

大数据专业本科生实践能力与科研素养协同培养模式研究

刘京昕¹, 彭 军^{1*}, 李 茜¹, 金尚柱², 寇喜鹏¹, 何 勇¹

¹重庆科技大学数理科学学院, 重庆

²重庆科技大学信息化中心, 重庆

收稿日期: 2025年7月7日; 录用日期: 2025年8月18日; 发布日期: 2025年8月27日

摘 要

随着大数据技术的快速发展, 培养兼具实践能力和科研素养的大数据专业人才已成为当下高等教育的一项重要任务。本文探索如何通过系统的实践技能和科研素养训练, 提高大数据专业本科生的综合能力。首先, 实践能力的培养涉及数据采集、处理、分析、可视化等核心能力的培养, 通过真实项目案例, 引导学生在实际问题中提升专业技能。其次, 科研素养方面重在培养学生的科研思维, 鼓励学生围绕大数据课题从选题、文献查阅、数据分析到论文撰写展开系统性训练, 以培养严谨的科研态度与创新思维。最后, 结合当前教学实践, 本文从课程设计、实践项目及科研活动等维度提出具体建议。提出的方案旨在优化大数据专业的教学模式, 培养具有实践能力和科研素养的应用研究型人才。

关键词

大数据技术, 实践能力, 科研素养, 高等教育

Research on the Collaborative Training Model of Practical Ability and Scientific Research Literacy of Big Data Major Undergraduates

Jingxin Liu¹, Jun Peng^{1*}, Xi Li¹, Shangzhu Jin², Xipeng Kou¹, Yong He¹

¹School of Mathematical and Physical Sciences, Chongqing University of Science and Technology, Chongqing

²Informatization Office, Chongqing University of Science and Technology, Chongqing

Received: Jul. 7th, 2025; accepted: Aug. 18th, 2025; published: Aug. 27th, 2025

*通讯作者。

文章引用: 刘京昕, 彭军, 李茜, 金尚柱, 寇喜鹏, 何勇. 大数据专业本科生实践能力与科研素养协同培养模式研究[J]. 创新教育研究, 2025, 13(8): 582-588. DOI: 10.12677/ces.2025.138635

Abstract

With the rapid development of big data technology, cultivating big data majors with practical ability and scientific research literacy has generated an important task of current higher education. This paper explores how to improve the comprehensive ability of big data major undergraduates through systematic practical skills and scientific research literacy training. Firstly, practical skills development focuses on core competencies such as data collection, processing, analysis, modeling, and visualization, utilizing real-world project cases to enable students to elevate their technical proficiency by solving practical problems. Secondly, scientific research literacy focuses on cultivating students' scientific thinking. Students are encouraged to engage in a systematic process encompassing topic selection, literature review, data analysis, and academic paper writing centered around big data research topics, thereby cultivating a rigorous research attitude and innovative thinking. Finally, drawing upon current teaching practices, this paper proposes concrete recommendations regarding curriculum design, practical projects, and research activities. The proposed program aims to optimize the teaching model for big data programs and cultivate applied research talents with practical ability and scientific research literacy.

Keywords

Big Data Technology, Practical Ability, Scientific Research Literacy, Higher Education

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 研究背景

大数据作为新型生产要素，已成为推动社会经济发展与科技进步的核心驱动力，在引发行业变革的同时，也对数据科学人才提出了更高要求。作为一门交叉学科，大数据专业涵盖了数据采集、存储、处理、分析和可视化等多个领域。为应对大数据技术的发展趋势和市场需求，培养具有较强实践能力和科研素养的专业人才成为当前大数据教育改革的紧迫任务。2021年11月，工业和信息化部发布《“十四五”大数据产业发展规划》，提出了一系列重点任务，包括加快数据要素市场的培育，充分发挥大数据的独特优势，夯实产业基础，构建高效稳定的产业链，打造繁荣且有序的产业生态，并强调要加强数据安全保障和数据治理。同时，规划指出应通过技术创新与产业融合，推动大数据价值转化，注重培养兼具信息技术与业务能力的复合型人才。2025年4月，国家发展改革委和国家数据局发布《2025年数字经济发展工作要点》，提出要以数据要素市场化改革为主线，推动隐私计算、区块链等数据流通与利用基础设施的先行试点，促进数字经济学科建设，优化高等院校大数据专业人才培养模式。近年来，全球范围内的大数据技术发展迅速，产业需求持续增长，各大高校相继开设了大数据专业课程，旨在培养具有计算机科学、统计学、数学和信息管理等多学科交叉知识的复合型人才。然而，现阶段的新工科教育模式仍面临一些亟待解决的问题，针对大数据专业来说，本科教育往往注重理论知识的传授，而对学生实际操作能力和科研创新能力的培养不足，导致很多毕业生在进入工作岗位后缺乏解决实际问题 and 独立开展科研课题的能力[1]。大数据专业本科生的实践能力和科研素养对于他们未来的学术研究、行业应用和创新创业非常重要。因此，如何协同培养大数据专业本科生的实践能力和科研素养，已成为大数据专业高等教育领域的一项重要议题。

2. 现状及挑战

2.1. 专业实践能力

大数据技术的应用范围涵盖医疗、金融、制造业、公共安全和政府管理等各个行业。为培养符合这一社会需求的专业性人才，众多高校针对大数据专业开设了数据分析、机器学习、人工智能等相关课程。这些的课程开设为学生基础理论知识的储备提供了必要的保障，但学生在专业实践能力的培养往往容易被忽视[2]。目前，大多数大数据专业的本科课程主要还是侧重于理论教学，尤其是针对大一大二的本科学生，他们主要学习数据结构、数据库原理、线性代数、概率论与数理统计等基础课程。这些课程主要集中于对计算机组成原理和数学背景的解释，很少涉及大数据技术及其代码编译的实际操作。这种“重理论、轻实践”的脱节现象，使得许多毕业生在进入工作单位后无法迅速适应实际的工作环境和需求，难以将所学的理论知识自然地应用到具体的工程项目中。另一方面，虽然部分高校针对大数据专业中开设有实验课或工程实践课程，但这些课程的内容大多较为简单，项目的规模和复杂度无法与现实中的大数据问题相匹配[3]。许多学生参与的实践活动往往局限于模拟数据或实验环境，缺乏对实际数据的处理与分析能力。这导致毕业生普遍缺乏实际项目经验，难以独立应对行业中真实的大数据挑战。

2.2. 科研素养

在大数据专业学生的本科教学中，科研素养的培养普遍较为薄弱。大多数大数据专业的本科生在完成课程学习后，并未参与到真正的科研项目中。大数据的学科特点决定其对学生跨学科知识储备的要求较高，不仅需要计算机科学的技术基础用于算法设计，还需要统计学的理论工具用于分析推理，包括人工智能和机器学习等方法用于模型构建。然而，当前大部分学生在本科阶段并未系统地接触到科研项目，也缺乏独立进行科研探索的机会[4]。尽管部分高校开设有数据分析方法、算法原理、科研写作等科研相关的课程，但此类课程的讲解深度和训练广度普遍不足，大部分仅停留在基础层面，缺乏对科研的系统训练[5]。大数据的研究领域涉及深度学习、自然语言处理、人工智能等前沿技术，这些技术的学习不仅要求学生掌握理论知识，更要求学生具备创新的思维和解决复杂问题的能力[6]。然而，当前大数据专业的本科课程依然主要局限于基础的理论课程教学，难以有效地激发学生的创新意识和科研兴趣。因此，如何扩大人工智能与大数据专业教师队伍，提升学生的综合素质，并培养其在大数据应用中的创新能力，已成为当前大数据人才培养面临的一项关键挑战。

3. 大数据实践能力和科研素养的重要性

3.1. 大数据专业对科研素养的要求

大数据的主要特征体现在容量大(Large Capacity)、速度快(Fast Speed)和多样性(Variety)，涉及对大量数据集的处理和分析。对于传统的数据处理工具来说，这些数据集往往过于复杂。当下社会各个行业和组织都依赖大数据做出明智的决策、预测趋势、优化运营，甚至创建全新的商业模式。从医疗到金融，从教育到城市规划，大数据都有可能改变现代生活的方方面面。因此，该领域对专业人员的技能要求不仅仅是单纯的数据处理和统计分析，还要求专业人员有开发大数据工具的基本能力。科研素养是涵盖批判性评估研究、设计新方法、追踪前沿进展及高效发现问题等能力的综合技能。对于大数据专业的本科生来说，科研素养的培养对于他们理解人工智能基础理论、了解大数据技术领域的最新发展都有着积极意义。

3.2. 提高批判性思维的训练

培养大数据本科生的科研素养显著优势之一是培养他们的批判性思维和独立解决问题的能力。相较

于传统学科，大数据领域的科研问题往往具有多维度、跨学科的特征，这要求学生必须突破单一视角局限，从技术实现、业务需求、社会影响等多维度拆解问题本质，并系统整合跨领域信息源中的有效内容。这一过程不仅有助于学生更好地理解大数据技术的应用价值，更促使他们对当前研究的局限之处展开深入的思考。在研究型课程中，应鼓励学生使用优于已有的方案去解决相应的数据挖掘与分析问题。他们的任务包括自定义研究问题、提出假设，并设计实验或模型来测试他们的想法，这个过程的锻炼是对他们创造力和创新能力的重要提升。对于大数据专业本科生，科研素养重在激发他们质疑假设、提出新问题、探索新方法的能力。此外，大数据研究经常涉及不确定性和模糊性。数据集可能是嘈杂的、不完整的或有偏见的，传统的方法可能并不总适用于当前挑战。通过科研素养的培养，学生应学会如何有效地应对这些挑战，应用适当的方法在面临不确定性时得出合理的结论。

3.3. 职业道路的铺垫作用

科研素养的重要性超越了学术研究本身，为大数据专业学生在学术界、工业界及政府部门等多元职业路径上的长远发展奠定坚实基础。在数据驱动行业中，大数据公司亟需具备科研素养的专业人才，要求其不仅能胜任数据分析，更能推动新算法工具的研发。具有科研素养的大数据专业人员能够更好地理解现有技术的局限性，并找到解决复杂问题的创造性方法。例如，从事医疗保健工作的数据专家可能需要设计研究医疗数据的方案，或为个性化的医疗诊断开发高效的算法。从事金融领域工作的数据工程师可能需要探索有关实时欺诈检测的新方法，或改进数据存储和数据检索系统。在此类场景下，科研素养都能使大数据专业人才为其组织的发展做出积极贡献，并在不断变化的技术领域保持领先地位。在学术界，科研素养对于希望攻读研究生学位或追求学术事业的本科学生来说是不可或缺的。设计实验和进行独立研究的能力是他们成功的基础。在本科阶段培养他们的科研素养非常重要，研究型的本科课程能为他们提供必要的技能和经验，引导性地让他们参加研究课题、研读相关前沿文献，激发他们的学术思维。另外，在公共政策方面，大数据研究正越来越多地被用于为公共健康、城市规划和环境可持续性发展等问题的决策制定。政府机构根据大数据专业人士的建议和改进方案来制定政策，并改善公共服务。严格的科研素养训练使得大数据专业学生能够切实体会到自身的成长，并基于实践和证据的研究结果为公共部门的相关部署提供可参考的方案。

4. 训练方案设计

大数据技术的快速发展改变了众多行业，未来该领域的专业人员不仅需要具备实用技能，同时也需要发现问题和开发算法解决问题的能力。为满足这一持续演进的行业需求，本培训方案旨在为大数据专业的本科生提供必要的知识、技能和工具，以在现实世界的应用和学术研究中茁壮成长。本训练方案旨在通过一系列的实践课程、项目和科研训练，全面提升大数据专业本科生的实践能力和科研素养，为他们步入职场和继续深造打下坚实的基础。总体的训练方案设计板块和结构安排见表 1。

Table 1. Overall training program course section
表 1. 总体训练方案课程板块

课程模块	主要内容	学时周期
数据采集与预处理	数据抓取、清洗、预处理技术	理论 20 学时，实践 40 学时
数据分析与建模	数据分析、模型选择与优化	理论 20 学时，实践 40 学时
大数据处理技术	Hadoop、Spark 等框架	理论 20 学时，实践 40 学时
科研与创新思维	科研方法与学术写作	理论 20 学时，实践 40 学时
团队协作与项目管理	项目管理、团队合作	实践 40 学时，展示与答辩 20 学时

4.1. 数据采集与预处理

大数据专业学生需要掌握不同数据源的采集方法,包括结构化数据和非结构化数据的获取方式。基于 Twitter API、Google Analytics API 等开放的 API 接口采集数据,学生可以通过 Python、Java、C++ 等编程语言编写代码获取在线数据,掌握数据采集的常用技术。通过对 Python 的 Scrapy、BeautifulSoup 等工具库的学习和使用,有助于学生对 HTTP 协议的理解,及时掌握静态和动态网页数据的技巧爬取技巧。对于物联网或实时监测类大数据项目,学生可以通过 ESP32、Raspberry Pi 等传感器和嵌入式设备获取实时数据,进行数据流的采集和存储。在数据采集过程中,需要对数据进行标准化,确保数据格式一致。对于多源数据,采用统一的 UTF-8 编码格式和 ISO 8601 格式的数据时间戳进行标记,方便数据分析与处理。

针对数据缺失问题,可采用插值法或删除法进行处理。对于数值型数据,根据已有数据进行插值填补。若缺失值占比较小,可删除相关记录。对于连续性的缺失数据,使用均值或中值填充。若数据属性中存在异常值,可通过箱型图、标准差以及 Z-score 值来判别。针对不同的属性值,偏离均值超过一定倍数的数据视为异常,Z-score 大于 3 或小于 -3 的样本也通常被视为异常值。

4.2. 数据分析与建模

数据探索性分析(Exploratory Data Analysis, EDA)是数据分析中的一个重要环节,旨在通过对数据的深入探索和可视化,发现数据的特征、结构和内在关系。学生需基于基本的统计分析与数据挖掘方法,快速获取数据的维度、缺失值情况、分布类型等基本信息。在此模块中,应引导学生如何借助直方图、箱型图、散点图等图表工具对数据进行初步分析,如何借助均值、标准差、偏度、峰度等描述性统计识别数据的潜在模式和趋势。

设计相应的数据可视化课程,帮助学生通过可视化手段理解数据的分布特性,进行数据标准化或归一化的处理。通过该模块的训练,学生应学会借助 Matplotlib、Seaborn 等工具绘制直方图、饼图、折线图,观测数据的变化趋势,以及如何借助热图、散点矩阵图等可视化结果分析变量间的相关性。

4.3. 大数据处理技术

大数据处理技术涵盖了从数据存储、处理、计算到分析等多个环节,要求学生掌握高效处理海量数据的技术和方法(主要环节及架构见图 1)。首先,该模块需要介绍 HDFS (Hadoop Distributed File System) 等分布式文件系统的基本原理,帮助学生理解大数据存储的可靠性、可扩展性和容错性。教授学生如何使用 HBase、Cassandra 等 NoSQL 数据库,理解它们与传统关系型数据库的区别,掌握如何根据数据特点选择合适的存储系统。

在数据管理环节,主要目的是让学生掌握如何使用 Zookeeper 分布式服务框架和 HDFS 分布式文件系统。同时,需指导学生掌握在 NoSQL 数据库中使用特定查询语言或 API 进行数据分析的方法。学生通过该模块的训练,需要掌握主流的大数据处理框架及其应用。相应的模块课程包括 Hadoop 生态系统和 Spark 与流式处理框架的使用训练。Hadoop 生态系统课程教授学生如何使用 Hadoop 的 MapReduce 编程模型进行数据处理,帮助学生理解 MapReduce 的工作原理、编程流程及优化技巧。通过 Hive 和 Pig 等高级数据处理工具,让学生熟悉 SQL 类语言在 Hadoop 上的使用,快速处理结构化数据。引导学生利用 HBase 进行数据存储和检索,学习 Spark 的内存计算优势与分布式处理能力。Spark 与流式处理框架课程重在讲解 Spark 的 RDD(弹性分布式数据集)与 DataFrame 等基本概念,以及如何通过 Spark 进行高效的批量数据处理,使得学生能够基于 Spark Streaming 在实时数据流的环境中进行大数据处理。

基于多线程和多进程的基本概念,引导学生在单机环境中实现数据处理的并行化。借助 MapReduce

模型，引导学生掌握分布式计算中任务分解、并行执行与结果整合的机制。该模块课程涉及数据划分、任务调度、节点协调等分布式计算的基本概念，引导学生如何利用分布式系统提高计算效率。通过 Apache Kafka、Flink、Storm 等流处理框架的学习，学生能够理解相关处理方式和状态管理方法，能够使用 Apache Druid、Elasticsearch 等实时分析工具实现对流数据的处理。

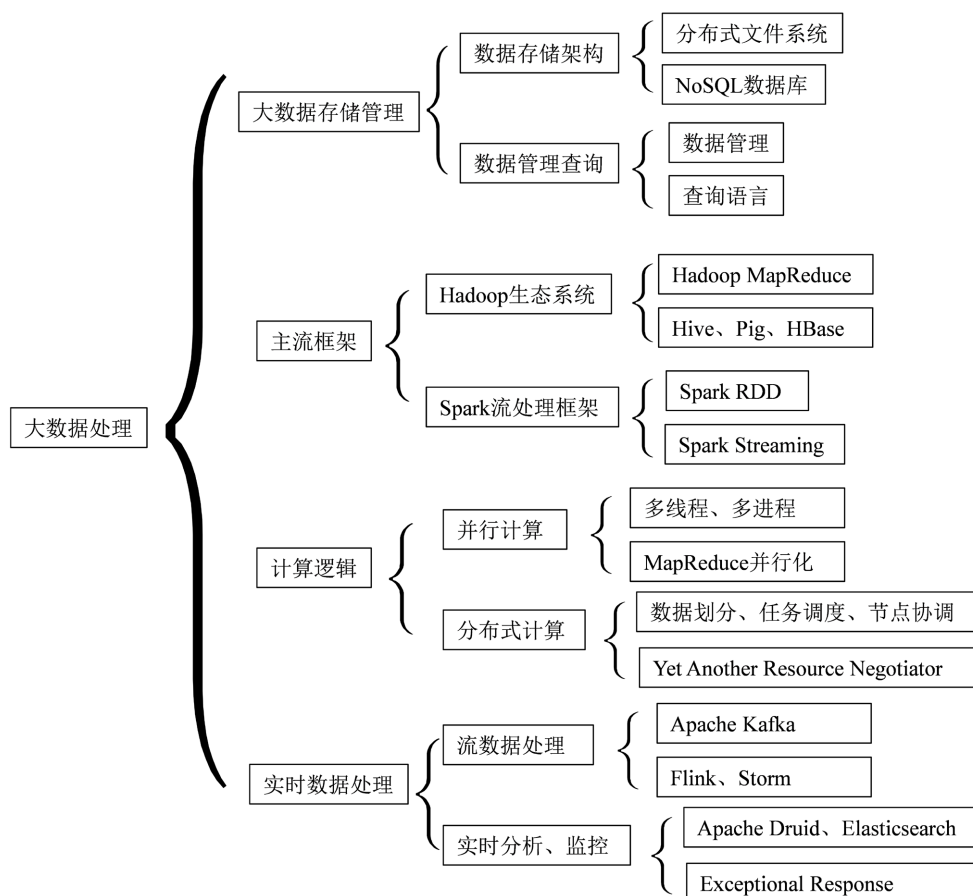


Figure 1. Main links and architecture of big data processing
图 1. 大数据处理主要环节及架构

4.4. 科研与创新思维培养

科研思维的培养需要从基础知识的系统学习开始，课程设置应覆盖大数据领域的基本概念、核心知识模块、技术框架及应用场景，尤其是在数据挖掘、机器学习、数据分析等核心内容上构建扎实的理论基础。同时，应注重课程的跨学科融合，强化学生对于数据科学、统计学、计算机科学以及人工智能的理解，从而提升学生的综合分析能力。该模块课程教授学生如何进行文献检索、如何阅读和分析已有的科研论文，并通过文献综述训练学生的批判性思维和创新思维。讲解科研论文的撰写技巧，包括如何编写研究背景、研究方法、实验设计、数据分析和结论部分。鼓励学生将科研成果撰写成论文，争取在学术会议或期刊上发表。学校可以组织定期的学术沙龙或论文答辩会，为学生搭建成果展示平台并提供专业反馈。

学术交流是科研能力提升的重要途径。学校应积极组织学生参加学术会议、论坛和科研活动等，培养学生的学术交流能力。通过与业内专家的接触，学生可以了解到前沿技术动态、获得专业的意见反馈，

从而拓展学术视野。此外，学校还应鼓励学生积极参与国内外科研合作项目，尤其是跨校、跨国的合作研究项目，为学生提供更广泛的学术资源，创造协同创新的机会。

4.5. 团队协作与项目管理

该模块课程设计主要通过组织团队合作项目，引导学生通过集思广益提出创新性科研构想，并通过协作共同推进方案实施。这有助于学生拓宽思维、加强团队协作能力，并培养其解决复杂问题的能力。鼓励学生与来自不同学科背景的同学合作，通过多角度思考问题，激发学生的创新灵感。

项目驱动的学习方式有助于学生科研实践能力和创新能力的提升。通过设置实际项目或科研课题，学生可以在实践中培养科研能力，发现问题并解决问题，培养创新思维。项目内容应根据学生的专业背景和兴趣进行合理设置，例如学生可以借助 Python、R、Hadoop 等工具进行数据分析、数据清洗和数据可视化，进一步提高科研能力和创新思维。围绕人工智能与机器学习算法有针对性地向学生们分发研究型项目作业，引导学生组队开展项目研究，开发创新性的算法或优化现有算法。同时，鼓励学生将科研成果撰写成学术论文的形式，争取在学术会议或期刊上发表。学校可以定期组织学术沙龙或论文答辩，增长大数据专业本科生的科学见闻，并为他们提供专业的指导。

5. 结语

在大数据时代背景下，社会各行各业对大数据专业人才的需求日益增长，特别是具有较强实践能力和科研素养的人才。所以，面向大数据专业本科生的高等教育务必注重实践能力的培养与科研素养的提升，这将有助于他们进入职场、有效开展个人工作和团队项目。实践能力的训练是大数据专业本科生培养过程中不可或缺的部分。学生在数据采集、处理、分析以及可视化方面的技能需要大量的实践操作来深化。科研素养的培养侧重于提升学生的学术思维和研究能力。学校可以通过开设科研课程，组织学生参加学术会议和科研项目，鼓励学生参与导师的科研课题，以此逐步激发科研兴趣并锤炼实践能力。大数据专业本科生的实践能力和科研素养的训练绝非孤立任务，而应与大数据课程体系、教学模式以及教育资源进行深度融合，实现协同推进。

基金项目

本文受以下基金项目资助：重庆市高等教育教学改革研究项目(232138)；重庆市高等教育学会高等教育科学研究课题(CQGJ23118C)；重庆市教育科学规划青年课题(K24YY2150009)；重庆科技大学本科教育教学改革研究项目：构建创新实验班探索大数据专业人才培养新模式(202308)。

参考文献

- [1] 付坤, 王瑞, 杨罕, 等. 高校本科生科研素养培养教育探索[J]. 实验室研究与探索, 2017, 36(3): 207-211.
- [2] 张乐, 陈岩松, 田英爱. 大数据专业本科生科技论文写作教学方法改革研究[J]. 创新教育研究, 2024, 12(8): 257-262. <https://doi.org/10.12677/ces.2024.128532>
- [3] 押男, 黄欣, 黄健. 新时代高校青年教师关键能力和素养模型构建及其应用[J]. 教师教育研究, 2024, 36(5): 90-97.
- [4] 蔡迎春. AI 驱动的科研范式变革: 跨学科视角下人工智能素养与教育培养策略研究[J]. 图书馆杂志, 2024, 43(11): 20-33.
- [5] 王宪平, 麦慧君, 朱国晶. 教师科学素养的内涵特征, 构成要素与发展路径[J]. 教师教育学报, 2024, 11(6): 8-17.
- [6] 欧阳峰, 白丽, 潘鹏. 应用型本科生数据素养能力综合评价——以大数据分析与应用相关专业课程为例[J]. 教育进展, 2023, 13(12): 10220-10229. <https://doi.org/10.12677/AE.2023.13121578>