数据可视化与可视分析课程设计

第6组设计文档

郑东林，王星力，王泽浩，王正

（上海交通大学电子信息与电气工程学院，上海 200240）

1. 系统框架初步设计
   1. **数据分析**

本次大赛使用的大气污染数据集为中国科学院大气物理研究所等单位发布的中国高分辨率大气污染再分析数据集，提供两种时间分辨率数据集，分别是日均值数据集和小时值数据集，两套数据集基本信息如表 1和表 2所示。

表 1 每日分析数据

|  |  |  |
| --- | --- | --- |
| 时间范围 | 2013 年-2018 年 | 共 72 个 zip 文件，每个文件以年月命名，包含当月的日均值数据。  采取 csv 文件格式，分隔符为逗号。  csv 文件以日期命名，包含当日污染物浓度和气象要素的日均值。 |
| 空间范围 | 中国 |
| 空间分辨率 | 15 公里 |
| 位置变量 | 网格中心纬度（Lat）、网格中心经度（Lon） |
| 污染物变量 | 细颗粒物（PM2.5）、可吸入颗粒物（PM10）、臭氧（O3）、一氧化碳（CO）、二氧化硫（SO2）、二氧化氮（NO2） |
| 气象变量 | 纬向风速（U）、经向风速（V）、温度（TEMP）、相对湿度（RH）、地面气压（PSFC） |

表 2 小时分析数据

|  |  |  |
| --- | --- | --- |
| 时间范围 | 2013 年-2018 年，每年 1 月份 | 共 6 个 zip 文件，分别为 2013~2018 年 1 月的小时数据。  采取 csv 文件格式，分隔符为逗号。  每个 csv 文件以所在的小时命名，包含该小时污染物浓度和气象要素值。 |
| 空间范围 | 中国 |
| 空间分辨率 | 15 公里 |
| 位置变量 | 网格中心纬度（Lat）、网格中心经度（Lon） |
| 污染物变量 | 细颗粒物（PM2.5）、可吸入颗粒物（PM10）、臭氧（O3）、一氧化碳（CO）、二氧化硫（SO2）、二氧化氮（NO2） |
| 气象变量 | 纬向风速（U）、经向风速（V）、温度（TEMP）、相对湿度（RH）、地面气压（PSFC） |

接下来，我们从数据的作用和适用的呈现方式来分析各数据段。

位置变量，即网格中心纬度（Lat）和网格中心经度（Lon）是其他数据展示的基础。依托经纬度的可视化，能有效的揭示数据的空间分布，展示数据的空间特征。直观地，数据可以通过绘制地图展示：将经纬度作为自变量，其他与位置有关的数据作为因变量，用热力图展示。

污染物变量，即细颗粒物（PM2.5）、可吸入颗粒物（PM10）、臭氧（O3）、一氧化碳（CO）、二氧化硫（SO2）和二氧化氮（NO2），可以表征特定地区在特定时刻关于特定污染物的污染状况。通过相应的计算公式，可粗略计算AQI，并确定首要污染物。在数据可视化方式上，可以使用热力图、分布统计图和雷达图等多种图表形式。热力图中，区域颜色代表该区域的不同污染物的污染状况。分布统计图可以展示某地区不同空气质量等级的分布。雷达图能够展示某地区不同污染物的指数。

气象变量可进一步细分为两类探讨。

对于纬向风速（U）和经向风速（V），由于其矢量的形式，可以在地图上标注箭头，箭头大小表示风速，箭头方向表示风向。我们可以探究风速和风向对污染物扩散的影响，并据此预测下一时间段的空气质量。

温度（TEMP）、相对湿度（RH）、地面气压（PSFC）三个变量，主要可用作探究各气象条件与空气质量、各污染指标之间的关系。同样可以在热力图中呈现，展示不同地区各气象变量的分布；也可以绘制一个模拟湿度计，展示某地区的各气象指标。

* 1. **主题分析**

对于可视化任务的不同主题，我们分别进行分析。

1) 大气污染源分析。识别主要大气污染源方面，可以进行大气污染源分成分分析，同时要考虑当地城建状况（如火力发电厂、化工厂等）与自然状况（森林、农田、沙漠等）。关键污染成因方面，可以考虑植被状况、城市建设、气候因素和异常事件等。

2) 大气污染时空态势分析。时间分布模式方面，可考虑每日变化趋势和污染高发季节；空间分布模式方面，可考虑污染高发地区、污染重心移动以及不同主要污染物的分布特点；演变态势方面，可考虑风速、风向、降水等以及当地地理特征。

3) 大气污染传输模式分析。分析各地大气污染物差异可采用雷达图可视化和分成分热力图可视化；分析大气污染传输模式可考虑污染重心随风向的变化、污染扩散速度随风速的变化以及其他气象因素对污染生成与扩散的影响。对于异常传输事件，可考虑自然环境因素和城市建设因素。

4) 大气污染预测。综合风速风向、污染物浓度、温度降水，城建情况、植被情况，异常事件，地理特征等关键信息，提炼关键隐藏特征，找到大气污染固定变化模式，预测污染物分布与浓度变化。

5) 大气环境的改善。对于大气治理过程中的大气环境状况，可展示每年一月空气质量指数的分布（下限）、每年优良空气质量天数（上限）和空气质量分布情况的变化（整体）。对于大气污染防治措施，可评价植树造林、绿色能源、经济发展方式转变等措施的作用。

* 1. **界面设计**

可视化界面的整体设计如图 1所示。点击地图上某地之后，会弹出相应地区的详细图表（如图 2所示）。

图 2 交互式详细弹出界面

图 1 界面整体布局

可缩放的地区热力图

选定时刻的污染物雷达图

标签：日均/时均分析模式切换

可缩放的地区热力图

标签：分析维度切换

可缩放、并且带有播放功能的时间轴

根据时间粒度与统计维度切换的分布统计图

折线统计图

（根据时间粒度与统计维度切换，与分布统计图联动）

温度

湿度

气压

直方图，用于展示当前时间地区的具体污染物数据

图 1中的“标签”处采用按钮选择不同的数据集。“时间轴”的设计类似于视频播放器底部可拖动的进度条，拖动时间轴可以改变热力图上呈现的数据所属的时间。“分布统计图”可展现某地区优良及污染天数的分布情况。温度/湿度/气压图以类似于湿度计的形式进行设计。

1. 任务分工

项目分阶段任务分工如表 3 项目任务分工所示。该任务分工为初步计划，具体分工以实际为准。

表 3 项目任务分工

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 项目阶段 | 工作任务 | 王星力 | 王泽浩 | 王正 | 郑东林 |
| 框架设计 | 数据分析 | √ |  | √ |  |
| 需求分析 |  | √ |  | √ |
| 界面设计 | √ |  |  | √ |
| 可行性分析 |  | √ | √ |  |
| 小组展示准备 | √ | √ | √ | √ |
| 代码实现 | 环境搭建与调试 |  |  | √ |  |
| 热力图实现 | √ | √ |  |  |
| 时间轴与交互实现 |  | √ | √ |  |
| 弹出窗口实现 |  |  | √ | √ |
| 组件工具实现 | √ |  |  | √ |
| 整体调试 | √ | √ | √ | √ |
| 设计调整 | 配色调整 |  | √ | √ |  |
| 布局调整 | √ |  |  | √ |
| 项目提交 | 作品说明文档编写 |  |  | √ | √ |
| 视频录制 | √ | √ |  |  |

1. 计划安排

项目计划安排如所示。该进度计划为初步计划，具体进度以实际为准。

表 4 计划安排

|  |  |
| --- | --- |
| 周数 | 任务 |
| 10 | 项目开题，完成系统框架初步设计，确定任务分工和计划安排。 |
| 11 | 完成开发环境的搭建和调试，编程知识的学习。 |
| 12 | 完成热度图主要框架的编程；完成具体地点交互界面的主要框架的编程。 |
| 13 | 完成热力图时间轴的添加；完成其他图表的编程。 |
| 14 | 各组件的调试完善。 |
| 15 | 系统整合调试、设计调整。 |
| 16 | 项目结题；完成报告和视频，小组汇报。 |