

README

西安交通大学医学部科研产出分析项目

License MIT python 3.12+

目录

[项目目的](#)

[项目预计受众](#)

[软件列表](#)

[项目结构](#)

[分析流程](#)

[快速启动指南](#)

项目目的

- 本项目旨在对**西安交通大学医学部** (Xi'an Jiaotong University) 在特定时间范围内 (2021-2025) 的科研文献产出进行文献计量学分析。我们通过从 **PubMed** 等数据库获取数据, 利用 **Python** 进行数据处理、分析和可视化, **VOSviewer**可视化, 以揭示其研究热点、合作网络、发展趋势等关键信息。
- 同时, 本项目作者在创作时间是大一到大二的暑假, **未曾学过专业课, 也未曾接受过任何科研训练**, 做这个项目, 一是自学一些科研基础工具, 锻炼科研思维, 对专业领域和学校有大致地了解。

项目预计受众

- 同为医学专业, 本校的学生
- 想要了解学习pubmed, python, VOSviewer等软件的基础使用
- 或者仅仅想要了解结论, 尽管结论不太准确

软件列表

- 数据来源: PubMed(.nbib 格式)
- 数据管理: Zotero
- 数据处理与分析: Python 3.12+ (NumPy, Pandas)
- 数据可视化: Python3.12+(Matplotlib, Seaborn, Plotly, NetworkX) ,VOSviewer

- 开发环境: Jupyter Notebook, Visual Studio Code
- 版本控制: Git, GitHub

项目结构

为了实现代码和数据的分离，本项目采用以下目录结构（核心部分）：

```
.
├── .gitignore                # 规定了哪些文件不被Git追踪
├── notebooks/                # 存放所有的Jupyter Notebook分析脚本
│   ├── data_combine.ipynb    # 处理数据的第一步Jupyter（Python）文
件                               件
│   ├── data_process.ipynb    # 处理数据的第二步Jupyter（Python）文
件                               件
│   ├── data_process_files    # 将notebook转换为latex时的图片
│   └── out
│       ├── data_combine.pdf   # 第一步数据处理文件的pdf版
│       └── data_process.pdf   # 第二步数据处理文件的pdf版
├── data
│   ├── raw                   # 从PubMed下载的原始数据
│   └── processed              # 可视化分析结果图片
├── environment_map_of_XJTU_MED.yml # 项目所需的Python依赖包列表
├── 研究日志.md                # 研究全部过程，项目最重要的一个文件
├── 研究日志.pdf               # pdf版本
└── README.md                  # 就是你正在阅读的这个文件
```

分析流程

本项目的标准分析工作流如下：

1. **数据获取**: 从 PubMed 检索文献，导出 .nbib 格式，共五个，按年份分，导入 Zotero 进行管理，并导出为 .csv 格式，同样有五个文件
2. **数据预处理**: 在 Notebook 中编写代码，合并 .csv 文件为一个，清洗数据（如处理缺失值、统一作者格式等），并转换为 Pandas DataFrame。
3. **深度分析**: 开展具体的专题分析，如网络构建、趋势预测等。
4. **结果可视化与报告**: 从 Python 中导出可以导入到VOSviewer中的 .txt 文件，导入后，分析结果图，并得出结论。
5. **进一步了解**: 阅读本目录下的 研究日志.md，或者 研究日志.pdf

快速启动指南

为了让您轻松启动本项目，我们提供了一个批处理脚本，它将自动为您设置运行环境。（仅限win用户）

重要前提：安装 Conda (Miniconda 或 Anaconda)

本项目依赖于 [Conda](#) (推荐使用轻量级的 [Miniconda](#)) 来管理Python环境和依赖库。

在尝试运行本项目的任何脚本之前，请务必先安装 Conda。

- **Miniconda 下载与安装:**
- 推荐下载: [Miniconda 官方下载页面](#)
- 请根据您的Windows系统版本（64位或32位）选择合适的安装包。
- **安装时请注意:**
- 在安装过程中，有一步会询问是否“Add Anaconda to my PATH environment variable”(将 Anaconda 添加到我的 PATH 环境变量)。**强烈建议勾选此选项**，这样您就能在任何命令行窗口中使用 `conda` 命令。
- 如果安装后命令行仍然无法识别 `conda` 命令，您可能需要重启CMD或PowerShell窗口，或者手动将 Conda 的安装路径添加到系统环境变量 PATH 中。

激活环境步骤

下载或克隆项目:

- **如果您使用 Git:** 打开 Git Bash 或命令提示符 (CMD)，执行以下命令:

```
git clone https://github.com/Tony0111/MEDscience_map_of_XJTU.git
```

```
cd MEDscience_map_of_XJTU
```

- **如果您不使用 Git:** 直接点击 GitHub 页面右上角的绿色按钮 "Code"，然后选择 "Download ZIP"，下载项目压缩包。解压到您电脑上一个方便的位置。

首次运行：创建并配置虚拟环境: 在项目根目录下（也就是包含 `environment_map_of_XJTU_MED.yml` 的目录），**双击运行 `start_project.bat` 文件。**

- ****如果您这是第一次运行此脚本**:** 脚本会检测到虚拟环境 ``map_of_XJTU_MED`` 不存在，并自动提示您如何创建它。 它会提示您需要手动执行以下命令来创建环境并安装所有依赖:

- ```
```bash
conda env create -f environment_map_of_XJTU_MED.yml
```
```

请在 `Anaconda Prompt` 或已配置好 `conda` 的命令行窗口中执行该命令。这个过程可能需要几分钟，请耐心等待。 \*\*创建成功后，请再次双击运行 `start\_project.bat` 文件。

- **如果环境已存在：** 脚本会直接激活虚拟环境。

**开始操作:** 成功激活环境后，会弹出一个新的命令行窗口。您会看到命令行提示符前面多了一个 (map\_of\_XJTU\_MED)，这表示虚拟环境已成功激活。

- 打开 /notebook/ 文件夹下的
  - data\_combine.ipynb 进行**第一步**数据处理
  - data\_process.ipynb 进行**第二步**数据处理
- 按照提示**直接**运行理解就行

## 研究日志

- 研究日志是这个项目的最重要的一个文件
- 没有具体的技术细节
- 直观呈现研究过程和研究结果