

**编者按:**“大数据”是近年国内外学界、业界所普遍关注的一个学术热点,据不完全统计,在2013年一年里,国内外以“大数据”为主题发表的学术论文总共超过3万余篇,而2012年尚不足千余篇,之前则更少,几乎微乎其微。

作为一个源自理工学科的专有术语,“大数据”已经随着4G网络技术、安卓4.0平台与交互性个人通讯终端的普及,逐渐成为广为大众所知且与大众日常生活息息相关的热门名词,并日渐影响到社会科学如新闻传播学、社会学与经济学等学科的研究路径与范式。然而,“大数据”与人文学科研究的关系研究却仍属空白。

本文由中国科学院博士后、中国作家协会会员韩晗在武汉大学120周年校庆“珞珈论坛”上的学术讲稿删改而成。作为一位年轻的80后学者,韩晗从“大数据”与人文科学研究转向这一富有挑战性、前沿性的话题入手,前瞻性地预测了未来人文科学研究在“大数据”语境下可能呈现出的嬗变趋势,因此,具有一定的启发意义。

值得一提的是,演讲稿由韩晗在“新浪微博”上以“长微博”的形式发表之后,在短短的5天时间里,被转发数百次,浏览近10万次。北京大学杨伯淑教授、武汉大学陈国恩教授等知名学者均积极参与了对该问题的探讨,产生了良好的社会反响,再度见证了“大数据”对人文学科研究的影响与意义。

基于如上考虑,本刊特编发此稿,以饕学界诸君,并期待与学界同仁争鸣讨论。

## 论“大数据”与人文研究的转向

韩 晗

(中国科学院自然科学史研究所,北京 100034)

**摘 要** 作为未来人类社会组织形态的“大数据”,导致人类社会呈现出了“数据化”的特征。这迫使人文研究的对象、方式都发生了新的转向,即对真实性的诉求。因此,在“大数据”的语境下,未来的人文研究或会呈现出及时性、务实性与多元性等新的特征。

**关键词** 大数据;人文研究;转向

中图分类号:G913

文献标识码:A

文章编号:1000-2987(2014)03-0022-04

问题,什么是大数据?

我相信,这个名词对于计算机专业的人来说并不陌生,大数据就是瞬息万变的海量数据,这个数据是客观存在的,但它又不确定。就像现在这间教室,不断有同学从正门进来,也不断有同学从后面出去,

“大数据”(Big data)是我近几年来一直关注的一个问题,也是本次讲座的核心概念。要谈“大数据”与人文研究的转向,则首先必须弄清楚一个关键

收稿日期 2013-12-11

**作者简介** 韩 晗(1985-),男,河北衡水人,中国科学院自然科学史研究所博士后,中国作家协会会员,武汉大学文艺生产与消费研究中心特邀研究员,四川大学符号传媒研究所兼职研究员,研究方向为现代科技文化思想史。

本文为作者在武汉大学120周年校庆“珞珈论坛”上的学术演讲,囿于篇幅限制,全文有删改。此外,感谢北京大学创意产业研究中心主任杨伯淑教授与武汉大学中文系主任陈国恩教授对该演讲稿成文时的建议,作者铭感至深,特此致谢。

教室里整体的人数是不固定的,好似一个小的世界。尽管不断变化,但这些数据又客观存在,这就是广义上的大数据。在这个人口增长、信息爆炸的年代里,大数据成为一个不可或缺的存在,已经受到学界的关注。

然而,目前学界对于大数据的研究,仍处于比较落后的阶段,关键在于许多学者仍然未将其与人文科学结合起来,始终还在工商管理、计算机与统计学领域内进行讨论,只有少数学者将其与新闻传播、社会学进行有限度的结合。甚至还有学者认为,大数据本身就是一个理工科的概念,与人文科学无关。

这就好比是20年前,一说到电脑、互联网,那必定是计算机业内的专有名词,是与Qbasic、C+语言、源代码息息相关的概念。而在今天这个计算机、互联网甚至是“云时代”里,我相信仍然会有很多整天拿着手机刷屏的朋友不懂什么是编程。所以说,现在不会有人再认为,计算机、互联网只是理工科的人才需了解的专有名词。

“大数据”也一样,在未来不远的日子里,我们每个人都是它的受益者或受害者。在我看来,“大数据”就是社会未来的组织形态,即数据化(datazation),这是我生造出的一个新单词。在这个人手一部手机、时时可以上网的年代,每个人既时常将他者数据化,也是被数据化的对象。

著名诗人卞之琳在《断章》里写道:“你站在桥上看风景,看风景的人在窗外看你。”其实,“大数据”也一样。我们一边收集数据,但在许多人看来,我们也是数据之一。因为这是一个把一切人、事与社会关系“数据化”的时代。

我们的身份证、银行卡、公交卡,乃至理发店、图书馆、健身房办的卡,都是数据卡。大家在里面都是一条数据,姓甚名谁、哪年出生、是男是女、家住何处,只要输入身份证或其他证件号码,一切一览无遗。通过网上购物,商家就能采集到购买者的家庭情况,是否有婴儿、是否有宠物、是否有车、消费能力如何等等。这样看来,“大数据”是不是只和理工专业有关?

现在有个观点,“大数据”就是在互联网上的总体数据,这个数据当然很大,而且在不断变换,所以很多研究大数据的人不知道该怎么办。因为任何人都不能穷尽网络时代的所有资源。事实上,这只是狭义的“大数据”概念。从广义上讲,大数据是一种社

会的组织形态,研究者面对的并非浩如烟海的数据,而是由这个数据所组织起来的社会结构。譬如,当年“甲流”泛滥时,国际卫生机构利用谷歌搜索所提供的“客户检索关键词”,迅速地在全球范围内第一时间定位了流感的发作情况,这就是社会形态的数据化。

## 二

下面我谈谈大数据与人文科学的关系问题。这几年我一直在研究中国当代文学的接受问题,但我并没有按照传统意义上的作家作品或思潮来归纳总结,而是从出版传播学的角度来切入。譬如就“莫言影响力大”这个命题而言,我所做的研究是:莫言的影响力究竟有多大?当然,这是需要一个数据来说话的,比如他的书在图书馆借阅量如何、他的名字在报纸上呈现次数如何、他的书在书店销量如何、他的言论被各种文章引用率如何,甚至他的名字在微博中出现的次数如何,等等。凡此种种,都是证明他影响力大小的工具。不能笼统地说莫言获得了诺贝尔奖,其影响力就大了,而没得奖就影响力小了。人文研究也是科学研究,科学研究的核心就是真实,真实来源于客观存在的数据。

我认为,青年学者在从事人文研究时,很容易犯一个错误:空泛。谈问题都是后现代、新史学、结构主义等宏大叙事的空泛概念,“巨大影响”、“绝大多数”等名词时常被使用,绝大多数是多大?巨大又是多大?

所以我认为,在大数据时代,人文研究必然会发生新的转向。人文研究要与大数据相结合,而大数据又依赖于社会学、统计学而存在。美国加州大学有一个叫爱蒙德·琼斯的青年学者,最近在做一个项目,就是利用大数据来进行人文研究。他利用电脑搜索引擎、统计软件以及发放网络问卷的形式,搜集到了美国近一大半公墓管理部门的材料,旨在解答一个问题:当下美国人怎样举办葬礼?答案当然千奇百怪。研究者把这个庞大的数据扔进了计算机软件,进行分类研究,最终化零为整,形成若干条结论,综合地反映了这个时代美国人对于生死的看法。

与此相对应的是,日本有一位名叫田中诚一的美术学学者,他并不拘泥于简单的画家、画作或思潮,以及时下最时髦的市场研究,而是找了4家美术学院的校医院,调出了50多万份电子档案,统计近3

年来美术学院学生与老师最容易患上的疾病,然后利用庞大的数据,最终指向一个学术问题:疾病与美术创作的关系。

我再举一个例子,伽达默尔、罗伯特·姚斯的接受美学是文艺学里的一个重要范畴,但接受美学如何实践?最近湖南卫视有档热播的节目,叫《爸爸去哪儿》,这个节目是模式引进的,节目引进就是把国外的电视模式搬到中国来,10年前的“超级女声”就是开先河之举,目前在国内电视节目制作领域内已经形成了一种趋势。近年来,“模式引进”已经成为各大电视台、媒体运营公司获得收视率的不二法门,进而成为广播电视艺术学研究所关注的热门问题。

实际上,这个“模式引进”在今天能够这样火爆、深入人心,就与对“大数据”的利用分不开。事实证明,负责模式引进的公司与媒体基本上不依托传统的收视率来获取信息,而是直接向百度、谷歌、PPTV、土豆网、微信与新浪微博购买实时的监控数据,看不同的网民如何评价这个节目,然后他们会根据市场与网民的反映对下一期节目作调整,这已经完全超越了先前“收视率”这个简单的概念。所以说,关注这个问题,就是接受美学最好的实践性研究。

由此可知,我们现在面对的整个社会已经“数据化”了,人文研究的对象是人、是社会,不从这样海量的大数据中获得信息,就会无路可走。所以我认为,人文研究一定要与社会学、统计学甚至计算机学科的知识相结合。只有这样,研究结果才会更可能地接近真实并富有生命力。

### 三

“大数据”语境下人文研究即将呈现的几个特征是我所要谈的另一个问题,这也是人文研究转向的重要标志。一直以来,我有个观点,“现代文学以史为主,当代文学以论为主,文学传播以证为主。”后者这个“证”,就是对于“大数据”实证研究的诉求。

第一个特征是“及时性”,这也是大数据的特征之一。众所周知,大数据源于发达的互联网技术,因此它具备互联网时代的某些特征。前面说,现在许多媒介机构用大数据来监控收视率,从而研究一部电视剧或一个电视节目的接受反馈情况,这就是对于及时性的诉求。现在社会瞬息万变,搞人文研究不能还坚守着几年甚至几十年前的老方法,必须要及时捕捉人在特定历史时代里的变迁,因为其关注研究

的对象是人。

这里我需要说明一下,人文研究对及时性的诉求并不代表急功近利或是浮躁,面对宏大的数据将其量化分析恰需要坐冷板凳的功夫。这是一个从量化到质化的过程,需要研究者有面对社会、“接地气”的研究立场。由此可知,从大数据与社会变革的时效性出发,将“及时性”融入到人文研究中并不是一个简单的事情。它需要研究者从跨学科的角度,面对宏观的时代背景与微观的海量数据,最终从定量到定性进行条分缕析的研究。

第二个特征是“务实性”。近十几年来,人文研究因为对西方理论的盲目依赖而患上了“失语症”。许多初出茅庐的研究者在研究论文中喜欢使用一些概念内涵不清甚至语法有误的语汇,譬如“话语漩涡”、“叙述杠杆”、“语言能量场”等等。全文通篇都是这样的话语,连许多资深教授都读不懂作者的想法。我对于这样的研究文字就4个字评价:人文气功。

所以说,不只是自然科学,人文科学里也有伪科学。所谓人文伪科学,就是通篇理论呓语,缺乏务实性的研究,而务实性研究则是真科学的标志。人文研究是必须要与社会对接的,现在的社会就是一个“大数据”的社会,要想获得信息,就必须对所处的社会——即自身的研究对象有清醒的认识。

在这里我举个简单的例子,以前企业为了营销而获得个人信息,需要去找中介公司购买不同人群的手机号码,然后挨个打电话推销。但现在是什么情况呢?百分之百我不敢讲,但至少有一半以上的企业,他们直接开发一个APP免费软件,譬如行程导航、音乐播放与图片修改等等。客户端的手机只要一下载,软件在后台会将客户的通讯录或其他有关隐私的内容全部复制,再用这个软件自带的程序在客户不知情的情况下发给APP开发商,最后客户手机里所有的信息都在开发商的掌控之下。

自2010年开始,欧美学界不断有学者关心这个问题,写了不少文章谈大数据与人类的道德、法律与伦理的各种冲突,用户下载了免费软件,就必须要向商家出卖本人的隐私,这是否合法?或者说是否符合基本的道德逻辑?这就是对大数据的搜集、整理与使用的反例。但也说明了一点,企业尚且如此,学术研究获取信息怎可还走老路?

在“大数据”到来之前的任何一个时代,都是技术先行、经济随后,人文研究最后尾随,这就形成了

人文研究长期以来的滞后性。但在“大数据”时代里,凡事瞬息万变,这种滞后性实际上已经严重威胁到了人文研究的学术意义。因此,如何利用“大数据”这一最迅捷、直接的方式来获取这个时代最重要、最有效的信息,构成了人文研究必然的转向趋势。

这是一个跨学科的时代,也是一个拆除不同知识壁垒与疆界的时代。人文研究既要有理论基础,也要关注文本、作品与历史,更要运用各种技术手段,不断进行观念的革新与转向,努力与时代对接,使其具备务实的时代价值与人文关怀,这才是人文研究的使命。

#### 四

“大数据”在很大程度上决定了人文研究的转向,并促使人文研究呈现出了一个新的特性:多元性。我认为,这是一个尤其不该忽视的问题。

多元性是当下人文研究一个重要的现状,同样也是未来的必然趋势。我一直认为,短暂地看,每一次科学技术的发展都会或多或少地导致人文精神的倒退,但人文精神一旦遭遇倒退之后,又会进行一次“反扑”,往往推动科学技术的进一步发展。因此宏观地看,每一次科学技术的重大变革,都带动了人文研究的转向,“大数据”也不例外。

以中国现代文学史研究为例,实际上其研究长期处于“答案既成,自圆其说”的状态,研究对象单一,研究方式单一,研究结论也单一。半个多世纪以来,所研究的经典作家总是鲁迅、郭沫若、茅盾、曹禺等少数几位,而研究方式也只是思潮、生平与作品研究,得出的结论就是:这些作家是经典作家。

坦率地说,我见过许多批判鲁迅的文章,但没有见过谁真的可以将鲁迅拉下中国现代文学的神坛,也没有看到哪篇研究成果可以将一位曾经默默无闻的作家重新挖掘为文坛巨匠。那么摆在面前的问题就是:这种研究的终极意义究竟何在?我认为,这就是人文学科在转向之前的局限性。既然已经通过研究证明了鲁迅是经典作家,那为何后来者还不断地运用各种手段、方法来重复证明?

如何打破这样的研究现状,就迫使研究者要直面现实的社会问题,让其研究变得多元化。这里所说

的多元化,是研究对象、研究方式、研究目的、研究策略的多样化。因此,研究者在面对“大数据”时,一定要学会甄别,取我所需,促使人文研究在“大数据”时代能够平稳、顺利地进行转向。

我认为,这种多元化主要表现在以下方面:

以前人文研究对象太单一,而如今是太复杂。图书馆里人文科学书架的书加起来,也不过大数据的冰山一角,连一个U盘都不到。现在研究者面对的是整个“大数据”,从前几年的网络文学、手机文学、微博文学到今天的APP阅读、微博热词生成、数据框架等等。面对这些海量数据,不要说穷尽一生心力,就是穷尽千秋万代,谁也无法完成,且不说数据本身浩瀚无边,更加之“大数据”每秒钟以几何级数暴增。因此,研究者谁也不能顾此失彼,要学会“海里淘金”,在“大数据”中筛选出自己需要的信息。

以前人文学科研究方法也相对单调,而“大数据”则要求研究者务必要掌握更多的研究方法。科学技术不但干预了我们的日常生活,而且对于人文研究有了新的要求。譬如说在20年前,计算机、互联网还是理工学科的专属工具,人文研究根本不必涉猎。但现在如果脱离了这两样东西,我们这个年代,还能停留在一些简单的方法上吗?

因为多元化成为趋势,所以时代要求我们在面对大数据时,必须要凭借新的研究对象、研究方式,进而得出新的研究结论。譬如美国麻省理工大学中国文化研究系的王瑾教授,近年来一直致力于互联网、多媒体、Web/Android与APP等技术及公民社会、中国文化心理关系的探求,取得了令人信服、瞩目的学术成果。作为一位人文学者,我认为王瑾教授的研究是有意义、“接地气”的,反映了“大数据”时代下人文研究的必然趋势。

“大数据”作为未来人类社会的组织形态,人文研究当然无法僭越其而独立存在。这一密切关系不但要被人文研究界所积极发现,而且要广为学界挖掘与应用,以推动未来10年里人文研究的顺利转向,进而促进技术与文化的进一步融合。

[责任编辑 马 艳]

## A Discussion of "Big Data" and the Diversion of Humanistic Studies

HAN Han

(Institute for the History of Natural Science, Chinese Academy of Sciences, Beijing 100034, China)