

网络大数据分析技术的 心理学方法论思考

赖凯声¹,马华维²,乐国安³

(1. 中山大学 传播与设计学院,广东 广州 510006;2. 天津师范大学 心理学系,天津市 300387;
3. 南开大学 社会心理学系,天津市 300350)

摘要:网络大数据已被成功应用于探索情绪心理学、人格心理学等诸多心理学研究议题。与以认知神经科学技术为例的现代研究方法,和以问卷法、行为实验法为例的传统研究方法相比,网络大数据分析在样本规模、数据客观性、时效性、生态学效度等方面均具有显著优势。具有浓厚学科交叉性质的认知神经科学技术和网络大数据分析技术,前者擅长于微观分析层面,后者擅长于分析宏观层面,且都是心理学适应现代化技术变迁、把握时代机遇的两大重要突破口。未来研究者可采取方法互补,数据驱动和理论驱动相结合等策略,从而更好地把握时代发展机遇。

关键词:网络大数据;心理学方法论;认知神经科学;因果性;社交媒体

中图分类号:B841 **文献标识码:**A **文章编号:**1673-9841(2017)03-0081-07

现代化的信息技术革命,使得人类步入了数据充裕的大数据时代。在以海量大数据为核心特征的信息化时代里,数据的重要性被形象地比喻为“石油”。通常认为,大数据“是一种无法在一定时间内用常规机器和硬件工具对其进行感知、获取、管理、处理和服务的数据集合”,并具有数据量大(Volume)、处理速度快(Velocity)和多样性高(Variety)以及价值大但密度低(Value)等特征^[1]。目前,大数据已经被广泛地应用于诸多领域的社会实践,包括金融交易(例如,情绪对冲基金)、社会治安、公共管理、政治选举、交通管理、气象监测、医疗卫生、企业战略等^[2]。

一、网络大数据概述

与基因、脑科学等产生的微观大数据有所不同,网络大数据主要指互联网平台产生的相对宏观的网络数据,并已成为大数据研究的重要组成部分。基于网络大数据分析技术开展的社会科学研究,对于把握政治、经济以及社会心理与行为规律具有重要意义^[3]。网络大数据分析技术主要具有以下优势:(1)样本规模和代表性:依托于网络平台的优势,可以实现覆盖大规模群体的测量,这种样本不断接近总体的特征和优势有利于解决传统研究方法代表性的问题。(2)时效性:网络大数据

收稿日期:2016-12-18

作者简介:赖凯声,中山大学传播与设计学院大数据传播实验室,副研究员。

通讯作者:马华维,天津师范大学心理学系,教授。

基金项目:国家社会科学基金重大项目“基于大规模网络实际测量的个体与群体行为影响分析研究”(12&ZD218),项目负责人:乐国安;教育部哲学社会科学重大课题攻关项目“大数据时代国家意识形态安全风险与防范体系建构研究”(16JZD006),项目负责人:乐国安;中国心理学会创新和服务能力提升工程优秀科技社团建设项目“‘互联网+’大战略中的心理学应用”,项目负责人:乐国安。

使得对大规模群体定期,甚至是实时地追踪测量成为可能。其追踪时间可以精细至每年、月、日、小时甚至每分每秒。(3)客观性:网络大数据基于网民客观的行为数据,例如搜索引擎的搜索和点击行为,社交网络的点赞、转发以及发帖内容,具备较好的证据客观性。(4)成本经济性:传统研究方法往往受限于人力、财力等研究成本,无法较好实现对大规模群体进行定期、实时测量,而网络大数据分析以网络爬虫、文本分析等技术支撑,使得以相对较低的成本获取海量数据不断成为可能^[4]。

基于海量大数据分析的研究方法引发了人们对于科学方法论的思考。研究无需直接接触研究对象,而通过直接分析和挖掘海量数据便可获得新的研究发现,这或许催生了一种新的科研模式^[1]。为此,图灵奖得主 Gray 把数据密集型科学从计算科学中区别出来,并描绘了基于数据密集型科学研究的“第四范式”^[5]。但也有研究者对大数据分析技术存在质疑,认为大数据并不能取代原有的研究方法。可见,我们关于如何看待网络大数据分析技术尚存在争议。

心理学也是网络大数据与社会科学结合的重要领域之一。^[2,6]目前网络大数据分析方法已经被较为广泛地应用于解决情绪心理学^[7-10]、行为经济学^[11-12]、人格心理学^[13-14]、健康心理学^[15]、政治心理学^[16]等诸多重要心理学议题。那么结合其他心理学实证方法来看,应当如何看待基于网络大数据分析技术的心理学研究?遗憾的是,目前关于网络大数据心理学方法论视角的系统性思考仍然非常少。

二、与其他心理学实证研究方法的对比

为阐述网络大数据分析技术与其他心理学实证研究方法的区别和联系,笔者将分别从现代研究技术和传统研究技术两大方面展开。其中,现代研究技术方面,将主要以认知神经科学技术为例;传统研究技术方面,将主要以问卷法和实验法为例。

(一)现代研究技术:认知神经科学技术

与现代研究技术的比较,尤以上世纪八九十年代兴起、由认知科学与神经科学结合形成的认知神经科学技术最为典型。同样处于迅猛发展期的认知神经科学研究,也引发了心理学研究者的广泛关注以及基于心理学发展视角的思考^[17-18]。认知神经科学研究,主要借助现代化的认知神经科学研究技术手段,例如脑电波(EEG)、功能性磁共振成像(fMRI)、脑磁(MEG)、经颅磁刺激(TMS)等技术,来揭示心理与行为的发生和发展规律。经过短短三十多年的发展,认知神经科学便成为当前最热的交叉研究领域之一。其研究问题主要包括认知行为的神经机制研究、认知行为心理学理论的神经科学研究以及基于脑神经激励构建认知行为心理机制理论模型的研究三大层面^[17]。例如,探索个体的感知觉、学习记忆、注意、言语、执行控制、思维、情绪等心理活动或行为的神经机制研究。

认知神经科学技术的优势主要包括:(1)在研究技术手段上:测量方法更加精细,研究证据的科学性更强;(2)在研究问题的层面上:可实现在分子、细胞、神经环路、脑机能系统等不同水平上揭示人类内部心理活动的“黑箱”,弥补行为实验只能测量外显行为的不足。但认知神经科学技术在分析水平、样本规模等方面也存在一些不足,相比之下网络大数据方法则可以在一定程度上弥补这些不足。具体来说:(1)在研究分析水平上,现代认知心理学主要侧重于人的内部心理过程。在当时行为主义占据统治地位的历史背景下,现代认知心理学在研究对象上的回归对学科发展具有重要的进步意义。但其主要建立在一套把人类比为计算机的信息加工理论基础,把人的活动作简单化、形式化和符号化的模拟。因而这种模拟存在种种局限,并不能完全说明人的全部心理活动。网络大数据则关注的是宏观层面的群体心理与行为规律。网络大数据可以借助众多数据来源渠道,充分考虑可能影响人类心理与行为的各种因素,包括情境、文化等因素。这可在一定程度上缓和认知神经科学技术关注于个体内部神经生理机制所面临的“还原论”风险。(2)在样本规模上,认知神经科学技术研究往往受研究成本等因素的限制,难以开展大规模样本的认知神经科学研究。fMRI

等仪器设备配备、使用以及维护等的成本皆不菲,因而大多数认知神经科学研究的样本量往往都非常小。小样本则带来了样本代表性以及结果可推广性等方面的风险。而网络大数据则在以较低研究成本获取大规模样本上具有较强优势。例如,Aral 和 Walker 开展的一项探索社交网络平台上究竟谁更容易受他人影响问题的研究,其分析的 Facebook 用户量高达 1 200 万^[19]。

虽然认知神经技术和网络大数据分析技术研究存在以上差异,但二者还是具有一定的共性。二者的产生和发展都是由现代化科学技术进步所推动。例如,计算机等信息科学技术手段在数据收集、存储以及分析方面的应用,是现代认知神经科学研究、网络大数据研究得以迅猛发展的重要基础。立足于微观个体内分析水平的现代认知神经技术,与立足于宏观个体外分析水平的网络大数据分析技术,共同推进了信息化时代背景下的心理学实证研究进展。

(二)传统研究技术:问卷法和行为实验法

1. 与问卷法的对比

问卷法是通过被访者针对一系列自陈式问卷题目采用自我报告的方式进行回答,并以此作为研究人们心理与行为规律的数据证据。该方法的主要优势有:(1)针对性强。通过设计与研究目的直接相关的结构化、标准化问题或者开放式问题,有针对性地搜集被访者关于该问题的一手数据。(2)快速获取较大人群的资料:问卷法通过标准化的操作流程,可以较为快速地获取较大规模人群的数据资料。与认知神经科学技术高昂的研究成本相比,问卷调查的成本相对较低,样本覆盖面也相对较大。

但问卷调查法同时在主观性偏差、样本规模、时效性等方面存在一定缺陷,而网络大数据分析技术则在这些方面相对有较大优势。(1)客观性差:问卷调查法由于采用自我报告方式,因此研究结果存在较强的主观性偏差,尤其以社会赞许性反应最为典型。而网络大数据分析技术依托于用户在互联网平台留下的客观行为数据,该证据具有较强的客观性,可以有效地弥补问卷调查在主观性偏差方面的不足。已有研究也有直接对比问卷调查和网络大数据的研究,结果也发现了一些有趣的现象。例如,Wojcik 等发表在《科学》杂志上的一项研究发现,美国保守主义者在问卷调查中报告了比自由主义者更高的主观幸福感,但从 Twitter 等社交媒体的客观数据来看,自由主义者却显著地比保守主义者表现出了更高的幸福感^[16]。另一项关于网络搜索的研究证据也显示了网络大数据在客观性上的优势。基于常识和自我报告等证据,人们普遍认为美国的保守主义往往与性开放度是负相关的。但 Maclnnis 和 Hodson 借助 Google 搜索引擎上与关于性相关内容的搜索数据却发现,对于保守主义倾向越强的州,该地区网民在匿名化的网络平台上对性相关内容的搜索量反而越高^[20]。(2)样本规模不够大:尽管问卷调查法在成本上相对认知神经研究方法具有较强优势,但如果涉及大规模人群的调查,在调查员培训,实际调查问卷发放、回收等流程上仍然成本高昂。例如,中国综合社会调查(Chinese General Social Survey, CGSS)作为目前中国最具代表性的全国性调查之一,其样本量也尚未超过 2 万人,且每次大调查均耗资巨大。而基于 Twitter、Facebook、微博等网络大数据的研究,则可较为轻松地实现百万级别以上样本规模的研究。例如,Bond 和 Messing 发表的一项尝试探索基于社交网络数据预测用户政治意识形态的研究,其 Facebook 用户样本量多达 600 万人^[21]。(3)时效性低:问卷调查由于成本和操作方法本身等因素限制,难以实现大样本的追踪测量。而网络大数据借助网络爬虫、计算机自动化分析等技术手段,可以实时、高效地进行测量。例如,Mitchell 等基于美国用户带有地理信息标签的 Twitter 数据,实现了对美国各地区的幸福感测量,并得到了美国幸福感分布的可视化地图^[22]。基于该技术原理,研究者建立了“Hedonometer”项目,实现了对美国各州、城市幸福感的实时测量。

2. 与行为实验法的对比

行为实验法主要由研究者通过人工操纵或设计不同的实验条件,并观察被试在不同实验条件下的行为结果差异,从而检验该实验条件是否对结果有显著的影响。该方法的主要优势有:(1)可

探讨因果性。由于实验法控制了其他无关干扰因素的影响,因此可以较为可信地判断行为结果的差异是否由实验条件(即自变量)造成,从而进行因果性推论。鉴于因果性在研究中的重要意义,实验法在科学方法论中也占据着重要的地位。(2)可重复性、可检验性。实验法的实验条件是由研究者预先设计的,因而研究者具有较强的主动性和控制性。尤其是行为实验都有着严格的实验设计要求和实施流程,关于实验的被试选取、所有指标或变量的测量工具和方法、具体的操作流程等也有详细的披露。因而实验法具有良好的可重复性和可检验性。

但实验法在生态学效度、样本规模等方面也存在一些不足,网络大数据则可以较好地弥补这些不足。(1)生态学效度上:在行为实验中,研究者的人为实验条件干预以及实验室特定的环境空间,都可能导致实验情境与被试的实际环境有所差异,从而对实验结果产生干扰。因此,行为实验法普遍面临生态学效度低的挑战。自然实验虽然可以提高环境的真实性,但在操作可行性、干扰变量控制等方面又存在挑战。而网络大数据则以用户真实的网络情境为背景,通过不介入用户行为活动的方式搜集客观数据,因而具有较高的生态效度。(2)样本规模小,代表性有限:行为实验受实验室客观环境、实验人员配备等因素限制,也难以开展涉及大规模人群的行为实验。而网络大数据样本庞大,能有效地解决样本规模小带来的代表性问题。例如,传统心理学采用实验室研究或自我报告等方法开展的个体情绪在日周期水平上的波动节律研究,一直未能得到较为一致的结果。这可能与研究样本存在抽样偏差,追踪测量存在较大误差等问题有关。为此,Golder 与其合作者 Macy 则充分发挥了 Twitter 在样本规模、数据客观性、长期追踪测量成本等方面的优势,抓取了来自全球 84 个国家 240 多万用户的 Twitter 数据,直接针对情绪的时间节律规律进行了跨文化、大样本的实证研究^[7]。

三、问题和挑战

(一)因果性问题

基于海量数据的大数据分析技术在挖掘相关关系方面具有得天独厚的优势,但是却面临难以解释变量之间因果关系的困境。一些大数据研究者认为,大数据研究擅长于探索相关性,因而可以放弃对因果性的寻求,即遵循“相关而不是因果”原则^[23]。但因果性作为社会科学的圣杯,在社会科学、自然科学界都具有举足轻重的地位。因此上述观点一直饱受争议。例如,Silver 认为,在数据混杂的大数据预测研究中,如果能以理论或关于事物根本原因的深入思考做支撑,这对于提高其统计推理的可靠性具有重要意义^[24]。谢志刚认为,大数据带来的“相关性和因果性”的争论,实质上是哲学认识论中的经验论和唯理论之争,并且可以追溯到 19 世纪中后期奥地利学派和德国历史学派之间的方法论之争^[25]。前者认为应该坚持以逻辑演绎为基本分析方法,而后者则主张通过搜集大量数据证据,从历史数据中发现经验规律,即关于演绎与归纳的方法论之争。李国杰和程学旗认为,对基于小数据的简单封闭的系统开展因果分析容易做到,但基于大数据开放复杂的巨系统,则可能面临互为因果、因果相互纠缠等困境,因而对相关性和因果性的关系不宜贸然下结论^[1]。王天思则从哲学视角阐述了大数据中的因果关系及其哲学内涵,认为“相关关系意味着对事物过程的定量描述模型,而因果关系则意味着对事物过程的定性描述模型,二者都不可或缺。以相关关系取代因果关系是短视;而以因果关系排斥相关关系则是固步”^[26]。综上,笔者认为相关关系和因果关系是揭示变量关系的不同层面。在没有充分证据彻底否认因果关系研究意义之前,不能抛弃对因果规律的探索。

从方法论的视角而言,网络大数据研究方法在解释因果机制上存在不足,但可以尝试与实验法结合等方式予以弥补。例如,人们的情绪是否会受到他人影响?尤其是在网络上人们仅仅依靠非言语线索进行交流互动的虚拟环境下又将如何?为此,Kramer 等人结合实验法的设计思路,对近 69 万 Facebook 用户接受到的动态信息进行了实验操纵:一组以积极情感为主,而另一组则以消极

情感为主^[27]。实验结果发现,人们的确会无意识地体验到与他人相同的情绪状态,即通过大规模的网络样本验证了情绪传染效应。当然,如何能够更好地解决网络大数据的因果解释机制问题,仍有待进一步探索。

(二)数据质量问题

数据量大毫无疑问是网络大数据分析技术的一大优势,但如何保证海量数据的质量,以及如何实现对海量数据进行清洗、管理和分析等问题,也成为网络大数据研究的一大技术难点。海量的网络大数据具有多源异构、交互性、时效性、突发性和高噪声等特点^[1],因而导致了网络大数据虽然价值巨大但噪音也大,价值密度低的特征。这对保证网络大数据分析研究中的数据质量则构成了巨大挑战。在问卷调查法中,正是凭借数据质量高而并不是样本规模大的特征,早期的盖洛普公司在与其他民意调查机构的竞争中脱颖而出。换言之,数据量大并不能一定保证数据质量高,因此也无法保证能够挖掘到更可信、更有价值的结果。因此探索如何在网络大数据研究中提高海量数据的质量,尽可能去除冗余噪音的干扰,是网络大数据研究亟需解决的一大难题和挑战。

(三)隐私问题

隐私问题也成为当前网络大数据研究面临的重大挑战之一。例如,2013年曝光的“棱镜门”事件激发了社会各界对于数据安全和数据隐私问题的广泛关注和疑虑。也有商业公司曾被指控利用搜集的用户个人信息,用于网络营销、购物推荐等商业目的相关的活动。甚至有关于网络大数据的学术研究也曾因隐私问题,一度陷入用户隐私问题的大争论中。例如,2014年Kramer等发表于《美国科学院院刊》的一项关于Facebook平台上69万用户情绪传染效应的实验研究^[27]。由于研究者人为的改变了用户接收到的动态消息,即按照实验处理的原则操纵为以积极情感为主的消息和以消极情感为主的消息,结果遭到了操纵网络信息、侵犯用户隐私的批评和质疑。

事实上,从方法论的角度而言,任何一种研究方法在数据收集的过程中都可能涉及到用户的隐私,甚至还可能涉及到其他一些伦理道德问题,例如采用实验法的“斯坦福监狱实验”。从其他研究方法解决隐私、伦理道德问题的经验来看,可尝试采取建立和完善伦理道德规范(例如,签署知情同意书),健全监督机制等手段予以解决。网络大数据研究亦可如此,相关研究者可逐步建立、健全数据使用规范,并倡导在充分保证用户个人隐私的条件下开展有价值的学术研究。例如,对用户信息进行加密、匿名化处理等方式,以更好地解决数据隐私困境^[2]。

四、未来研究展望

因果解释困境同样存在于认知神经科学研究领域,但却不足以撼动它作为“21世纪最有发展前景的自然科学前沿研究领域”的地位^[17]。网络大数据也是借助现代化科学技术手段进步发展起来的新兴研究领域,有着巨大的发展潜力和广阔的成长空间。从心理学方法论的视角来看,认知神经科学技术和网络大数据分析技术,前者擅长于挖掘微观层面的心理与行为规律,后者则更加擅长于探索宏观层面的规律,并且二者都有着浓厚的学科交叉性质。这无论是对于心理学研究向内的深化,还是向外的拓展和延伸,都有着重要而深远的学科内涵。因此,笔者认为认知神经科学技术和网络大数据技术都是心理学适应现代化技术变迁、把握时代机遇的两大重要突破口。认知神经科学技术在心理学中的地位和意义已经得到了广泛的重视,但相比之下网络大数据分析技术的认识和重视则相对较少。国内一些研究者已经尝试从学科体系建构^[28]、中国人社会心理与行为规律挖掘^[2,9,14]等角度积极探索了如何将网络大数据应用于中国心理学发展中。在未来研究方面,研究者应该理性看待网络大数据研究技术,客观地认识其作为一种研究方法的优势和不足。具体而言,有以下两点发展建议:

(一)跨方法优势互补

未来研究中,一方面可以结合具体的研究问题需要,考量研究主题和方法的契合性,以最大限

度地发挥该研究方法的优势而同时降低其劣势带来的风险。另一方面,在研究条件允许的情况下,研究者可尽可能地考虑多种方法相结合实现跨方法优势互补,从而增强研究的科学性。例如,目前网络大数据分析的技术、模型也被尝试应用于解决认知神经科学问题。针对海量的基因、神经生理大数据,机器学习等网络大数据分析方法和算法被直接应用于脑神经科学等心理学领域^[29]。关于问卷法与网络大数据的结合,例如开展大规模的网络问卷调查研究,近年来也日益受到研究者青睐。例如,Rentfrow, Gosling 和 Potter 基于美国 60 多万居民的网络人格问卷调查数据开展了美国州水平的人格差异研究^[30]; Rentfrow, Jokela 和 Lamb 基于英国近 40 万居民、覆盖 380 个地区的人格网络调查数据开展了英国地区人格差异研究^[31]。在实验法与网络大数据相结合方面,有例如 Kramer 等基于 Facebook 用户开展的情绪传染实验研究^[27]。

(二)数据驱动与理论驱动相结合

数据是网络大数据分析方法优势的集中体现,数据驱动也因此成为网络大数据分析的重要特征之一。传统心理学实证研究的思路往往是从已有的理论和研究证据出发,推导出相对明确的研究假设,进而通过调查或者实验等方法搜集研究证据对假设进行证实或者证伪,最终形成研究结论。而数据驱动下的网络大数据分析,则可通过直接对获取的海量网络数据进行整理和分析,进而挖掘出具有价值的信息。因此,在数据驱动的研究过程中可能并没有明确的研究假设,这可以有效地降低研究者先验经验对研究的束缚。该优势在探索新领域和新议题上表现得尤为明显。例如,群体动力学属于心理学的重要研究领域之一,但传统研究方法难以实现对大规模人群进行的追踪测量,因而该领域长期处于直接的心理实证研究相对缺乏的状态。Lu 和 Brelsford 则以日本地震为例,基于 Twitter 网络大数据分析人们在突发事件下的社交网络结构及其演化规律^[32]。

事实上,理论驱动也能在网络大数据研究中发挥重要的作用。从现有的一些研究成果来看,不乏一些直接受经典理论启发或者指导的网络大数据研究。例如人类学家邓巴提出的邓巴定律认为,由于智力的限制,允许人类拥有稳定社交网络的人数是 148 人。那么在互联网平台上,人们的社交网络结构是否仍然符合邓巴定律呢?为此,Gonçalves 等人基于 170 多万 Twitter 用户数据的研究结果发现:在网络社交中稳定朋友关系上限也在 100~200 之间,与邓巴数字基本一致^[33]。

因此,未来的网络大数据研究需要处理好数据和理论之间的关系。一方面,数据驱动证据可以验证或修正既有理论,并且随着数据驱动证据的不断积累,也有望进一步提炼出具有创新性的理论。而另一方面,新的理论又可以进一步为后续的实证研究提供指导。数据驱动与理论驱动相结合,有利于实现理论和数据相互促进的良性发展。

本文为“广州大数据与公共传播研究”人文社科重点研究基地成果之一。

参考文献:

- [1] 李国杰,程学旗. 大数据研究:未来科技及经济社会发展的重大战略领域——大数据的研究现状与科学思考[J]. 中国科学院院刊,2012,27(6):647-657.
- [2] 乐国安,赖凯声. 基于网络大数据的社会心理学研究进展[J]. 苏州大学学报(教育科学版),2016,4(1):1-11.
- [3] 陈浩,乐国安,李萌,等. 计算社会科学:社会科学与信息科学的共同机遇[J]. 西南大学学报(社会科学版),2013,39(3):87-93.
- [4] 朱廷劭,汪静莹,赵楠,等. 论大数据时代的心理学研究变革[J]. 新疆师范大学学报(哲学社会科学版),2015,36(4):100-107.
- [5] HEY T, TANSLEY S, TOLLE K M. The fourth paradigm: data-intensive scientific discovery[M]. Redmond, WA: Microsoft Research, 2009.
- [6] 薛婷,陈浩,赖凯声,等. 心理信息学:网络信息时代下的心理学新发展[J]. 心理科学进展,2015,23(2):325-337.
- [7] GOLDER S A, MACY M W. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures[J]. Science, 2011, 333(6051): 1878-1881.
- [8] 乐国安,董颖红,陈浩,等. 在线文本情感分析技术及应用[J]. 心理科学进展,2013,21(10):1711-1719.
- [9] 叶勇豪,许燕,朱一杰,等. 网民对“人祸”事件的道德情绪特点——基于微博大数据研究[J]. 心理学报,2016,48(3):290-304.
- [10] 董颖红,陈浩,赖凯声,等. 微博客基本社会情绪的测量及效度检验[J]. 心理科学,2015,38(5):521-528.

- [11] BOLLEN J, MAO H, ZENG X. Twitter mood predicts the stock market[J]. *Journal of Computational Science*, 2011, 2(1): 1-8.
- [12] 赖凯声,陈浩,钱卫宁,等. 微博情绪与中国股市:基于协整分析[J]. *系统科学与数学*,2014,34(5):565-575.
- [13] WU Y, KOSINSKI M, STILLWELL D. Computer-based personality judgments are more accurate than those made by humans [J]. *Proceedings of the National Academy of Sciences*, 2015, 112(4): 1036-1040.
- [14] LI L, LI A, HAO B, et al. Predicting active users' personality based on micro-blogging behaviors[J]. *Plos One*, 2014, 9(1): e84997.
- [15] EICHSTAEDT J C, SCHWARTZ H A, KERN M L, et al. Psychological language on Twitter predicts county-level heart disease mortality[J]. *Psychological Science*, 2015, 26(2): 159-169.
- [16] WOJCIK S P, HOVASAPIAN A, GRAHAM J, et al. Conservatives report, but liberals display, greater happiness[J]. *Science*, 2015, 347(6227): 1243-1246.
- [17] 张卫东,李其维. 认知神经科学对心理学的研究贡献—主要来自我国心理学界的重要研究工作述评[J]. *华东师范大学学报(教育科学版)*,2007,25(1):46-55.
- [18] 李其维. “认知革命”与“第二代认知科学”刍议[J]. *心理学报*,2008,40(12):1306-1327.
- [19] ARAL S, WALKER D. Identifying influential and susceptible members of social networks[J]. *Science*, 2012, 337(6092): 337-341.
- [20] MACINNIS C C, HODSON G. Do American States with more religious or conservative populations search more for sexual content on Google? [J]. *Archives of Sexual Behavior*, 2015, 44(1): 137-147.
- [21] BOND R, MESSING S. Quantifying social media's political space: estimating ideology from publicly revealed preferences on Facebook[J]. *American Political Science Review*, 2015, 109(1): 62-78.
- [22] MITCHELL L, FRANK M R, HARRIS K D, et al. The geography of happiness: connecting twitter sentiment and expression, demographics, and objective characteristics of place[J]. *Plos One*, 2013, 8(5): e64417.
- [23] MAYER-SCHÖNBERGER V, CUKIER K. Big data: a revolution that will transform how we live, work, and think[M]. Boston: Houghton Mifflin Harcourt, 2013.
- [24] SILVER N. The signal and the noise: why so many predictions fail-but some don't[M]. New York: Penguin, 2012.
- [25] 谢志刚. 大数据再掀经济学方法论之争[N]. *中国社会科学报*,2015-09-17(2).
- [26] 王天思. 大数据中的因果关系及其哲学内涵[J]. *中国社会科学*,2016(5):22-42.
- [27] KRAMER A D, GUILLORY J E, HANCOCK J T. Experimental evidence of massive-scale emotional contagion through social networks[J]. *Proceedings of the National Academy of Sciences*, 2014, 111(24): 8788-8790.
- [28] 喻丰,彭凯平,郑先隽. 大数据背景下的心理学:中国心理学的学科体系重构及特征[J]. *科学通报*,2015,60(5):520-533.
- [29] YARKONI T. Psychoinformatics: new horizons at the interface of the psychological and computing sciences[J]. *Current Directions in Psychological Science*, 2012, 21(6): 391-397.
- [30] RENTFROW P J, GOSLING S D, POTTER J. A theory of the emergence, persistence, and expression of geographic variation in psychological characteristics[J]. *Perspectives on Psychological Science*, 2008, 3(5): 339-369.
- [31] RENTFROW P J, JOKELA M, LAMB M E. regional personality differences in great britain[J]. *Plos One*, 2015, 10(3): e0122245.
- [32] LU X, BRELSFORD C. Network structure and community evolution on Twitter: human behavior change in response to the 2011 Japanese earthquake and tsunami[J]. *Scientific Reports*, 2014, 4: 6773.
- [33] GONÇALVES B, PERRA N, VESPIGNANI A. Modeling users' activity on Twitter networks: validation of Dunbar's number[J]. *Plos One*, 2011, 6(8): e22656.

责任编辑 曹莉

网 址:<http://xbbjb.swu.edu.cn>