

· 改革开放40周年专题: 数字经济 ·

大数据与人工智能的伦理挑战(上)



□周 涛^{1,2}

[1. 电子科技大学 成都 611731; 2. 成都新经济发展研究院 成都 610094]

[摘要] 一个以大数据为原材料,以人工智能为引擎的新科技时代的到来不可阻挡。大数据和人工智能在给人类社会带来巨大利益的同时,也带来了诸如个人隐私、数据独裁、新型智能生命等让人担忧的问题。回顾了大数据和人工智能伦理研究的背景、意义和现状,着重从中立性、时效性、导向性三个方面介绍大数据于人工智能发展带来的具体伦理挑战。续篇将介绍边界问题、隐私问题和责权问题,并简述当前有效的应对策略以及未来开放性的政策和技术问题。

[关键词] 大数据; 人工智能; 伦理; 算法

[中图分类号] B82-057

[文献标识码] A

[DOI] 10.14071/j.1008-8105(2018)-1018

Ethical Challenge Brought by Big Data and Artificial Intelligence (Part 1)

ZHOU Tao^{1,2}

(1. University of Electronic Science and Technology of China Chengdu 611731 China;

2. Institution of New Economic Development Chengdu 610094 China)

Abstract A new technical era with characters of big data as raw material and artificial intelligence as engine is unstoppable. Big data and artificial intelligence have brought human beings great benefit, meanwhile they also have brought problems such as personal privacy, data autocracy, and new type of intelligent life. This paper overviews the background, significance and present status of ethical researches on big data and artificial intelligence, especially focusing on the ethical challenges brought by them from the aspects of neutrality, timeliness, and guidance. Part 2 followed by this paper will introduce the boundary problem, privacy problem and responsibility problem, and sketches effective countermeasures and future open policy and technological problems.

Key words big data; artificial intelligence; ethic; algorithm

引言

随着互联网、移动互联网、物联网和各种遥感探测技术的发展,一个“一切都被记录,一切都被分析”的数据化时代已经到来——现在全球数据存储量已经达到了20ZB左右,也就是2000万PB。得益于可用于算法训练的海量数据的积累和以深度学习为代表的机器学习算法的发展^[1]。近几年来,人工智能在若干应用场景达到甚至超过了人类专家的

水平——大家记忆犹新的就是2016年AlphaGo以4:1的比分战胜围棋世界冠军李世石^[2]。越来越多的证据显示,一个以大数据为原材料,以人工智能为引擎的新科技时代正在到来^[3-4]。

大数据和人工智能的技术发展与应用实践,已经为我们的生活带来了显著的贡献^[4-5]。在医疗领域,大数据和人工智能被用于精准识别医疗影像中早期的病灶,定位致病基因并开展相应的靶向治疗,实时监测评估健康状况、提前预警重大健康风

[收稿日期] 2018-09-07

[基金项目] 国家自然科学基金重点项目(61433014)。

[作者简介] 周涛(1982-)男,博士,电子科技大学互联网科学中心教授,成都新经济发展研究院执行院长。

险等^[6]；在交通领域，海量数据和预测算法的结合能够帮助我们进行更高效的交通导航，尽可能减少拥堵时间^[7]；在教育领域，学生发展过程中学习和生活数据的分析，可以帮助了解学生行为和学业表现之间的关系，设计更高效的个性化教育方案，提前对不利于学业发展的异常行为进行干预^[8]等。

在带给我们可感知的贡献的同时，大数据与人工智能也带来了一系列让人忧虑的问题，譬如个人关键隐私信息的大面积泄漏问题，又如人工智能对若干劳动岗位的冲击，再如大数据和人工智能深度结合形成的新型数据独裁问题，等。这些问题要求我们在伦理意义上仔细审视大数据和人工智能的发展。简而言之，伦理问题就是判断好与坏、对与错的问题，大数据与人工智能的伦理研究是站在人类整体利益的角度去审视判断大数据与人工智能哪些技术、何种应用对人类是有利的，哪些是危险的，哪些是不应触碰的，其最终的目的是要找到路径、梳理规范，以保证大数据和人工智能的发展给人类带来的利益远大于伤害。

大数据与人工智能的伦理问题引起了全球学界、业界和政府广泛的关注。2016年，奥巴马政府发表了题为《为人工智能的未来做好准备》的报告^[9]，强烈建议要评估人工智能技术和产品带来的风险；2017年，欧洲经济和社会委员发表了题为《大数据伦理——在欧盟政策背景下，实现大数据的经济利益与道德伦理之间的综合平衡》的报告，对大数据的伦理问题进行了概括，特别强调了数据采集和使用带来的安全和隐私问题^[10]；2018年，英国议会发表了《英国人工智能的准备、计划和能力》的报告，呼吁政府为人工智能研发和应用设定伦理原则^[11]；美国电气与工程师协会、美国未来生命研究所等研究组织和微软、谷歌等科技公司也陆续发布了相关的研究报告。其他国际上相关的报告例子还很多，不一一举出。

中国也开始意识到这些伦理问题的重要性，例如我国2017年发布的《新一代人工智能发展规划》中指出制定人工智能发展的伦理规范是促进人工智能发展的重要保证措施。2017年12月8日，中共中央政治局就实施国家大数据战略进行第二次集体学习，习近平主席主持学习并发表了讲话，在提出实施国家大数据战略加快建设数字中国要求的同时，特别强调了数据安全、数据隐私、数据确权、数据版权、数据治理政策储备和治理规则等与数据伦理相关的问题。与欧美等国相比，我国学界、业界和政府在大数据与人工智能伦理问题上系统性的深入

研究还较少，目前尚未形成具有国际影响力的伦理研究报告和政策实施方案，与习近平主席提出的“要加强国际数据治理政策储备和治理规则研究，提出中国方案”这一要求尚有较大差距。

因为大数据和人工智能的发展是高度融合的，所以本文不拟特别区分大数据的伦理问题和人工智能的伦理问题，而是结合讨论从数据采集、存储和应用，一直到数据挖掘、机器学习等典型人工智能技术发展所带来的一系列伦理问题。限于篇幅，本文及后续文章将集中讨论中立性、时效性、导向性、边界问题、隐私问题和责权问题六个方面。本文尽量采用通俗的语言和生动的例子来说明这些问题，既可以作为政府相关管理人员以及学者和从业人员了解大数据与人工智能伦理问题的入门材料，也可以作为一般读者的科普读物。

一、中立性

表面上客观的数据和理性的算法，也可以产生非中立性的结果。事实上，数据和算法导致的歧视往往更难发现也更难消除。数据和算法对中立性的破坏，可能来自三方面的原因：一是采集数据或设计算法的相关人员蓄意为之；二是原始数据本身就存在偏见，因此相应的数据驱动算法的结果也会有偏见；三是所设计的算法会导致有偏见的结果。第一种原因归根到底是人的问题，在任何时代和任何环境中都可能存在，数据和算法不过是他们利用的工具罢了。因此本文着重分析后面两种情况。

即便数据是人类社会客观中立的记录，如果人类社会本身就存在偏见、歧视和不公平，那么相关数据自然也会带入我们社会的不公。例如，互联网求职的简历数据显示，在职场中身高和性别的歧视都非常严重：平均身高越高，平均收入越高；在同等学历条件和行业背景下，女性要多工作5~10年才能获得和男性相当的薪水^[12]。显然，使用这类简历数据进行职位的推荐时，其结果必然自带歧视。卡内基梅隆大学的Datta等人最近的研究就显示，Google广告系统的人工智能算法在推送职位招聘信息的时候，同等教育背景和工作经历下的男性要比女性以高得多的频率收到高收入职位招聘信息^[13]。普林斯顿大学Caliskan等人使用常见的纯统计机器学习模型，在万维网的标准文本语料库上进行训练，发现计算机可以“学会”沉淀在人类语言记录中隐含的偏见——既包括一些无关道德也无伤大雅的偏见，例如昆虫让我们联想到不愉快而花朵则常

与愉悦的事情相伴，还包括一些严重的偏见，包括来自性别和种族的歧视^[14]。实际上，有些我们人类自己都没有注意到的潜在的歧视，计算机也能通过机器学习捕捉到。这些数据上存在的偏见，会通过算法表现为带歧视的结果，这些结果可能进一步加大歧视，从而新的数据包含的偏见有可能被加剧，造成恶性循环。比如说数据中显示每十个前1%高年薪的高端职位只有一位女性，于是“性别为女性”这个特征值在获得高端职位推荐中将是一个负面的因素，算法也将避免给女性推荐高端职位信息。在没有基于大数据和人工智能的招聘信息服务的情况下，男性和女性获取高端职位信息的数量可能相差不大，这种情况下女性真正获聘高端职位的可能性也远低于男性。如今，计算机的自动服务在源头上就让女性更少知道信息，所以可以预期女性获得高端职位的比例将进一步降低，而这又再次加强了数据的偏差，从而让算法更少向女性推荐高端职位。这种恶性循环，会进一步放大原本就有的不公。

算法的设计和学习过程，也可能带来偏见甚至歧视。个性化推荐算法是目前在大数据和人工智能领域应用最为广泛的算法之一^[15]，其目的是根据用户过往的浏览、点击、收藏、购买等记录，向用户推荐他需要或者喜欢的资讯、商品等。淘宝的猜你喜欢和今日头条的个性化新闻推荐就是推荐算法非常典型的应用。在电商网站上，同一个品类下商品的价格差别巨大，如果某目标用户以前在电商网站收藏和购买的商品，在相关品类中价格排名都特别靠后，那么算法可能会在一个名为“价格敏感度”的特征维度上给该用户标上高分。于是，当该用户搜索一个关键词后，如果自己不做调整，可能从前到后翻十页，他看到的都是便宜货。尽管算法的初衷是提高该用户的点击率——这被认为对用户和商家都好——但是这事实上形成了同类商品展示对低收入消费者的歧视。试想如果在一个购物商场中部分人被告知因为收入过低不允许进入LV的店面，或者因为收入过低而要排队数小时才能入店，而高收入者有专享通道。这种歧视肯定是我们不能容忍的，而在互联网上，数据和算法带来的偏见与之类似但更为隐蔽。

随着数据化浪潮的进一步发展，个人获取和处理信息的难度会进一步加大，我们会更加依赖各种各样的信息中介。例如我们到一个陌生的城市，如何规划一条一日游的路线？又比如我们如何通过在线教育，选择若干课程，进入一个新的领域？再比如我们怎么在读书的时候就规划和选择自己的职业

道路？高度发达的人工智能会充分考虑包括家庭、性别、民族、消费水平等关于你的各种数据，给出“最适合你的选择”。于是，不同收入和不同家庭背景的人会抵达城市中不同的角落、下载不同的课程、规划不同的人生。在大数据的时代，不同出身的人所获取到的信息差异，可能比现实世界的差异还大，因此很可能更早就形成截然不同的视野、格局和能力，从而加剧而不是减少阶级的固化。

二、时效性

与人类相比，计算机有一种特别突出的能力，就是能够保真存储和分析很多年以前的数据，并通过算法对未来某些事件给出远远比人精确的预测。

Mayer-Schönberger曾讨论过一个未来的可能情形，就是当智能机器预测到你将在某时某地犯罪，就可以在你还没有实施犯罪的时候逮捕你^[3]。幸运的是，这个比Orwell笔下的世界^[16]还可怕的情景，只是Mayer-Schönberger设想的一种极端情况，实际上并没有发生过。Dormehl讲了一个温和得多但已经现实发生的版本^[17]，就是美国政府根据姓名、出生地、宗教信仰、历史行为数据——例如所有旅行数据、人脸图片数据等，利用机器学习算法，对每一位航空旅客是恐怖分子的嫌疑度进行打分。一些无辜的人因为疑似恐怖分子或者近期有较高从事恐怖活动的可能性，而经常在机场被羁留检查，甚至多次错过飞机。通过数据和算法，机器可以预测一个人犯罪的概率，如果这个概率足够大，那么在他还没有实施犯罪行为时，就已经实质上受到了惩罚，尽管他可能是完全无辜的。从功利主义的角度讲，这些智能化的方法能够降低犯罪率，但是这里面一个核心的伦理问题就是“我们是否应该为尚未发生的一种可能性付出代价”。

反恐的例子距离普通人的生活较远，下面我们来看一个现实生活中的例子——酒驾。对于机动车驾驶员来说，饮酒驾车是违法，醉酒驾驶可入刑。这条法律受到了数据有力的支撑：“当驾驶者血液中酒精含量达80 mg/100 mL时，发生交通事故的几率是血液中含酒精时的2.5倍；达到100 mg/100 mL时，发生交通事故的几率是血液中含酒精时的4.7倍；即使在少量饮酒的状态下，交通事故的危险度也可达到未饮酒状态的2倍左右。”事实上，这条法规的出台，有效降低了严重交通事故的发生率。这条法律的基本精神是惩处可能的尚未发生的更严重的犯罪。这当然无可厚非，例如危害公共安

全罪背后的法理出发点也是类似的。当数据的采集能力和算法的分析能力进一步增强之后,我们肯定能够发现更多可能危害公共安全的因素,并把它们写进法律中。几十年后,如果自动驾驶的车辆占有相当的比例,很可能主要的交通事故都是由人类操纵驾驶车辆引起的。那个时候的法律,可能就不是针对饮酒和吸毒后驾车了,但凡活人开车,都得抓了。因为人驾驶车辆本身,就是一种对公共安全的危害。那么,今天能接受酒驾入刑的人,明天能够接受手驾入刑吗?在使用数据和算法预测能力的时候,是否也需要有一个限度?

计算机让人类更加望尘莫及的,是记录和回顾历史的能力。因为大量过往数据被记录下来,所以我们今天的错误可能会带来很长时间难以消除的影响。中学期间一次冲动的打架斗殴,大学期间因为攀比借了款又没有及时还款……这些记录可能会在十年后呈现给潜在的雇主和金融机构,影响一个人职场的发展,降低他申请车贷房贷的成功率,甚至让他的创业融资遭遇阴影。我们年少轻狂时在自己博客、微博、微信或者其他论坛、贴吧上留下的冲动言论,又或者我们脑残时光上传的裸露、暴力或者炫富的照片,都可能在未来成为我们感情和事业再上一个台阶时的定时炸弹。

中国有句古话,叫做“浪子回头金不换”。父母亲友也许会选择原谅甚至忘记,但大数据和人工智能不会遗忘。Mayer-Schönberger曾经强烈呼吁应该给我们的数据一个“被遗忘的权利”——除了出于科学研究、档案管理、公共安全或其他特殊的既非盈利也不面向一般公众的目的,个人和企业不应该存储、分析、使用和传播超出一定年限的数据^[18]。一些科技企业已经开始了类似的尝试,例如著名的社交网站Snapchat起家的时候就是推出了“阅后即焚”的功能——你发给朋友的信息和图片,对方阅读后一段时间内自动消失。2017年,腾讯投资了Snapchat,实际上腾讯在微信的私信助手手中也有“阅后即焚”的功能,还可以设定多长时间后删除消息。但是,你和你的好友虽然看不到了,Snapchat和腾讯原则上可以保存和利用这些信息。另外,社交媒体产生的数据,相比于一个人所有被记录下来的行为数据,毕竟只是很小的一部分,而绝大部分与我们紧密相关的数据都不是我们主动上传或发送的。越来越多的企业认识到了数据的价值,因此主动删除或者放弃某些数据的使用权,似乎并不是一家盈利性企业优先考虑的问题。

三百年前,瑞典哲学家、神学家Swedenborg曾

感慨:“上帝赋予我们大脑以便让我们具备遗忘的能力”。如今,面对电脑,上帝也无法让它们遗忘。尽管Mayer-Schönberger要求被遗忘权利的呼吁已经过去了五年多,我们还没有看到在这个方面任何实质性的进展。

我们每个人都应该尊重历史,都应该怀揣梦想,但更多的时候我们生活在当下。当一个人每做一件事情,都可能承受因过往历史的影响,都必须考虑对未来发展的影响,那这样的人生是不堪重负的。如果数据的使用时限不受控制,这种机制下最安全的选择只能是沿着数据和算法所青睐的某种智能时代的主流道路前进。人类与生俱有的自由猖狂和特立独行恐怕最终都不得不服从机器制定的优化后的规则。

三、导向性

算法根据其目标函数所产生的“优化的结果”对受众,尤其是一般客户端用户产生的引导效果,可能与用户真实的需求大相径庭,也可能虽然能够满足用户短期的需求,但是带来中长期的负面效果。我们把这类问题归纳为导向性问题,其短期、中期和长期的具体表现有所不同。

短期表现为算法给出的结果与用户真实需求不符合,用户立刻就有负面的感知。可以说,算法诱骗用户在其需求没有获得有效或充分满足的情况下,帮助算法完成了它拟实现的优化目标。举个例子,在给用户进行个性化新闻推荐或者自动生成新闻的时候^[19],如果算法的目标是最大化推荐新闻或者自动生成新闻的点击次数,那么计算机自己就会变成一个“标题党”,选择甚至自己生成一些耸人听闻的标题,又或者与当前新闻热点紧密联系的标题,但新闻的实际内容却和标题毫无关系或只有极其浅薄的联系。用户点击进入新闻页面,发现与其需求和期望严重不符合,立刻就会有上当的感觉。但是用户的点击行为实际上帮助完成了算法的既定目标,而且还让计算机认为“推荐这类新闻或者生成这些标题是正确的”,因此算法可能会变本加厉,越来越标题党。对于这类问题,如果算法的设计和使用人员本身不存恶意,其实是较好解决的。因为我们可以通过设计更好的目标函数来提升用户体验——毕竟任何一个目标函数都不可能完全覆盖用户的意图,所以优化目标函数以提升用户体验是一个长期的过程。举个例子,一篇新闻打开后用户停留阅读的时间,或者一个微信打开后被转发和点

赞的比率，都可以纳入到目标函数中来。但如果算法的设计和使用人员本身存有恶意，或者完全被短期商业利益绑架，这就可能变成一个难以舒缓的矛盾。譬如百度的魏则西们，搜索病情相关知识的时候，被漫天遍野真假难辨的广告诱骗采用一些不适病情的治疗方案，最终人财两空；又或者如作家木木吐槽在百度搜索上海美国领事馆官网的地址，出来一堆广告却几页翻不出真身。如何有效打击这类“恶人利用算法作恶”的问题，就不是简单调整目标函数可以完成的任务了。

中期表现为算法能够满足用户即时需求，但其累积效果与用户的发展利益不一致。简而言之，就是短期虽然满足了用户，但这种“满足”对用户而言或许并非好事。一些大型的游戏开发公司，可以对海量用户的行为数据进行深入挖掘分析，了解用户容易退出游戏的时间点和前兆行为，分析哪些场景、关卡和境遇最易让用户退出游戏，从而通过赠予装备，下发新任务等方式尽最大可能让用户获得更好的游戏体验，延长游戏时间。一些视频网站，尤其是短视频的网站，正在尝试通过个性化推荐，为用户创造一种沉浸式的美好体验，让用户连续不断地点开视频，停不下来。这些算法很好地满足了用户的需求，用户也不会有负面体验，但是时间长了，对于用户自身的健康成长和学业工作或许会起到负面的作用^[20]。某些企业创造经济价值的同时也带来环境的破坏，后者被称为外部成本。类似地，这些累积产生的负面效果也可以看作是“算法的外部性”。由于用户和商家都在这类算法中直接获益，所以这种外部性问题的解决，还需要外部的力量来干预。举例而言，2018年8月30日，教育部官网发布八部门关于印发《综合防控儿童青少年近视实施方案》的通知，其中国家新闻出版署明确将对网络游戏实施总量调控，采取措施限制未成年人使用时间。之后腾讯立刻提出将推进最严格的实名游戏制度，避免未成年人过度沉迷于王者荣耀等游戏。需要特别强调的是，对用户行为数据的分析挖掘这一技术本身是没有善恶的，类似的技术可用于判断在线课程中哪些地方是学习者普遍认为的难点，从而帮助他们更高质量完成课程学习^[21]。

长期表现为算法累积的效果与人类社会整体发展的目标和价值不一致。譬如我们很早就知道，一个人的家庭出身和这个人未来学业和事业所能取得的成绩在统计上是高度相关的^[22-23]，因此如果我们能够掌握充分的数据，聪明的算法会在成长的早期和学业工作关键抉择点上倾向于把最优质的资源推

荐给家庭背景比较好的人。就一些可以客观量化的指标，如经济发展指标而言，这个算法也许表现优异，但它可能会大幅度加剧阶级的固化，使得原本贫穷、缺少教育或因种族肤色等原因受到歧视的家庭，改变命运的机会变得更加渺茫。如果我们只看数字利益，阶级的流动并不一定带来比阶级的固化更好的经济表现。但是人类整体发展的目标和价值，不是一群人奴役压迫另一群人而获得更高的效率，所以经济上的成功，数目字上的成功，并不一定代表整个人类社会的成功。又比如大规模电子商务网站（如阿里巴巴、京东等）和大规模内容网站（如腾讯视频、爱奇艺等）的出现，使得所有流量和关注向少数几个平台汇聚。平台在运营流量的时候，会采用以个性化推荐^[15, 24]为代表的一系列算法来吸引用户更多点击、浏览、收藏和购买。尽管这些算法都考虑了用户个性化的特征，所给出的推荐结果也不尽相同，但事实上受算法驱动的用户整体的行为所表现出来的多样性会降低，因为推荐小众对象的风险毕竟要大得多^[25]。长久而久之，小众的喜好和小众的商品都会失去萎缩，人类社会的多样性以及非主流的个性和创造力，都有可能在信息高度丰富的时代反而降低，这对人类社会的长远发展也可能是不利的。

四、小结

本文介绍了大数据和人工智能所面临的伦理挑战，以及全球范围内相关的进展情况。特别地，我们用生动的语言向读者展示了大数据与人工智能在中立性、时效性和导向性方面可能带来的负面的问题。这种负面的效应，既可能针对个人，也可能影响整个人类社会；既可能是即刻反馈的，也可能是长期积累导致的。通过本文的介绍，希望读者对于大数据和人工智能的伦理问题有直观的认识和具象的感知。

在本文的续篇中，我们将继续深入讨论三个问题：算法的边界问题、数据的隐私问题和人工智能的责权问题。这些都是当前学界充满争议的问题。续篇的最后将简要介绍当前有效的应对策略并讨论未来开放性的政策和技术问题。

参考文献

- [1] LECUN Y, BENGIO Y, HINTON G. Deep Learning[J]. *Nature*, 2015, 521: 436.
- [2] SILVER D, HUANG A, et al. Mastering the game of

Go with deep neural networks and tree search[J]. *Nature*, 2016, 529(7587): 484-489.

[3] 维克托·迈尔-舍恩伯格, 肯尼思·库克耶. 大数据时代: 工作、生活与思维的大变革[M]. 盛杨燕, 周涛, 译. 杭州: 浙江人民出版社, 2013.

[4] 吴军. 智能时代: 大数据与智能革命重新定义未来[M]. 北京: 中信出版集团, 2016.

[5] 周涛. 为数据而生: 大数据创新实[M]. 北京: 北京联合出版公司, 2016.

[6] 埃克里·托普. 颠覆医疗: 大数据时代的个人健康革命[M]. 张南, 译. 北京: 电子工业出版社, 2014.

[7] LU Y, DUAN Y, et al. Traffic flow prediction with big data: A deep learning approach[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2015, 16(2): 865-873.

[8] CAO Y, GAO J, et al. Orderliness predicts academic performance: Behavioral analysis on campus lifestyle[J]. *Journal of the Royal Society Interface*, 2018, 15(146): 20180210.

[9] Executive Office of the President (Obama). National Science and Technology Council Committee on Technology[R]. Preparing for the future of artificial intelligence, 2016.

[10] European Economic and Social Committee. The ethics of Big Data: Balancing economic benefits and ethical questions of Big Data in the EU policy context[R]. 2017.

[11] House of Lords in UK. AI in the UK: ready, willing and able[R]. 2018.

[12] YANG X, GAO J, et al. Height conditions salary expectations: Evidence from large-scale data in China[J]. *Physica A: Statistical Mechanics and its Applications*, 2018, 501: 86-97.

[13] DATTA A, DATTA A, et al. Discrimination in Online Advertising: A Multidisciplinary Inquiry[J]. *Proceedings of Machine Learning Research*, 2018, 81: 20-34.

[14] CALISKAN A, BRYSON J, NARAYANAN A.

Semantics derived automatically from language corpora contain human-like biases[J]. *Science*, 2017, 356: 183.

[15] LU L, MEDO M, et al. Recommender Systems[J]. *Physics Reports*, 2012, 519(1): 1-49.

[16] 乔治·奥威尔. 一九八四[M]. 傅惟慈, 董乐山, 译. 沈阳: 万卷出版公司, 2010.

[17] 卢克·多梅尔. 算法时代: 新经济的新引擎[M]. 胡小锐, 钟毅, 译. 北京: 中信出版集团, 2016.

[18] 维克托·迈尔-舍恩伯格. 删除: 大数据取舍之道[M]. 袁杰, 译. 杭州: 浙江人民出版社, 2013.

[19] LIU J, DOLAN P, PEDERSEN E R. Personalized news recommendation based on click behavior[C] // The 15th International Conference on Intelligent User Interfaces, Hong Kong, China: ACM Press, 2010: 31-40.

[20] AKHTER N. Relationship between internet addiction and academic performance among university undergraduates[J]. *Educational Research Review*, 2013, 8: 1793-1796.

[21] RAI L, CHUNRAO D. Influencing factors of success and failure in MOOC and general analysis of learner behavior[J]. *International Journal of Information and Education Technology*, 2016, 6: 262.

[22] WHITE K R. The relation between socio-economic status and academic achievement[J]. *Psychological Bulletin*, 1982, 91(2): 229-248.

[23] GONZALES N A, MARI C A, et al. Family, peer, and neighborhood influences on academic achievement among African-American adolescents: One-year prospective effects[J]. *American Journal of Community Psychology*, 1996, 24: 365.

[24] JOACHIMS T, LI H, et al. Learning to rank for information retrieval Now publishers[J]. *Foundations and Trends® in Information Retrieval*, 2009, 43(2): 41-45.

[25] ZHOU T, KUSCSIK Z, et al. Solving the apparent diversity-accuracy dilemma of recommender systems[J]. *PNAS*, 2010, 107(10): 4511-4515.

编辑 何婧