

Python数据可视化在数据挖掘课程教学中的应用

王升超, 余小东, 张志强

成都大学计算机学院, 四川 成都

收稿日期: 2025年6月30日; 录用日期: 2025年7月29日; 发布日期: 2025年8月6日

摘 要

大数据时代已然到来, 针对大数据专业的数据挖掘课程变得尤为重要。然而, 目前数据挖掘课程的传统教学方法面临诸多问题, 教学过程无法直观展示日益更新的数据挖掘算法和流程, 极大增加了学生的理解难度。Python拥有多种可视化方法, 可以有效解决数据挖掘课程中的直观展示问题, 为数据挖掘课程教学提供了新的途径。本文探讨了Python数据可视化在数据挖掘课程教学中的应用, 包括其优势、具体的教学应用方法以及教学效果评估, 旨在提高数据挖掘课程的教学质量和学生的学习效果。

关键词

Python, 数据可视化, 数据挖掘, 教学设计

The Application of Python Data Visualization in Data Mining Course Instruction

Shengchao Wang, Xiaodong Yu, Zhiqiang Zhang

School of Computer Science, Chengdu University, Chengdu Sichuan

Received: Jun. 30th, 2025; accepted: Jul. 29th, 2025; published: Aug. 6th, 2025

Abstract

The era of big data has arrived, and data mining courses for big data majors have become particularly important. However, the traditional teaching methods of data mining courses currently face many problems, and the teaching process cannot intuitively display the increasingly updated data mining algorithms and processes, greatly increasing the difficulty of students' understanding. Python has multiple visualization methods that can effectively solve the problem of intuitive display in data mining courses, providing a new approach for teaching data mining courses. This article

explores the application of Python data visualization in the teaching of data mining courses, including its advantages, specific teaching application methods, and teaching effectiveness evaluation, aiming to improve the teaching quality of data mining courses and students' learning outcomes.

Keywords

Python, Data Visualization, Data Mining, Instructional Design

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. 引言

数据挖掘作为一门融合了计算机科学、数学、统计学等多学科知识的交叉学科，其课程内容具有较强的理论性和抽象性。学生在学习过程中往往面临诸多困难，如对复杂算法的理解不够深入、难以将理论知识与实际应用相结合等。Python 作为一种广泛应用于数据处理和分析的编程语言，其数据可视化库如 Matplotlib、Seaborn、Plotly 等能够将数据以直观易懂的图表形式展示出来，这为数据挖掘课程教学带来了新的契机[1]-[4]。

2. Python 数据可视化在数据挖掘课程教学中的优势

2.1. 增强知识直观性

数据挖掘中的许多算法，如相关性计算、聚类分析、关联规则挖掘等，其过程和结果通常较为抽象。通过 Python 数据可视化，可以将这些复杂的数据关系和挖掘结果以图形化的方式呈现，如散点图、热力图、箱线图、提琴图等。这些图片的使用使学生能够更直观地理解算法的原理和运行机制，降低学习难度。

2.2. 提高学习兴趣

传统学习过程中，对原理和算法的讲解复杂而枯燥，学生容易失去兴趣导致注意力不集中的情况。与传统的文字和公式教学相比，数据可视化具有更强的趣味性和吸引力。生动形象的图表能够激发学生的学习兴趣 and 积极性，使他们主动参与到数据挖掘课程的学习中，极大的提高学生的学习积极性，提高学习效果[5]-[7]。

2.3. 促进实践能力培养

在数据挖掘领域，实践能力至关重要。Python 数据可视化工具的使用需要学生具备一定的编程基础和数据处理能力，这促使学生在学习过程中不断实践，将理论知识应用于实际操作，从而提高他们的实践能力和问题解决能力。

3. Python 数据可视化在数据挖掘课程教学中的应用方法

3.1. 课程内容讲解

在讲解数据挖掘的基本概念和算法时，结合 Python 数据可视化进行演示。例如，在介绍 K-Means 聚类算法时，利用 Matplotlib 绘制聚类过程中的数据点分布图和聚类中心变化图，帮助学生理解聚类的迭代

过程和算法的原理。通过这种方式，使抽象的知识变得形象化，便于学生理解和记忆[8] [9]。

3.2. 实践教学环节

设计基于 Python 数据可视化的实践教学项目，让学生在实际操作中掌握数据挖掘技术和数据可视化方法。例如，要求学生使用 Python 数据挖掘库(如 seaborn 库)进行数据预处理、挖掘分析，并运用数据可视化工具展示挖掘结果。在项目实施过程中，引导学生分析和解决实际问题，培养他们的创新思维和实践能力。

4. Python 数据可视化在数据挖掘教学中的应用案例

以 seaborn 库中的 tips 数据集为例，该数据集统计了影响就餐消费金额的各种因素，对该数据集进行数据预处理、并运用 Python 数据可视化工具展示挖掘结果。使用 seaborn 库中的热力图、箱线图、提琴图展示挖掘结果，有利于学生直观的了解挖掘出的重点信息。

4.1. 热力图

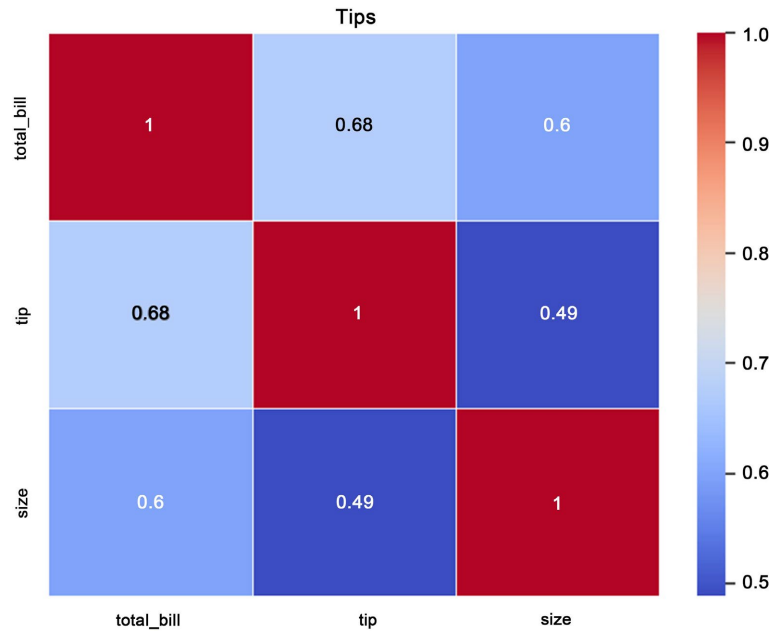


Figure 1. Heat map
图 1. 热力图

图 1 的热力图主要展示不同数据之间的相关性，在热力图的网格中的值代表两元素之间的相关系数大小，从图中可以看出，小费(tip)和消费人数(size)以及账单总金额(total bill)之间的相关性，得出小费和消费金额的相关性程度较大的结论。

4.2. 箱线图

图 2 箱线图是一种用一组数据的分位数绘制的图形。箱子的中间横线表示数据的中位数，也就是将数据从小到大排序后位于中间位置的数，它反映了数据的中心位置。箱子的顶部和底部边缘分别表示数据的第三四分位数和第一四分位数。从箱子顶部和底部延伸出来的两条线段，通常用来表示数据的上下界，异常值位于两条之外的单个数据点，可能是数据录入错误、特殊的极端情况等。图 2 的箱线图展示

了不同日期账单总金额的分布情况。直观展示了数据的离散程度和分布区域。

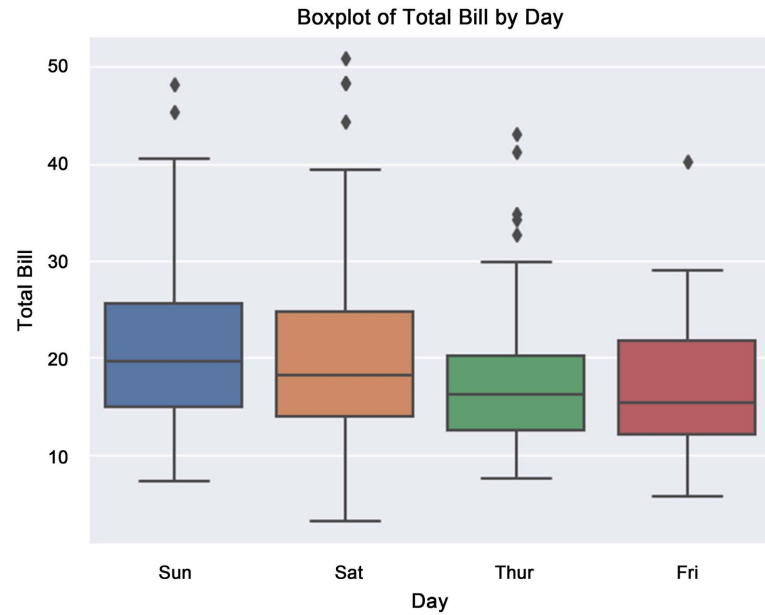


Figure 2. Box plot
图 2. 箱线图

4.3. 提琴图

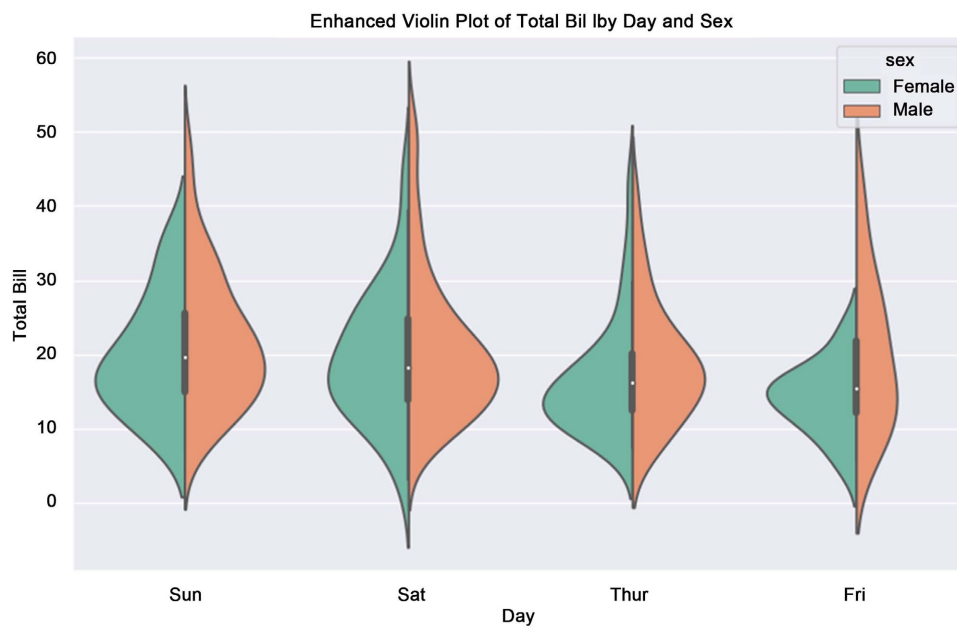


Figure 3. Violin plot
图 3. 提琴图

图 3 提琴图(Violin Plot)是在箱线图的基础上增加了数据的密度信息,综合了箱线图和密度图的特点,用于展示数据的分布及其概率密度。相比箱线图,提琴图通过密度曲线提供了更详细的数据分布信息,

能够展示数据的形状，比如是否为单峰、多峰分布，以及峰值的位置等，有助于更全面地理解数据的特性。图 3 中的提琴图可以对不同性别用餐时的账单金额的变化，直观的显示性别的影响。

4.4. 散点图

在 seaborn 中，鸢尾花数据集是一个经典的案例，可以帮助学生有效理解相关性计算。该数据集包含了 150 个样本，分为 3 个品种的鸢尾花，每个样本有 4 个特征，分别是花萼长度(Sepal Length)、花萼宽度(Sepal Width)、花瓣长度(Petal Length)和花瓣宽度(Petal Width)。样本特征之间会有一定的相关性，为了直观的让学生理解相关计算，可以首选通过散点图来展示变量间的相关性。以花萼长度(Sepal Length)和花瓣宽度(Petal Width)为例，通过 python 获得散点图如下。

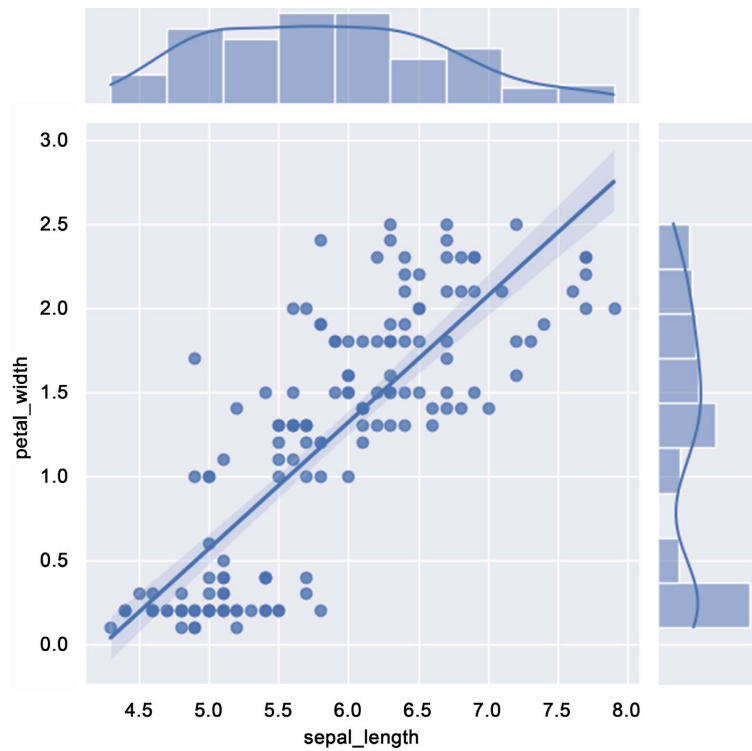


Figure 4. Scatter plot
图 4. 散点图

图 4 可以清晰的看出变量之间存在正相关关系，为了进一步使高密度区域显现，还可以使用蜂窝图。当数据点较多且集中在某些区域时，蜂窝图可以将数据点所在的区域划分成六边形网格，每个六边形的颜色深浅代表该区域内数据点的密度。这样可以更直观地看出哪些区域数据点密集，哪些区域数据点稀疏，而回归线的散点图对于大量重叠的数据点无法清晰地展示这种密度差异。图 5 展示了蜂窝图的效果。图 5 减少视觉杂乱，避免了点与点之间的重叠，使图像更加简洁清晰，便于学生观察和分析。

5. Python 数据可视化对学生影响的理论分析

在教学过程中，Python 数据可视化工具的应用，不仅为数据挖掘课程提供了新的教学手段，更为学生的认知过程、学习动机以及实践能力带来了积极影响。从教育心理学和学习理论的视角进行深入分析，这一影响具有重要的教育价值。

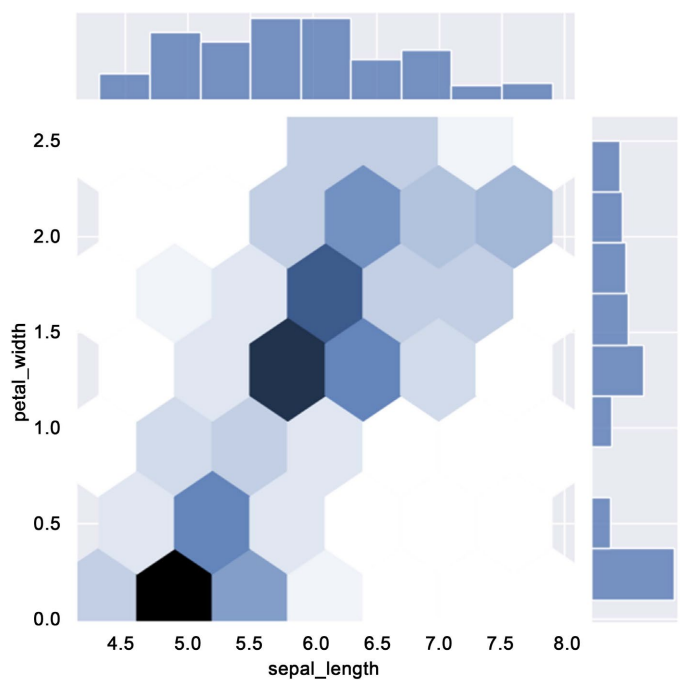


Figure 5. Hexagonal lattice
图 5. 蜂窝图

5.1. 对学生认知过程的影响

根据认知主义学习理论，学生的学习过程是通过将新知识与已有知识经验进行相互作用，从而构建新的认知结构。Python 数据可视化所呈现出的直观图表，为学生提供了一种更为有效的信息加工方式。在数据挖掘课程中，复杂的算法过程和数据关系往往难以通过文字和口头讲解让学生理解。而数据可视化工具将抽象的数据转化为直观的图形，例如热力图清晰地展示了数据之间的相关性，箱线图直观地呈现了数据的分布特征和离散程度，提琴图则综合了箱线图和密度图的特点，使学生能够更深入地理解数据的特性。

5.2. 对学生学习动机的影响

从自我决定理论的角度来看，学生的学习动机主要由自主性、胜任感和归属感三个基本心理需求驱动。Python 数据可视化工具的运用，在很大程度上激发了学生的自主性和胜任感，进而提高了学生的学习动机。

此外，当学生通过数据可视化成功地呈现出复杂的数据关系和挖掘结果时，他们会获得一种成就感和胜任感。例如，学生在完成一个基于 Python 数据可视化的实践项目后，看到自己绘制的精美且富有洞察力的图表，会对自己在数据挖掘领域的能力产生信心，这种正面的情感体验会进一步激励他们投入更多的学习和实践活动中。

5.3. 对学生实践能力的影响

从体验式学习理论出发，学生在实践中通过亲身体验和反思，能够更好地掌握知识和技能。Python 数据可视化工具的使用，为学生提供了丰富的实践机会，促进了学生实践能力的培养。

数据挖掘领域强调学生的实践能力，而 Python 数据可视化工具的运用要求学生具备一定的编程基础和数据处理能力。在学习过程中，学生不仅要理解数据挖掘算法的原理，还需要通过编写代码实现数据

的可视化展示。这一过程促使学生不断地练习编程技能，提高数据处理和分析能力[10]-[12]。

6. 总结

Python 数据可视化在数据挖掘课程教学中的应用取得了显著的教学效果，为解决传统教学中的难题提供了有效的途径，其提供的如热力图、箱线图、提琴图等不仅增强了课程内容的直观性，有助于学生了解数据之间的内在联系，还促进了教学质量和教学效果的提升。在未来的数据挖掘课程教学中，应进一步深化 Python 数据可视化与课程教学的融合，不断探索创新教学方法和手段，以适应大数据时代对人才培养的需求。同时，也应加强对 Python 数据可视化技术的研究和应用，为数据挖掘课程教学提供更加丰富和先进的技术支持。

基金项目

本研究收到以下项目和基金的资助：

教育部产学研协同育人项目：基于人工智能动作识别的体育课研究与实践(241204213194946)；

成都大学第三批“三全育人”示范团队项目——“数据驱动、智创未来”数智育人教学团队；

2022 年成都市教育局教育科研规划一般课题“应用型地方高校双创教育与专业课程的‘专创深度融合’研究”(CY2022Y052)。

参考文献

- [1] 丁毅涛,任水利. 基于 Python 语言的数理统计课程应用改革探索——以应用统计学专业为例[J]. 科技风, 2021(28): 25-27.
- [2] 何小年,段风华. Python 在统计学课程实训中的应用探索[J]. 软件, 2021, 42(9): 32-34.
- [3] 郭鹏. Python 在应用统计学课程教学中的应用[J]. 现代信息科技, 2024, 8(14): 183-189.
- [4] 何鑫. 基于 Python 软件的统计课程教学探索与实践[J]. 教育进展, 2025, 15(5): 921-928.
- [5] 秦华妮. 数据分析与挖掘课程多维教学改革探讨[J]. 高教学刊, 2023, 9(32): 111-115.
- [6] 惠蓉. 实践教学在应用型本科人才培养中的重要意义——评《Python 数据分析与挖掘实践(第 2 版)》[J]. 中国教育学刊, 2023(11): 148.
- [7] 王文强,张启民,孙智星. 《供应链大数据分析与应用》课程思政元素的挖掘与育人研究[J]. 物质科技, 2023, 46(24): 170-172+176.
- [8] 杜江,戴君,曹瑞元. 人工智能与数字技术背景下统计学专业实践教学体系优化及设计[J]. 高教学刊, 2024, 10(9): 115-118.
- [9] 李毕祥,李兰芳,朱琳. 标准化统计分析方法在 Python 教学中的应用与课程设计[J]. 中国标准化, 2024(10): 235-238.
- [10] 陈华. Python 在数据分析中的应用[J]. 上海轻工业, 2024(3): 108-110.
- [11] 袁欧,何山. 基于 python 的 Z 检验法和 T 检验法研究[J]. 大众标准化, 2022(15): 174-176.
- [12] 王家武,姜庆. Python 在医学文献检索与统计学课程教学实践中的应用[J]. 现代医药卫生, 2023, 39(9): 1590-1595.