

基于 Jupyter 交互式平台的灾害大数据与人工智能课程教学改革实践

符伟 李博南 范小平

(南京工业大学交通运输学院, 江苏 南京 211816)

【摘要】灾害大数据与人工智能课程是防灾减灾科学与工程专业必修专业基础课程之一, 其旨在培养学生利用新一代信息技术进行高效灾害管理和应急处置的能力。然而该专业学生计算机相关理论知识以及程序设计原理方面的基础较为薄弱, 探索一种入门简单、易用性强且交互性能好的编程语言与编程环境, 有助于学生专注于智能化灾害信息管理与分析的应用学习。本研究利用 Jupyter Notebook 强大的交互性能及其低廉的学习成本, 开展交互式课程编程实践。分析了灾害大数据与人工智能课程中的编程需求分析, 并通过案例展示 Jupyter 在课堂教学与实践中的优势, 目的是在防灾类课程建设中, 启发教师的备课思路, 提升学生的学习兴趣, 拓展教学研究的广度和深度。

【关键词】Jupyter 交互式平台; 大数据与人工智能; 灾害; 防灾减灾科学与工程; 地震灾害

一、引言

如今, 随着“第三次信息化浪潮”的来临, 大数据与人工智能技术向社会各领域迅速渗透(林子雨等)。以防灾减灾科学与工程专业为例, 开设的灾害大数据与人工智能课程旨在培养学生进行智慧感知、互联、处理和协调的能力, 利用新一代信息技术建立高效的灾害管理和应急处理方法。

然而, 人工智能、物联网、云计算、大数据等信息技术涵盖面较广, 且以面向应用为主。对于防灾减灾科学与工程这一非计算机类专业而言, 其在计算机相关理论知识以及程序设计原理方面的培养较为薄弱, 但又需要学生具备利用编程语言解决灾害防治与管理过程中的海量数据处理与决策的能力。因此, 选择一种对计算相关理论知识要求低、入门简单、学习成本低且交互性能好的编程语言与编程环境, 对促使学生将主要精力放在专业知识的应用和实践上具有重要意义。同时, 友好的交互方式也将提高教师在课堂上进行编程演示的效率。

Jupyter Notebook 可以让学生以浏览器远程访问的方式进行 Python 的交互式编程, 支持包括 Python 在内的 40 多种编程语言, 其作为一种扩展性高、交互方式友好、学习成本低的编程演示与实践方式, 已被广泛用于各类计算与非计算机专业的课堂

教学活动中(薛煜阳, 2018; 王茂发等, 2022; 刘征等, 2020; 马良伟, 2022)。

本文以防灾减灾科学与工程专业基础课之一的《灾害大数据与人工智能》课程为例, 结合该课程的课程大纲利用 Jupyter Notebook 交互式编程平台探索一种便于理解、操作简单的目标导向性教学方法。

二、灾害大数据与人工智能课程中的编程需求分析

灾害大数据与人工智能课程是一门针对防灾减灾科学与工程本科专业的专业基础课。其旨在加深防灾减灾专业理论知识的理解, 培养利用人工智能、物联网、云计算、大数据等信息技术, 进行智慧感知、互联、处理和协调的能力, 为日后从事灾害防治与管理相关工作积累理论和实践基础。在教学过程中, 需要教师进行课堂演示或学生进行上机实验的编程内容主要包括:

(一) 网络爬虫

互联网是获取灾害大数据从而进行各类灾害的有效预测与科学管理的重要途径之一。对于灾害大数据与人工智能而言, 灾害大数据的互联网来源主要有两种类型, 一是政府或科研组织公开的各类观测数据, 如国家数据科学中心、美国地震学研究联合会等平台公布各类与灾害相关的观测数据, 其数

据获取方式因数据源的不同而各有差异；二是各类门户网站、社交平台上发布的新闻报道、专题报道以及与灾害相关的社交信息等，如新华网（<http://www.news.cn/>）、中国地震台网中心（<https://www.cenc.ac.cn/>）、新浪微博（<https://weibo.com/>）等。通过对这类信息的分析可以为灾害的有效决策提供参考（周义棋等，2022），通常以网页的形式存在互联网中，网络爬虫正是获取这类信息的绝佳手段。因此，在灾害大数据与人工智能课程中，网络爬虫技术是其中一项需要开展理论与实践教学的内容，以便让学生掌握网络爬虫的实现过程并具备一定的从特定网页中自动获取灾害信息的能力。

（二）数据库

数据库可以用于收集、整合和存储各种与灾害相关的数据，包括灾害概况、气象数据、地质数据、社会经济数据、受灾区域的地理信息等。主要体现在：

1. 通过数据库的结构化存储和管理，可以方便地对海量数据进行汇总和共享，提高数据的利用效率。
2. 数据库提供了快速查询和分析的能力，可以帮助决策者及时了解灾情，做出准确的决策。
3. 数据库可以记录灾害发生后的情况和灾后重建的进展，能够提供灾后评估和规划的数据支持。因此，掌握和理解数据的理念与建设方法，并进行相应的实践操作也是本门课程的重要目标之一。尽管近年来非关系型数据库在大数据的数据存储中占据着越来越大的比重，但灾害类数据的来源或使用往往与政府、科研院所、事业单位密切相关，并主要用于社会管理与决策中，其对数据的规范性和安全性仍存在较高的要求。因此，在本科课程中的仍教授和实践以关系型数据为主的数据库类型。其可以提供数据组织和存储、数据一致性和完整性、数据查询和分析、数据共享和协作以及数据安全和备份等功能，为灾害预警和应急响应提供有力支持。主要演示和实践内容包括，数据库的创建、写入、查询、导出等。

（三）机器学习与人工智能算法应用

机器学习与人工智能算法是进行灾害大数据分析的重要手段，其旨在分析获得大数据集中数据模型（规律）。其需要利用各类有效算法挖掘出数据中隐含的有用信息，从而达到分析推理与预测的效果，实现预定的高层次数据分析需求。主要表现在如下方面：

1. 灾害预测和预警。通过对历史灾害数据

的分析和模式识别，这些算法可以识别出潜在的灾害风险和趋势，提前进行预警，帮助减少灾害造成的损失。

2. 灾害数据的分类和识别。通过对不同类型的灾害数据进行训练和分类，这些算法可以自动识别和标记不同类型的灾害事件，提高数据处理的效率和准确性。
3. 灾害数据的模拟和预测。通过对灾害数据进行建模和仿真，这些算法可以预测灾害事件的发展趋势和可能影响的范围，为决策者提供科学依据。
4. 灾害数据的应急响应和救援。通过对实时灾害数据的监测和分析，这些算法可以帮助快速评估灾情、优化资源调度和指导救援行动，提高灾害应急响应的效率和准确性。因此，基于以上方面，本课程需要讲授和实践的机器学习与人工智能算法包括：回归、聚类、支持向量机、决策树、人工神经网络、深度学习等算法。在教学过程中将对以上算法的具体实现或开源函数库的调用进行编程演示及实现。

（四）数据可视化

灾害大数据的可视化利用人眼的感知能力对数据进行交互的、可视的表达，来增强认知的一种技术，可以提高数据识别的效率，传递有效的信息。对于灾害大数据的可视化方法主要分为两大类：一是科学可视化，主要是来源于与灾害直接或间接相关的观测数据，包括各类标量场、向量场、张量场数据；二是信息可视化，包括从各类平台中获取的文本数据、视频数据、时空信息数据等。对于以上两大类的数据，本课程以 Python 编程为基本手段进行可视化教学，并在教学过程中适当调用信息图工具（如 Google charts、Echarts、Tableau 等）、地图工具（如 Generic Mapping Tools、Basemap、Folium 等）以及时间线工具（如 Timetoast、Timeline 等）等进行演示和实践。

三、Jupyter 交互式平台的灾害大数据与人工智能课程教学中的应用

本文以“Python 爬虫获取地震灾害事件与可视化分析”这项上机实验为例，详细展示如何利用 Jupyter 为载体进行灾害大数据与人工智能课堂教学与实践。本实验可在课堂理论授课中按章节进行分步演示，也可在理论授课完毕之后集中进行实践教学，后者占用学时约 8 个课时，演示和实践过程如下：

（一）实验基础讲授

在实践开展之前，进行必要的实验背景介绍，并进行简单的课堂理论教学内容回顾。利用 Jupyter

Notebook 支持 Markdown 语法的优势，可直接在授课脚本中添加实验背景、实验技术路线和实验原理的描述。如图 1 所示，案例中介绍了利用 Python 爬虫获取网络上的灾害信息的社会背景以及目的，并且交代本次实验的主要目标。

《灾害大数据与人工智能实践课程》

实践一：Python 爬虫获取地震灾害事件与可视化分析

主讲教师：符伟 南京工业大学

一、实验背景

在灾害发生前后，政府相关部门门户网站、社交媒体、新闻媒体等互联网平台会产生海量、多源、异构的灾害大数据，尤其是社交媒体对这些灾害数据进行更新的速度非常快，进行大数据的提取和挖掘具有重要的意义。为了能够及时地获取与在灾害相关的海量互联网数据，高校应进行数据挖掘、分析与可视化，可为政府灾害防治与应急管理决策提供参照。

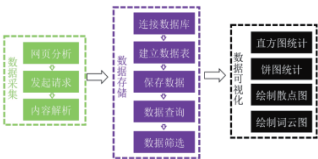
本实验以地震灾害为例，从中国地震台网中心门户网站爬取2022年以来记录到的地震信息，将其保存在关系型数据库中，同时进行一定的可视化分析，以此巩固本课程中灾害大数据采集、存储、分析与可视化的相关内容。



图 1 在 Jupyter Notebook 中展示实验背景及 Markdown 文本

如图 2 所示，在 Jupyter 实验脚本中添加本次实验的技术路线，有助于在实验开始之前对实验整体思路进行把控，同时添加网络爬虫、关系型数据库、可视化方法等实验原理的描述，其目的在于对课堂理论内容进行回顾。

二、实验原理



(一) 网络爬虫的基本步骤

- ① 发起请求：通过 HTTP 向目标站点发起请求，若服务器正常接收并处理请求，会返回一个 Response（响应），该 Response 即为获取的页面内容，Response 可以以 HTML（超文本标记语言）、JSON（对象表示法）字符串、二进制数据等数据类型。
- ② 解析内容：利用正则表达式、网页解析库对 HTML 进行解析，将 JSON 数据转为 JSON 对象进行解析，保存为我们需要的二进制数据（图片、视频等）
- ③ 保存数据：可将爬取并解析后的内容保存为文本、或存入数据库。

(二) 关系型数据库的相关概念及常用操作

1. 关系型数据库的相关概念

- ① 元组（记录、行）：表中的一行即为一个元组，或称为一条记录
- ② 属性（字段、列）：数据表中的每一列称为一个属性（字段）
- ③ 属性值：行中的属性对应的值称为属性值
- ④ 主键：是表中用于唯一确定一个元组的属性或属性组合
- ⑤ 域：属性的取值范围



图 2 在 Jupyter notebook 中展示实验原理及 Markdown 文本

(二) 实验准备

在实验准备过程中，提供较详尽的准备过程描述（图 3），包括必要的文字和图片描述，有助于帮助学生迅速进入实验状态，减少实验之前不必要的计算机环境调试。本次准备过程给出从 Windows 和 Linux 系统配置实验环境的过程，使学生能够在个人电脑上开展实验，提高实验的可重复性，以期使学生能够灵活使用实验知识并为后续学习和研究工作提供参照。实验的准备包括：

1. 编程环境的配置：Jupyter 软件的安装方法，爬虫、数据库管理、数据可视化等实验所需第三方软件库的安装；
2. 第三方软件平台的配置：MySQL、Navicat 数据库软件的安装与配置、Generic Mapping Tools 地学绘图软件的安装与配置等。

三、实验准备

- 安装 Jupyter 运行环境：
- ① 打开命令窗口（win+R，输入cmd，回车）
 - ② 输入命令 pip install jupyter -i <https://pypi.tuna.tsinghua.edu.cn/simple>
 - ③ 安装完成请在命令窗口输入：jupyter notebook 打开 Jupyter
- 安装依赖库：（爬虫中的网页解析库、中文分词库）
- ① pip install requests -i <https://pypi.tuna.tsinghua.edu.cn/simple>
 - ② pip install beautifulsoup4 -i <https://pypi.tuna.tsinghua.edu.cn/simple>
 - ③ pip install lxml -i <https://pypi.tuna.tsinghua.edu.cn/simple>
 - ④ pip install jieba -i <https://pypi.tuna.tsinghua.edu.cn/simple>
- 安装和配置 MySQL
- ① 下载并安装默认方式安装 MySQL Community 版本：<https://dev.mysql.com/downloads/mysql/>
 - ② 运行 MySQL Configurator 配置服务器参数，其中：服务器端口（Port）和 X Protocol Port 修改为：3307 和 33070，其余按照默认安装即可。
- 安装和配置 Navicat 中文版
- ① 下载并安装默认方式安装 Navicat Premium 版本：<https://www.navicat.com.cn/download/navicat-premium>
 - ② 启动 Navicat Premium，执行“文件”->“新建连接”->“MySQL 数据库”，在“MySQL 新建连接”对话框中输入连接名“MySQL”，用户名“root”和密码“root123”，将端口修改为 3307，其余设置按照默认即可。

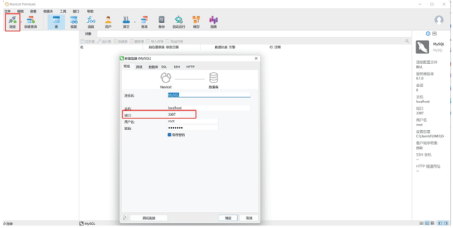


图 3 在 Jupyter Notebook 中对实验准备过程进行描述

(三) 实验过程

本实验过程分为数据爬虫、数据存储、数据可视化三个环节，并且在每个环节中分步骤开展，利用 Jupyter notebook 强大的交互功能，在每一步骤中实时数据输出结果，以便增强学生对该知识点的理解力。具体实验过程如下：

1. 数据爬虫。本实验利用 Python 第三方爬虫库对中国地震台网中心实时发布的地震事件进行下载，首先设置查询条件，根据查询条件构造数据网址，利用 request 函数并对该网址发起请求，若请求成功则范围函数值为 200，利用 BeautifulSoup 函数对网页内容进行解析，截取包含地震事件的内容并转换

为 json 数据格式,该实验过程如图 4 所示。通过设置网页页码循环构造数据网址,可实现该查询条件下所有地震事件信息的下载。

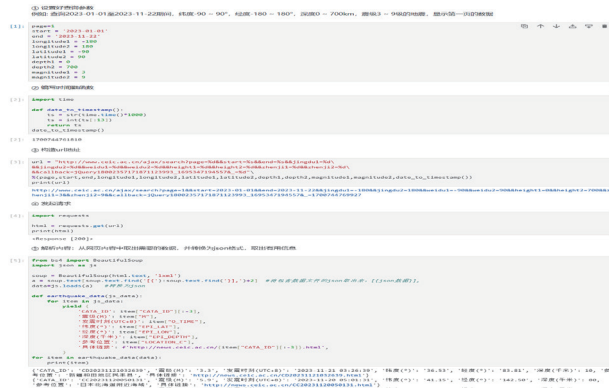


图 4 在 Jupyter notebook 中利用 python 第三方爬虫库爬取地震事件信息

2. 数据存储。本实验利用关系型数据库中的 MySQL 数据库对上一步获取的地震数据进行存储。在实验过程利用 Python 第三方数据库 pymysql 调用和执行 SQL 语句,从而实现 Jupyter 平台统一运行,该过程中的主要程序如图 5 所示,并可在 Navicat 中较为方便地查看,如图 6 所示。

```
sql="INSERT INTO earthquake (ID, 发震时间, 震级, 纬度, 经度, 震源深度, 震中位置, 链接, 国家, 地区)\nVALUES(id,time,magnitude,lat,lon,depth,location,link,country,district)"
try:
    cursor.execute(sql)          # 执行 SQL 语句
    db.commit()                 # 提交到数据库保存
    count = count+1
    # print('插入成功!')
except:
    db.rollback()               # 如果发生错误,则回滚
    # print('插入失败!')

cursor.close()                 # 关闭游标
db.close()                     # 关闭数据库连接
```

图 5 在 Jupyter notebook 中利用 Python 第三方库向 MySQL 数据库写入地震数据

ID	发震时间	震级	纬度	经度	震源深度	震中位置	国家	地区
CC20230110014736	2023-01-10 01:47:33	7.6	-7.2	130.1	100	印度尼西亚海	印尼	班达海
CC20230110132454	2023-01-10 13:24:54	5.8	-14.95	166.8	120	瓦努阿图群岛	瓦努阿图	群岛
CC20230111163957	2023-01-11 16:39:57	5.8	-14.95	166.55	10	瓦努阿图群岛	瓦努阿图	群岛
CC20230112141910	2023-01-12 14:19:09	5.4	-30.65	-71.4	40	智利	智利	智利
CC20230115200421	2023-01-15 20:04:21	5.5	-14.85	166.95	60	瓦努阿图群岛	瓦努阿图	群岛
CC20230116062959	2023-01-16 06:29:59	6	1.9	97.9	30	印度尼西亚群岛北部海域	印尼	苏门答腊岛
CC20230116124053	2023-01-16 12:40:52	5.8	29	139.55	400	日本小笠原群岛地区	日本	小笠原群岛
CC20230118083446	2023-01-18 08:34:45	6.1	-0.1	123.3	160	印度尼西亚群岛西部海域	印尼	苏拉威西岛
CC20230118140615	2023-01-18 14:06:14	7	2.8	127.1	40	印度尼西亚群岛	印尼	马鲁古
CC20230118180813	2023-01-18 18:08:13	5.6	38.5	45.05	10	伊朗	伊朗	
CC20230119193041	2023-01-19 19:30:41	5.4	35.4	71.4	150	阿富汗	阿富汗	
CC20230120192340	2023-01-20 19:23:39	6.2	16.15	-62.15	170	海内群岛	海内	群岛
CC20230121060938	2023-01-21 06:09:38	6.6	-26.8	-63.2	600	阿根廷	阿根廷	
CC20230124101317	2023-01-24 10:13:17	5.5	2.85	127	30	印度尼西亚群岛	印尼	马鲁古
CC20230124165832	2023-01-24 16:58:32	5	29.55	81.5	10	尼泊尔	尼泊尔	
CC20230125023659	2023-01-25 02:36:59	6.3	-26.75	-63.15	600	阿根廷	阿根廷	
CC2023012503706	2023-01-25 03:07:06	5.8	35.65	28.55	50	地中海	地中海	

图 6 利用 Navicat 查看 MySQL 数据库中的地震数据

(四) Jupyter 交互式平台在教学中的优势分析

基于 Jupyter 开展的教学实践,有异于传统的教学方法,对于灾害大数据与人工智能课程而言,主要体现出如下优势,如:

1. 简单易用性:在有一定 Python 语言基础的前提下,Jupyter Notebook 可以通过命令的方式一键安装与启动,无需额外的学习基础;同时 Jupyter Notebook 支持 Markdown 语法,可以便捷地添加注释文本、公式、图片等,使得算法讲解和代码演示同时进行,提高了教学效率;同时其支持生成 html、pdf 等格式,方便教学演示和实践结果的保存和交流。

2. 数据可视化:Jupyter 提供了交互式的数据可视化功能,可以将灾害大数据以图表、地图等形式展示,帮助学生更直观地理解和分析数据。通过可视化,学生可以更好地理解灾害数据的特征、趋势和关联性,从而更好地应对和预测灾害。

四、结语

灾害大数据与人工智能是防灾减灾科学与工程专业基础课程之一,要求学生具备应用新一代信息技术手段管理和分析灾害防治与管理中的大量数据,培养学生进行智慧感知、互联、处理和协调的能力,为日后从事灾害防治与管理相关工作积累理论和实践基础。如何高效、低成本地开展面向应用的编程计算一直是非计算机专业理论与实践教学的一个痛点问题。本文从防灾减灾工程专业中灾害大数据与人工智能课程中的编程需求出发,提出利用 Jupyter 平台进行交互式编程演示与实践教学。以“Python 爬虫获取地震灾害事件与可视化分析”为例,展示了 Jupyter Notebook 在进行实际教学应用中的优势,其主要包括:学习成本低、交互性好、可视化能力出众、实践结果易于保存、协作能力强等特点。

参考文献:

- [1] 林子雨. 大数据技术原理与应用 第三版 [M]. 人民邮电出版社,2021
- [2] 杨忠宝, 余向飞. 大数据与人工智能 第一版 [M]. 北京大学出版社,2022
- [3] 马伟良. Jupyter Notebook 平台在 Python 教学中的应用 [J]. 数字通信世界,2022(05):82-84.

国家自然科学基金青年科学基金项目 (42104099): 贺根山—黑河缝合带北段壳—幔精细结构研究——深地震反射剖面的揭示, 2022.01-2024.12