

面向数据科学应用：大数据专业 C 语言课程教学改革与实践探索

袁磊^{1,*}, 高德鹏¹

1. 南通理工学院, 用友数智学院, 江苏 南通, 226002

摘要: 随着大数据时代的到来, 数据科学与大数据技术专业应运而生。C 语言作为一门底层、高效的编程语言, 在该专业的知识体系中扮演着奠基者的角色。然而, 传统的 C 语言教学普遍存在与专业背景脱节、重语法轻应用、缺乏数据思维导向等问题, 导致学生难以建立 C 语言与后续大数据核心技术 (如 Hadoop、Spark 等) 的关联, 学习动力不足。本文旨在分析数据科学与大数据技术专业的特点及其对 C 语言课程的内在需求, 指出当前教学中存在的痛点, 并系统性地提出一套以数据科学应用为导向的教学改革方案。该方案涵盖教学目标重构、教学内容重塑、教学方法创新、实践环节强化以及考核体系改革, 旨在将 C 语言教学从单纯的语法学习转变为培养学生计算思维、数据抽象能力和系统级编程能力的重要载体, 为其后续专业课程学习及解决复杂数据问题奠定坚实基础。

关键词: 数据科学与大数据技术; C 语言; 教学改革; 计算思维

Towards Data Science Applications: Teaching Reform and Practical Exploration of C Language Courses in Big Data Majors

Lei Yuan^{1,*}, Depeng Gao¹

1. Nantong Institute of Technology, School of Yongyou Digital Intelligence, Jiangsu, China, 226002

Abstract: With the advent of the big data era, the Data Science and Big Data Technology major has emerged. As a low-level, efficient programming language, C plays a fundamental role in the knowledge base of this major. However, traditional C language instruction often suffers from disconnection from the professional context, an emphasis on syntax over application, and a lack of data-driven thinking. This makes it difficult for students to connect C with subsequent core big data technologies (such as Hadoop and Spark), leading to low motivation to learn. This paper analyzes the characteristics of the Data Science and Big Data Technology major and its inherent demand for a C language course, identifies the pain points in current instruction, and systematically proposes a teaching reform plan oriented towards the application of data science. This plan encompasses restructuring teaching objectives, reshaping teaching content, innovating teaching methods, strengthening practical sessions, and reforming the assessment system. It aims to transform C language instruction from simply learning grammar into a key vehicle for cultivating students' computational thinking, data abstraction and system-level programming skills, laying a solid foundation for subsequent professional coursework and solving complex data problems.

Keywords: Data science and big data technology; C language; Teaching reform; Computational thinking

数据科学与大数据技术是一门跨学科的复合型专业, 它融合了计算机科学、统计学、数学和领域专业知识, 旨在从海量、高维、多源的数据中提取有价值的信息^[1]。该专业要求学生具备扎实的

编程基础、强大的数据处理能力和系统的算法知识。在众多编程语言中, C 语言以其接近硬件、执行效率高、指针操作灵活等特性, 在系统编程、高性能计算、算法实现等领域具有不可替代的地位^[2]。许多大数据生态系统的底层核心组件(如 HDFS 和 Spark 的部分核心模块)均使用 C/C++ 开发, 理解这些系统的内部机制, 往往需要深厚的 C 语言功底。

国内绝大多数高校的数据科学与大数据技术专业都将 C 语言程序设计列为必修的专业基础课。其教学目标不应仅仅是让学生掌握一门编程语言, 更应是培养学生底层的计算逻辑思维, 以及解决实际数据问题的初步能力^[3]。然而, 反观当前该专业的 C 语言教学现状, 大多仍沿袭计算机科学与技术专业的传统模式, 存在明显的问题。教学内容与数据科学的应用场景严重脱节, 学生陷入繁琐的语法细节和脱离实际的练习题中, 无法感知 C 语言在大数据技术栈中的价值, 从而产生学习 C 语言有何用的困惑, 严重影响了教学效果和学习积极性^[4,5]。

1 传统 C 语言课程教学存在的问题分析

1.1 教学目标与专业需求脱节

传统教学大纲通常将熟练掌握 C 语言语法作为核心目标, 强调对变量、控制结构、函数、数组、指针等语法点的孤立掌握。对于大数据专业而言, 这一目标显得狭隘。学生还需要理解如何用 C 语言高效地处理数据、C 语言在数据流水线中的哪个环节能发挥优势, 而传统目标并未体现这种专业导向性。

1.2 教学内容缺乏数据科学语境

教材和课堂案例多集中于功能实现(如字符串复制)、游戏(如猜数字)或简单的管理系统(如学生成绩管理系统)。这些案例虽然经典, 但距离数据科学中的数据采集、清洗、存储、分析、可视化等核心环节相去甚远, 学生无法将所学知识与大数据这一宏大的专业背景建立有效连接。

1.3 实践环节薄弱, 脱离真实应用场景

上机实验题目往往是小而碎的语法练习, 如用指针实现字符串反转、用结构体存储学生信息。缺乏综合性、项目式的实践任务, 导致学生只见树木、不见森林。他们可能熟练掌握了指针的语法, 却不知道如何利用指针去构建一个高效的数据结构(如哈希表、B 树)来处理海量数据的索引。

1.4 教学方法单向灌输, 忽视计算思维培养

教学多采用教师讲、学生听的单向模式, 侧重于语法规则的讲解和代码的逐行分析。这种模式忽视了对学生计算思维的培养, 即如何将一个复杂的数据处理问题分解、抽象、模式识别并设计算法来解决。而这恰恰是数据科学家最核心的能力之一。

1.5 考核方式单一, 难以评估综合能力

期末考核通常由笔试或简单的上机编程构成。这种考核方式强调小知识点和某些程序的记忆容易导致学生考前突击、死记硬背, 无法真实反映学生运用 C 语言解决实际数据问题的综合能力、代码调试能力和项目协作能力。

2 以数据科学应用为导向的 C 语言课程教学改革方案

2.1 教学目标的重构

在知识层面上不局限于传统 C 语言教学的零散知识点, 更多地让学生熟练掌握 C 语言的核心语法与特性, 特别是与数据操作密切相关的部分, 如数组、结构体、指针、文件 I/O、动态内存管理等。

在能力层面上，要培养学生能够使用结构体或指针构建复杂的数据类型，以模拟现实世界中的数据对象的数据抽象能力；能够熟练编写经典算法并对各种程序进行时空间复杂度分析的算法实现与分析能力；理解程序在内存中的布局，培养对内存管理和性能优化的敏感性，为后续学习分布式系统原理打下基础的系统思维能力；能够针对给定的、小规模的数据处理任务，进行问题分析、算法设计、编码实现和测试验证的问题解决能力。在素养层面，激发学生对底层技术的兴趣，培养严谨、规范的编程习惯，建立软件工程的基本意识，并初步形成数据驱动的思维模式。

2.2 教学内容的重塑与案例设计

2.2.1 数据表示与基本操作（对应基础语法）

核心知识点： 变量、数据类型、运算符、控制流程。

数据科学案例：

案例（数据类型）：读取一个包含数值型和分类型数据的简单 CSV 文件，讨论如何用 int、float、char array 来正确表示这些数据。

2.2.2 数据聚合与批量处理（对应数组与函数）

核心知识点： 一维/二维数组、函数定义与调用、参数传递。

数据科学案例：

案例（数组与函数）： 编写一个程序，读取一个包含大量数值的文件到数组中，然后通过函数实现数据的集中趋势度量（计算平均值、中位数、众数）和离散程度度量（计算方差、标准差）。

2.2.3 数据记录的抽象与管理（对应结构体与指针）

核心知识点： 结构体、指针、动态内存分配。

数据科学案例：

案例（结构体与文件 I/O）： 定义一个 Student 结构体，包含学号、姓名、多门成绩等字段。实现从文件加载学生记录到结构体数组、计算每个学生的平均分和总分、再将结果保存回文件的功能。这模拟了数据 ETL（抽取、转换、加载）的简单过程。

2.2.4 高效数据处理的基石（深入指针与算法）

核心知识点： 指针数组、函数指针、经典算法。

数据科学案例：

案例（构建哈希表）： 利用结构体、指针和动态内存管理，实现一个简单的链式哈希表，用于快速查找学生信息。此案例极具挑战性，但能让学生深刻理解大数据技术中分区、键值对等核心概念的底层实现。

2.2.5 综合课程设计（项目式学习）

在课程末期，设立一个为期 2-3 周的综合课程设计，让学生以小组形式完成一个小型项目。项目选题紧扣数据科学，例如：传感器数据模拟与特征提取。模拟生成一批时间序列的传感器数据（如温度、湿度），然后计算数据的移动平均、最大值、最小值等特征，并进行简单的异常检测。

2.3 教学方法与手段的创新

2.3.1 项目驱动学习（PBL）

以综合课程设计为核心，将整个学期的知识点串联起来。让学生在做中学，在解决真实问题的

过程中主动探索和整合知识。

2.3.2 线上线下混合式教学

利用在线教学平台（如 MOOC、SPOC），将语法基础、代码演示等内容制作成短视频，供学生课前预习。课堂时间则更多地用于案例讨论、难点答疑、小组协作和代码评审，实现课堂翻转。

2.3.1 可视化辅助工具的应用

引入内存可视化工具（如 C Tutor）或图形化调试器，帮助学生直观地理解指针、内存分配等抽象概念，化解学习难点。

2.3.1 代码评审环节

在项目实践中引入简单的代码评审机制。学生之间互相审查代码的风格、可读性、效率和潜在错误。这不仅能提高代码质量，更能促进学生之间的交流和学习。

2.4 考核评价体系的改革

建立多元化的、过程性的考核评价体系，降低期末笔试的权重，增加实践能力和项目贡献的考核。

平时表现（10%）：包括出勤、课堂互动、在线学习记录。

实验作业（30%）：评价每个模块的编程作业完成情况，重点关注代码的正确性、规范性和创新性。

综合课程设计（30%）：评价项目成果，包括项目报告、源代码、程序演示。考核维度包括：需求实现度、代码质量、算法效率、团队协作、答辩表现。

期末考试（30%）：笔试内容从单纯的语法考察，转向更多结合场景的应用题和算法设计题，考察学生对知识的理解和综合运用能力。

3 讨论

本文分析了 C 语言在大数据专业中的重要地位与传统教学方法中的不足之处，同时突出指出了在大数据这一新兴专业下 C 语言课程应该具有的作用，从目标内容到教学方法与考评的改革可以促进学生通过 C 语言课程理解大数据专业的核心思想，通过引入数据处理等真实应用场景，能够激发学生的学习兴趣和内在动力。项目驱动的综合实践，则能有效锻炼学生实际解决复杂问题的能力。经过改革后的 C 语言课程必然会更加贴合大数据专业，为学生后续学习，乃至未来的职业生涯，打下坚实而有益的基础。

4 结论

（1）C 语言课程对于数据科学与大数据技术专业的学生而言，是培养学生底层计算思维、理解计算机系统工作原理、并最终驾驭庞大复杂的大数据生态系统的关键基石。

（2）本文提出的以数据科学应用为导向的教学改革方案，旨在打破传统教学的窠臼，通过重构教学目标、重塑教学内容、创新教学方法和改革考核体系，将 C 语言的学习与专业背景深度融合。

（3）未来的工作将集中于具体教学案例库的精细化建设、在线教学平台的完善以及教学效果的长期跟踪与评估。

参考文献

- [1] 张平, 武新乾, 石念峰. 数据科学与大数据技术专业人才培养模式研究——以河南科技大学为例[J]. 教育进展, 2023, 13(11): 8462-8467.
- [2] 基于计算思维培养的 C 语言课程教学改革实践. (2025). 教育学刊, 193-198.
- [3] 毛镠. 分组策略——高校“计算机基础”课程实验教学改革初探[J]. 海南大学学报(自然科学版中英文), 2009, 27(1): 100-103.
- [4] 赵瑶池, 胡祝华, 陈明锐, 彭金莲. 以计算思维为导向的大学“计算机基础”课程教学改革研究[J]. 海南大学学报(自然科学版中英文), 2014, 32(4): 383-388.
- [5] 耿张燕. 新工科视域下高校计算机教学改革策略[J]. Educational Theory and Research, 2025, 3(10): 91-93.

^{1,*} 作者简介: 袁磊 (1997-), 男, 硕士研究生, 研究方向: 数据科学与大数据技术。E-mail: 20239259@ntit.edu.cn。