

心理学研究与大数据技术携手并进——大数据的应用领域越来越广泛

刘明子，卫桐圻

越来越多的心理学家正在学习如何构建和利用大数据，从而获得对人类行为的新见解。为此，他们正在学习分析技术与工具使用，特别是人工智能和机器学习，以便与大数据携手并进。

对于行为科学家来说，大数据的来源十分广泛，从传统的大型数据库到医疗记录，再到从手机、社交媒体和可穿戴技术收集的非结构化数据。新闻媒体公司 Axios 的社会心理学家和高级数据科学家 Sean Wojcik 博士表示，这些数据和相关的方法使研究人员比以往任何时候都能研究更多类型的构念和变量，并且比传统方法更具有探索性——采用推导生成假设，而不是假设检验。Wojcik 博士说，对研究人员来说，精通这些技能有很多好处，因为这些数据中可以学到很多东西。

以美国国立卫生研究院 (NIH) 的青少年脑认知 (ABCD) 研究为例，这是美国有关脑发育和儿童健康的最大规模长期研究。该研究“正在通过自陈问卷、认知测试、大脑成像和生物标本收集被试的大量数据。”在美国国家药物滥用研究所工作的 ABCD 项目主任 Gaythri Dowling 博士说，“它还引入了来自 FitBit¹和手机应用程序等来源的数据，以及基于被试居住地的现有数据集的环境数据。”Dowling 博士进一步表示，这个丰富的数据集将帮助我们更好地理解影响发展轨迹（从个人到家庭，再到社区和社会）的不同因素。

像 ABCD 这样的研究例证了如何科学地、遵守诚信地使用大数据。例如，ABCD 的监管者促进研究人员采用最佳实践方案，包括使用恰当的统计模型，解释研究结果时考虑伦理问题等等。但对于包括商业领域在内的其他企业来说，大数据的广阔前景有时会促使人们“走上歪路”。该领域的工作人员指出，总体而言该领域缺乏充分的监管。心理健康领域尚未解决的伦理问题，包括诸如对患者自主权和隐私权的尊重以及确保公平性。

“私人信息是隐私，这就是为什么它可能与心理健康功能如此密切相关，对研究人员和心理健康提供者如此有帮助的原因。”在线医疗服务提供商 Meru Health 的临床研究科学家 Benjamin W. Nelson 博士说，“因此，需要围绕这些数据的使用制定强有力的法规和伦理规范。”

¹ 该公司主要研发可穿戴设备和健康数据记录 APP。

为什么大数据如此热门?

尽管对大数据的使用存在一些担忧,但因为大数据方法有可能帮助包括心理学家在内的科学家提高现有发现的准确性和广度并挖掘新发现,所以获得了越来越多的支持。亚利桑那州立大学研究机器学习技术的教授 Kevin Grimm 博士表示,通过这些技术,研究人员可以获取大量的数据集并揭示其复杂的关联性,包括非线性和交互式的关联。例如机器学习,他说,“你不需要预先设定太多(你想检验的关联性或变量)——此模型在本质上可从数据中学习其关联性。” Grimm 提到,这些技术还可以加强传统研究。他解释说:“一旦你使用传统技术分析了数据,你就可以使用机器学习技术来发现你可能没有意识到的其他关联性。这可能会引出新的假设和理论,然后可以用一组新的数据来验证。”

Wojcik 指出,另一个潜在的好处是:大数据有助于解决心理学的可重复性问题。“针对较小样本的心理学研究的有效性,以及研究人员是否能够复制一些众所周知的实验效应已经开展了许多讨论。” Wojcik 说,“有了更大的数据量,你通常可以对你正在测量的实验效应的有效性和可靠性更有信心。”他说,“这增加了这些数据可能适用于更广泛人群或实验未涉及人群的可能性。”

抓住机遇

鉴于大数据应用的潜力,许多心理学家正在寻找创造性地利用大数据及相关方法的方式。例如,在 Axios 公司, Wojcik 正在使用机器学习技术来分析语言模式,目的是将新闻内容传播给更多人。同样,他也曾在 Upworthy 工作过几年,那是一个数字化的“正面故事讲述”平台。在那里,他与盖茨基金会(Gates Foundation)合作,利用社会科学和数据科学帮助全球健康、贫困和其他重要话题相关的报道走红。

其他人正在将这些技术应用于传统的研究领域,既出于商业目的,也为了扩大他们在解决重要问题方面的发现。俄勒冈大学的临床心理学教授 Nick Allen 博士,是一个青年心理健康研究人员。例如在使用大数据分析来解决心理健康护理问题中,他看到了一个常见问题:孩子经常在疗程间隙经历心理健康危机,而此时他们最需要帮助却缺乏服务提供方。在一个由程序员和数据科学家组成的团队的帮助下,他研究出了如何通过青少年的手机传感器捕捉他们在疗程间隙提示心理健康问题的潜在指标,比如睡眠模式、身体活动和社交互动的变化。这项工作最终促成了 Allen 博士的公司 Ksana Health 的成立,该公司提供的产品允许研究人员和心理健康临床医生通过手机数据和大数据分析,将持续的行为监测融入到他们的项目和实践中。该研究软件已经被全球的研究人员广泛使用,包括 ABCD 研究。这款临床软件将在今年年初推出,目前正在与一些临床合作伙伴进行测试。

与此同时，Nelson 正在帮助 Meru Health 使用大数据技术，与研究团队和数据科学家一起，为数千名正在经历抑郁、焦虑和倦怠的人提供循证护理。与 Allen 的平台类似，Meru Health 收集个人数据和其他自陈报告指标，以整合患者症状和治疗 App 使用情况的信息。Nelson 解释说，这些信息反过来帮助 Meru Health 的执业治疗师了解何时应该干预患者，并为患者提供与他们病情相关的数字化提醒。鉴于大数据在该企业中的作用，利用大数据改善医患日常互动直接支持了 Meru Health 实现其远大目标，即“到 2027 年治疗一千万人”。

基础研究人员也在充分利用大数据。儿童和青少年精神病学家、国立精神卫生研究所 (NIMH) 发育神经基因组学部门的负责人 Armin Raznahan，利用诸如英国生物样本库、健康大脑网络和 ABCD 研究等大型数据集来分析行为、基因、以及神经成像的数据，以更好地理解健康个体的脑组织如何随着性别、基因型和行为而变化。然后，他利用这些研究发现来阐明造成心理健康困境的风险因素。

由于 NIMH 的所内研究项目为 NIH 的研究人员提供了大数据分析方面的统计帮助，Raznahan 已能够收集和深入研究来自约 1 万名 17 岁以下的年轻人的多模态数据。例如，最近 2000 多份大脑扫描结果的分析强有力的证明了人类脑区的体积存在性别差异——这一发现之前只在老鼠身上出现过 (*Proceedings of the National Academy of Sciences*, Vol. 117, No. 31, 2020)。Raznahan 指出，这些研究发现对我们理解认知、行为和精神疾病风险方面已经确立的性别差异具有重要意义。Raznahan 补充说，这些发现如此值得信赖的一个原因是，研究人员能够将他们的发现与来自英国生物样本库的数千张不相关数据集的图像进行比较，发现它们具有高度一致性。这是大数据的另一个潜在优势。

加入大数据的应用行列

虽然一些心理学家对大数据越来越熟悉，但对另一些人来说，大数据似乎是一座需要掌握新概念和技术、高得令人生畏的山。除了学习使用大数据集，大数据分析还包括学习获取、管理和分析这些数据的多种技术。

幸运的是，有很多可用的资源来学习这些技术。一方面，专科和大学正在这些领域提供更多的课程、项目，甚至学位。虽然心理学系并不经常提供这样的课程，但学生们通常可以在其他系学习到这些内容。Wojcik 说“这有帮助他们完成更多想做的事情，掌握更多可用工具。”Wojcik 指出，学习编程语言，例如 Python、R 或 SQL 能够迅速收集数据，可以成为这项工作的基础支持。

与此同时，远程教育提供了丰富的培训机会，包括 Neuromatch、Statistical Horizons 和 Stats Camp 等夏令营，以及其他培训和认证项目。你也必须准备好进行大量的自学，Grimm

补充道，但在线论坛的社区支持可以帮助解决自学中碰到的各种问题。

Wojcik 说，心理学家无论是自学还是与数据采集专家合作，对大数据感兴趣都是明智的（选择），因为大数据有潜力改善研究和整个心理学领域。“作为一名心理学家，我发现这些来源丰富的、有意义的行为数据也是十分有趣的，至少与我在实验室收集的常规调查数据相比是同样有趣的。”

原文标题：

Big data ups its reach

原文链接：

<https://www.apa.org/monitor/2022/01/special-big-data>

中科院心理所潘毅图书馆