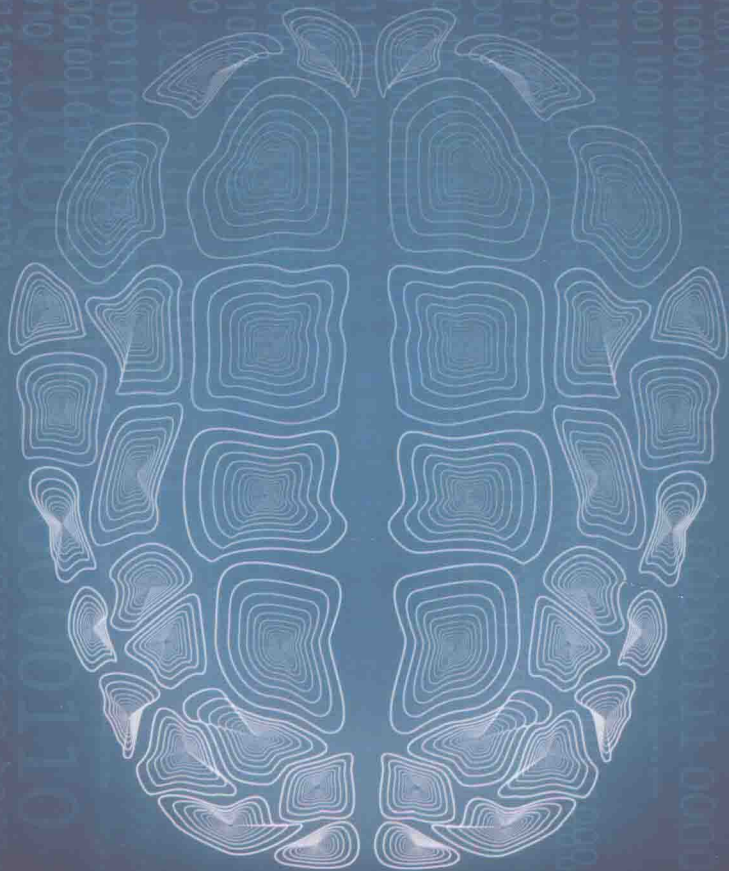




大数据 时代的人机关系

◎ 闫奎铭 著



北京航空航天大学出版社
BEIHANG UNIVERSITY PRESS

大数据时代的人机关系

闫奎铭 著

北京航空航天大学出版社

内 容 简 介

大数据为何如此聪明？它是否会超越人类理性的控制？本书选取人机关系这一切入点作为分析视角，综合以往哲学界做出的相关论述，阐释了新的概念与理论体系下大数据如何在人机交互中构建和涌现智能。

本书以大数据时代作为总的分析背景，进行了“人机关系的内涵分析”“人机关系与心物关系问题的关联”“大数据的构建与其意向性的负载”“大数据的可建构性研究”“人机关系视域下的大数据的伦理问题”等方面的论述，在跟随时代热点的同时也对“心物关系”和“他心知”等经典的哲学问题做出了视角较为新颖的回应。

本书知识性、系统性较强，可作为科技哲学领域研究的专家学者进一步从事人工智能哲学和技术哲学等领域研究的参考用书。

图书在版编目(CIP)数据

大数据时代的人机关系 / 闫奎铭著. -- 北京 : 北京航空航天大学出版社, 2016.4

ISBN 978-7-5124-2098-4

I. ①大… II. ①闫… III. ①人-机系统—研究
IV. ①TB18

中国版本图书馆 CIP 数据核字(2016)第 060893 号

版权所有，侵权必究。

大数据时代的人机关系

闫奎铭 著

责任编辑 杨 昕

*

北京航空航天大学出版社出版发行

北京市海淀区学院路 37 号(邮编 100191) <http://www.buaapress.com.cn>

发行部电话:(010)82317024 传真:(010)82328026

读者信箱: bhpess@263.net 邮购电话:(010)82316936

北京市同江印刷有限公司印装 各地书店经销

*

开本: 710×1 000 1/16 印张: 12 字数: 179 千字

2016 年 7 月第 1 版 2016 年 7 月第 1 次印刷

ISBN 978-7-5124-2098-4 定价: 29.00 元

前 言

大数据的问世及其所关涉的哲学问题,是凸显于当代并很有可能随着科技发展而受到持续关注的问题。大数据时代的到来,意味着人类将从追求数据的准确性与结构性转向追求数据的全面性与相关性,也正是这种转变,让大数据具备了可以做出与人类相类似的预测和判断的能力。

大数据为何如此聪明?它是否会超越人类理性的控制?本书选取人机关系这一切入点作为分析视角,综合以往哲学界做出的相关论述,阐释了新的概念与理论体系下大数据如何在人机交互中构建和涌现智能。

本书主要从以下几方面对大数据时代的人机关系进行了阐述:

(1) 结合哲学层面上关于心物范畴的研究,以及人机工程和人机交互等相关领域的研究可知,可以按照功能属性来对人与机器的组成要素进行划分。就像人可以划分为心灵与身体两个功能范畴一样,与之对应机器也可划分为软件(包括程序与数据)与硬件两个功能范畴,并且大数据时代所面临的最为主要的问题是心心层面的人机关系问题,即人的心灵与计算机的程序与数据的沟通和构建的关系问题。

(2) 结合心灵哲学方面对于意向性的研究可知,随着信息技术的飞速发展,机器对于人类的行动的识别与表征的精度逐渐提升,人类的行为在本质上属于对人类心灵(或称心理系统)中的意向性的表征,因此作为数据集合的大数据,实质上就是存储了丰富的人类的意向性的数据集合。由于大数据在数据的收集与挖掘方面不仅严格要求数据间

的因果性,而且要求数据样本就应该是数据全体,因此它对于人类心灵的表征也就更加全面,甚至可以表征一些类似于集体无意识的行为。正因为其能够对于人类心灵做到相对更加完整的表征,才使其可以获得更加接近于人类心灵的预测结果;但由于大数据并不会完全等同于任何一个人类个体的心灵,因此这样一个人工智能领域的产物更应该被视为一种人工他心。

(3) 与一般心灵哲学意义上的他心问题不同的是,大数据涌现的智能机制,在本质上就是基于数据的聚类 and 拟合分析的数据挖掘机制,并且大数据从数据源头上,其所追求的数据间的相关性可以通过人为捆绑社会实践中的人类行为而被人为建构;并且所谓的数据全体其实并不可能是一切的所有的数据,而是在一个由经济、技术和政治等要素人为划定的权力范围内的全部数据,这更让其在一般的用户那里成为数据权力壁垒下的黑箱。在这种意义上,对大数据的应用应该采取一定的谨慎态度。

(4) 由于大数据存在诸多与人的相似性,因而有必要像处理人与人的关系一样处理人与大数据的关系。由于大数据具有一定的预测能力,使其可以作为决策主体出现,并且可能做出一些和道德相关的判断,但是这些判断本质上是被人赋予了道德价值的;由于构成要素方面,大数据系统中无论在物理构成还是数据构成上都具有水乳交融的人的要素,因此当大数据作为认知与实践客体时可能会涉及道德层面的判断,但是由于数据不会明确指向任何道德原则,因此大数据具备的是潜在的道德客体地位。

本书知识性、系统性较强,可作为科技哲学领域研究的专家、学者进一步从事人工智能哲学和技术哲学等相关领域研究的参考用书。

致 谢

感谢我的导师尚智丛,他不仅是我学术道路上的老师,也是我人生成长过程中的父亲,作为一个学生我深知自己在学术功底上的薄弱、在学术研究过程中的不足。但是,他一直没有放松对我的管理,并对我投入了很多的时间和精力,帮助我、指导我,才让我有信心相信其实我也可以做到更好。作为一个在职的博士生,在我的成长过程中遇到问题、困难时,他也帮我仔细分析,用心帮我解决困难,让我能够快速适应生活、更好地工作。所以他不仅是我的导师,还是我人生的引航者!

感谢我的研究生导师孙雍军,感谢他在我攻读博士学位期间对我的帮助和支持!感谢他对我付出的心血和指导!

感谢人文学院我的老师——任定成教授、胡志强教授、张增一教授、王大洲教授、王大明教授。此外,刘立教授、肖显静教授在我论文撰写过程中提出了宝贵意见,在此一并致谢。

感谢我的父母,感谢他们 32 年来对我的培养和无私的帮助!感谢他们在我攻读博士学位期间最煎熬、最困难的阶段给予我的精神支持,是他们赋予我无穷的力量!

感谢同师门、同学院的每一位包容我的同学。感谢任安波师姐,杨萌、王慧斌师弟对我的辅助;感谢班长毕丞在我读书期间对我的照顾和理解;感谢周敏、张芳喜、杨晶、王翠平等同学在我求学路上的一路陪伴。但

愿毕业后我们之间的友谊会地久天长。

最后,感谢我的爱人李澜,感谢她在我攻读博士学位期间,对我默默的付出,尤其感谢她在怀孕期间独自一人去体检、独自一人打理家庭事务,她给我带来的不仅是优越的学习环境还有温馨的家庭,谢谢她不仅扛起家庭重担,还给我带来了聪明乖巧的女儿,让我体会到做父亲的喜悦与责任!

闫奎铭

于 2015 年 11 月 1 日凌晨

目 录

第一章 绪 论	1
第一节 研究背景及意义	2
一、研究背景	2
二、研究意义	4
第二节 国内外相关研究综述	5
一、信息技术领域的相关研究	5
二、哲学领域的研究文献	8
三、文献述评	20
第三节 研究方法	21
第四节 本书的主要架构	22
第二章 人机关系的内涵分析	23
第一节 人机关系的四重内涵分析	24
第二节 人机工程为人机关系提供操作层面的支持	26
第三节 信息技术实现以平台操作为中介的人机交互	32
第四节 本章小结	36
第三章 人机关系与心物关系问题的关联	41
第一节 人机关系问题与心物范畴	42

第二节 人与机器可能都是物质性的存在吗?	43
第三节 可能存在一个同时支配人与机器的心灵吗?	48
第四节 人机关系问题中的心物范畴辨析	51
第五节 物物层面与心心层面的人机关系问题的界定	55
一、物物层面的人机关系	56
二、心心层面的人机关系	59
第六节 大数据时代的人机关系内涵分析	64
一、人机的交互让人作为机器的载体,机器作为人的网络	64
二、大数据时代实现人机合一	68
第七节 本章小结	69
第四章 大数据的构建与其意向性的负载	73
第一节 大数据系统构建的技术基础	75
第二节 大数据时代人机交互的基本形式及其理论与价值渗透	81
第三节 人化的大数据及其意向性负载	89
第四节 大数据的人工他心本质分析	96
第五节 作为人工他心的思考方式的数据挖掘	105
第六节 他心难题决定大数据将缺乏情感	110
第七节 本章小结	113
第五章 大数据的可建构性研究	117
第一节 作为技术黑箱的大数据	118
第二节 相关性的可建构性研究	122
一、“啤酒+尿布”案例概况	122
二、大数据的相关性的语境负荷	123
三、人为构建大数据的相关性的可能性	125

四、从垄断捆绑到相关性的固化	127
五、从显见垄断到隐性垄断再到对大数据的人为建构	131
六、从大数据的相关性的可建构性看其应用局限	133
第三节 数据全体的人为改造的可能性研究	135
一、大数据的全体性的价值与利益负载	136
二、数据全体与相关性的构建之间的关系	139
三、数据全体的构建策略探究	141
四、数据全体的意义的有限性	144
五、从可建构的大数据到被建构的人工他心	147
第四节 本章小结	149
第六章 人机关系视域下的大数据的伦理问题	153
第一节 大数据的非完全道德主体性分析	154
第二节 大数据的潜在道德客体性分析	156
第三节 本章小结	160
参考文献	163
后 记	174

第一章

绪 论

第一节 研究背景及意义

一、研究背景

自近现代哲学产生以来,心物关系或身心关系问题一直都是哲学家们探讨的重点,尤其在计算机科学产生以后,相应的问题又有了新的哲学载体,即人工智能哲学或心灵哲学问题。

相关的问题进一步就可以转化为,通过人类的计算机科学技术能否构建出像人类一样思考的机器,以至于在人机对话的过程中,人们无法分辨出被提问者是人还是机器;除了在本体论层面更多地探讨心灵的本源究竟是不是物质以外,也涉及了心与身之间的交互如何达成,他心是否存在,心与他心之间能否沟通,以及如何沟通的问题。

上述这些问题的共同指向就是人类的理性如何产生,以及能否人工创造或复制人类理性的问题。那么很明显当计算机科学技术产生一定的进步,使得机器运算处理问题的方式或结果更加趋近于人时,就往往会对心物关系或身心关系问题的解答产生影响。

目前方兴未艾的大数据科学技术就对人类的理性如何产生,以及能否复制的问题产生了不小的影响。所谓大数据(Big Data)是具有 4V 特点的大型数据集,所谓的 4V 特点就是“1. 数据体量(Volumes)巨大:大型数据集,从 TB 级别跃升到 PB 级别。2. 数据类别(Variety)繁多:数据来自多种数据源,数据种类和格式冲破了以前所限定的结构化数据范畴,囊括了半结构化和非结构化数据。3. 价值(Value)密度低:以视频为例,连

续不间断监控过程中,可能有用的数据仅仅一两秒。4. 处理速度(Velocity)快:包含大量在线或实时数据分析处理的需求,1秒定律。”^[1]并且,借助于现代的数据挖掘研究可以发现,大数据具有比较明显的可量化、可自演进的特征。如 Google 公司的 Flu Trends 流感预测工具通过收集网络信息,对流感疫情做出了预测,其结果与美国国家疾控中心的数据基本吻合;微软研究院以 David Rothschild 为首的研究团队通过对入围影片进行数据分析,成功地预测了第 85 届奥斯卡除最佳导演奖以外的 13 项大奖的归属。

大数据在上述案例中取得的成功不容小视,因为其说明大数据不仅可以对于客观世界进行准确的反映,还可以做出与人的主观判断相近的预测结果。那么结合哲学史上的心物关系以及身心关系问题来看,这样一个能够相当准确地演绎和预测世界正在或正要发生的变化,甚至可以顺利演绎人类带有主观意向的判断的科技产物,是否可以被视为科技对于人类心灵的模仿甚至再造?或者说大数据得出演绎或预测结论的模式和人类的思考模式是否相似甚至相同?大数据的构建需要通过收集人类的操作数据而获得,得出相应的演绎结论则又需要通过人类的挖掘而获得,既是数据来源,又是数据的处理者和使用者的人类,在大数据的构建中究竟具有怎样的地位和作用?

基于以上分析,可以发现,在哲学层面上,能够准确地对客观世界和人的主观思想做出相当准确预测的大数据的产生,将带来以下几个需要解决的人机关系问题:

(1) 人和以大数据为代表的计算机科学技术可以做出相似的判断,那么是否可以认为这种功能上的相似反映了二者某种结构上的相似,那么基于大数据的原理与实质,能否窥探人类理性的本质?

(2) 大数据的相关功能的实现,有赖于人类对于一系列数据终端的

使用,因此在大数据做出相关预测结果的过程中,人与机器之间存在着多重、多种的交互模式,相应的交互形式是怎样的,相应的交互的本质又是什么?

(3) 大数据的预测能力,超出了一般人的理性驾驭范围,并且由于原始数据的出处,以及数据最终的处理结果的使用者都是人类,人与机器之间,在社会功能范畴上、社会价值范畴上的地位如何评判?

二、研究意义

相关问题中涉及了多方面的人机关系问题。其中问题(1)涉及人与机器之间的本体论层面的关联,追问人是什么、人的思想是什么、机器是什么,以及人和机器的本体论基础是否相同等;问题(2)涉及人机之间的认识论关系,即需要回应,在认识论层面上,人与机器在整体的操作和使用进程中,彼此交流的反馈机制是什么的问题;问题(3)涉及人机在价值层面的比较关系,以及认识论和本体论层面的地位与作用在社会历史的视野中的地位与作用关联。

由此可见,研究相关问题在哲学层面上,一方面可以对传统哲学意义上的身心问题和心物问题做出具有时代特征的思考与回应,另一方面可以结合科学哲学和技术哲学的研究视角,探讨科学技术的价值赋予与价值负载;在实践层面上,大数据作为一个方兴未艾的科技产物,对其哲学本质以及实践价值的研究,将有助于人们更好地制定对于大数据的发展策略和应用策略。总体而言,研究大数据时代的这种人机关系问题具有重大意义。

第二节 国内外相关研究综述

一、信息技术领域的相关研究

Nature 在 2008 年推出了 *Big Data* 专刊,探讨大数据形成和发展的可能的理论基础^[2],计算社区联盟 *Computing Community Consortium* 在 2008 年发表了报告 *Big data: computing Creating revolutionary breakthroughs in commerce, science and society*,阐述了解决和处理数据所带来的问题需要的技术手段及面临的挑战^[3]。*Science* 在 2011 年 2 月推出专刊 *Dealing with the data*,主要从科研实践的数据问题入手,说明对大数据的适当处理将有助于科研活动的进行^[4]。

可以说大数据是一个非常前沿的课题,同时也因为它刚刚兴起,相关研究可能还没有形成较好的范式,因而显得不很成熟。结合国内外学者的相关研究可以发现,信息技术领域对于大数据的研究主要集中在以下三个方面:

1. 大数据的核心技术研究

大数据的产生、收集与挖掘需要一系列的核心技术作为支持,主要集中在以下几类核心技术中。

(1) 云计算技术

云计算(Cloud Computing)是分布式计算(Distributed Computing)、并行计算(Parallel Computing)、效用计算(Utility Computing)、网络存储

技术(Network Storage Technologies)、虚拟化(Virtualization)、负载均衡(Load Balance)、热备份冗余(High Available)等传统计算机技术和网络技术发展融合的产物^[5]。

目前,Google 云计算已经拥有 100 多万台服务器,Amazon、IBM、微软、Yahoo 等的云均拥有几十万台服务器,上述公司是目前进行云计算研究与应用的巨头,可见美国在云计算的研究和应用方面处于世界的前列。近年来我国对于云计算的研究和应用也取得了一定进展,但与美国相比仍有较大的差距。

云计算技术包括文件系统、数据库系统、索引和查询技术等。在这些领域中上述几家美国公司均处于领先地位。

(2) 分布式处理平台

Hadoop 平台是目前最为流行的大数据处理平台,它包括三部分,分别是 Hadoop Distributed File System(HDFS)、Hadoop Map Reduce 分析模型,以及 Hadoop Common 数据集成模型。

美国其他一些网络巨头也基于 Hadoop 平台,根据自身需求构建了自己的处理平台,其主要针对的是对于分析模型的修改,如 IBM 的 System S 模型、微软的 Dryad 模型和 Facebook 的 NameNode 等。

在上述模型中,其主要的区分在于对数据的处理,是采取流处理、批处理还是二者结合的方案。其中 Map Reduce 和 Dryad 倾向于批处理,而 System S 则倾向于流处理。

目前,我国在相关领域的研究主要还是对于处于开源状态的 Hadoop 平台进行分析改良和应用^[6]。

2. 大数据的应用前景研究

从前文的分析梳理中可以发现,大数据的应用处理研究一直由 Google 等大型的信息技术公司所掌控,美国政府在意识到大数据的研究

价值后,于2012年3月发布了大数据研究和发展倡议,计划投资超过2亿美元并正式启动大数据发展计划,计划在科学研究、环境、生物、医学等领域利用大数据技术进行突破。

除此之外,大数据在传统的市场营销^[7]、图书馆信息管理^[8]等领域也都有所应用。

大数据的概念诞生于美国,美国也拥有目前世界上对于大数据应用处理最为权威的公司和研究机构,可以说美国在与大数据相关的信息技术研究领域处于领先地位。相比之下,我国的相关研究仍属于追随和引进介绍国外研究成果的阶段。据悉,以百度集团和阿里巴巴集团为首的中国企业目前已经大力投入到大数据的研究当中。在巴西世界杯期间,百度集团通过应用大数据工具准确预测了世界杯淘汰赛阶段的所有比赛结果^[9],可见我国企业在大数据应用方面已经取得了一定的进展。另外,百度公司还透露,近期将会上线百度公司研发的传染性疾病预测工具“百度疾病预测”^[10],其表现值得期待。

除此之外,埃里克·西格尔所著的《大数据预测》重点展望了大数据时代,人类对于气候、消费者购买偏好和经济发展等人类社会行为的预测能力;埃里克·托普的《颠覆医疗》结合案例说明了大数据对于人类疾病的准确诊断和有效治疗方面的重要意义;戈登·贝尔的《全面回忆》以个人大数据(将个人生活全面社会化后所获得的大数据集)对人们社会生活的影响进行系统论述。此类书籍的特点是,烘托了大数据在一些具体领域的应用价值,并且容易让人们联想到人类生活被全面数据化后,人和大数据系统之间的关系问题——是人类以自己的目的为核心构建大数据系统,还是大数据系统自身左右了人类的生活方向。

我国学者也有相关的书籍出版,如涂子沛所著的《大数据》和《数据之巅:大数据革命,历史、现实与未来》,车品觉的《决战大数据》,赵国栋的