

# 融合数据伦理的数据科学与大数据技术专业培养分析

李实于凤

(东北林业大学 黑龙江 哈尔滨 150000)

**【摘要】**在国家教育部的指导下和人力资源市场的需求下,数据科学与大数据技术及相关专业已经在600余所高等学校开设,对于这一新兴专业的培养模式已成为高等教育的研究热点。而数据伦理问题是培养方案中一个亟待提升的领域,刚开始受到高等教育学者和工程教育专家的关注。本研究首先分析了数据伦理的一般问题,然后结合国内外文献和案例提出了在课程体系中建立数据伦理相关课程思路,并且提出了在培养模式中融合数据伦理理念的策略。

**【关键词】**数据科学与大数据技术;数据伦理;课程设置;培养方案

## 一、引言

数据伦理学是数据科学学科中一个快速发展但尚未真正起步的研究领域<sup>[1]</sup>。数据科学和大数据领域的研究人员和教学人员越来越意识到道德问题至关重要。随着数据源数目的增长,以及数据格式的多样性增加,数据生成、获取、分析、处理以及输出将呈更为夸张的几何级数增长。但我们在大数据热潮中也需要冷思考,特别是正确认识和应对大数据技术带来的伦理问题,以更好地趋利避害<sup>[3]</sup>。目前,大数据引起的伦理问题已经在公共领域出现了很多影响巨大的事件。例如2016年Facebook利用数据影响选举的传闻导致了数家社交媒体平台的政策发生了实质性变化,数个知名公众人物被禁。

作为未来大数据行业的从业者,数据科学与大数据技术专业的学生培养应该为即将到来的数据伦理发展做好准备。在前面研究中,Elliott等人<sup>[4]</sup>探讨了如何教授数据科学伦理。一些数据科学教科书已经开始包括伦理章节<sup>[5]</sup>。美国国家科学院、工程院和医学院数据科学高等教育圆桌会议讨论了如何在统计学、数据科学和计算机科学中教授伦理学。Grosz等人描述了道德教育是如何融入哈佛大学计算机科学课程的<sup>[6]</sup>。

本文正是在以上研究基础上,首先分析大数据技术中的常见伦理问题,然后针对这些伦理问题探讨如何在数据科学与大数据技术专业增加数据伦理教育内容。

## 二、大数据技术中的常见的伦理问题

目前大数据技术中常见的伦理问题依据文献<sup>[2]</sup>,主要分为以下几种:数字身份、隐私、访问、信息安全、数据鸿沟。

数字身份和隐私问题经常在一起讨论。围绕数字身份有两个伦理问题<sup>[2]</sup>:一是身份盗用的事件层出不穷。例如美国国土安全局广泛监控、窃取和收集国内外公民的海量信息。二是获取数据及数据处理能力的技术发展下,数字身份越来越可追溯。例如现在社交网络上的“网暴”现象。这种情况如果不加以管理,人们可能会为了保护自己,拒绝实名信息或退出网络,这样对于网络和信息的使用是一种倒退。数字身份的盗用往往伴随着隐私的泄露。大数据技术具有随时随地保真性记录、保存、还原等强大功能。个人的身份信息、行为信息、位置信息甚至情感与社交关系等隐私信息,都可能被记录、呈现<sup>[7]</sup>。2020年的新冠疫情中,很多地区使用了行程码,这些信息亟待需要各种

周全的政策进行维护。大数据时代,个人隐私的非常容易丧失,甚至在被合理采集后,被二次甚至多次利用<sup>[8]</sup>。

互联网的迅猛发展,随之而来的也有垃圾信息泛滥,包括诈骗、色情、虚假广告等信息充斥着互联网的各个角落。这就涉及到了访问权利的问题。一方面是如何确保使用者信息访问的正当权利,另一方面如何防止恶意访问,包括垃圾邮件、色情广告、兜售药品等十分严重的问题,已引起业界和政府的关注。

习近平总书记强调指出,没有网络安全就没有国家安全,没有信息化就没有现代化。某些行业的信息比如金融数据、医疗信息以及政府情报等都可能因有保密措施不完善引起的信息安全问题,大数据的分析和应用会催生一些新的、需要考虑的信息安全问题。

在所有的国家,有一些人拥有社会提供的最好的信息技术。他们有最强大的计算机、最快的网络服务,也受到了这方面的教育。另外一部分人,他们出于各种原因不能接入最新的或最好的计算机、最快最方便的网络服务。这两部分人之间的差别,就是所谓的“数字鸿沟”<sup>[9]</sup>。例如,在2020年的新冠疫情中很多老人因为不会使用微信二维码,造成了出行困难。运用大数据技术要求必须拥有海量数据、设备和人才,这是普通人无法做到的标准。并且运用大数据技术获取利益较大者也并没有实现差额对较小获利者进行应有的补偿,商业大数据技术使用者在搜集消费者的个人数据时往往没有给予消费者应有的知情权,反而将消费者们的个人隐私数据进行泄露与贩卖。

## 三、将伦理和隐私问题纳入数据科学教育

为了消除大数据引起的伦理风险,目前除了构建完善的法制体系,还需要加强技术创新和技术控制。毕竟解铃还须系铃人,大数据技术带来的伦理问题,最有效的解决办法就是推动技术创新。作为数据科学与大数据技术专业的教育从业者,有责任培养技术过硬,伦理清晰的优秀数据人才,有必要在让学生毕业前掌握在未来的职业应用中应该具备伦理思考的能力,特别对于工程类的人才这也是满足“新工科”培养的需要。

美国国家科学院、工程院和医学院(2018年)包括数据科学本科课程建议如下:

(1) 伦理学是一个主题,鉴于数据科学的性质,学生应该在整个教育过程中学习理论,并加入实践。专业培养应确保伦

理从一开始就融入到数据科学课程中,并贯穿始终。

(2) 数据科学界应该采纳一个道德准则;这样的准则应该由专业协会成员确认,纳入专业发展计划和课程,并通过最终的教学培养方案进行实施。

(3) 代码应该经常根据新的发展情况重新评估。

(4) 数据科学与大数据技术专业中,我们可将伦理学的学习融入到高年级的必修课,课程中约25%的内容涉及数据科学伦理学。

目前对于大数据相关专业如何教授数据科学伦理的成功模式还很少。事实上,在数据科学教学方面的相关工作很少提到伦理<sup>[10]</sup>。这种现象的主要原因是教育工作者通常在人类受试者研究的伦理方面训练有素,但很少有人有算法偏见,甚至一般伦理哲学方面接受过明确的培训。伦理问题其实不是一个技术问题,而是“技术对社会影响的一般性问题”<sup>[11]</sup>。我们可以通过合作来弥补培训的不足,例如与哲学家和伦理学家一起开发一个适合专业培养的伦理课程。结合文献<sup>[12]</sup>借助于美国高校案例分析探讨了如何在数据科学专业中融入伦理教育。

本文结合中国高校的实际情况,提出了包含六个模块的数据科学伦理教学举措,力图辅助各数据科学与大数据技术专业将数据伦理充分整合到培养方案中:

(1) 建议在大数据相关专业的培养目标或者课程目标中增加以下相关内容:以知情的方式评估基于数据的研究、分析和技术的伦理影响。利用专业指南、机构审查委员会和已发表的研究等资源,明确信息伦理道德责任。

(2) 将道德融入所有的课程,在高年级课程中让学生充分接触数据伦理学习。一方面可以安排相应的课程,例如直接开设由数据科学专家和伦理学家共同研发的数据伦理课程。另一方面可以结合课程思政教学改革,在专业课授课过程中,融合数据伦理相关案例和思想。

(3) 支持学生参与涉及数据科学伦理的课外实践活动,例如参与各种创新或科研项目,在其中进行数据的获取、分析等活动时,考虑数据伦理分析。现在的大学生对于数据科学中的伦理问题是非常感兴趣的。他们经常被称为“互联网的原住民”,本身就有一些数据伦理的意识,例如在网上共享个人数据的安全意识,获取资源时候考虑版权等等。在实践类课程中融入伦理学是培养学生对该学科兴趣的另一种方法。将数据科学中的伦理问题界定为未解决的专业问题,有助于学生想象自己在某种程度上对该领域做出了有意义的贡献。

(4) 邀请各类专家进入校园,可以是数据伦理理论研究者,也可以是社会企业人士,进行大家所关注的伦理问题的演讲。

(5) 对于应聘专业教师的人员,要求参与数据科学伦理的能力培训,并将培训结果作为招聘标准。

#### 四、结论

作为大数据时代技术和工程人员的摇篮,数据科学与大数据专业的学习和研究离不开数据的应用,为了这以专业和学科长期健康的发展,我们需要完善这一学科的各方面知识和技能。现在数据科学应用领域可见的偏见,道德失误或对道德的严重漠视都会导致数据科学产品的应用对社会造成伤害。我们的学

生是新一代数据科学家和数据工程师,他们将解决这些问题,恢复人们对于数据驱动技术发展的信心。为了做到这一点,我们作为教育从业者,专业人员培育者需要把数据伦理考虑作为这一学科发展过程中不可或缺的一部分。希望本文所做的分析和建议可以实现数据科学技术和人文社会和谐共生,实现科技与人文的统一。

#### 参考文献:

[1] Wender, B. & Kloefkorn, T. (2017), Roundtable on Data Science Postsecondary Education: A Compilation of Meeting Highlights, The National Academies of Sciences, Engineering, and Medicine.

[2] 邱仁宗, 黄雯, 翟晓梅, 大数据技术的伦理问题. 科学与社会, 2014(01)

[3] Davies, H. (2015), Ted Cruz using firm that harvested data on millions of unwitting Facebook users, The Guardian., 2015(11)

[4] Elliott, A. C., Stokes, S. L. & Cao, J. (2018), Teaching ethics in a statistics curriculum with a cross-cultural emphasis, The American Statistician 72(4), 359-367.

Hardin, J., Hoerl, R., Horton, N. J., Nolan, D., Baumer, B. S., Hall-Holt, O., Murrell,

[5] P., Peng, R., Roback, P., Temple Lang, D. & Ward, M. D. (2015), Data science in statistics curricula: Preparing students to think with data, The American Statistician 69(4), 343-353.

[6] Grosz, B. J., Grant, D. G., Vredenburg, K., Behrends, J., Hu, L., Simmons, A. & Waldo, J. (2019), Embedded EthiCS: Integrating ethics across CS education, Commun. ACM 62(8), 54-61.

[7] 杨维东, 有效应对大数据技术的伦理问题, 人民日报, 2018年3月23日

[8] 刘晓凡, 大数据背景下工程师的伦理责任分析, 电脑知识与技术, 2018 11(14): 11-13

[9] 陆攀攀, 信息技术的负面影响, 合作经济与科技, 2011(12)

[10] Kaplan, D. (2018), Teaching stats for data science, The American Statistician 72(1), 89-96.

[11] Bruce, K. B. (2018), Five big open questions in computing education, ACM Inroads 9(4), 77-80.

[12] Benjamin S. Baumer and Randi L. Garcia and Albert Y. Kim and Katherine M. Kinnaird and Miles Q. Ott, Integrating data science ethics into an undergraduate major, 2020.6, arXiv: {2001.07649}

#### 作者简介:

李实, 副教授, 东北林业大学 信息与计算机工程学院。

黑龙江省高等教育教学改革资助项目(SJGY20190031); 东北林业大学教育教学改革资助项目(DGY2019-38)