

大数据背景下大学生数据素养教育探讨

马晓慧

(山西大学商务学院, 山西 太原 030031)

摘 要: 在大数据背景下, 大学生作为未来社会数据使用的主体, 其数据素养水平将关乎整个社会的发展。本文讨论了数据素养的能力框架, 包括能力指标、能力项目、对应的学习内容和需要掌握的技术工具, 并在此基础上从教育内容和教育路径两方面对大学生数据素养教育进行了探讨。通过数据素养的能力框架的构建和数据素养教育路径的研究, 为数据素养教育提供思路借鉴, 为大学生数据素养教育实践提供参考。

关键词: 数据素养; 能力指标; 教育路径

中图分类号: TP399 **文献标识码:** A

Discussion on Data Literacy Education of Undergraduates under Big Data Background

MA Xiaohui

(Business College of Shanxi University, Taiyuan 030031, China)

Abstract: In the context of big data, as the major group of future social data users, the data literacy level of college students is closely related to the development of the whole society. This paper discusses the data literacy framework, including the ability indicators, ability projects, corresponding learning content and tools to be mastered. On this basis, it explores the data literacy education of college students from the perspectives of educational content and educational path. Through the construction of the data literacy framework and the research of data literacy education, this paper provides reference for data literacy education and college students' data literacy education practice.

Keywords: data literacy; ability index; educational path

1 引言(Introduction)

近年来, 在开放数据运动的背景下, 公共领域中统计、科学和技术性数据的数量呈指数增长。另一方面, 日常生活中产生的数据(如: 通过电子媒介)也越来越多, 影响着人们的行为和决策。数据素养正日益成为衡量社会及个人加工处理数据的基本意识和基本能力标准。高校学生作为未来社会数据使用的主体, 其数据素养水平将关乎整个社会的发展。在此背景下, 本文探讨了数据素养的定义、内涵、能力指标和学习框架, 以期高等教育中的数据素养教育的开展提供参考。

2 数据素养(Data literacy)

素养(Literacy)一词的传统含义是一个人的读写能力。现在“素养”一词有了更广泛的含义, 经济合作与发展组织(OECD)将其定义为特定领域的知识和能力, 包括使用语言、数字、图像、计算机等基本手段了解、沟通, 获得有用的知

识, 解决问题, 通过技术评估复杂背景的能力^[1]。

目前关于数据素养(Data Literacy)的定义, 基本是从数据管理、数据利用、数据意识等角度出发进行梳理, 有以下几种代表性看法: Mandinach等提出数据素养包含了数据的解读和运用技巧, 能够有效地利用信息指导决策制定^[2]。Prado等人认为, 数据素养包含个人获取、理解、批判性地评估和管理数据的能力, 并在使用数据的过程中遵守道德规范^[3]。郝媛玲等人认为数据素养是具备数据意识和数据敏感性, 能够有效且恰当地获取、分析、处理、利用和展现数据, 并对数据具有批判性思维的能力^[4]。虽然各定义出发视角不同, 但可以看出, 数据素养的概念是信息素养在大数据背景下的发展与延伸, 是收集、管理、评估和合乎规范的应用数据的能力。

国内外的一些组织机构都提出了在不同教育阶段进行数据素养教育的思路, 将提高公众收集数据、分析数据、处理

数据和利用数据的能力作为当前的一项重要工作。高校作为教育的一线阵地，在开展大学生的数据素养教育，进而提高公众的数据素养方面起着重要的作用。

3 数据素养的能力指标和内容框架(Ability index and content framework of data literacy)

在大数据背景下，起决定因素的不是数据的数量，而是搜索、聚合、分析、交叉引用数据集的能力。数据素养教育不仅存在于自然科学中，也存在于社会科学、人文科学等领域中。它对所有学科、所有学习环境，以及各个层次的教育具有共性。在目前本科教育中，由专业知识和专业技能构成的专业基础素养影响度远远高于数据素养，但是，随着对数据素养重要性的认知逐步深入，数据素养也逐渐引起了学者和教育机构的重视。开展有效的数据素养教育，首先需要对接数据素养的核心能力和学习内容框架进行分析和鉴别。

Prado和Marzal认为数据素养包括理解数据、发现和获得数据、解释和评估数据、管理数据、利用数据等几个方面的能力，并对每个能力做出了具体阐释。Carlson等人结合ACRL信息素养能力标准，将数据素养能力划分为：数据的发现和获取、数据管理和组织、数据变换与互操作性、数据质量保证、元数据、数据的重用、数据保存、数据分析、数据可视化，数据伦理等^[5]。国内学者刘彩娥、胡卉等人从数据的意识层面、知识层面、应用能力层面，以及伦理与道德层面对数据素养的内涵进行了解读^[6,7]。从现有研究成果可以看出，学者们对数据素养应包含的能力达成了普遍的共识，但对应的能力指标和培养内容并没有统一的标准，评估方法也有所不同。

3.1 数据素养能力指标

本文综合了国内外文献对数据素养能力、内涵的分析，结合数据素养的教育实践，将数据素养能力指标分为划分为两个层面，分别为数据意识层面和数据使用层面，每个层面对应的能力指标如图1所示。

数据意识			数据使用				
数据需求意识	数据使用意识	数据法律意识	数据获取能力	数据整理能力	数据分析能力	数据呈现能力	数据应用能力

图1 数据素养核心能力指标

Fig.1 The key ability indexes of data literacy

3.2 数据素养内容框架

在数据意识层面，着重建立对数据的需求意识、价值意识、法律意识和使用意识。要求能明确个人的数据需求并能够清晰描述，能根据需求寻求数据源并进行采集和使用，使用过程中要合乎法律规范，加强数据保护意识和版权意识。数据意识模块指标中对应的内容框架见表1所示。

表1 数据意识层面的内容框架

Tab.1 Content framework at the level of data awareness

指标	对应的能力描述	内容及技术工具
数据需求意识	能够描述对数据的需求	能描述需求数据的类型、采集工具、时间范围、格式、用途等
	能够了解数据的来源	能够了解数据源有哪些，可以根据数据需求去寻找对应的数据源，比如行业指数、数据监测、研究报告和官方发布数据等
数据使用意识	意识到数据的地位和价值	数据能够减少信息不对称、改善客户服务、提供决策支撑等
数据法律意识	建立对知识产权、隐私保密的基本认知，能够合乎规范的使用数据	用户拥有他们对自己数据的控制权，数据发布及处理时要采用必要的技术保护用户隐私，如：数据匿名发布技术、数据随机扰动、平移翻转技术、数据的角色访问、加密访问控制技术

在数据使用层面，按照数据的使用流程可以分别为数据获取能力、数据整理能力、数据呈现能力、数据应用能力。每一种能力指标对应的学习内容和工具技术如表2所示。

表2 数据使用层面的内容框架

Tab.2 Content framework at the level of data use

指标	对应的能力描述	内容及工具
数据获取能力	熟悉数据库和搜索引擎，能够熟练获取各类数据	能够根据搜索引擎、数据库、综合数据平台、专题数据平台、第三方统计数据、专业调研机构报告、市场调查或自写爬虫得到所需数据
	能够对不同形式的数据(格式、类型)进行解读	类型：表格、点集、时间序列、图像、视频、网络数据等 数学结构：度量结构、网络结构、拓扑结构、函数结构等 格式：CSV、JSON、XML等
数据整理能力	能够按照条件过滤原始数据、剔除错误和无效数据、组织与简化数据	缺失值处理、垃圾信息清理、特征变量转化、特征选择和维度变化(升维或者降维)、标准化\归一化\稀疏化
	能够转换不同数据格式，以便于后续加工工具处理	根据分析需求掌握相关的数据格式转换技术，如：通用数据库格式转换工具、GIS数据格式转换工具、JSON、XML、CSV数据格式转换等
数据分析能力	熟悉本学科的基本分析工具，利用软件进行分析给定数据	数据透视表、SPSS、SAS、Hive、Hadoop、R、Python、Java
	准确表述结论	提炼结论、理解业务、懂算法
数据呈现能力	使用统计量描述数据	集中趋势：均值、中位数等 离散程度：极差、方差、标准差、变异系数等 分布形态：二项分布、多项分布、超几何分布、指数分布、偏度系数、峰度系数等
	使用图表表征数据	饼图、柱形图、条形图、折线图、气泡图、散点图、假色图、平行左边图等
	熟悉本学科的基本可视化工具	Tableau、FineBI、Plotly、Echarts、Power Map、Visual.ly、WolframAlpha、Google Charts等
数据应用能力	能够依据数据结果支撑论点、将结果应用于决策	使用数据结果优化决策

4 数据素养教育路径(Educational path of data literacy)

数据素养教育最重要的目标之一是培养数据思维,使我们能够理解数据,解读数据,远离数据陷阱。数据素养在不同的学科领域具有不同的表现形式,不同的学科和专业对学生的数据素养能力要求也不尽相同。结合我国数据素养教育的现状,大学生的数据素养教育应以数据意识为先导,数据通讯技能为基础,提高个体的数据道德观念和获取、处理、应用能力。

4.1 数据素养教育内容通识化

对大学生群体的数据素养教育应该从通识层面入手,教育内容可以围绕数据意识和数据使用技能两个方向来展开。

(1)数据意识的培养

数据意识是数据素养的前提,包含了对数据需求、数据使用、数据安全和数据伦理的敏感度。数据意识的强弱决定了使用者是否能有效地发现数据,高效的处理和利用数据,积极地传播和分享数据成果,能否在数据的检索、验证、复制、引用时树立正确的法制观念,增强数据安全意识和数据隐私保护意识。除此之外,数据意识的培养教育还应着力强化学生数据的价值观念、自主学习观念,引导学生善于利用数据提高专业能力。

(2)数据使用技能的培养

大学生群体对数据的需求和使用主要通过课程论文、学年论文、社团活动、创新项目训练、学科竞赛等形式体现。学生参与的社会实践活动、创新训练项目、数学建模、电子设计大赛等竞赛,对数据素养能力也有较高的要求。所以,在数据使用技能层面,虽然学生的专业背景可能不同,但基本应当掌握以下一些数据使用技能:了解不同数字对象有不同的数据结构、理解数据收集方法和传播渠道、能够寻求满足不同信息需求的数据源、理解数据的存储策略、能够基本判断数据的质量、能够使用相关软件简单的分析给定数据集并给出可视化图例,使结果应用于决策。

4.2 数据素养教育途径线上化

提高大学生的数据素养,应积极尝试数据素养教育实践。目前,学校、公共图书馆和高校图书馆越来越多的参与到了数据素养的教育领域,进行专业的参考咨询服务或数据能力培训^[8,9]。除了传统的线下教育,国外的高校在具体的实践过程中也引入了优秀的在线课程,充分利用线上的开放资源结合学生的专业背景展开数据素养教育。

美国的NNLM(National Network of Libraries of Medicine)网站上开放了专门的数据素养课程资料、数据素养指南和参考等资源。表3列出了NNLM的数据素养教育部分的在线课程资源。除此之外,还有对其他学科专门设置的数据课程资源。

表3 NNLM部分在线课程资源

Tab.3 Some online course resources on NNLM

课程名称	地址	主要内容
研究数据管理基础(华盛顿大学)	https://canvas.uw.edu/courses/889213	为科学、健康科学和工程学科的本科学和研究生水平的学生开发。课程设计为一系列七个一小时课程模块,基于NECDMC课程。包括学习目标、演示文稿、阅读和练习
新英格兰协作数据管理课程	https://library.umassmed.edu/resources/necdmc/index	基于案例的数据管理课程,用于教授健康科学,科学和工程领域的学生,包含六个学习模块
研究数据管理(俄勒冈州立大学)	https://figshare.com/articles/GRAD_521_Research_Data_Management_Syllabus_and_Lesson_Plans/1003834	帮助学生在日益复杂的数字研究环境中为数据规划、文档编制、保存和共享提供超出资助者要求的准备。向所有学科的学生开放
翻转教室RDM课程	https://sites.google.com/a/umn.edu/data-management-workshop-series-instructor-site/	包括五个解决常见数据管理问题的方案,由明尼苏达大学的图书馆员设计
研究数据MANTRA(爱丁堡大学)	https://mantra.edina.ac.uk/	MANTRA是一个免费的在线非评估课程,包括八个在线单元和四个可下载的数据处理实用程序:R、SPSS、NVivo和ArcGIS

德国的Stifterverband基金会会对高等教育数据素养教学路径进行了研究,并于2018年9月发布了《未来技能:高等教育中数据素养的培养方法》^[10],其中也列出了数据素养教育的诸多在线资源,具体见表4所示。Qlik的在线学习平台提供了数据素养的免费在线课程和学习软件。

表4 数据素养教育的部分在线课程资源

Tab.4 Some online course resources of data literacy education

名称	课程提供方	链接
大数据分析的在线课程	哈索普列特纳研究所	https://open.hpi.de/courses/bigdata2017
数据素养和数据可视化	俄亥俄州立大学/iTunes U	https://itunes.apple.com/us/course/data-literacy-and-datavisualization/id693097601
数据科学精要	EDX/微软	https://www.edx.org/course/data-science-essentials-microsoftdat203-1x-5
数据科学在线研讨会	华盛顿数据科学中心	https://www.datasciencecentral.com/video/video/listFeatured
专业化数据训练	StackFuel	https://stackfuel.com/de/
不同的在线教育模块	Data One @ University of New Mexico	https://www.dataone.org/education-modules
谷歌认证专家-数据工程师	谷歌	https://cloud.google.com/certification/data-engineer
微软数据科学专业证书	微软	https://www.edx.org/microsoft-professional-program-data-science
数字变换在线课程:	哈索普列特纳研究所	https://mooc.house/courses/msg-mdt1
应对大数据的挑战	麻省理工学院	https://mitxpro.mit.edu/courses/coursev1/MITProfessionalX+6.BDx+2016_T2/about

(下转第41页)