

新工科下数据采集课程的校企 协同教学改革与实践^{*}

唐海涛^{1,2**} 李锐¹

1 桂林航天工业学院理学院, 桂林 541004
2 桂航大数据技术应用研究中心, 桂林 541004

摘要 针对当前《数据采集与网络爬虫》课程中存在的理论教学与实践应用脱节、校企协同参与不足等问题, 本文基于“新工科”背景与产业需求导向, 提出“三阶四维·校企·院际”协同驱动的模块化教学改革方案, 通过构建“知识点-企业案例实践拆解”三阶递进和“理论-实验-企业资源-考核”四维协同模式, 成功引入企业真实项目案例、视频 35 个, 企业导师参与课时占比从 0% 提升至 20%。实践结果显示, 2024 级学生 (相较于 2023 级) 平均成绩从 64.84 提升至 74.71 (+15.2%), 及格率从 74.21% 提升至 86.55%, 课程评教分数从 91.22 提升至 95.37, 最后, 深度融合航天遥感数据采集任务完成率显著提高, 配套企业云端开发环境使用率达 100%, 进一步提升了学生工程实战能力, 为新工科大数据相关专业课程改革提供了可参考的实践范例。

关键字 新工科, 数据采集, 网络爬虫, 教学改革

Collaborative Teaching Reform and Practice of Data Acquisition Courses in the New Engineering Paradigm

Haitao Tang^{1,2**}

lirui¹

1. School of Science
Guilin University of Aerospace Technology,
Guilin Guangxi 541004, China;
2. Research Center for Big Data Technology Application in Guat
Guilin University of Aerospace Technology
Guilin Guangxi 541004, China

Abstract—Addressing the gap between theoretical instruction and practical application and the lack of collaboration between industry and academia in the current course on data collection and web-crawling, this paper proposes a modular learning reform plan based on a three-level, four-layer, interinstitutional and inter-faculty collaborative framework. Built on the New Engineering paradigm and industry-driven needs, the plan introduces a three-level step structure (knowledge points - business case practice decomposition) and a four-level model (theory - experiment - business resource - assessment). A four-dimensional synergy model, this paper successfully integrates 35 real-life business project cases and videos, increasing the participation of business mentors from 0 to 20 percent of total class hours. Practical results show that the 2024 cohort achieved an average score of 74.71 (an increase of 15.2 percent from 64.84 in 2023) and that the pass rate increased from 74.21 to 86.55 percent. Course scores increased from 91.22 to 95.37. Finally, the completion rate of deeply integrated remote sensing tasks for aerospace has increased significantly, with a 100 percent use of the supporting enterprise cloud development environment. This has further increased the practical engineering skills of students and provided a reference practical model for curriculum reform in the new discipline of engineering related to big data.

Keywords—New Engineering disciplines, Data acquisition, Web crawlers, School reform

1 引言

大数据与人工智能时代, 数据采集与网络爬虫技术已成为获取多源信息的重要手段, 在智能体、金融互联网科技、舆情监测等领域发挥着基础性支撑作用。而当前存在部分高校在数据采集等相关课程的教学上

常存在教学内容滞后、工程化应用能力不足等问题^[1,2]。特别是在复合交叉专业型领域, 传统“按章教学”的模式以理论讲授为主, 应用实践内容通常局限于书本案例或单一问题模拟解决导向, 难以满足多行业对具备多源信息采集与分析综合能力的复合型人才需求。

针对上述问题, 本文借助课程教学改革背景, 结合垂直领域特色, 探索基于校企协作^[3]的“数据采集与网络爬虫”课程教学改革。通过与企业、学院的三方共同合作, 共建教学案例库, 将企业在舆情监控、招聘信息抓取、商品价格跟踪等数据采集任务中的实际问

*** 基金资助:** 本文得到 2023 年度桂林航天工业学院校企合作示范课程-数据采集与网络爬虫(SF2024010), 桂林航天工业学院校级教改基金项目(KX201914001, 2024JB14, JX201914002)资助。

**** 通讯作者:** 唐海涛 tanghaitao@guat.edu.cn。

题转化为项目式教学内容,推动学生在真实任务驱动下掌握核心技术技能。企业工程师参与项目讲解、技术指导与成果评审,进一步提升了教学的真实性与专业性。

本文提出的改革方案具有三个突出特点:首先,采用模块化设计重构课程内容,将通用爬虫技术与垂直领域数据采集等专业应用深度融合;其次,引入企业真实项目和数据资源,特别是偏热门领域(如电商价格比对、视频音频采集等)的特色数据集;最后,建立动态化评价机制,邀请行业专家参与教学评估。通过学院两届数据科学与大数据技术专业的学生中的开展教学改革实施对比。新模式较大提升了学生数据采集的实践能力与工程化素养,拓展了校企协同育人^[4-6]的广度与深度,为计算机、大数据类课程构建多元融合的实验教学体系提供了改革探索支撑。

2 数据采集与网络爬虫课程

2.1 本课程基本情况

数据采集与网络爬虫课程是计算机科学与技术、数据科学与大数据技术专业的专业必修课程,是技术实践性和应用性极强的专业课程。数据采集与网络爬虫技术是数据驱动决策和信息挖掘的重要手段,在互联网、电子商务、金融、媒体等领域有着广泛应用。无论是在大规模数据采集、信息整合,还是在特定领域数据分析中,数据采集与网络爬虫技术在推动信息智能化处理、提升数据价值、支持决策优化等方面都起着不可或缺的作用。

2.2 本课程教学与育人目标

(1) 课程教学目标

教学目标 1:掌握 HTTP 协议和 Web 请求基本原理,熟练使用 Python 库(如 Requests)实现数据请求与响应处理。

教学目标 2:具备 HTML/XML 数据解析能力,能运用 BeautifulSoup、lxml 等工具精准提取目标信息。

教学目标 3:掌握多模态数据存储技术,包括 MySQL、MongoDB 等数据库的存取与管理操作。

教学目标 4:能开发高效爬虫脚本并处理反爬机制(如 Scrapy 框架应用、代理设置及验证码破解)。

(2) 课程育人目标

育人目标 1:培养数据思维与工程实践能力,通过实战项目强化复杂问题的分析与解决能力。

育人目标 2:建立法律与伦理意识,在数据采集过程中恪守隐私保护与信息安全规范

育人目标 3:激发创新协作精神,通过团队项目

锻炼沟通能力与系统性思维。

育人目标 4:塑造持续学习理念,为数据科学领域的进阶研究奠定方法论基础。

3 融合企业真实案例的教学改革体系设计

3.1 课程教学改革体系设计总体思路

本次教学改革以“三阶四维”为核心框架,通过校企深度融合企业真实案例重构课程体系。“三阶”聚焦教学流程优化,将传统课堂拆分为知识点精讲、企业案例实践、即时反馈评价三个递进阶段(重点推进企业案例实践),确保理论到实践的闭环转化;“四维”强调资源整合,通过理论教学、实验设计、企业资源、动态考核四个维度协同发力,构建“学-练-评”一体化培养模式。改革以模块化知识点为最小单元,依托企业真实项目案例库和线上实验平台,实现教学内容与产业需求的精准对接,最终提升学生的工程实践能力和就业竞争力。具体课程教学改革体系设计结构如图 1 所示。

(1) 三递进阶段

① 阶段一:知识点讲授

知识点讲授阶段采用“模块化拆分+精讲互动”模式。每次课程(90分钟)的第一小节(45分钟)划分为三部分:前5分钟通过企业实际场景(如电商爬虫反爬案例)引入知识点背景;中间30分钟精讲2-3个核心知识点(如XPath选择器、动态页面抓取),每个知识点讲解10-15分钟,辅以代码片段实时演示;最后10分钟组织3-4人小组讨论,针对知识点应用中的难点(如IP封禁应对策略)展开头脑风暴。通过模块化设计,学生每节课均可掌握可落地的技术单元,避免传统教学中“满堂灌”的知识碎片化问题。

② 阶段二:企业真实案例实践

企业案例实践阶段依托校企共建的真实项目库展开。第二小节(45分钟)中,前15分钟学生观看企业提供的实战视频(如舆情监控系统爬虫开发流程);中间25分钟基于企业数据集(如链家房产API接口)完成编码任务,教师与企业导师联合答疑;最后5分钟通过企业实验管理平台提交报告,系统自动评测代码规范性(如PEP8标准)和功能完成度。案例覆盖多行业(金融、电商、政务),确保学生掌握跨领域数据采集能力。例如,某次实验要求学生爬取京东商品评论并集成至企业数据库,模拟真实数据分析需求。

③ 阶段三:即时讨论评价

即时反馈评价贯穿教学全过程。课堂讨论环节主要通过“问答”方式进行,具体操作则按照学期初所

有学生默认是100总分，然后学期中随机点名问答，答对不减分，答错一次减10分，以此类推，获得即时评价得分（占10%），实验报告通过企业平台自动生成性能指标（如爬取效率、数据清洗准确率，占20%）。企业导师每周审核项目成果，从工程实用性（如反爬策

略鲁棒性）角度给出评分（占20%），教师则侧重理论掌握度考核（占50%）。动态评价数据实时同步至教学管理系统，用于调整后续知识点讲解重点。例如，当80%学生在“动态页面爬取”实验中得分低于阈值时，下节课将增加Selenium框架的专项训练。

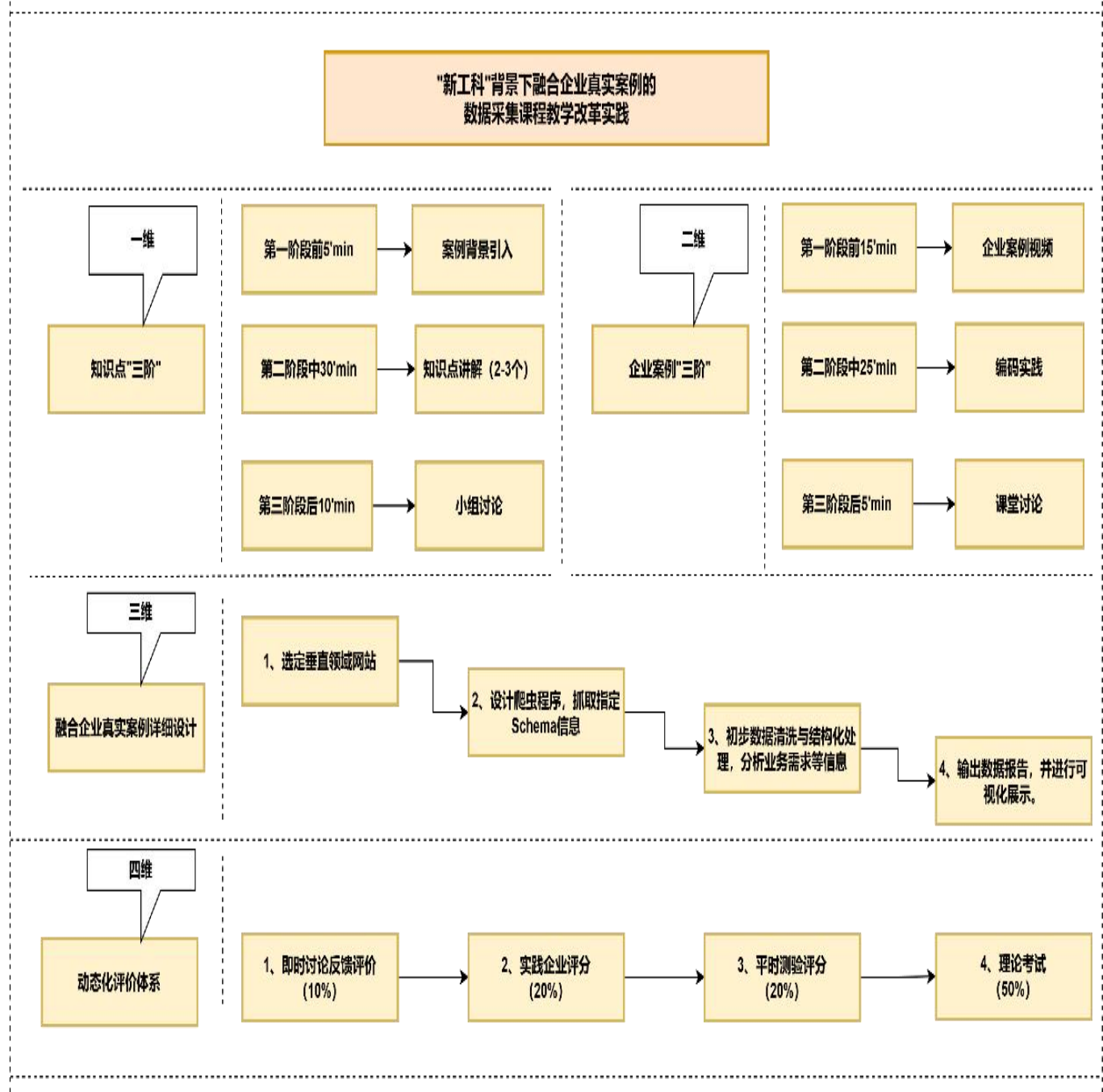


图 1 课程教学改革体系设计结构

(2) 四维度

① “理论”维度

理论教学维度强调“轻量、聚焦、可迁移”。课

程将传统教材内容重构为15个知识模块（如HTTP协议、Scrapy框架、反爬机制），每个模块匹配1-2个企业应用场景。例如，在讲解“User-Agent 轮换”时，结合携程旅游价格监控案例，分析其如何规避反爬策

略。理论考核采用“微答辩”形式，学生需针对随机抽取的知识点（如 Ajax 数据抓取）进行 5 分钟技术陈述，由企业专家与教师共同评分。

② “实验”维度

实验设计遵循“真实性、阶梯性、开放性”原则。基础实验（如静态网页爬取）使用企业脱敏数据；综合实验（如多源数据集成）直接对接企业生产环境（如智慧城市交通数据API）。实验难度逐级提升：L1级（单任务爬虫）→L3级（分布式爬虫系统）。开放实验允许学生自选行业数据集（如医疗挂号信息），最终成果可纳入企业案例库。某学生团队开发的“新冠疫情数据爬取与分析系统”被某大数据公司采用，作为其内部培训案例。

③ “企业资源”维度

企业资源深度嵌入教学全链条。合作企业提供三类支持：技术资源（如代理IP池、验证码破解API）、数据资源（如二手房交易数据集）、平台资源（如阿里云爬虫实训环境）。企业导师每月参与2次线上工作坊，讲解行业最新技术（如Playwright自动化工具）。学生可通过企业平台申领“悬赏任务”（如爬取指定电商平台

台SKU价格），优秀成果可获得实习推荐资格。

④ “考核”维度

考核体系采用“过程性+终结性”双轨模式。过程性考核（60%）包括实验报告（企业评分30%+教师评分20%）、课堂表现（10%）；终结性考核（40%）为综合项目开发，学生需组队完成从需求分析到部署的全流程，企业专家根据项目商业价值（爬虫系统稳定性）评分。

3.2 融合企业真实案例的教学模式设计

真实案例1：在获取企业在招聘信息监测与分析方面的真实需求，课程教学设计了“基于招聘网站的数据采集与分析”教学案例（见下图2）。该项目源于合作企业人力资源部门在岗位数据分析中的实际痛点，即难以高效获取不同平台上的岗位需求信息，进行职位趋势预测与竞争分析。

真实案例2：在获取企业在智能电商分析领域的业务需求，旨在通过定期采集主流电商平台（如京东、淘宝等）上的商品价格信息，实现商品价格波动趋势的监控与分析，为企业的定价策略与市场竞争分析提供数据支持（见下图3）。

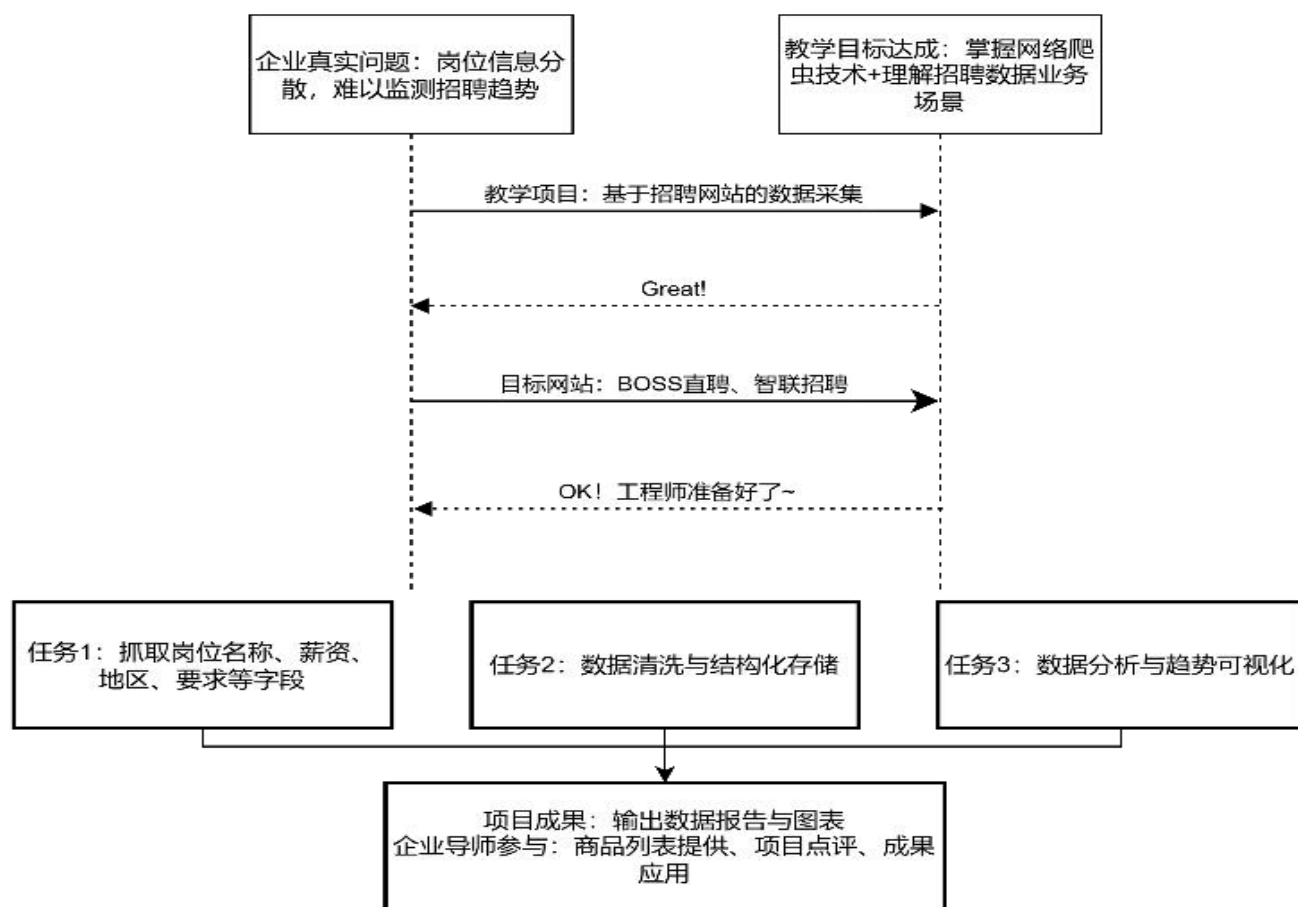


图2 融合企业真实案例1的教学模式设计

(1) 协作机制教学设计

本次课程教学改革重点构建了“校企-院际”多轨协同机制，融合多层次合作实现教学资源的优化配置。在校企协同方面，邀请有需求学院专任教师参与数据科学与大数据技术专业课程大纲制定，重点引入各垂直领域的日志分析、学科数据API接口规范等工业级需求，共建包含企业通用案例（如淘宝电商销售数据、社交舆情爬取）与垂直案例（如无人机遥感元数据采集、低空遥测数据集成）的双轨实验库。在跨学科协作方面，与院际建立深度合作关系，共享各垂直领域数据集，共同开发“垂直领域通用数据智能采集”实验模块，形成大数据技术与其他专业交叉融合的教学团队。这种协作机制既保证了技术教学的普适性，又凸显了专业数据采集特色。

(2) 分层实验体系优化

课程采用“基础-应用-综合”的三层次架构，本研究设计了融合通用垂直领域数据采集的实验体系。在基础层，除传统网页爬虫技术外，新增各垂直领域

认知实验，指导学生解析包括但不限于无人机元数据的XML/JSON结构特征。应用层采用双轨任务设计：一方面保留IT数据API集成等企业常规任务，另一方面兼顾各领域特色任务，如基于USGS/ESA开放API的遥感影像索引爬取、机械故障日志分析等。综合层则组织学生开展跨学科项目实战，既可选择开发传统电商爬虫系统等通用项目，也可参与多源数据自动化采集平台等特色项目，所有成果均需企业提供的线上平台接受三方指导老师的综合评审。

(3) 分层实验体系优化

本改革实施了创新的“三师协同”教学模式：由校内教师负责爬虫技术原理讲解，企业导师演示工业级应用案例，各学科专业教师指导相应数据集特性分析。在项目实施环节，学生可自主选择通用或各领域方向，但需完成交叉性任务要求。同时为支持教学实施，搭建了包含企业云端爬虫沙箱、垂直领域数据处理中间件在内的资源共享平台，并建立了双方案例库的持续更新机制。这种实施路径充分整合了多方资源，形成了“技术通用化、应用特色化”的教学特色。

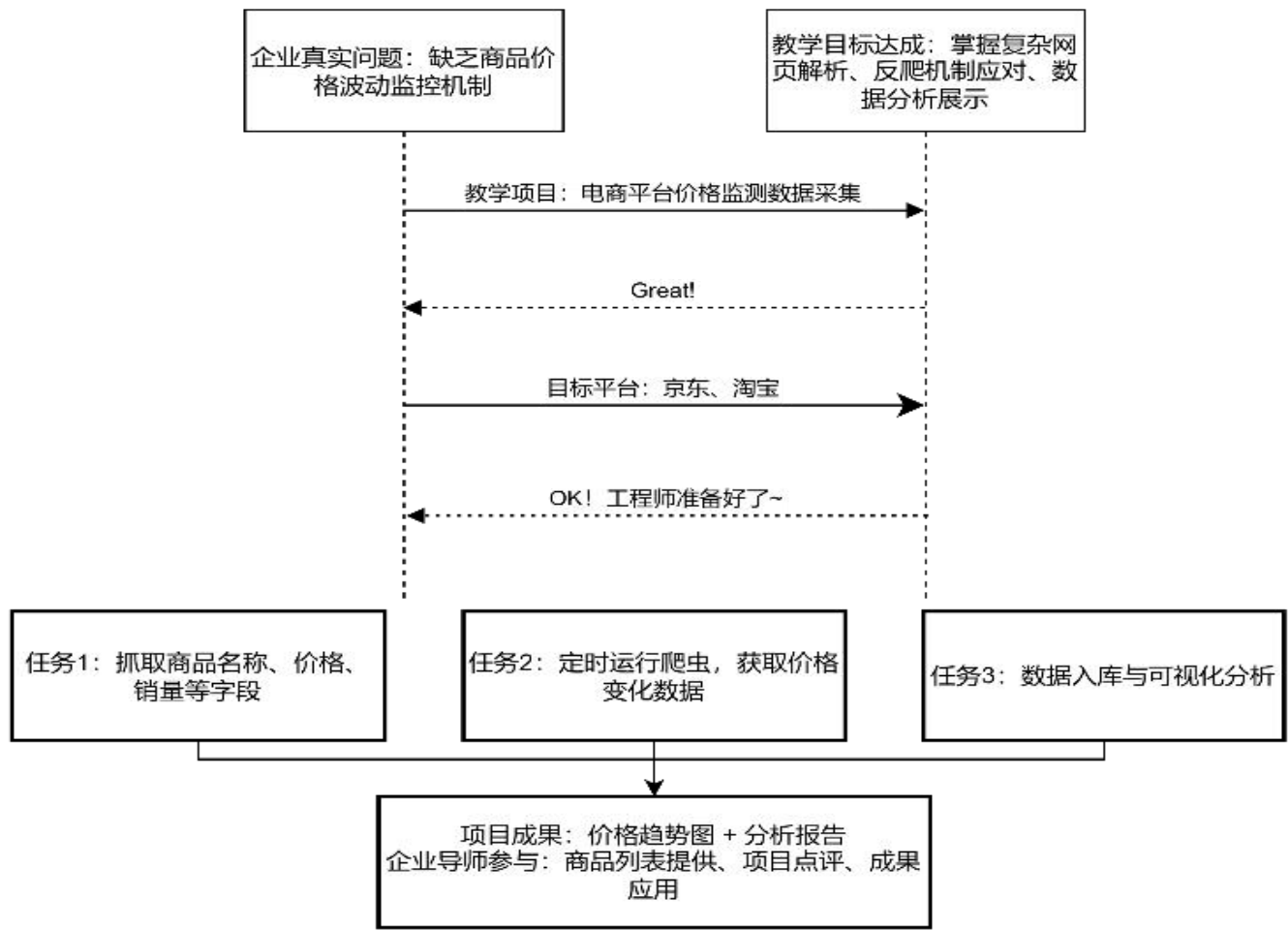


图 3 融合企业真实案例 2 的教学模式设计

4 课程教学改革分析及效果
(23 年 vs. 24 年)

4.1 成绩对比分析

系统导出两届数据科学与大数据技术专业课程《数据采集与网络爬虫》真实成绩为数据基石（具体可见下表 1）。统计数据显示：课程教学改革在提升整体教学质量的同时，也改善了成绩分布结构，虽然学生个体差异仍然存在，但各层次学生的成绩均获得实质性提升，特别是及格率的大幅提高和平均成绩的显著增长，进一步验证了“企业真实案例+学科交叉”教学模式的有效性。

4.2 预期效果

课程教学改革在 2024 级学生中实施后，预期目标的达成情况如下：

首先，在能力培养方面，自课程引进企业协作后，透过企业真实项目训练，大部分学生能独立完成复杂爬虫系统开发，尤其航天遥感垂直领域的数据采集任务的完成率得到极大提升，以此夯实了工程实践能力；另外企业提供的真实的实训项目融入了空间数据处理元素（如卫星影像坐标解析），丰富了传统教学模式，实现了跨学科应用能力的锻炼。

同时，教学资源建设方面，经本次教学改革实践成功引入企业真实项目案例 20 个（新增），教学视频 15 个（新增），线上考核平台 1 套（新增），企业提供的云端开发环境使用率达 100%（新增，2023 级仅实验

室本地环境），经线上线下^[7-9]结合训练后，多数学生具备了独立开发复杂爬虫系统的能力，其中多源数据采集任务的完成率较好，充分体现了专业特色与工程实践的有机结合。

第三，校企协同深度显著提升，企业导师参与课程大纲修订以及线上指导，各学院提供的相关数据集 6 份，同时，企业导师参与课时占比提升至 20%（2023 级 0%，未引入企业参与课程建设），航空宇航学院提供的遥感数据集调用频次达 4.7 次/人·学期（数据来源：2023-2024 学年本课程的过程形成性档案材料之达成度报告）。有效促进了跨学科知识融合。

第四，通过对两届学生的教师评教系统评分数据分析发现，改革后学生的学习体验以及课程评教综合分数明显改善。数据显示，学生对课程评教分数从 91.22 提升至 95.37（由学校教务系统统一导出），与此同时，从学生能力发展维度看，改革成效同样显著。结合企业导师评估（由企业导师提供企业级别的开发课程设计文档），学生在需求分析、系统设计、代码规范等维度的能力方面均有明显进步，尤其是在解决相对较难的实际工程问题能力方面。

综上，通过对两届数据科学与大数据技术专业学生的《数据采集与网络爬虫》实际课程教学实践的考核评估数据、企业评价反馈及教学存档材料的综合分析，本课程改革在以上四方面均取得了显著成效。这些成效充分证明，“新工科”背景下融合企业真实案例的教学改革模式有效促进了产教深度融合，实现了人才培养与产业需求的精准对接，为同类院校的工科专业课程改革提供了可参考的实践范例。

表 1 课程教学改革前后学生成绩表现对比数据一览

评价指标	2023级（改革前）	2024级（改革后）	提升幅度	提升百分比
平均成绩	64.84	74.71↑	+9.87	+15.2%
及格率	74.21%	86.55%↑	+12.34%	+16.6%
最高成绩	79.0	85.0↑	+6.0	+7.6%
最低成绩	51.0	46.0↓	-5.0	-10.9%
成绩标准差	7.00	7.51↑	+0.51	+7.3%

5 结束语

本文基于“新工科”背景，构建了“三阶四维·校企·院际”协同驱动的数据采集课程教学改革体系，通过融合企业真实案例与跨学科协作，探索了产教融合的新路径。本课程教学改革实践表明，该模式有效提升了教学质量，强化了学生的工程实践能力与解决复

杂工程问题能力。

下一步，本课程将在以下三方面持续迭代优化：一是拓展行业合作广度与深度，构建更开放的数据采集课程的产教融合生态；二是完善校企资源共享机制，推动教学资源动态更新；三是建立毕业生发展追踪体系，形成闭环改进机制。进一步完善和丰富本课程教学改革方案，满足以及促进本课程的本科教育教学质量更优质发展。

参考文献

- [1] 郭玉臣,卫志华,谢雨霞,等.基于大数据专业贯通的数据采集与集成课程实验体系设计[J].计算机教育,2023,(01):126-131.DOI:10.16512/j.cnki.jsjy.2023.01.020.
- [2] 陈英.BOPPPS模式下的数据采集与预处理课程思政培养机制探索[J].计算机教育,2022,(05):112-115.DOI:10.16512/j.cnki.jsjy.2022.05.006.
- [3] 刘平,许家铭.基于BIM及校企合作的《土木工程施工》课程教学改革探索[J].砖瓦,2025,(04):171-173.DOI:10.16001/j.cnki.1001-6945.2025.04.016.
- [4] 吴晓雪,何东钢,李响,等.“互联网+实验教学”的线上线下混合教学模式研究[J].高教学刊,2018(9):115-117.
- [5] 刘继芳.校企协同育人背景下饲料企业“三创型”人才培养研究[J].中国饲料,2024,(12):161-164.DOI:10.15906/j.cnki.cn11-2975/s.20241241.
- [6] 彭第,李小华,段建,等.基于工程项目协作的校企合作模式的探索——以土木工程专业为例[J].常州工学院学报,2024,37(01):85-90.
- [7] 杨权,黄俊,邱怀娜.线上线下结合的翻转课堂在医学免疫学实验教学中的应用与体会[J].中国免疫学杂志,2025,41(10):2505-2508.
- [8] 邓立苗,李洪霞,李绍静.“数据库系统原理与应用”线上线下混合式教学模式探索与实践[J].计算机时代,2025,(10):86-89+99.DOI:10.16644/j.cnki.cn33-1094/tp.2025.10.016.
- [9] 姚忠平,王群,宋英,等.表面物理化学线上线下“双螺旋”混合教学模式探索与实践[J/OL].大学化学,1-8[2025-10-28].<https://link.cnki.net/urlid/11.1815.O6.20251009.1732.002..>