

基于知识蒸馏的“大数据可视化技术”课程教学体系研究

张悦云

湖南信息学院 湖南省长沙市 410151

摘 要: 本研究创新性地将知识蒸馏引入大数据可视化技术课程教学体系。从学生难以完成复杂可视化、重工具轻设计、滞留于模仿层三大痛点出发,在分析学科培养目标和学生情况的基础上,构建数据可视化双通道知识蒸馏体系、提出“知识解构-分层训练”教学范式并设计“变异模仿训练-错题重蒸馏”流程,旨在提升教学质量与教学效果,实现教师经验向学生的有效迁移。

关键词: 数据可视化;知识蒸馏;教学研究

引言

在数字化浪潮推动下,数据分析人才已成为企业核心竞争力的关键来源。据 2023 年中国企业数据分析人才需求指数报告显示,超八成的雇主将可视化设计能力列为招聘必备技能^[1]。本校开设的大数据可视化课程已成为大数据专业学生的重要课程之一。笔者作为该课程的主讲人,通过近年来的授课经验以及对学生进行问卷调查发现本课程普遍存在两极化困境:教师精通 Matplotlib 高级技巧,但学生仅能机械复现基础图表,主要问题表现为学生复现复杂作品能力不足、存在“重工具轻设计”思想、长期滞留于模仿学习阶段等。研究表明,传统“演示-模仿”教学模式导致代码复用率高达 60% 以上,学生独立设计能力薄弱^[2]。笔者在对相关资料分析的基础上,提出该类课程基于知识蒸馏的教学改革新思路,思考如何将教师经验高效传递,对如何上好这门课程进行初步探索。

1. 大数据可视化课程教学现状以及存在的问题

1.1. 学生无法独立复现复杂可视化作品

笔者在教学实践中发现,课堂上 80% 左右的学生无法完整实现包含 3 个以上视图联动的仪表盘,在视图切分阶段就已经显示出对 subplots 函数掌握不足。例如在大型超市看板项目中,学生能够在前期独立完成单个视图,如月度营业额折线图。但在看板实现时,大部分学生的代码出现函数回调错误、数据逻辑混乱等问题。学生无法复现复杂仪表盘,体现出知识整合部分掌握不够,究其根本在于基础知识把握不足。

1.2.“重工具轻设计”思维倾向

笔者在前期预定教材阶段就已分析发现,大部分可视化

教材编排分为两种,一种是纯粹的可视化概论,此类专门针对非大数据专业学生,采取的可视化工具也以简单绘图工具为主;第二种针对大数据专业学生,为满足对数据处理的高需求而选择采用 python 对数据进行采集、预处理及可视化分析,同时也符合本门课程的培养目的。但由于第二类教材设置普遍重代码轻理论,导致学生依赖工具而忽视了图表选择的理念。具体表现为:在作业提交中,仅有少数学生能够在代码注释中说明其可视化设计为何选择柱状图而非饼图;部分学生一味使用基础图表展示数据,使用率前三的图表有折线图、柱状图、饼图,甚至出现使用折线图展示家庭日常消费分布的情况;缺乏基本视觉设计意识,存在色彩滥用问题,如使用高对比度的彩虹色系、与背景色同色系等,降低了读者挖掘数据内涵的速度;部分作品图表未设置合适的负空间,导致视觉拥挤,如条形图间隙大于条形宽度,大大降低了数据的可读性。有学生反馈:“知道怎么画图,但不确定画什么样的图最能说明问题”。

1.3. 学生停留在模仿层,长期滞留于模仿学习阶段

传统教学模式导致学生陷入低阶重复,经过 16 周学习后,最后一次阶段性测试,仍有超半数的作业与教师示例的相似度超过 75%,甚至更多。而面对未教授过的可视化类型,仅两成的学生能通过查阅文档自主实现。实验课学习行为观察发现,大部分学生在遇到问题时,首选策略是寻找类似案例复制,而非分析问题本质。深层原因在于教学过程中缺乏有效的经验传递机制。例如,学生虽然能模仿教师使用 subplots() 函数创建多图布局,但 72% 的作业未根据数据特性调整子图间距,仍保持默认间距参数,表明学生仅生搬硬

套语句而未能真正内化设计决策逻辑。

知识蒸馏原是机器学习中压缩复杂模型的技术,其核心思想是通过模仿“教师模型”的决策逻辑,训练出轻量但高效的“学生模型”^[3]。这一理念与教育教学中教师向学生进行知识传递的过程高度契合。本研究创新性地将知识蒸馏迁移至数据可视化教学,构建分层蒸馏体系,将教师图表设计经验分解为可阶梯式模仿的软知识,使学生既能快速上手实践,又能理解底层逻辑。

2. 基于知识蒸馏的“大数据分析可视化技术”课程教育改革措施

2.1. 采用“逆向蒸馏”策略

为提高知识传递效率,加深知识理解,可将复杂知识细化,分阶段进行知识解构。将复杂作品解构为知识原子,解构相当于教师模型的特征提取,将复杂知识分解为可迁移的知识原子^[3],重组是学生的渐进式模仿,通过逐步重组建立神经模式泛化能力,最后通过错误预植入对重组训练进行负样本增强,强化学生模型的决策边界。整体策略如下图 1 所示:

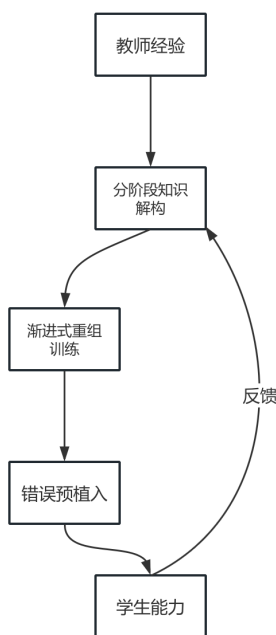


图 1

(1) 分阶段知识解构。仪表盘设计是本课程典型的重难点之一,以该知识点为例,将其解构进行知识传递,该知识解构如下图 2 所示:

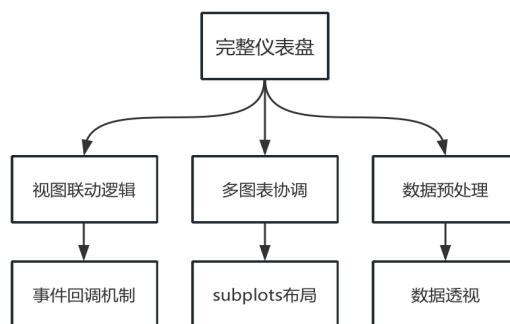


图 2

(2) 渐进式重组训练。其对应知识蒸馏中的分阶段知识释放,通过控制信息量逐步提升任务复杂度^[3]。针对上述知识原子可设计三级重构任务:第一级骨架复现软目标蒸馏,教师可提供 80% 保留了核心部分的代码,让学生填空关键参数。传统的教学设计当中针对学生绘制柱形图,会直接给出 `plt.bar(x, y, 0.8)` 语句,而骨架复现操作下,教师仅提供部分代码如: `plt.bar(x, height=, width=)`, 要求学生填写柱子宽度和高度。此方式下,学生既看到整体框架理解参数意义,又能降低学生初始认知负荷,由于必须读懂上下文方能填空,促进了学生强制理解代码逻辑链。第二级模块替换,教学过程中可要求学生替换图表类型但保持数据流,如仅修改特定视图类型,添加误差线;教师绘制柱状图,要求学生整合修改成为具有相同的 data 输入和色彩配置的堆叠柱形图。第三级数据迁移,数据迁移能力包括理解原理并适配新问题、技术栈转换与工具替代、自主设计解决方案等^[4]。知识蒸馏中高温蒸馏和低温蒸馏是控制知识传递过程的两个关键参数^[3],可对应数据可视化中的数据迁移阶段中。低温蒸馏的核心特征是精确模仿教师模型的输出,学生需严格遵循教师提供的模式,仅替换最表层变量,直接复用原有数据处理流程。如精确复现核心功能:作业中要要求必须包含趋势线,同时适配新数据集;仅更换图例中的坐标轴列名、行名。而高温阶段可适配多种数据集,鼓励学生探索不同方案,如在展示销售额随时间变化的简单折线图基础上,要求学生挑战自主选择至少 3 种可视化形式的组合以此揭示指标间的潜在关系。高、低温由温度参数 T 进行界定,针对本门课程,温度参数 T 可设定如下。

$$T = \alpha \cdot S_{\text{level}} + \beta \left(1 - \frac{C_{\text{task}}}{C_{\text{max}}} \right)$$

上述公式中 S_{level} 表示学生当前能力水平, 计算方式为历史作业平均分 $\times 0.6 +$ 迁移能力测试 $\times 0.4$; C_{task} 表示任务复杂度, 基于 TCI 模型计算, $TCI = 0.6 \times (\text{代码行数} / 50) + 0.4 \times (\text{新概念数} / 5)$; C_{max} 表示课程最高复杂度阈值, 采用的是本课程全过程中设置的复杂度最高的某次任务的 TCI 值。是调节系数, 基于本门课程属于设计思维依赖任务复杂度的课程特性, 可将 α 设置为 0.4, β 设置为 0.6。但同时也可针对实际情况进行动态调整^[5], 笔者某次教学过程中, 计算结果满足设置高温任务, 但最终学生完成率仅 35.4%, 推测 β 过高导致伪高温, 低能力学生获取到了高温任务, 后续将系数调整为 $\alpha = 0.7$ 以此增加学生能力权重, 更精准匹配学生当前水平, 设置 $\beta = 0.3$ 降低任务复杂度影响。修改后重新发放作业任务, 高温任务完成率提升至 62%, 学生创新数量增加。

引入温度参数的本质是控制教学约束的松紧度。低温则精确复制知识, 保证学生掌握核心技能; 高温则积极进行创造性知识迁移, 培养真实问题解决能力, 笔者建议在数据迁移训练中, 建议从低温开始, 随学生能力提升逐步调高温度, 最终达到创新训练目的。根据温度参数 T 设置测试分配如下表 1 所示:

表 1

T 参数范围	阶段	适用场景	测试示例
T=0.1 ~ 0.5	低温	基础巩固, 相似情境直接应用	将销售数据图表迁移运用到库存储数据; 替换数据文件路径。
T=0.5~1.0	中温	中等难度, 增加条件适配	时间序列数据改为多变量对比
T=1.0 ~ 2.0	高温	开放性问题的创新解决, 发展解决问题能力	用全新可视化形式展示人口老龄化社会关系

(3) 错误预植入训练。该训练是知识蒸馏中的对抗性学习机制, 通过负样本来优化模型鲁棒性^[3]。在课堂中可通过故意植入典型错误代码运行报错, 引导学生主动回答问题并修复代码, 从而深度理解正确模式。其本质是从失败中蒸馏知识, 能有效保障学习效果。

知识解构是经验传递的起点, 渐进重组通过知识掌握技能, 错误预植入是从技能的掌握升华到了素养的保持。这三个环节构成了“分解-建构-强化”的完整学习闭环。

2.2. 构建双通道知识蒸馏体系

双通道知识蒸馏体系是一种将技术实现与设计思维同步传递的教学框架, 其核心是通过技术通道与设计通道两个并行通道, 实现知识的完整迁移。基于此, 可构建可视化双通道知识蒸馏体系如下^[6]:

(1) 设计思维编码化。在正式学习绘图编码、讲解图

表代码知识点前期, 分配 1 课时对学生进行数据可视化分析中图表设计的讲解。引用可视化图例的发展史, 如从统计学图表时期到动态可视化看板的发展过程。如数据可分为时间连续型数据与时间离散型数据, 连续型数据可选择折线图进行可视化, 离散型数据可选择散点图进行可视化。还可在后期课程中针对新引入的图形讲解其适用的数据类型。本门课程设置在大三下学期, 先修课程为 python 程序设计等, 此时学生已熟练掌握编码思维且编码兴趣较高, 故可在图表选择分类上建立可视化设计模式库, 使用符合学生接受度的方式, 同时还能加深图表设计思维印象。如:

```
design_patterns = {
    '趋势展示': {
        '推荐图表': ['折线图', '面积图'],
        '视觉要素': ['平滑曲线', '趋势线标注'],
        '避错指南': ['避免过多数据系列']
    },
    '分布对比': {
        '推荐图表': ['箱线图', '小提琴图'],
        '视觉要素': ['四分位标注', '密度显示']
    }
}
```

(2) 设计思维显性化蒸馏。采用双核注释教学法, 在代码中同步呈现技术实现与设计原理, 将抽象设计原则转化为可执行的代码注释。如:

```
{
    # 色彩选择: 使用色盲友好配色, 饱和度差异控制 30%-50%
    colors = ['#377eb8', '#ff7f00', '#4daf4a'] # 蓝-橙-绿色盲安全色

    # 柱间距 (width) 应控制在 0.5-0.8 倍柱宽, 确保视觉分组效果, 避免产生错误关联
    plt.bar(x, height, width=0.6)
}
```

(3) 设计决策训练。设计思考题强化学生设计思维。当展示某电商平台季度销售数据时, 提问最优可视化方案是什么, 同时要求学生提交答案说明文档, 由此加深学生印象。

2.3. 设计变异模仿训练, 错题重蒸馏

当前可视化教学中, 教师单向灌输、学生缺乏错误反

思机制导致学生自身普遍陷入机械模仿的恶性循环,基于知识蒸馏理论,可构建变异模仿训练强化学生自主编码能力。变异模仿训练是围绕在标准示例基础上生成的变体任务,如将原坐标轴转换为对数坐标,增加同比数据系列等。同时教师还可设计双重编码规则要求学生用注释标注每个代码块对应的理论依据,教师后续检查注释匹配度。构建错题重蒸馏流程收集学生作业高频错误,生成针对性训练任务。能力加强阶段则采用渐进式撤脚手架测试,从提供完整代码讲解到隐藏非关键参数,如设置蓝色时隐藏 blue 属性仅提供“color=”;最后测试升级至仅给出函数名要求学生进行可视化设计。

实践发现,上述教学方案基于知识蒸馏的核心思想,系统性地解决了数据可视化教学痛点。教师作为“知识蒸馏器”将专业经验转化为适温的学习内容传授给学生,帮助学生在以往的模仿基础上实现能力跃迁,有效提高学生的学习效率,学生参与课程互动时更加积极主动。

3. 结语

本研究基于知识蒸馏理论,针对大数据可视化技术课程的教学痛点,提出了一套创新性的教学改革方案。通过构建双通道知识传递体系,将抽象的可视化设计思维转化为可执行的代码注释,有效解决了学生重工具操作轻设计思维的问题。研究的任务复杂度量化模型和分层训练框架,实现了教学内容的精准分级,使复杂仪表盘任务的完成率显著提升。

变异模仿训练、错误模式萃取和针对性训练成功帮助学生突破模仿学习的局限。未来研究将着重拓展该体系在交互式可视化教学中的应用,并探索与智能教育技术的深度融合。本研究的应用价值不仅在于解决了具体的教学问题,更在于建立了一套可推广的技术课程教学模式,力争为培养新时代数据可视化人才打下坚实基础。

参考文献

- [1] 中国商业联合会数据分析专业委员会. 中国数据分析行业人才指数报告(2023)[R]. 北京:中国商业联合会, 2023.
- [2] 邓晓霞. 指向深度学习的高中英语读后续写教学策略[J]. 校园英语, 2024,(51):146-148.
- [3] Geoffrey E. Hinton, Oriol Vinyals, Jeffrey Dean. Distilling the Knowledge in a Neural Network.[J]. CoRR, 2015, abs/1503.02531
- [4] 黄震华, 杨顺志, 林威, 等. 知识蒸馏研究综述[J]. 计算机学报, 2022,45(03):624-653.
- [5] 王萍. 带自适应隐私保护的动态知识蒸馏研究[D]. 曲阜师范大学, 2025.
- [6] Qu Meng, Yoshua Bengio, Jian Tang. GMNN: graph Markov neural networks[C]//International Conference on Machine Learning, New York, USA:ACM, 2019:5241-5250.