greenfield (2015)发现 1970 到 2008 年之间中文图书中集体主义相关词频率下降,这可能与美国同时期的经济改革有关。

使用谷歌图书数据可以研究其他价值观变迁。P. Kesebir 和 S. Kesebir (2012)使用 Google Ngram Viewer 考察了美国图书中 1900 年到 2000 年之间与道德特质和美德概念的变化趋势,发现 10 个一般的道德词(如 virtue, decency)和 50 个具体的道德

词(如 honesty, patience)中大多数词的出现频率呈下降趋势,与年代成负相关。他们推论道德概念相关词频率下降和文化中个体主义的逐年升高有关。Acerbi, Lampo, Garnett 和 Bentley (2013)通过 6 类情绪词在最近 1 个世纪的英式英语和美式英语中的出现的频率,研究情绪表达随时间的变化。研究发现不同类型情绪词的峰值出现在不同的历史时期,美式和英式英语中情绪词使用都有所减少。研究者还通过男女代词频率变化研究女性地位(Twenge, Campbell, & Gentile, 2012b);通过 happy nation 和 happy person 这两个短语的频率变化考察美国人对幸福理解的变化(Oishi, Graham, Kesebir, & Galinha, 2013);通过将自尊相关词(self-esteem, self-control)变化趋势作为因变量,将从个人到文化三个层次的客观调查数据作为预测变量,进行回归分析来考察影响自尊的因素(Konrath & Anderson, 2011)。

使用谷歌图书还可研究其他文化现象的变迁。Juola (2013)使用谷歌图书 Ngram 数据库中的词对(2-grams)数量研究美国文化的文化复杂性(cultural complexity)随时间的变化。结果发现美国文化的复杂性一直在稳定增加。Bochkarev, Shevlyakova 和 Solovyev (2012)使用 Google Ngram 提供的词语(n-grams)集合数据对近 200 年来的英语和俄语图书中的单词长度进行了统计分析,结果发现两种语言的单词长度都在 19 世纪缓慢增长、20 世纪迅猛增长、20 世纪末到 21 世纪初开始减少。进一步分析发现这种变化趋势主要与社会发展相关词语的使用变化,以及人称代词的使用变化有关。Roth (2014)对谷歌 1800~2000 年间不同语言的数据进行分析,结果发现,社会功能系统(如政治、经济、科学、艺术、宗教等)的关联性在社会系统中具有不均衡的分布,不同社会都表现出了较为稳定的世俗化、政治化、经济化、媒体化的趋势。Acerbi 和 Bentley (2014)追踪了谷歌英文图书中颜色词语使用、美国婴儿名字使用、音乐网站 Last.fm 用户音乐偏好的变化,结果发现文化演变并非完全遵循中性模型(neutral model),而是会受到基于内容的偏差(内容吸引力)和基于情境的偏差(从众与反从众)的影响。Virues-Ortega 和 Pear (2015)分析了 1900~2008 年的谷歌图书中与“行为”和“认知”有关词语的频率,发现在 1930 年左右“行为”相关词频率持续上升,直到 1970 年

左右达到开始下降;而“认知”相关词则从1970年开始持续上升。这种变化模式支持了根据“认知革命”研究做出的推测。

除了利用谷歌图书进行文化纵向历史研究,也有研究者关注其他互联网平台反映出的当代文化变迁趋势与新特点。Facebook、YouTube、Twitter及Google+等平台具有更高的实时性与更强的互动性,因此是研究当代文化发展的绝佳阵地。Church (2013)通过对Facebook上的死者的悼念主页上的悼文做文本分析发现,死者家属通常通过悼文与死者直接沟通,悼文的可见性增加了生者之间的共同经验。线上悼念相对于传统线下悼念持续时间较短,但生者与死者的直接沟通似乎打破了死亡带来的永恒隔阂。互联网技术改变着人们悼念的传统,帮助人们超越死亡“永存”。Goode等(2011)对主流新闻媒体发布的YouTube视频收到的用户评论进行了内容分析,认为YouTube以其不受约束的特点,成为了人们观点激烈交锋的平台,影响了公共领域文化的发展。有研究者发现Twitter上包含消极情绪的推文将获得更多转发(Naveed, Gottron, Kunegis, & Alhadi, 2011)。类似地,Heimbach, Schiller, Strufe和Hinz (2015)使用文本挖掘方法在不同社交网络平台(Twitter, Facebook和Google+)上研究文章内容特点对文章传播的影响,结果发现文章内容的情感特征(如悲伤)、文章组成特点(如含有图片的数量)、作者特点(如名望)以及对读者注意的竞争能力(如文章所在位置)都会影响文章在社交网络上的传播。李彪(2013)对2010年到2013年间的6025条热门微博进行了内容主题分析发现“休闲心情”类微博最多,并进一步根据热门微博作者和转发者的地域、性别特征提出,微博话语权力仍存在男性主导、精英群体主导、经济发达地区主导的不平等现象。

3.2.2 多媒体数据

图片与影像全景记录了人们生活,也反映了文化发展的历程。然而由于多媒体资料可得年限较短,所以纵向积累较少。因此使用网络多媒体数据进行的文化变迁研究主要集中在使用图片及视频研究当代文化特点与发展趋势方面。

Ribisl, Lee, Henriksen和Haladjian (2003)对提倡吸烟文化和生活方式的网站进行内容分析发现,这些网站不仅允许青少年浏览、缺少吸烟健康警示,而且包含烟草品牌及促销、吸烟图片、性内

容等鼓励吸烟的信息,对青少年有潜在的不良影响。Lee和Barton (2011)对中国和西班牙用户在图片分享平台Flickr上发布内容进行探索性内容分析,并结合线上问卷与线上焦点访谈研究了用户的多语言使用及其对全球化认同的影响。研究发现多语言者在Flickr上的语言使用受到语言环境、假想观众和图片内容的影响。Flickr用户的多语言使用一方面是全球化的表现,增强了用户的全球化身份认同,另一方面可能也使用户在向全球观众介绍本土文化的过程中增强了本土化身份认同。

Morris和Anderson (2015)通过对YouTube上英国最受欢迎的男性用户所发布视频进行解释性分析和对其中一名用户的访谈发现,这些受欢迎的男性普遍具有一种更为温和的包容性男子气(Inclusive Masculinity),不同于传统的具有攻击性、大男子主义、恐惧同性恋、贬低女性的异性恋男子气。研究者认为包容性男子气在当代青少年文化中占据主导地位是文化中同性歇斯底里(homophobia)减弱的反映,男性不再需要通过表现得阳刚来避免同性恋污名。同样利用Youtube平台, van Zoonen, Vis和Mihelj (2010)对与一部反伊斯兰视频Fitna有关的视频进行了量化和质性分析,发现这些视频内容大都是用户发表自己对于Fitna的意见,包括向伊斯兰表示歉意、向伊斯兰表达讽刺等不同的类型。这说明,Youtube把不同宗教和政治身份的用户连接起来,促进了公民表达意见和行使公民权。高崇和杨伯淑(2016)对“秒拍”上的高转发微视频进行人工内容主题分析,发现生活、搞笑和娱乐类视频占到绝大多数,这体现了国内当前文化中崇尚分享与快乐的元素。

3.2.3 社会网络关系数据

社交网络平台是人们发布与传播信息的平台,因此被不少研究者用来研究社会网络的当代文化特征。Conover等(2011)使用Twitter研究了美国2010年竞选季时的政治极化现象,即右倾用户倾向于只转发来自其他右倾用户的微博,左倾用户倾向于只转发其他左倾用户的微博,因此挖掘出了社交网络中因价值观不同而产生的社群划分。Quercia等(2012)利用Twitter研究在线社会网络的规律与现实世界的社会网络规律是否一致,结果发现充当社交经纪人(social broker)、为整个社会网络带来了更丰富联结关系的用户,通常会意见领袖(opinion leader)。他们的推文包含话题更

广,传播地理范围更大,表达了更丰富的情感内涵。另外,表达相似情感(积极或消极)的人更容易聚集成小社群。王嘉、戴建军和王智宇(2012)利用社会网络分析软件 UCINET 6.0 对人人网上当代中国大学生的“网络意见领袖”进行研究,发现大学生对娱乐休闲的关注度会随着年级升高而增加,并且网络意见领袖和现实意见领袖具有一定相关性。

3.2.4 互联网使用行为数据

人们在搜索引擎上的搜索行为可以记录文化变迁。国内研究者基于用百度搜索上 2011 年至 2016 年城市词语的搜索指数,发现中国主要城市知名度的历史变迁可能受到经济、政治、文化资源等多方面因素的影响(柳建坤等, 2016)

人们在社交网络上的点击链接、点击“喜欢(likes)”和评论等行为可以作为文化现状的刻画与文化发展趋势的风向标。Borra 和 Weber (2012)发现 Twitter 用户对最受欢迎的 155 条美国政治博客的点击情况可以用来推断这些用户的政治倾向。Araújo, Corrêa, da Silva, Prates 和 Meira Jr (2014)对 Instagram 上的 25 万用户的行为进行分析发现,用户倾向于在周末和一天结束时发布照片;并且会支持那些已经获得了很多喜欢(likes)和评论的照片,这导致了“富者越富”的现象。

个体参与维基百科编辑词条的操作也被记录下来,成为当代人际互动与群际互动研究的对象。Ferron 和 Massa (2011)利用维基百科中相关词条编辑情况(如创建时间、编辑次数、编辑者等)研究了北非动乱中的群体记忆实时构建的过程。Yasseri 等(2014)对 10 种不同语言版本(如英语、法语、阿拉伯语)的维基百科上的最有争议的词条进行了量化和可视化分析,结果发现宗教相关词条被编辑次数最多,这个结果具有跨文化一致性。

4 局限与展望

从上面的综述可以看出,互联网数据因其开放易得、生态高效、种类丰富、规模巨大、时间跨度长等特点,促进了跨地区的文化差异研究和跨时间的文化变迁研究。当前研究在取得以上丰硕成果的同时,也表现出了研究效度、技术、理论方面的局限性。这些局限提示了未来研究完善与发展的方向。

第一,所有应用互联网数据进行的研究,其推广效度、内部效度和理论效度都可能受到质疑。

首先,网民有偏抽样和网络情境特异性造成结果推论困难。一方面,网民在社会阶层、民族、性别、年龄、学历、地区等人口学变量方面与国民总体有显著差异(Suarez-Balcazar, Balcazar, & Taylor-Ritzler, 2009),因此使用互联网数据研究全体大众的心理与行为是一种有偏抽样。另一方面,线上网络环境与线下现实环境相比具有匿名性等特征,因此个体在线上的表现(如自我展露)也会与在线下时有所不同(Nguyen, Bin, & Campbell, 2012),使用互联网研究文化可能也带有某种特异性。其次,额外因素导致难探因果。互联网数据大多以日志、记录形式存在,研究者难以就实验目的对额外变量进行控制、对研究被试进行随机化,因此研究者运用互联网数据进行的研究多是描述性、相关性的,难以探明其中的因果关系。最后,新数据挑战指标与概念对应合理性。目前研究中文化心理学概念与互联网数据指标建立对应关系的合理性,可能基于理论推测而缺乏效度数据的支持。

因此,未来的文化心理学研究应注意:(1)如果研究者要将基于线上数据的研究结果推广到线下,应当注意有偏样本和特殊情境对于研究问题涉及总体的代表性。当样本特征及情境特征与预测变量对因变量的影响不存在交互作用时,研究结果的推广效度不会被降低;(2)基于互联网数据研究共变关系的原理时,可以考虑使用统计方法事后控制额外变量,以及时间序列相关分析等准因果分析方法辅助因果推断;(3)基于新的互联网数据指标做研究时,为验证指标的理论效度,可以计算互联网数据指标和传统概念指标的效标关联效度,来考察这种互联网数据指标的合理性。研究者使用多种有效新指标的尝试应当受到鼓励,因为通过使用多种指标相结合的方法,可以增加文化心理研究结果的聚合效度。

第二,目前文化心理学研究者对计算机前沿技术利用程度较低。首先,当前研究中对文本数据的分析大多集中在词语层面,较少涉及语段和语篇层面;其次,多媒体数据的分析仍采用人工编码和诠释学方法,很少使用模式识别等自动化分析;再次,文化心理学研究者对社会网络关系数据分析的基本技术掌握程度较低,很少能够独立实施相关研究;最后,互联网使用行为数据多采用前端用户行为统计的方法,缺少对后端服务

器日志的挖掘。

因此,未来的文化心理学研究应注意:(1)增加技术敏感度,综合运用最新的信息技术分析互联网数据,尤其是在如信息提取、主题模型、文本摘要、分类、聚类、答疑和意见挖掘等自然语言处理技术方面,以及多媒体摘要、注解、索引和检索、多媒体事件监测等计算机辅助的多媒体分析技术方面;(2)加强与计算机科学家的合作,探索跨学科合作科研模式。由文化心理学家充当领域专家提供文化相关理论与研究思路,并对研究结果进行理解与解释,计算机科学家充当技术专家使用前沿技术实施研究;(3)加强与互联网平台的合作,探索商界与学界互促发展模式。这种合作不仅可以帮助文化心理学研究者全方位获取后台服务器数据、获得分析技术支持,并且可以促进研究成果的转化。同时,对获取的所有互联网数据都要做匿名化处理,防止泄露用户隐私。

第三,当前研究多是通过用户在互联网上的表现来验证已有的文化心理学理论,较少理论发展与创新。各项研究应用的文化心理学理论出现了一定的同质化倾向。Hofstede 的文化维度模型较多被用于解释文化差异(Chatzithomas et al., 2014; Dotan & Zaphiris, 2010; Marcus & Gould, 2000; Morio & Buchholz, 2009; Pfeil et al., 2006; Reinecke et al., 2013)。城市化理论较多被用来解释文化变迁(Greenfield, 2013; Grossmann & Varnum, 2015),而其他理论应用较少。致力于从互联网海量数据中发现新的隐含模式的研究数量更少,因此很少有研究能够大幅推动文化心理学理论发展与创新。

因此,未来的文化心理学研究者要注意:(1)提升数据在研究中的地位,有意识以数据为中心来思考、设计和实施科学研究,不拘泥于理论假设涉及的指标,而是尽可能多地获取数据指标,通过对海量数据的收集和分析处理来发现新规律;(2)树立学科意识,注意研究结果积累。借鉴已有学者提出的计算社会科学(Computational social science) (陈浩等, 2013; Lazer et al., 2009)和心理信息学(Psychoinformatics) (薛婷, 陈浩, 赖凯生, 董颖红, 乐国安, 2015; Yarkoni, 2012),建立注重收集和分析各种心理信息的文化心理信息学,以及使用建模方法模拟心理过程进行研究的计算文化心理学(Hey, Tansley, & Tolle, 2009),以系统积

累文化差异与文化变迁研究的材料与经验。

参考文献

- 陈浩, 乐国安, 李萌, 董颖红. (2013). 计算社会科学: 社会科学与信息科学的共同机遇. *西南大学学报(社会科学版)*, 39(3), 87-93.
- 丰怡, 蔡华俭, 施媛媛. (2013). 文化产品研究——文化心理学的独特视角. *心理科学进展*, 21, 326-335.
- 高崇, 杨伯淑. (2016). 微视频的内容主题发展趋势分析——基于对新浪微博官方短视频应用“秒拍”上高转发微视频的研究. *新闻界*, (12), 47-50.
- 李彪. (2013). 微博中热点话题的内容特质及传播机制研究——基于新浪微博 6025 条高转发微博的数据挖掘分析. *中国人民大学学报*, 27(5), 10-17.
- 柳建坤, 吴愈晓, 刘伟峰. (2016). 中国城市国内知名度的变迁和机制——基于海量书籍和互联网搜索引擎的大数据分析. *学术论坛*, (6), 113-121.
- 王嘉, 戴艳军, 王智宇. (2012). 大学生“网络意见领袖”研究——基于人人网. *中国青年研究*, (7), 79-85.
- 王娱琦, 余震坤, 罗宇, 陈杰明, 蔡华俭. (2015). 利用网络进行心理学研究: 西方与中国概况. *心理科学进展*, 23, 510-519.
- 薛婷, 陈浩, 赖凯生, 董颖红, 乐国安. (2015). 心理信息学: 网络信息时代下的心理学新发展. *心理科学进展*, 23, 325-337.
- 袁军鹏, 朱东华, 李毅, 李连宏, 黄进. (2006). 文本挖掘技术研究进展. *计算机应用研究*, (2), 1-4.
- 张引, 陈敏, 廖小飞. (2013). 大数据应用的现状与展望. *计算机研究与发展*, 50, 216-233.
- 赵志裕, 康莹仪. (2011). 什么是文化? 见 赵志裕, 康莹仪 (编), *文化社会心理学* (刘爽 译). 北京: 中国人民大学出版社.
- 中国互联网络信息中心. (2016). *中国互联网络发展状况统计报告*. 2016-06-01 取自 <http://www.cnnic.net.cn/hlwfzyj/hlwzxbg/hlwjtjbg/201601/P020160122444930951954.pdf>
- Acerbi, A., Lampos, V., Garnett, P., & Bentley, R. A. (2013). The expression of emotions in 20th century books. *PLoS One*, 8, e59030.
- Acerbi, A., & Bentley, R. A. (2014). Biases in cultural transmission shape the turnover of popular traits. *Evolution and Human Behavior*, 35, 228-236.
- Araújo, C. S., Corrêa, L. P. D., da Silva, A. P. C., Prates, R. O., & Meira, W., Jr. (2014, October). It is not just a picture: Revealing some user practices in instagram. In *Proceedings of ninth Latin American web congress* (pp. 19-23), Minas Gerais, Brazil.
- Barber, W., & Badre, A. (1998) Culturability: The merging of culture and usability. In *Proceedings of the fourth*

- conference on human factors and the web, Atlanta, GA, USA. Retrieved June 2, 2016, from <http://zing.ncsl.nist.gov/hfweb/att4/proceedings/barber/>
- Bochkarev, V. V., Shevlyakova, A. V., & Solovyev, V. D. (2012). Average word length dynamics as indicator of cultural changes in society. *arXiv preprint arXiv: 1208.6109*. Retrieved June 2, 2016, from <https://arxiv.org/pdf/1208.6109.pdf>
- Borra, E., & Weber, I. (2012). Political insights: Exploring partisanship in Web search queries. *First Monday*, 17. Retrieved June 2, 2016, from <http://firstmonday.org/ojs/index.php/fm/article/view/4070/3272?sa=X&ved=0CC0Q9QEwDwoVChMIwbup1-n4xgIVRr5yCh1Yhw-3>
- Chatzithomas, N., Boutsouki, N., Hatzithomas, L., & Zotos, Y. (2014). Social media advertising platforms: a cross-cultural study. *International Journal of Strategic Innovative Marketing*, 1, 74–90.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14, 464–504.
- Church, S. H. (2013). Digital gravescapes: Digital memorializing on Facebook. *The Information Society*, 29, 184–189.
- Conover, M. D., Ratkiewicz, J., Francisco, M. R., Gonçalves, B., Menczer, F., & Flammini, A. (2011, July). Political polarization on twitter. In *Proceedings of the fifth international AAAI conference on weblogs and social media* (pp. 89–96). Barcelona, Spain: AAAI.
- DeAndrea, D. C., Shaw, A. S., & Levine, T. R. (2010). Online language: The role of culture in self-expression and self-construal on Facebook. *Journal of Language and Social Psychology*, 29, 425–442.
- Deuze, M. (2006). Participation, remediation, bricolage: Considering principal components of a digital culture. *The Information Society*, 22, 63–75.
- Dotan, A., & Zaphiris, P. (2010). A cross-cultural analysis of Flickr users from Peru, Israel, Iran, Taiwan and the UK. *International Journal of Web Based Communities*, 6, 284–302.
- Ferron, M., & Massa, P. (2011, October). Collective memory building in Wikipedia: The case of North African uprisings. In *Proceedings of the seventh international symposium on wikis and open collaboration* (pp. 114–123). Mountain View, California: ACM.
- Goode, L., McCullough, A., & O'Hare, G. (2011). Unruly publics and the fourth estate on YouTube. *Participations: Journal of Audience and Reception Studies*, 8, 594–615.
- Greenfield, P. M. (2013). The changing psychology of culture from 1800 through 2000. *Psychological Science*, 24, 1722–1731.
- Grossmann, I., & Varnum, M. E. W. (2015). Social structure, infectious diseases, disasters, secularism, and cultural change in America. *Psychological Science*, 26, 311–324.
- Heimbach, I., Schiller, B., Strufe, T., & Hinz, O. (2015, September). Content virality on online social networks: Empirical evidence from Twitter, Facebook, and Google+ on German News Websites. In *Proceedings of the twenty-sixth ACM conference on hypertext & social media* (pp. 39–47). Guzelyurt, Northern Cyprus: ACM.
- Heine, S. J. (2010). Cultural psychology. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., Vol. 2, pp. 1423–1464). Hoboken, NJ: John Wiley & Sons.
- Heine, S. J., Lehman, D. R., Peng, K. P., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales?: The reference-group effect. *Journal of Personality and Social Psychology*, 82, 903–918.
- Hey, T., Tansley, S., & Tolle, K. (Eds.). (2009). *The fourth paradigm: Data-intensive scientific discovery*. Redmond, WA: Microsoft Research.
- Hochman, N., & Schwartz, R. (2012). Visualizing instagram: Tracing cultural visual rhythms. In *Proceedings of the workshop on social media visualization (SocMedVis) in conjunction with the sixth international AAAI conference on weblogs and social media* (pp. 6–9). Dublin, Ireland: AAAI.
- Hofstede, G. (1983). National cultures in four dimensions: A research-based theory of cultural differences among nations. *International Studies of Management & Organization*, 13, 46–74.
- Inglehart, R. (1997). *Modernization and postmodernization: Cultural, economic, and political change in 43 societies* (Vol. 19). Princeton, NJ: Princeton University Press.
- Juola, P. (2013). Using the Google N-Gram corpus to measure cultural complexity. *Literary and Linguistic Computing*, 28, 668–675.
- Kahn, J. H., Tobin, R. M., Massey, A. E., & Anderson, J. A. (2007). Measuring emotional expression with the Linguistic Inquiry and Word Count. *The American Journal of Psychology*, 120, 263–286.
- Kesebir, P., & Kesebir, S. (2012). The cultural salience of moral character and virtue declined in twentieth century America. *The Journal of Positive Psychology*, 7, 471–480.
- Konrath, S. H., & Anderson, P. A. (2011). A century of self-esteem. In S. De Wals & K. Meszaros (Eds.), *Handbook on psychology of self-esteem* (pp. 157–174). New York: Nova Science Publishers.
- Kraidy, M. M. (2002). Hybridity in cultural globalization. *Communication Theory*, 12, 316–339.

- Kroeber, A. L., & Kluckhohn, C. (1952). *Culture: A critical review of concepts and definitions*. Cambridge, MA: Peabody Museum.
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A. L., Brewer, D.,... & van Alstyne, M. (2009). Life in the network: The coming age of Computational social science. *Science*, 323, 721–723.
- Lee, C. K. M., & Barton, D. (2011). Constructing glocal identities through multilingual writing practices on Flickr.com®. *International Multilingual Research Journal*, 5, 39–59.
- Lee, F., Hallahan, M., & Herzog, T. (1996). Explaining real-life events: How culture and domain shape attributions. *Personality and Social Psychology Bulletin*, 22, 732–741.
- Lesthaeghe, R. (1983). A century of demographic and cultural change in Western Europe: An exploration of underlying dimensions. *Population and Development Review*, 9, 411–435.
- Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A., & Christakis, N. (2008). Tastes, ties, and time: A new social network dataset using Facebook.com. *Social Networks*, 30, 330–342.
- Marcus, A., & Gould, E. W. (2000). Crosscurrents: cultural dimensions and global Web user-interface design. *Interactions*, 7(4), 32–46.
- Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., The Google Books Team,... & Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, 331, 176–182.
- Morio, H., & Buchholz, C. (2009). How anonymous are you online? Examining online social behaviors from a cross-cultural perspective. *Ai & Society*, 23, 297–307.
- Morris, M., & Anderson, E. (2015). 'Charlie is so cool like': Authenticity, popularity and inclusive masculinity on YouTube. *Sociology*, 49, 1200–1217.
- Naveed, N., Gottron, T., Kunegis, J., & Alhadi, A. C. (2011, June). Bad news travel fast: A content-based analysis of interestingness on twitter. In *Proceedings of the third international web science conference* (Article No. 8). Koblenz, Germany: ACM.
- Nguyen, M., Bin, Y. S., & Campbell, A. (2012). Comparing online and offline self-disclosure: A systematic review. *Cyberpsychology, Behavior, and Social Networking*, 15, 103–111.
- Oishi, S., Graham, J., Kesebir, S., & Galinha, I. C. (2013). Concepts of happiness across time and cultures. *Personality and Social Psychology Bulletin*, 39, 559–577.
- Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). *Linguistic inquiry and word count: LIWC* [Computer Software]. Austin, TX: LIWC. Net.
- Pfeil, U., Zaphiris, P., & Ang, C. S. (2006). Cultural differences in collaborative authoring of Wikipedia. *Journal of Computer-Mediated Communication*, 12, 88–113.
- Poblete, B., Garcia, R., Mendoza, M., & Jaimes, A. (2011, October). Do all birds tweet the same?: characterizing twitter around the world. In *Proceedings of the Twentieth ACM international conference on information and knowledge management* (pp. 1025–1030). Glasgow, Scotland, UK: ACM.
- Quek, C. Y., & Mitchell, T. (1997). *Classification of world wide web documents* (Unpublished master's thesis). Carnegie Mellon University, Pittsburgh.
- Quercia, D., Capra, L., & Crowcroft, J. (2012, June). The social world of twitter: Topics, geography, and emotions. In *Proceedings of the sixth international AAAI conference on weblogs and social media* (pp. 298–305). Dublin, Ireland: AAAI.
- Reinecke, K., Nguyen, M. K., Bernstein, A., Näf, M., & Gajos, K. Z. (2013, February). Doodle around the world: Online scheduling behavior reflects cultural differences in time perception and group decision-making. In *Proceedings of the 2013 conference on computer supported cooperative work* (pp. 45–54). San Antonio, Texas, USA: ACM.
- Ribisl, K. M., Lee, R. E., Henriksen, L., & Haladjian, H. H. (2003). A content analysis of Web sites promoting smoking culture and lifestyle. *Health Education & Behavior*, 30, 64–78.
- Robbins, S. S., & Stylianou, A. C. (2002). A study of cultural differences in global corporate Web sites. *Journal of Computer Information Systems*, 42(2), 3–9.
- Roth, S. (2014). Fashionable functions: A Google ngram view of trends in functional differentiation (1800–2000). *International Journal of Technology and Human Interaction*, 10(2), 34–58.
- Schimmack, U., Oishi, S., & Diener, E. (2005). Individualism: A valid and important dimension of cultural differences between nations. *Personality and Social Psychology Review*, 9, 17–31.
- Schmid-Isler, S. (2000, January). The language of digital genres. A semiotic investigation of style and iconology on the world wide web. In *Proceedings of the thirty-third Hawaii international conference on system sciences* (Vol. 3, p. 3012), Washington, DC, USA.
- Shweder, R. A. (1995). Cultural psychology: What is it? In N. R. Goldberger & J. B. Veroff (Eds.), *The culture and psychology reader* (pp. 41–86). New York: Cambridge University Press.
- Suarez-Balcazar, Y., Balcazar, F. E., & Taylor-Ritzler, T. (2009). Using the Internet to conduct research with culturally diverse populations: Challenges and opportunities.

- Cultural Diversity and Ethnic Minority Psychology*, 15(1), 96–104.
- Tsai, W. H., & Men, L. R. (2012). Cultural values reflected in corporate pages on popular social network sites in China and the United States. *Journal of Research in Interactive Marketing*, 6, 42–58.
- Twenge, J. M., Campbell, W. K., & Gentile, B. (2012a). Increases in individualistic words and phrases in American books, 1960–2008. *PLoS One*, 7, e40181.
- Twenge, J. M., Campbell, W. K., & Gentile, B. (2012b). Male and female pronoun use in U.S. books reflects women's status, 1900–2008. *Sex Roles*, 67, 488–493.
- Twenge, J. M., Campbell, W. K., & Gentile, B. (2013). Changes in pronoun use in American books and the rise of individualism, 1960–2008. *Journal of Cross-Cultural Psychology*, 44, 406–415.
- Uz, I. (2014). Individualism and first person pronoun use in written texts across languages. *Journal of Cross-Cultural Psychology*, 45, 1671–1678.
- van Zoonen, L., Vis, F., & Mihelj, S. (2010). Performing citizenship on YouTube: Activism, satire and online debate around the anti-Islam video *Fitna*. *Critical Discourse Studies*, 7, 249–262.
- Virues-Ortega, J., & Pear, J. J. (2015). A history of “behavior” and “mind”: Use of behavioral and cognitive terms in the 20th century. *The Psychological Record*, 65, 23–30.
- Wang, X. W., Xu, S. M., Peng, L., Wang, Z., Wang, C. L., Zhang, C. B., & Wang, X. B. (2012). Exploring scientists' working timetable: Do scientists often work overtime?. *Journal of Informetrics*, 6, 655–660.
- Waters, R. D., & Lo, K. D. (2012). Exploring the impact of culture in the social media sphere: A content analysis of nonprofit organizations' use of Facebook. *Journal of Intercultural Communication Research*, 41, 297–319.
- Wellman, B., & Berkowitz, S. D. (1988). *Social structures: A network approach*. New York: Cambridge University Press.
- Yanai, K., Yaegashi, K., & Qiu, B. Y. (2009, April). Detecting cultural differences using consumer-generated geotagged photos. In *Proceedings of the second international workshop on location and the web* (Article No. 12). Boston, Massachusetts, USA: ACM.
- Yarkoni, T. (2012). Psychoinformatics: New horizons at the interface of the psychological and computing sciences. *Current Directions in Psychological Science*, 21, 391–397.
- Yasseri, T., Spoerri, A., Graham, M., & Kertész, J. (2014). The most controversial topics in Wikipedia: A multilingual and geographical analysis. In P. Fichman & N. Hara (Eds). *Global Wikipedia: International and cross-cultural issues in online collaboration* (pp. 25–48). Lanham, MD: Rowman & Littlefield.
- Zeng, R., & Greenfield, P. M. (2015). Cultural evolution over the last 40 years in China: Using the Google Ngram Viewer to study implications of social and political change for cultural values. *International Journal of Psychology*, 50, 47–55.

Cultural difference and cultural change in internet data

LI Yongshuai¹; HUANG Zhenwei²; TAN Xunyun³; LIU Li¹

(¹ Beijing Key Lab of Applied Experimental Psychology, School of Psychology, Beijing Normal University,

Beijing 100875, China) (² Beijing Administration Institute, Beijing 100000, China)

(³ Institute of Sociology, Chinese Academy of Social Sciences, Beijing 100732, China)

Abstract: In the Internet era, information on the Internet reflects individuals' psychological processes and cultural characteristics. These huge and diverse Internet data provide new perspectives for the study of cultural psychology. First, Internet data has shown several strengths on investigating cultural difference and cultural change, which are the two approaches of studying cultural psychology. Second, there are four types of Internet data, and accordingly, cultural psychologists utilize text analysis, multimedia analysis, social network analysis, and online behavior analysis. Third, based on Internet data and corresponding analytic methods, research on cultural difference and cultural changes has obtained fruitful results. Finally, there are limitations of validity, technique, and theory in current research. Future research should pay attention to sampling, tests of indicators' validity, quasi-causal analysis, and data driven analysis, etc. to improve the validity of methods, the depth of results, and the diversity of theories.

Key words: Internet data; big data; cultural psychology; cultural change; cultural difference