

# 西南科技大学

Southwest University of Science and Technology

## 本科毕业设计（论文）

题目名称： 基于多源数据融合的城市气象污染  
可视分析系统设计与实现

学 院 名 称	<u>计算机科学与技术学院</u>
专 业 名 称	<u>软件工程</u>
学 生 姓 名	<u>王琛</u>
学 号	<u>5120190931</u>
指 导 教 师	<u>王松 副教授</u>

二〇二三年六月

# 西南科技大学

## 本科毕业设计（论文）学术诚信声明

本人郑重声明：所呈交的毕业设计（论文），是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

作者签名：

日期：        年    月    日

# 西南科技大学

## 本科毕业设计（论文）版权使用授权书

本毕业设计（论文）作者同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权西南科技大学可以将本毕业设计（论文）的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本毕业设计（论文）。

保密☐，在\_\_\_\_年解密后适用本授权书。

本论文属于

不保密☐。

（请在以上方框内打“√”）

作者签名：

指导教师签名：

日期：      年    月    日

日期：      年    月    日

---

# 基于多源数据融合的城市气象污染可视分析系统设计与实现

**摘要：**随着社会经济的高速发展，人类社会的物质生活条件越来越好，对于天气污染状况也越来越重视。国家工业化建设的加速发展使得大气污染对人们的生产生活造成了严重危害，逐渐成为了当前较为严重的环境问题之一。但是随着大气污染监测站点数量的增加，所产生的数据量也越来越大，数据维度越来越多源化，如何更加的科学处理数据以提供良好的城市气象污染数据的信息化及预测服务问题迫在眉睫。

本课题通过构建城市大气污染数据可视化系统，将气象属性数据和污染物浓度监测数据进行可视分析和可视化展示。在满足人们对基本大气环境状况以及污染物可视化分析的需求的基础上，极大方便了人们的生产生活。本课题的主要工作内容如下：

（1）探索城区空气质量与污染物相关属性的时变过程，将时间序列与其他气象影响因素相结合进行可视化分析，分析大气污染物内部及其与外部气象因素的相关程度，探究其中的时空演变规律。

（2）分析各空气污染监测站点的气象污染分布情况与污染差异性，结合机动车流量数据探究影响大气污染时空分布模式的因素，分析某一时段内的大气污染扩散传输关系。

（3）利用 LSTM 长短时序记忆神经网络模型，并结合降雨量、风速、机动车流量等属性，准确分析大气污染时空变化规律，并进行污染气体的精确预测。

（4）基于气象数据的关键词提取算法，使用 Text-Rank 算法输入向量化关键词，实现气象文本关键词的分析。

（5）分析监测站点的污染物浓度与风场轨迹变化、交通卡口机动车车流量迁移的关联性，并提出面向气象污染的城市交通整治方案，例如货车限行、交通卡口车流量管辖等。

最后，系统通过公开的气象污染数据集辅助从多维度、多层次分析大气污染对城市不同区域的影响程度，从 PM10、PM2.5、SO2、O3、NO2 污染物以及 AQI 等多维气象属性之间挖掘关联关系。

**关键词：**多源数据；大气污染；可视化；关联分析

# Design and Implementation of a Visual Analysis System for Urban Meteorological Pollution based on Multi-source Data Fusion

**Abstract:** With the rapid development of social economy, the material living conditions of human society are getting better and better, and more and more attention is paid to the weather conditions. With the accelerated development of national industrialization, air pollution has caused serious harm to people's production and life, and has gradually become one of the more serious environmental problems. However, with the increase of the number of air pollution monitoring stations, the amount of data generated is also increasing, and the data dimension is becoming more and more multi-source. How to process the data more scientifically to provide good information and prediction services for urban meteorological pollution data is imminent. Through the construction of urban air pollution data visualization system, the meteorological data and pollutant concentration monitoring data are visually analyzed and visually displayed. On the basis of meeting the needs of people for visual analysis of basic atmospheric environmental conditions and pollutants, it greatly facilitates people's production and life. The main work of this subject is as follows:

(1) Explore the time-varying process of urban air pollution attributes, combine the time series with other meteorological factors for visual analysis, analyze the correlation between internal and external meteorological factors of air pollutants, and analyze the spatial and temporal evolution of them.

(2) Analyze the distribution of meteorological pollution and the difference of pollution at each monitoring station, explore the factors affecting the spatial and temporal distribution pattern of air pollution combined with the data of motor vehicle flow, and analyze the diffusion and transmission relationship of air pollution in a certain period of time.

(3) Using LSTM long and short time series memory neural network, combined with rainfall, wind speed, motor vehicle flow and other attributes, accurately analyze the spatial and temporal variation of air pollution, and accurately predict the pollution gas.

(4) Keyword extraction algorithm based on meteorological data, using Text-Rank algorithm to input vectorized keywords, to achieve the analysis of meteorological text keywords.

(5) Analyze the correlation between the pollutant concentration of monitoring stations and the change of wind field trajectory and the migration of motor vehicle flow at traffic checkpoints, and propose urban traffic control schemes for meteorological pollution, such as truck restriction, traffic flow jurisdiction at traffic checkpoints, etc.

Finally, the system analyzes the impact of air pollution on different areas of the city from multiple dimensions and levels through the open meteorological pollution data set, and mines the correlation between PM10, PM2.5, SO2, O3, NO2 pollutants and AQI and other multi-dimensional meteorological attributes.

**Key words:** Multi-source data; Air pollution; Visualization; Association analysis

# 目 录

第 1 章 绪论.....	1
1.1 概述.....	1
1.2 国内外研究现状及存在问题.....	2
1.2.1 研究现状 .....	2
1.2.2 存在问题 .....	3
1.3 系统目标.....	4
1.4 论文结构.....	4
第 2 章 需求分析与可行性分析.....	6
2.1 数据处理可行性.....	6
2.1.1 关键词提取 .....	6
2.1.2 空气污染预测模型构建 .....	7
2.2 可视化表达可行性.....	7
2.3 功能需求.....	8
2.3.1 系统功能结构 .....	8
2.3.2 系统功能简述 .....	8
2.4 人机交互需求分析 .....	11
2.5 本章小结.....	12
第 3 章 可视分析系统设计.....	13
3.1 系统设计.....	13
3.2 数据处理.....	14
3.2.1 数据来源 .....	14
3.2.2 数据预处理 .....	17
3.3 数据特征提取.....	18
3.3.1 基于 Text-Rank 的天气词云.....	18
3.3.2 基于 LSTM 长短时记忆网络模型的空气质量预测分析 .....	21
3.3.3 结合车流量数据与污染物之间及污染物与 AQI 指数的关联分析.....	23
3.4 可视化布局.....	24

3.5 可视化交互方案设计.....	27
3.6 本章小结.....	28
第 4 章 可视分析系统实现.....	29
4.1 系统总体架构.....	29
4.2 系统整体布局构建.....	29
4.3 可视化模块实现.....	30
4.3.1 气象属性分析展示模块 .....	30
4.3.2 大气污染物与 AQI 数据对比分析模块.....	34
4.3.3 机动车流量与污染物相关性分析模块 .....	38
4.3.4 污染物之间相关性与数据预测综合分析模块 .....	41
4.3.5 中心地图可视化绘制分析模块 .....	44
4.4 本章小结.....	45
第 5 章 系统评估.....	46
5.1 案例分析.....	46
5.1.1 案例一：降水量对污染物浓度及 AQI 指数的影响分析.....	46
5.1.2 案例二：城市不同监测站点污染物对比及污染物与 AQI 关联分析....	48
5.1.3 案例三：交通数据对各类污染物的影响分析 .....	50
5.1.4 案例四：监测站点污染传播关系与风速风向对空气污染的影响分析 ..	53
5.2 系统测试.....	57
5.2.1 系统视图测试 .....	57
5.2.2 系统交互测试 .....	58
5.3 本章小结.....	60
总结与展望.....	61
致谢.....	62
参考文献.....	67



# 第 1 章 绪论

## 1.1 概述

近年来随着社会经济生活的快速发展，空气污染问题已然成为了群众、政府部门与相关研究学者等各方日益关注的问题，且人们对于天气污染状况问题的重视程度也越来越高。2013 年由国务院相关部门印发的《国务院关于印发大气污染防治行动计划的通知国发〔2013〕37 号》文件中，充分体现了国家对于大气污染改善问题的重视，目前空气污染治理工作已成为了政府相关部门需要解决的关键问题之一，特别是针对雾霾天气频发问题对环境和人体健康影响的严重损害，大气污染监测与防治工作成为环保部门当前最重要的任务之一。

以山东省聊城市为例，该市各区县都建立起了大气污染监测站点，对所管辖区域进行空气质量监测和污染源监测工作，因此每年会产生大量气象数据和空气污染浓度排放数据，这些数据对于环境监管部门深入了解该地区空气污染状态，改善空气质量，提升生态环境治理水平提供了重要保障。与此同时，收集到的气象数据也具有量大、多维度、时序变化等特征，对于这些气象数据的综合分析应用还不完善，如何利用此类数据分析理解大气污染成因及传输模式，对其实际应用的深入开发和历史数据的挖掘分析工作进行拓展，指导环保工作部门推动空气污染监管治理工作效率的提高。

利用大数据分析技术和可视化方法，通过将各种抽象数据类型转变为以图形信息进行展示，并利用不同类型的视图交互技术，将数据处理、视觉表征展示与智能交互分析相结合，使机器智能和人类智慧深度融合、优势互补，如图 1-1 所示，快速感知大气污染的时变规律，辅助工作人员因地制宜地制定防治策略，为大气污染治理防护工作的分析、指导和决策提供有效手段。同时也为大气污染治理的信息化与数字化提供了技术支持。

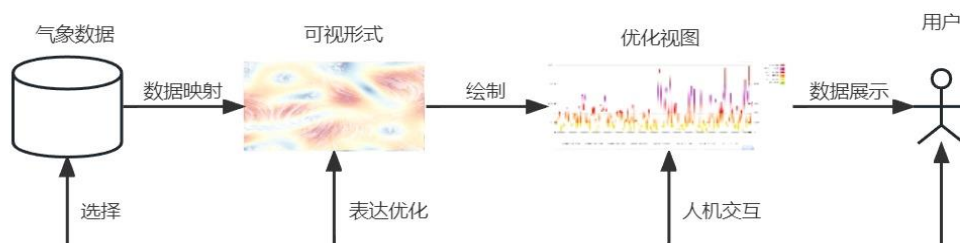


图 1-1 大气污染数据可视化流程图

## 1.2 国内外研究现状及存在问题

### 1.2.1 研究现状

气象污染可视分析是对大气污染治理工作进行评估分析的重要组成部分,随着城镇化进程的加快,相关部门在大气环境污染领域收集了大量历史数据,信息化技术的快速发展也为可视化分析系统的开发提供了技术保障。国内外相关学者也对其进行了大量学术研究。

国内方面,屈华民<sup>[1]</sup>等人在香港空气污染问题方面通过可视化分析技术进行了深入研究。蒋志方<sup>[2]</sup>等人结合大气污染物浓度数据,采用通过圆环段技术(Circle Egments)对其进行了可视化分析。马春燕<sup>[3]</sup>等人对空气污染的时空分布规律进行了探究,并将各个城市的空气污染数据信息进行了对比展示,反应了城市气象污染随时间的变化趋势。孙国道<sup>[4]</sup>等人对城市群区域污染协同防治策略的空气污染动态演化过程进行了探索,并提出了一种基于城市群区域布局的大气污染监测分布可视化系统。雷宝<sup>[5]</sup>等人利用大数据分析技术与可视化技术对具有时序特点的空气污染数据进行了探索性分析。孙浩鑫<sup>[6]</sup>等针对山东省内收集的大气污染数据的时间序列特征进行了探究,发现了一种基于主题河流图(ThemeRiver)的时间序列聚类算法的可视化方法,能够让用户更好的掌控监测站点数据之间时序特征的关联性与相似性。Guyu Zhao<sup>[7]</sup>等人利用复杂网络社区的特点,反映了空气质量的局部性和传播性特点,为大气污染的防治和预测奠定了基础。张涛<sup>[8]</sup>通过分析石家庄市的质量监测数据,采用 Echarts、Hight Charts 等工具提供的图表对数据进行了可视化分析,结合数据挖掘技术和 BP 神经网络算法模型对监测数据进行了有效预测,构造了大气污染预测模型。Zhou et al<sup>[9]</sup>提出了基于京津冀、珠江三角洲以及长江流域三大经济区的空气污染数据时空区域图,主要采用了日历图与热力图相结合的形式呈现了每日空气质量情况,并提出了一种分析研究散点区域空气污染状况与时空分布模式的方法。王义祥<sup>[10]</sup>等借助 ArcGIS Engine 二次开发组件库对大气污染时序特征以及空间传播趋势进行了可视化分析试验。赵桢<sup>[11]</sup>通过分析哈尔滨市的空气质量数据实现了污染物相关预测,提出基于粗糙属性的数据神经网络预测和聚类分析功能。

国外方面,例如, Du et al<sup>[12]</sup>创造了一种集成了日历图、填充地图、折线图等功能特点的大气污染检测办法。 Shaadan et al<sup>[13]</sup>针对马来西亚国家的 O<sub>3</sub> 污染物浓度数据进行了进一步研究,采用可视化技术分析 O<sub>3</sub> 浓度标量特征。Choi et al<sup>[14]</sup>提出了基于 Python 语言开发的空气污染物分析工具,可实现各类污染物之间相关性研究与可视化展示。Li

et al<sup>[15]</sup>提出了一种能够快速建立包含时序特征的空气污染数据的可视化分析系统。Jiang et al<sup>[16]</sup>探究了新南威尔士(NSW)背景下的大气污染数据与多监测站点气象的聚类分析的自组织地图可视化办法,用以对气候气象条件和空气质量污染因子之间影响程度的评估。Zhang et al<sup>[17]</sup>建立了一种基于虚拟地理空间环境(Virtual Geospatial Environment, 简称 VGE)的系统集成多维度气候气象数据和空气质量管理决策模型。Meindnardi<sup>[18]</sup>等人通过分析大气污染物与交通相关数据之间的关系,得到了交通流量的增加和其排放尾气的增加是城市气象污染浓度急剧增加的主要原因。Baker and Pan<sup>[19]</sup>通过在 MONET 开发和应用程序中使用模块化的 python 包,实现了在时空维度上对监测数据进行配对以及相关数据的可视化分析。An et al<sup>[20]</sup>构建了一种可视化分析工具来支持微观尺度中的空气质量(MSAQ)关联分析系统,用户可以针对多种不同控制决策办法的实现效果进行比较。

### 1.2.2 存在问题

以上研究成果都为本题研究的顺利完成提供了有效参考与借鉴,然而在大多现有研究中,相关研究人员所建立的可视化分析系统都是在其规定的功能和特殊环境需求上构建的,照搬或者在原有系统上改进此类技术方法并不能完全满足本题研究的需要,并且在以下方面还存在较大的提升空间:

(1) 目前大部分现有系统都只针对大气污染数据的可视化分析,缺乏区域经济生产生活、交通数据及人口变化等反映空气质量外部影响的数据集。此类数据的可视化展示也将极大的帮助管理者从多维度、多视觉的分析大气污染成因,为大气环境保护、车辆管控、生活生产方面的政策制定提供重要依据。

(2) 对站点积累的历史数据研究分析需要进一步加强,针对目前已存储的庞大的历史空气污染数据而言,固有的数据预测分析模型已很难充分挖掘其数据中蕴含的信息,深入挖掘这些数据在时空上的关联性也不失为有效治理城市气象污染问题的基础,这对于相关部门及人员从全局出发制定区域针对性治理方案具有重要意义。

(3) 近年来,许多学者对层次结构以及具有时序的特征数据进行了深究并取得了许多研究成果,然而对于二者相结合的研究成果还不多,如不同区域之间的空气质量在时空上的变化趋势,这也有助于从全省层面分析大气污染的地域相关性,直观系统的掌握大气污染的演变规律,成为环保相关部门工作中的一项关键内容。

### 1.3 系统目标

以城市大气污染数据为分析基础，抽取相关数据特征，实现相关数据可视化分析工作。对本系统的设计方案有如下期望：

（1）在功能实现上此程序可以帮助用户分析大规模复杂多源气象污染数据集，最终快速发现气象异常，并且分析大气污染问题及成因、跟踪监测大气污染发展趋势并确定关联情况。

（2）在数据挖掘上可以从不同维度深入研究大气污染因子异常属性，并分析出大气气象污染出现异常的具体原因。

（3）在感官呈现上针对多源城市气象污染数据进行可视化展示与视觉感官优化，使用户能够在不进一步了解其污染特性如何影响的情况下，实现气象污染因素对比。

（4）在用户使用上该系统的分析图表具有良好的协同交互能力，用户可以高效操作系统达到分析研究的目的。

基于以上目标，本课题将围绕需求分析、可视化模块选取与构建、数据预处理、交互设计、案例分析和系统测试六个方面进行具体阐述。

### 1.4 论文结构

本文主要为设计并实现一种基于多源数据融合的城市气象污染可视分析系统进行探讨，整体组织结构如下：

第一章为“绪论”，阐述了城市气象污染可视化研究提出的背景与意义，同时简要回顾了国内外研究现状，并对本课题的主要研究工作进行简明总结。

第二章为“需求分析与设计”，该章节将结合本课题的研究意义确定可视化系统的相关需求，最终确定系统相关功能、数据运用以及开发运行环境，并根据具体的需求分析设计出城市气象污染分析可视化系统的总体结构。

第三章为“可视化表达设计与研究”，对上述部分描述的系统设计方案中可能遇到的问题及其解决办法进行详细阐述，最终得到本课题中涉及到的可视化分析模型与进一步的优化方案。

第四章为“可视化系统设计与实现”，主要对系统开发实现过程中所遇到的问题以及相关解决方法逐一详细的进行解释，并对系统的实现、运行等过程进行描述。

第五章为“可视化系统评估”，主要对前面实现的可视化系统进行案例分析，以达到预期目标的效果，并对系统总体实现程度进行评价。

最后为“总结与展望”部分，对本课题的工作成果进行总结归纳，对未来课题研究工作做出展望。

## 第2章 需求分析与可行性分析

### 2.1 数据处理可行性

#### 2.1.1 关键词提取

在研究人群情感意见时,主要以文本提取和分析作为研究目标,也称为意见抽取(Opinion Extraction),构成了文本分析研究的重要范畴。试图通过民意数据和天气信息对进行主观探索,以获取关键线索。

近年来,机器学习、人工智能及神经网络的快速发展,使得关键词的准确提取成为可能,目前市场上出现了许多传统提取方法与机器学习相融合的创新型关键词提取方法。其研究方法对海量数据集进行挖掘和训练,在关键词抽取领域发挥出了巨大作用<sup>[21]</sup>。

目前的机器学习方法主要有无监督学习、监督学习和半监督学习。主要分析方向为:面向任务、认知模拟和理论分析<sup>[22]</sup>。对监督学习与无监督的学习方法提取的关键词进行比较分析可以得知,有监督学习方法抽取出的关键词大多数依赖于语料库的好坏程度。在关键词文本抽取过程中,有监督学习的规则建立往往是基于庞杂的语料库,文本语义提取结果相比于无监督的学习方法也更加的精准。而无监督学习方法较多,并且也不需要人工标注的语料库,是近年来机器学习研究的热门话题,但文本自身具有复杂性,存在文本提取语义准确度较低,词语冗余度较高等问题。

目前,在已有的文本分析方法的基础上,结合本课题所使用的数据集提出了基于监督学习模型的关键词提取办法,主要描述如下所示:

(1) 数据分割:以 7:3 的分割比例将数据分为训练集与测试集,此时的数据分割通常为无序分割,并且数据存在于标记的样本之中。

(2) 样本特征提取:中文文本比英文文本更为复杂难懂,中文文本中存在很多拟声词、虚词等无实际意义的词语。故需对特定特征的数据进行建模来创建、量化词库,采取 Word2Vec 作为分词算法策略,选择获取到的特征向量并根据其影响程度进行权重排序,最后形成具有特征维度的词库模型。

(3) 输入模型,提取出准确的模型才能表达气象数据关键词的正确含义,目前常用的提取方法主要为 Text-Rank 算法、TF-IDF 算法等。

(4) 总体评估,借助评价指标进行分析是对于构造成的模型最有效的评估方式,目前常见的评价指标:Recall、Precise、F1-Score 等。

### 2.1.2 空气污染预测模型构建

空气污染监测数据是典型的时间序列数据类型,且具有维度高、数据量大、复杂度高的特点,传统预测方法多为简单时序平均数法<sup>[23]</sup>、移动平均法<sup>[24]</sup>、加权移动平均法<sup>[25]</sup>、指数平滑法<sup>[26]</sup>、季节性趋势预测法<sup>[27]</sup>等,但是预测结果多数较差。随着大数据分析技术的日渐成熟,针对时间序列特征的预测模型构建,循环神经网络(Recurrent Neural Networks, 简称 RNN)表现出了良好的预测属性<sup>[28]</sup>。本文主要介绍 LSTM 长短时记忆神经网络的工作特点和算法流程,为后续的预测模型构建工作提供理论基础。

简单来说 LSTM 也同样是一种循环神经网络,并采用不断向后传播进行相关训练来克服梯度消失等问题,空气污染数据除了具有典型时间序列特征之外,也具有随机性、周期性、非线性等特点,而 LSTM 神经网络在这一方面具有较好的适应性<sup>[29]</sup>。

## 2.2 可视化表达可行性

在该可视化系统中,主要采用了 Echarts.js 作为首要组件, G2.js、G6.js 作为支撑, L7.js 则为地图部分的底层组件。其中 Echarts 是一款遵循 Apache 2.0 开源协议的免费 JavaScript 图表库。最新版本的库具有更高数据量级的可视化渲染能力,通过 SVG 与 Canvas 双引擎实现多图表联动数据分析<sup>[30]</sup>。

另外 Echarts 也支持多维数据和时空数据的可视化表达,并且拥有丰富的可视化图表库和灵活的交互能力,用户使用便捷,配置简单,且具有可扩展性、易用性以及高性价比等特点。

D3 (Data-Driven Documents) 是一种不需要依赖任何其他开发框架且严格遵循 Web 标准的图表库,兼容市场上的各大浏览器。同时也会出现因为数据刷新过于频繁而造成渲染速度变慢的情况。

Highcharts 是基于 HTML5 技术的开源图形库,通过两个 JS 文件即可运行,故其不需要安装任何插件。并且有着动态交互性、多坐标轴、时间轴、缩放、钻取等功能,但不支持 WebGL 组件。

系统前端使用了 Vue 框架搭建而成,自底向上增量开发的设计模式使其成为了一套基于用户界面所构建的渐进式框架。Vue 核心库只对应图层,并且本身也具有响应式编程和组件化开发的优点。

本系统使用 python 中的 pandas、numpy、jieba 等库函数进行数据预处理。Python 库具有语法简洁、维护方便、库函数调用方便、可移植性强等优点，相比于其他开发语言，python 在数据的处理、分析、探索性计算及可视化方面有较大优势。

## 2.3 功能需求

### 2.3.1 系统功能结构

本系统功能主要关注挖掘气象污染数据属性之间的关联关系（如降雨量、空气质量、能见度等），分析各监测站点的气象污染分布情况和污染物差异性，探索监测站点空气污染变化与卡口车流量迁移、风场轨迹的关联性等方面的内容。系统所设计的技术和主要功能需求如图 2-1 所示。

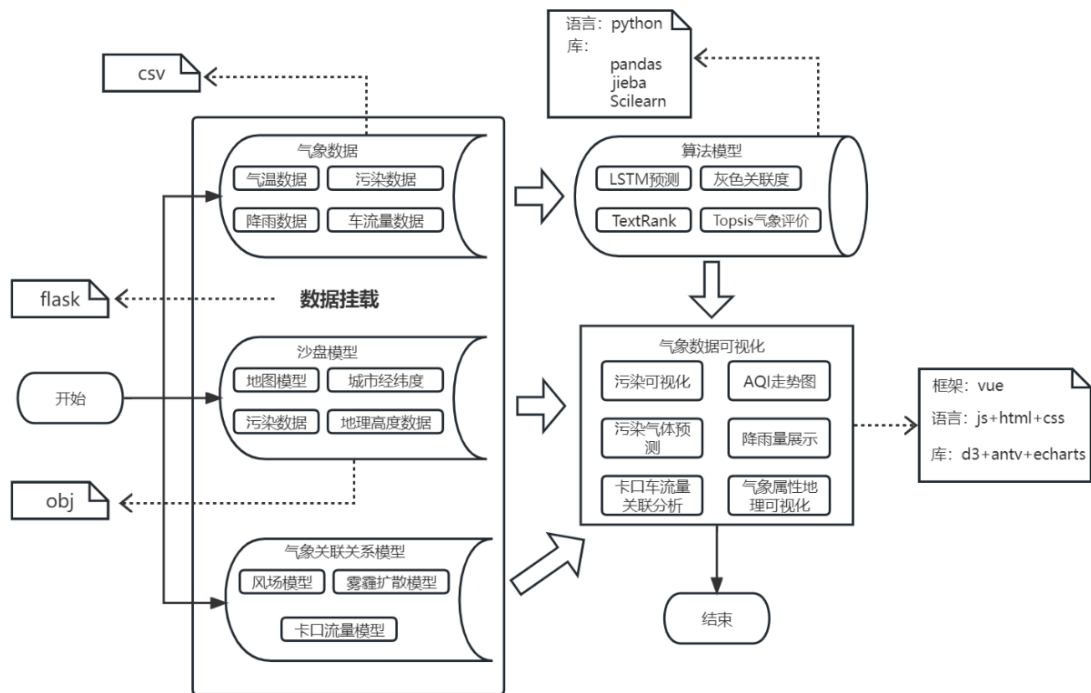


图 2-1 主要功能需求图

### 2.3.2 系统功能简述

针对本课题所研究的主要内容，设计系统的主要功能模块，如下表 2-1 所示，下面将对各个模块的具体功能进行简要概述。



表 2-1 功能需求汇总

序号	模块	功能点	备注
1	气象数据关联关系	天气属性查询与时间选取	
		AQI 与降雨量关联分析	
		风速风向、天气情况关联分析	
		AQI 等级与时间序列关联分析	
		天气现象词云	
2	污染物数据分析	城市污染物聚类分析	
		各类污染物浓度趋势对比分析	
		AQI 与各类污染物关联关系	
		不同时段污染物浓度对比分析	
		城区之间污染物异常特征分析	
		城区之间污染物特征对比	
3	机动车流量数据	交通卡口各时段车流量分析	
		车流量与污染因子含量关联关系	
		各机动车型车流量管控建议	
		机动车区域之间流动与污染物扩散关联关系	
4	空气质量预测模型	AQI 预测分析	
		监测站点污染物相似度对比	
		大气污染浓度与机动车流量拟合度分析	
		污染物预测误差分析	

## 1、气象数据关联关系分析

(1) 天气属性查询与时间选取。对天气属性相关数据(例如温度、降水量、风速等)进行直观数据的分析,同时可进行时间片段的自由选取。

(2) AQI 与降水量关联分析。对 AQI 与降水量进行统一的绘图分析,使用不同颜色的折线来表示二者的关系,也可以使用时间轴进行不同时间段的筛选。

(3) 风速风向与天气情况的关联分析,使用不同颜色的箭头和图标来表示不同的风速风向与天气情况,使用户更好的了解某一时间段内的天气变化情况,进而更好的分析污染物的产生情况。

(4) AQI 等级与时间序列的关联分析。主要分析 AQI 在某一周、某个月或者某一季度中的大小分布情况,使用户更好的掌握城市总体的污染情况,在时间上更有针对性的进行治理。

(5) 天气现象词云。对某段时间内的天气现象进行关键词提取,可以更直观的分析出天气变化情况,并与污染物的变化进行关联分析。

## 2、污染物数据分析

(1) 城市污染物聚类分析。通过构建聚类模型,将不同监测站点的污染物特征进行比对,可以对不同区域的污染问题有针对性采取治理措施,从而达到有效加快治理改善生态环境的目的。

(2) 各类空气污染物浓度趋势变化对比分析。将污染物含量进行等级分类,通过对不同种类的大气污染物趋势的变化分析,设计得到主要污染物的季节性差异视图。

(3) AQI 大小与各类污染物关联关系。AQI 根据 PM<sub>2.5</sub>、PM<sub>10</sub>、SO<sub>2</sub>、NO<sub>2</sub>、O<sub>3</sub>、CO 六项污染因子综合参数得出的空气污染程度的表述。可通过对不同污染物类型浓度的改变来确定各类污染物对 AQI 的影响程度。

(4) 监测站点不同时段污染物浓度对比分析。对各区域不同污染物在不同时间的最大最小值以及均值进行计算并进行绘图设计展示,得出污染物在不同月份或者季节的变化情况。

(5) 城区间污染物异常特征分析。针对某些区域中发现某种污染物的突然沉积或释放的情况,则应将其设计为污染物明显变化的视图。

(6) 城区间污染物特征概况对比分析。当某一区域的某一污染因子在一定时间内发生较为剧烈波动时,用户需要及时检测查询不同地区的不同污染物浓度,针对性的分析气象污染出现的具体原因,故需要实现满足上述需求的可视化编码。

### 3、机动车流量与大气污染物浓度关联分析

（1）监测站点各时段车流量分析。针对各个监测站点中的各类车型在不同时间段内的通过程度，设计合适的可视化视图来满足车流量变化情况。

（2）车流量与污染因子含量关联关系。针对机动车排放对污染物浓度高低的关联关系，设计可视化图表来展现各类型机动车流量对某一些特定污染因子的影响情况，有助于相关部门更好的将机动车管控与环境保护相结合，进而改善生态环境。

（3）各类车型车流量管控建议。依据污染物浓度的变化情况，对某一卡口的通车情况进行分析，并对不同卡口设计车流量管控建议并展示。

（4）各类车型城区之间流动与污染物变动关联关系。考虑车辆的空间分布情况，通过污染物数据研究车辆污染废气排放对环境的影响程度。

### 4、空气质量预测分析

（1）AQI 预测分析。通过建立预测模型，结合降雨量、污染物数据对 AQI 数据进行预测，并进行合理的可视化呈现。

（2）监测站点污染物相似度对比。对不同监测站点的同一污染物进行相似度对比，结合机动车流量、经济生产、植被变化情况等因素对空气污染的关联情况进行评价。

（3）AQI 与车流量大小拟合度分析。对于 AQI 本身来讲，不同污染物之间存在复杂且有序的关联关系，但在某些污染物和机动车流量的背景下也可能存在负增长的关联关系。

（4）污染物预测误差分析。对预测的数据进行小范围误差分析，最终结合实际数据为大气污染防治工作的分析、指挥和决策提供有效依据。

## 2.4 人机交互需求分析

可视化系统始终需要秉承以用户为中心的设计理念（UCD），针对不同应用场景，将数据呈现的艺术性和功能性并重，通过丰富、精确、恰当的展示方式、交互方式，充分挖掘数据背后蕴含的信息。在人机交互功能设计上，充分考虑用户的业务需求和使用场景，制定合适的人机交互方案。对于本课题，对交互方面的要求如下：

（1）多视图协调联动：将显示区域划分为多个视图或图层，并将不同种类的视图组合起来，降低分析数据的复杂性。主要包括使用同一编码方式编码多个细分数据集的多组视图，以及采用不同编码方式编码同一数据集的多样式视图。通过多视图关联协调，可对视图之间进行动态联动，挖掘隐含的关联关系。

（2）自定义选择：通过鼠标左右键单击、滚轮滑动或拖动等交互方法（例如滑动时间轴、鼠标点击事件等）实现对数据查询的自定义选择。

（3）缩聚移位：因为数据规模比例和展示的详细程度主要由时间跨度确定，故当数据无法一次性全部呈现给用户时，通过拖动横纵坐标或缩放数据的时间段是最合适的选择。

## 2.5 本章小结

本章节对系统研究中提到的相关专业技术、数据处理方式和功能需求评估的分析任务系统地进行了可行性分析。依据本课题所提及的分析任务，主要从气象数据关联关系、污染物数据分析、机动车流量监测数据、空气质量预测模型四个部分对可视化系统功能进行详细的划分与阐述。并从交互性能、数据需求等方面对系统的优化提升进行了进一步改进设计。

## 第 3 章 可视分析系统设计

### 3.1 系统设计

可视化系统的核心部分是设计并改造合适的可视化呈现方式，并添加适宜的人机交互功能，以此满足用户对交互方式的需求和天气数据、污染数据分析等方面的需要。考虑到本次所开发的系统主要面向政府大气污染管理部门和其他相关部门的工作人员和管理人员。旧有的数据处理方式较为孤立且效率低下，不足以作为政府相关部门分析城市气象污染的依据，进而无法帮助决策者更为高效的改善治理方案与政策。

如图 3-1 给出了本课题可视化系统框架示意图，该结构具有用户端、系统平台层和数据处理与计算层。

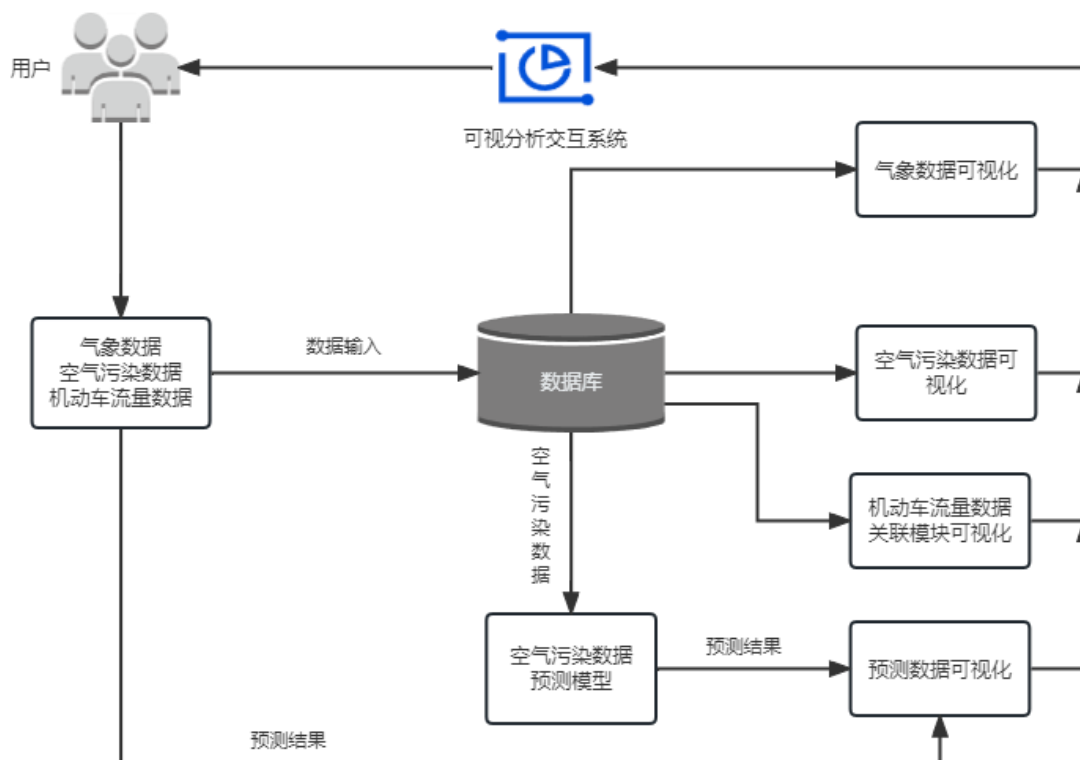


图 3-1 系统总体框架结构图

其中在系统中使用可视化技术与表达方式如下表 3-1 所示。

表 3-1 典型绘图技术总结

分类	可视化技术	维度	数据类型	用例与特点
简单图	仪表盘	1 维	数值型	展现气温、风速、颗粒物、降水量等气象数据数值
	多轴折线图	2 维	时序数值型	展现 AQI 与降水量走势关联关系
	矢量折线图	2 维	时序数值型	展现风速风向以及天气状况
	日历图	1 维	时序数值型	展现 AQI 在某时段内的变化趋势
常规图	散点热力地图	1 维	数值型	展示监测站点污染物浓度分布情况以及车流量监测情况
	横向条形图	2 维	时序数值型	展示某一污染物浓度与机动车流量的增减关系
	聚合散点图	1 维	数值型	构建城区污染物类型聚类模型，分析各区域污染物含量异同点
	气泡图	1 维	数值型	展示各城区 PM10、CO 等污染因子与 AQI 的关联关系
新颖图	词云	1 维	文本型	探究某一时间段内的主要天气情况
	环形关系图	1 维	数值型	展示交通卡口车流量和污染物的关联关系，用以探究机动车管控
	渐变堆叠面积图	2 维	时序数值型	不同区域的各类污染物预测趋势分析

3.2 数据处理

3.2.1 数据来源

1. 数据来源与规模

数据来源于山东省第三届数据应用创新创业大赛聊城赛道创新应用赛题以及山东省公共数据开放网，数据为环境污染数据，其中包含了城市不同站点的小时天气数据、天气现象数据、小时大气污染监测数据、监测站点经纬度数据、车流量监测站点信息、车流量小时监测数据等内容，编码格式为 UTF-8 无 bom 编码格式，存储格式为 csv，总大小 76.2MB。

## 2. 原始数据说明

数据字段描述如表 3-2、表 3-3、表 3-4 所示。

表 3-2 城市污染数据

序号	数据类型	字段名	字段含义
1	Date	Examine_Time	监测时间
2	String	Examine_City_Point	监测站点坐标
3	Float	NO2_Number	当前站点 NO2 浓度
4	Float	SO2_Number	当前站点 SO2 浓度
5	Float	O3_Number	当前站点 O3 浓度
6	Int	PM10_Number	当前站点 PM10 浓度
7	Int	PM2.5_Number	当前站点 PM2.5 浓度
8	Int	AQI_Index	空气质量

表 3-3 城市气象数据

序号	类型	字段名	字段含义
1	Int	ID	唯一序号标识
2	Float	Temperature	气温
3	Int	Wind_Direct	风向(0~360°)

续表 3-3 城市气象数据

序号	类型	字段名	字段含义
4	Float	Wind_Speed	风速(km/s)
5	String	Weather_Phe	城市小时天气现象
6	Float	Humidity	城市小时湿度

表 3-4 机动车流量数据

序号	类型	字段名	字段含义
1	String	Point_Name	监测站点名称
2	Date	Date_ID	日期
3	String	Period_Time	监测时段
4	Int	Car_Number	车流量
5	String	Car_Type	车辆类型

以最新空气质量指数标准为参考，AQI 主要分为六个等级，表 3-5 列出了不同空气质量指数（AQI）等级、颜色标记以及相关信息。

表 3-5 AQI 等级标准参考

空气质量指数	空气质量 指数级别	空气质量 指数类别	颜色标识	对健康状况影响	生产生活相关建议
0~50	一级	优	绿色	空气质量令人满意，极少出现空气污染	各类人群可正常进行生产生活
51~100	二级	良	黄色	空气质量可接受，存在部分污染物对少数人敏感群有较小影响	较少数的敏感人群建议减少外出活动



续表 3-5 AQI 等级标准参考

空气质量指数	空气质量 指数级别	空气质量 指数类别	颜色标识	对健康状况影响	生产生活相关建议
101~150	三级	轻度污染	橙色	易感人群症状可能出现轻度不适，健康人群可能出现刺激症状	婴幼儿、老人及患有呼吸系统疾病、心脏病的患者建议减少长时间的室外活动
151~200	四级	中度污染	红色	易感人群不适症状加剧，健康人群心脏、呼吸系统受到影响	婴幼儿、老人及患有呼吸系统疾病、心脏病的患者建议减少长时间、高强度的户外锻炼，其他人群应减少户外活动
201~300	五级	重度污染	紫色	心肺疾病患者病情明显加重，运动耐受力下降，健康人群开始出现呼吸不畅等症状	婴幼儿、老人及患有呼吸系统疾病、心脏病的患者需待在室内，停止一切室外活动，健康人群减少室外活动
>300	六级	严重污染	褐红色	健康人群的运动耐受力明显下降，出现明显刺激性症状，严重者可引发某些疾病	婴幼儿、老年人和其他病人需停留在室内，一般健康人群应避免长时间户外活动

### 3.2.2 数据预处理

在设计可视化分析系统时，需要对数据的属性进行详细分析，这样可以对后续的问题解决有较大帮助，从而更有针对性的解决问题。

在气象污染数据分析处理中，主要使用到了 Pandas、Numpy 和 jieba 第三方库相结合的方式，Pandas 库在大规模数据集的处理上有更大优势。数据处理流程主要为以下步骤：

（1）首先是相关数据整理合并，即在保持时间序列特征的前提下合并气象数据和污染物浓度数据，为后面的关联分析做准备。

（2）第二步为数据预处理，对于不规范、错位异常甚至缺失等情况的数据进行适当处理，以保证接下来系统开发测试工作开展的高效性。

（3）第三步为数据分类，根据监测站点的坐标以及时间序列特征进行分类处理，使数据之间保持关联性。

（4）最后为构建模型，根据数据特征和目标建立相关模型，并计算预测数据结果。

最后将数据设定为 csv 格式，帮助服务后续的数调用、数据呈现等需求。

在机动车流量数据处理中，主要使用了 Maplocation 地名经纬度查询对车流量监测站点进行经纬度的获取，并对卡口站点车流量进行均值计算。

在大气污染物的迁移轨迹研究分析过程处理中，通过风力相关字段数据将风向格式化为[N,NNE,NE,ENE,E,ESE,SE,SSE,S,SSW,SW,WSW,W,WNW,NW]中的其中一种，最后在地图中进行绘制。风速数据单位为 km/s（记为  $v$ ），风向数据设定为 0 到 360°（记为  $\theta$ ），其中 0° 和 360° 风向重合，且同时设定为东。同时将风向余弦与风速  $v$  相乘得到下一时刻的风向经度（记为  $l_1$ ），将风向正弦与风速  $v$  相乘得到下一时刻的风向维度（记为  $l_2$ ）。以这种方式不停迭代可以计算出某一时间段内的坐标集合。具体计算方法如下所示：

$$l_1 = l \cdot \cos \theta \cdot v \quad (3.1)$$

$$l_2 = l \cdot \sin \theta \cdot v \quad (3.2)$$

### 3.3 数据特征提取

#### 3.3.1 基于 Text-Rank 的天气词云

在本系统中，将处理好的数据输入到 Text-Rank 算法模型中进行关键词的抽取，其中 Text-Rank 算法是基于图权值的排序算法，主要依靠不断构建两个或多个词语节点边上的权重，并将重新计算的权重迭代下去。直到达到平衡状态，权重值不再发生改变，此时获取排列靠前的词语作为关键词的提取结果。如图 3-2 所示。

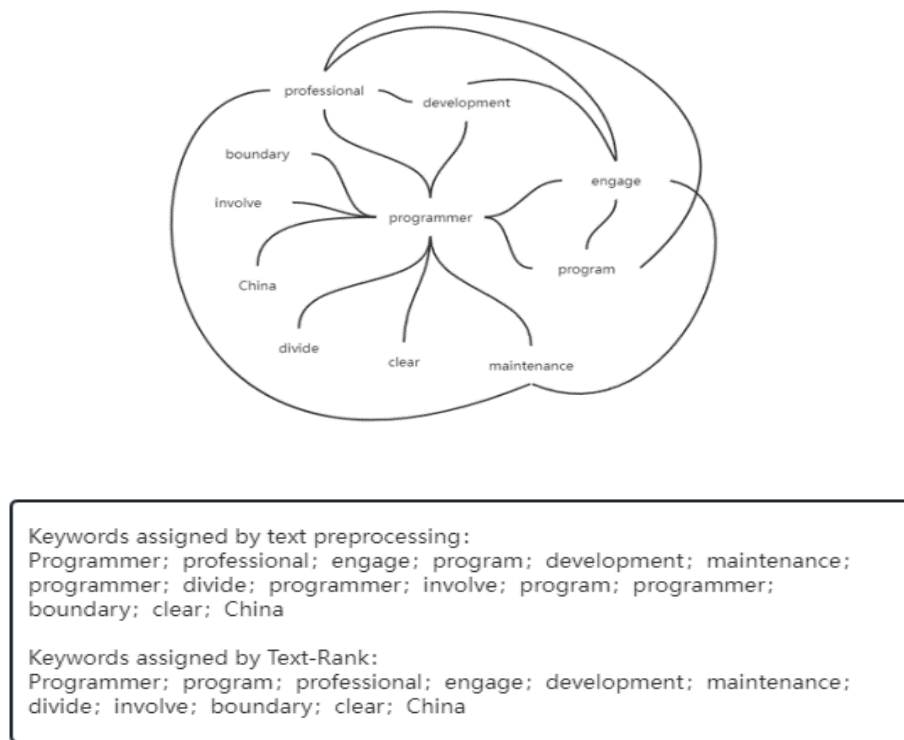


图 3-2 文本示意图

因为该算法在使用之前不需要对其进行预训练，极大减少了运行时间且实现代码简洁高效，基本思想来源于谷歌的 PageRank 算法<sup>[31]</sup>。在文本关键词提取、搜索引擎优化处理等方面具有很大的发展前景，基本流程如图 3-3 所示。

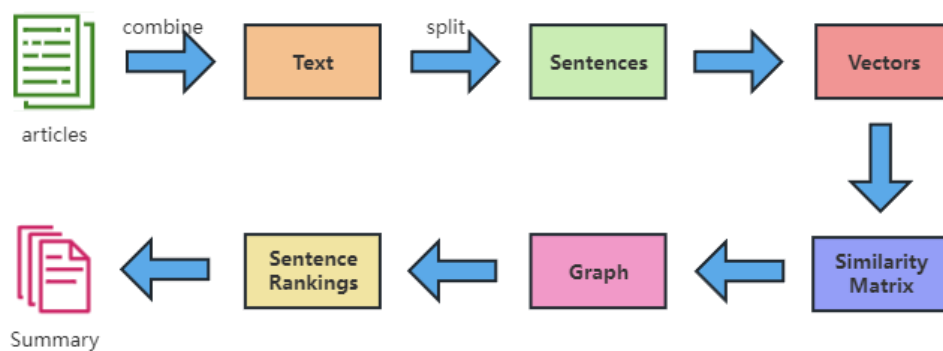


图 3-3 Text-Rank 算法流程图

Text-Rank 算法基本计算流程如下<sup>[32-35]</sup>:

（1）文本预处理：通过 jieba 库将给定的文本信息  $T$  中的语句进行分词处理，同时剔除掉非关键性词语，每个句子  $S_i \in T$  都包含： $S_i = [\omega_{i,1}, \omega_{i,2}, \dots, \omega_{i,n}]$ ，其中  $\omega_{i,j} (j \in [1, n])$  表示为分词计算结束后的候选关键词。

（2）构造节点集合：集合  $V$  表示相关候选关键词，边集  $E$  表示候选关键词节点矩阵，从而构造出带权关键词候选图模型  $G=(V, E)$ ，对于给定的任意节点  $V_i$ ， $\ln(V_i)$  表示指向节点的集合，即节点入度， $\text{Out}(V_i)$  表示节点指向的集合，即节点出度。节点  $V_i$  的 TextRank 值计算公式如下所示：

$$WS(v_i) = (1-d) + d * \sum_{V_j \in \ln(V_i)} \frac{\omega_{ji}}{\sum_{V_k \in \text{out}(V_j)} \omega_{jk}} WS(V_j) \quad (3.3)$$

（3）收敛计算：根据上述公式计算节点对应数值直至收敛，其中  $d$  为阻尼系数，取值在区间  $[0, 1]$  上，若数值过大会造成算法迭代次数增加且排序不稳定，过小则造成迭代过程效果不明显，通常情况下取  $0.85^{[36]}$ 。候选关键词如图 3-5 所示。

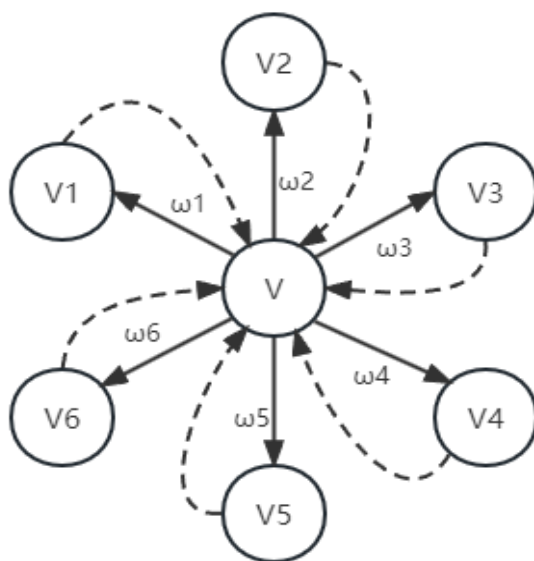


图 3-4 TextRank 候选关键词图

（4）关键词合并：将第三步节点的最终计算结果进行倒序，将前  $n$  个气象关键词作为候选词。在原文中标记的相邻关键字下的条目合并为最终结果。

### 3.3.2 基于 LSTM 长短时记忆网络模型的空气质量预测分析

研究空气污染预测应考虑多方面因素，结合多源数据进行综合分析，本课题将利用气象和交通数据来预测空气污染的波动情况，分析各类污染物的发展趋势，打破以往仅从历史监测数据进行预测的限制。

针对空气污染数据的时间序列特征，循环神经网络（Recurrent Neural Networks, RNN）在自然语言处理、语音识别、机器翻译、图像生成等方面有着较为广泛的应用<sup>[37]</sup>。RNN 允许网络中出现环形结构，即可以将部分中间层神经元的输出反馈再一次作为输入信号，该过程可以使网络在  $t$  时刻的输出结果与  $t$  时刻的输入信号、 $t-1$  时刻的状态相关，从而可以有效处理时间序列的变化，其基本结构如下图 3-5 所示。

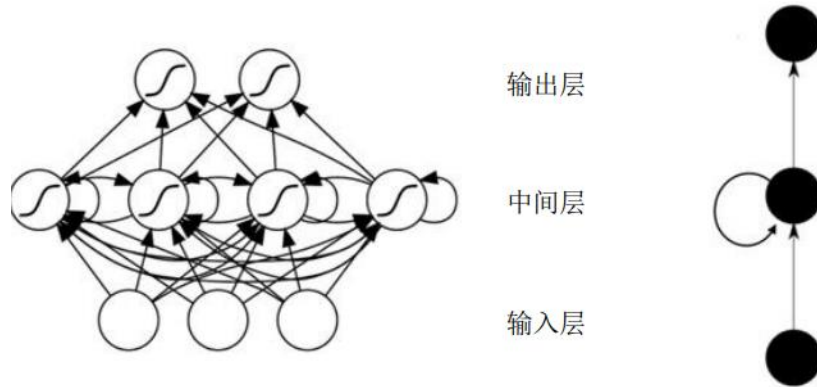


图 3-5 RNN 基本结构示意图

循环神经网络（RNN）在  $(t+1)$  时刻的结果  $Output(t+1)$  是此时的输入状态及历史状态共同作用的结果。其中间层计算方法示意图如下所示。

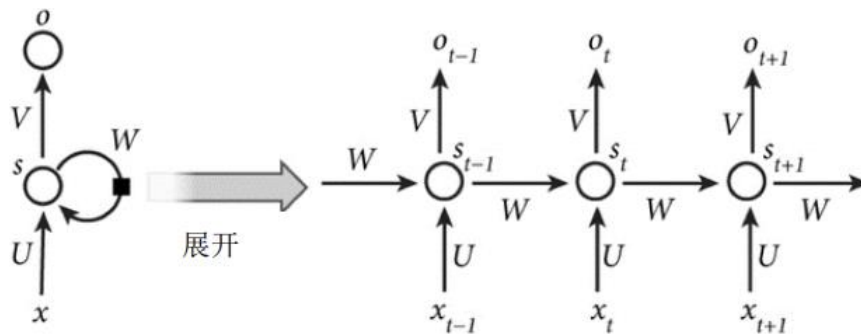


图 3-6 RNN 中间层展开计算示意图

其中： $x_t$  表示在  $t$  时刻的输入， $s_t$  表示  $t$  时刻中间层状态值， $s_t = f(Ux_t + ws_{t-1})$ ， $o_t$  是在  $t$  时刻的输出值， $o_t = \text{soft max}(V_{s_t})$ 。

但是经过调查研究，循环神经网络在迭代计算以此更新权重值的过程中，因为在时间梯度上传播数层后计算梯度极易出现消失或者爆炸的状况，所以在实际应用的预测过程中存在无法有效利用时间粒度这一特征影响因素的问题。后续为解决这一对时间序列存在长期依赖而引起的梯度消失问题，有研究者提出了长短时记忆单元（Long-Short Term Memory，简称 LSTM）算法<sup>[38]</sup>，通过对记忆门的开关控制防止了梯度的消失，实现了时间记忆功能。

LSTM 相比于 RNN 添加了遗忘门，除了使用单一网格层之外，还通过特殊方式将四个网格层组合在一起。下图为 LSTM 神经网络的基本结构。

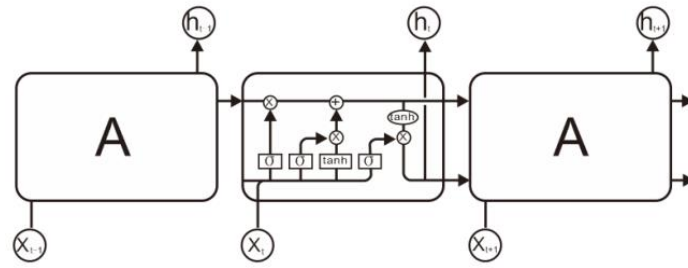


图 3-7 LSTM 神经网络结构图

LSTM 一共包含遗忘门、输入门、输出门三种门(gates)进行状态改变和信息流的选择，其中每一个门都实现了 sigmoid 函数来控制判定相关触发机制。

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + b_f) \quad (3.4)$$

$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i) \quad (3.5)$$

$$g_t = \sigma(W_{gx}x_t + W_{gh}h_{t-1} + b_g) \quad (3.6)$$

$$O_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + b_o) \quad (3.7)$$

$$st = gt \cdot it + s_{t-1} \cdot f_t \quad (3.8)$$

$$h_t = \phi(st) \cdot ot \quad (3.9)$$

其中：

$f_t$ 表示遗忘门， $i_t$ 表示输入门， $g_t$ 表示输入节点， $O_t$ 表示输出门， $st$ 表示本单元状态， $h_t$ 表示本单元输出。 $x_t$ 表示某一时刻向量的输入。 $W_{fx}$ ， $W_{fh}$ ， $W_{ih}$ ， $W_{ix}$ ， $W_{gx}$ ， $W_{gh}$ ， $W_{ox}$ ，

$W_{oh}$ 表示权重矩阵,  $b_f, b_i, b_g, b_o$ 是对应权重的偏置值,  $\sigma$ 表示 sigmoid 函数,  $\phi$ 表示 tanh 函数,  $\cdot$  表示点积。  $h_t$ 是  $t$  时刻以及之前时刻存储了所有有用信息的隐状态向量<sup>[39]</sup>。

下面将 LSTM 神经网络融合到空气污染预测模型中, 大致步骤如下:

(1) 在输入层中将原始气象污染数据进行预处理, 使其满足预测模型网络输入条件。

(2) 在输入数据后进入隐藏层进行网络训练, 将处理完毕的数据记为  $X$ , 将  $X$  输入隐藏层, 经过计算后将输出值与理论值进行对比分析并计算相应损失函数, 以损失函数最小为目标来设置训练迭代次数和学习率, 进而不断更新优化网络权重, 确定最后的隐藏层网络。

(3) 使用训练好的模型进行预测, 设定最后一行的数据为理论值, 将  $n+1$  时刻的预测值记为  $P_{n+1}$ , 并与理论数值合并为新数据来预测  $n+2$  时刻的预测值, 以此类推, 最后经过输出层的反标准化得到预测实际值, 计算出预测精度和误差。

### 3.3.3 结合车流量数据与污染物之间及污染物与 AQI 指数的关联分析

#### (1) 车流量数据与污染物之间相关性分析

通过皮尔逊相关系数 (Pearson Correlation Coefficient, PCC) 度量各类大气污染物之间以及机动车流量与污染物间的相关性。该系数反映了污染物内部之间相关转化或抑制的关系, 以及机动车流量数据对于各污染物迁移、扩散的影响。

协方差可以反映两个随机变量之间的相关关系, 如式 (3.10) 所示。

$$\text{cov}(X,Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1} \quad (3.10)$$

其中,  $x = (x_1, x_2, \dots, x_n)$ ,  $Y = (y_1, y_2, \dots, y_n)$ 。

皮尔逊相关系数可以定量地衡量变量之间的相关关系, 其定义如下公式 (3.11) 所示:

$$r(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}[X]\text{var}[Y]}} \quad (3.11)$$

其中,  $\text{cov}(X,Y)$  为  $X$  与  $Y$  的协方差,  $\text{var}[X]$ 、 $\text{var}[Y]$  分别为  $X$ 、 $Y$  的方差。相关系数为 +1 或 -1 表示所有离散点全部分布在一条直线上, 系数的正负主要取决于直线斜率  $k$  的正负; 若相关系数等于 0, 则证明  $x$  和  $y$  之间无相关性。相关系数数值与相关程度之间具体关系如表 3-6 所示:

表 3-6 相关系数关系表

相关系数	相关程度
$ r $	不相关
$0 <  r  \leq 0.3$	微弱相关
$0.3 <  r  \leq 0.5$	低度相关
$0.5 <  r  \leq 0.8$	显著相关
$0.8 <  r  < 1$	高度相关
$ r  = 1$	完全相关

### （2）污染物与 AQI 指数关联关系分析

结合 PM<sub>10</sub>（可吸入颗粒物）、PM<sub>2.5</sub>（细颗粒物）、O<sub>3</sub>、SO<sub>2</sub>、CO、NO<sub>2</sub> 六项污染物的实际监测值计算得到空气质量分指数（IAQI），然后取各污染物空气质量分指数（IAQI）中的最大值即为空气质量指数（AQI），具体计算公式如下（3.12 所示）：

$$IAQI_P = \frac{IAQI_{H_i} - IAQI_{L_O}}{BP_{H_i} - BP_{L_O}} (C_P - BP_{L_O}) + IAQI_{L_O} \quad (3.12)$$

其中  $C_P$  为污染物的浓度值， $BP_{H_i}$  为污染物浓度值限高位值； $BP_{L_O}$  为污染物浓度限低位值； $IAQI_{H_i}$  为  $BP_{H_i}$  对应的空气质量分指数； $IAQI_{L_O}$  为  $BP_{L_O}$  对应的空气质量分指数。

## 3.4 可视化布局

该可视化系统包含五个大模块，21 个视图，主要包括仪表盘、双轴折线图、词云、日历热力图、循环图、风向图、散点图、气泡图、趋势折线图、桑基图以及城市污染物关系对比图等可视化图表，系统整体布局设计见图 3-9 所示。



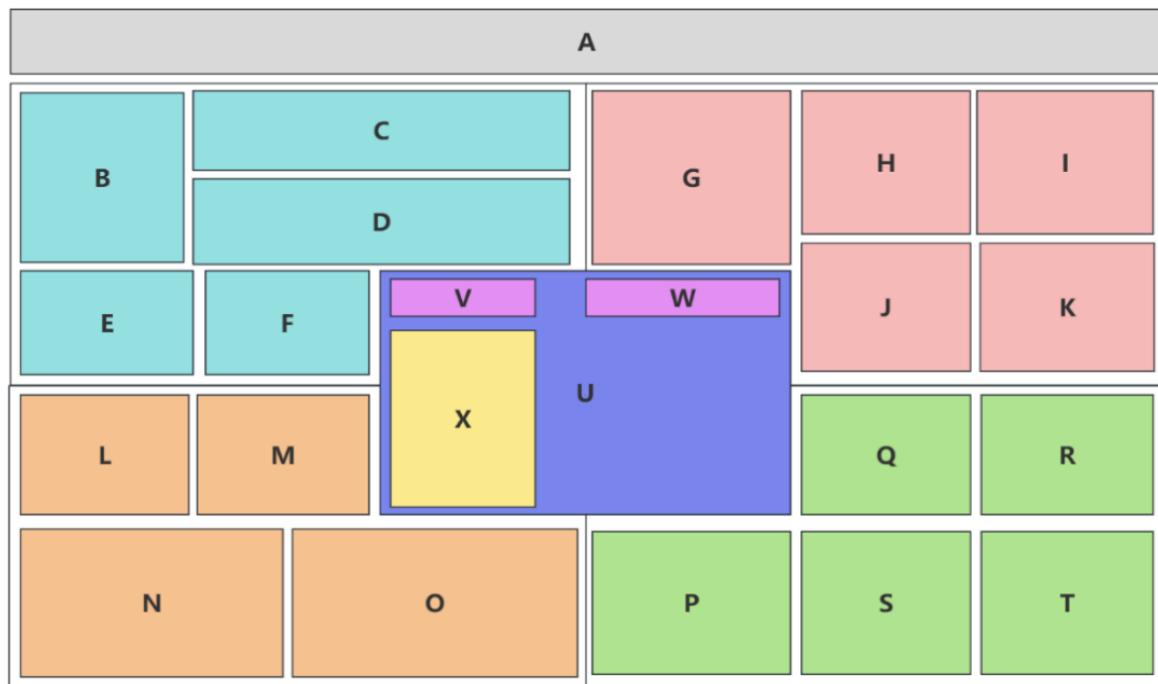


图 3-8 系统界面设计布局

其中：

(1) A 区域为系统标题，位于系统居中位置，名为“基于多源数据融合的城市气象污染可视分析系统”。

(2) 城市气象数据分析模块(B-F 区域)：

B 区域：主要呈现各时间段内的气温、降水量、风速、AQI 等气象特征，此部分使用 elementui 插件、仪表盘图表以及文字实现。

C 区域：展示某一时段内整个聊城市城区的 AQI 与降水量的关系，用户可以通过选择时间轴的跨度查看当前时段内的 AQI 与降水量的相关程度，也可以点击区域内的按钮实现单独查看某一项数据。

D 区域：展示某一时段内的风速、风向、气温以及天气状况，由双轴折线图和天气图标构成，天气阴晴状况和风速大小使用不同图标表示，用户可以直观的查看具体某一天内的相关气象状况以及某一时段内的整体趋势变化。

E 区域：主要为日历热力图，通过对 AQI 划分等级标注不同颜色，展示一定时间内空气质量的好坏情况，用户可以拖动鼠标查看具体一天内的空气质量。

F 区域：天气现象关键词分析，此部分使用词云来展现，表现出某一时间内主要天气情况。

(3) 城市气象污染因子数据分析模块(G-K 区域):

**G 区域:** 主要分析各类污染物因子对空气质量指数 AQI 的影响程度,使用气泡图搭配不同颜色实现,用户可以点击区域内的按钮实现不同监测站点的筛选。

**H 区域:** 用来分析不同污染指标在不同时间段内的聚类划分结果,用户可以通过此视图观察各类污染物变化趋势在时间上具有的空间分布特征。

**I 区域:** 各类污染物在某段时间内的具体变化趋势,折线图使用不同颜色进行污染物严重程度的等级划分,用户也可以点击区域内的按钮筛选和查看某一天的污染物严重程度。

**J 区域:** 主要展示各类污染物在某一时间段内的周期变化,主要包括最值和均值,用户可以点击按钮自由选择需要查看的数据。

**K 区域:** 展示了不同污染因子和 AQI 的变化趋势,主要由两种颜色的双轴折现走势图构成,用户可以通过该图表直观研究得到各类污染物对空气质量的影响大小。

(4) 机动车流量与污染物关联分析模块(L-O 区域):

**L 区域:** 主要展示某一时间段内的不同站点机动车流量变化趋势,采用堆叠面积图,帮助用户了解机动车流量变化情况,结合大气污染数据分析结果更好的对机动车流量管控。

**M 区域:** 主要用以展示各类车型和污染气体的关联程度,准备采用环形布局的关系图,用户可以通过点击任意一种车型或者污染气体来与其他图表产生联动查看具体数据及趋势,用圆的大小代表关联程度的深浅,线条的丰富度代表关联关系的复杂性。

**N 区域:** 针对污染物和机动车流量的关联分析,给出机动车流量配置建议,使得相关工作人员更好的对机动车进行管控,主要使用横向堆叠条形图实现目标效果。

**O 区域:** 表示各类车流量转移分析,即用户可以通过该图表了解到各类机动车的在不同区域的流动情况,配合上述的污染数据分析,可以更好更有针对性的进行大气污染改善,通过对比研究确定使用桑基图实现该目标。

(5) 空气污染数据预测及异常情况分析模块(P-T 区域):

**P 区域:** 对各监测站点的污染物进行相似度对比,挖掘各城区之间污染物的异同点,有利于用户更好的整体掌握城市污染信息,针对不同区域的污染物因地制宜的制定防治策略,用户可以对污染物种类进行切换筛选。

**Q 区域:** 主要用来展示不同站点污染物的分布规律,挖掘监测站点内的污染物异常情况,用户可以筛选不同污染物和监测站点来分类分析。

R 区域：主要对各类污染物和各类机动车型进行拟合度分析研究，即通过各类型机动车流量的增减分析造成各类污染物增长的原因。

S 区域：主要展示预测结果数据的趋势，采用堆叠面积图，用户可以通过点击来切换污染物。

T 区域：展示 AQI 预测结果趋势分析，采用单一大数据量面积图展示不同时间段的 AQI 变化情况。

(5) 中心地图模块 (U-X 区域)：

U 区域：主要为填充地图，用于呈现风场传播、机动车流量迁移轨迹的空间分布情况。

V、W 区域：填充地图操作控制台，主要用于污染物种类的筛选和地图中相关视图的优化。

X 区域：主要各监测站点污染物排序，即针对污染物浓度进行从大到小的排序操作。并且分析站点各类污染物占比情况。

### 3.5 可视化交互方案设计

本系统欲从视图自身与视图间多角度、多属性的联动设计交互方案，从用户认知的角度出发，提升系统本身的可交互性、可选择性，增加了用户与系统界面交互的体验设计，配合用户完成数据分析的目的。目前可视化系统中的主要交互方式有以下几种：

(1) 单击操作：用户选择并单击目标数据，例如应用于时间轴、图标、日历、预览图像、填充地图中。在系统中用户可以单击选择所需时间范围或查看具体数据信息；

(2) 复选操作：通过拖动时间线，鼠标拖动选中等交互，可查看某一时段的污染数据信息或选取图标中的某一区域进行研究分析。主要应用在降水量和 AQI 的关联分析中。

(3) 刷新操作：主要对多维坐标进行刷新，可以对图表中未选中的区域进行淡化处理，对选中区域进行色彩加强处理，增加用户的感知度。

(4) 悬浮：用户将鼠标指针悬停在图表或填充地图区域中，可以展示当前视图的具体内容，可实现与其他图表中的信息相比较，悬浮操作属性不会阻碍其他用户思考和分析思维建立的过程。

(5) 缩聚操作：通过鼠标滚轮滑动，填充地图整体展开或变窄，并且允许用户独立查看不同区域的数据信息。

（6）切换：不同的视图具有不同的分析功能，允许用户自由切换填充地图或者图表来分析污染数据的时空分布情况。

（7）过滤：通过设置约束条件实现信息的动态查询，对离散类型数据进行枚举，对连续型数据圈定选择范围，例如单选框、滑块、文本框等。

（8）对比：在地图或部分视图中，对不同的时间范围、空间范围进行对比，也可以自定义维度进行对比，或者将总体数据和局部数据进行对比。

（9）协同交互：当数据之间的关系较为复杂时，通常在研究过程中对数据信息进行多角度分析，即在统一的主题分析的基础上，使用视图之间的联动交互，挖掘更深层次的数据关系，以得到更加准确的结果。

### 3.6 本章小结

本章从系统结构设计、数据处理、可视化布局、可视化编码以及人机交互设计五个方面详细阐述了本系统的设计方案，通过描述各部分的设计理念，为后续的可视化视图编码的使用、人机交互方式的设计提供了理论上的支持，能更加有效的评估系统本身的呈现效果。

## 第4章 可视分析系统实现

### 4.1 系统总体架构

系统数据处理部分主要为 Python（版本号：3.10.11）语言，后端主要采用 Flask 框架搭建。第二部分为基于 Vue 框架的前端界面开发，为了简化开发逻辑，提高代码复用程度以及降低系统耦合性，本系统采用了模块化与组件化的开发模式。并通过推特公司推出的 Bootstrap 包工具实现模块堆积方式来完成可视化界面的整体布局。于此同时使用 Echarts、Antv L7、Antv G6、Antv G2plot 等作为可视化分析系统模型的基础视图支撑，其中 Bootstrap 版本为 4.5.3，Antv G2plot 版本为 2.4.29，Antv G6 版本为 4.8.8，Antv L7 版本为 2.15.0，echarts 版本为 5.4.2。

本系统主要适配的浏览器主要有火狐 Firefox、谷歌 Chrome 以及微软 Edge，经过不同环境的显示测试，针对本系统推荐使用谷歌浏览器。

系统总体架构如下图所示：

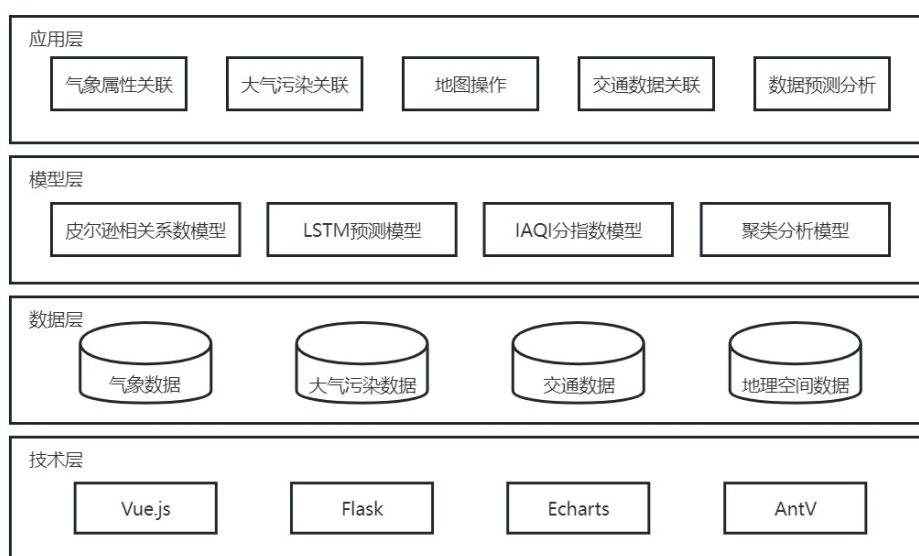


图 4-1 系统总体架构图

### 4.2 系统整体布局构建

本系统从大气污染浓度值变化、污染物及其气象因素的相关性、污染物的时空演变趋势、预测大气污染发展趋势等层面入手，分析了大气污染时空分布模式，监测大气污染时空上的演变趋势。最终完成的系统界面如图 4-2 所示，实现了气象数据分析、污染



视图设计:

天气属性与日期查询模块对日期选择(图 4-3 A)采用文本展示,对气温、降水量、风速、AQI 指数、PM10、PM2.5 使用仪表盘(图 4-3 B)呈现,使用渐变颜色样式表示各类指数的大小。

交互设计:

用户通过点击操作选择需要查询数据的时间段,实现数据的动态更新。其次,其他视图也会跟随时间跨度的变化而刷新。

(2) 降水量与 AQI 关联分析模块:



图 4-4 降水量与 AQI 关系图



图 4-5 单一降水量查询



图 4-6 单一 AQI 查询

视图设计:

采用双轴折线图展现 AQI 与降水量的关联关系,如图 4-4 所示,蓝色代表降水量变化趋势,绿色代表 AQI 指数变化趋势,颜色所占面积可代表此段时间内降水量或者 AQI 的高低,时间轴可由日期查询模块选定后自动刷新。

交互设计:

用户将鼠标悬停在降水量或者 AQI 的某一具体时间的坐标点处，会对单一数据信息进行筛选（如图 4-5、4-6 所示），另一折线灰度化，使得用户更好的单独研究某一气象属性。

### （3）风速、风向趋势分析图：

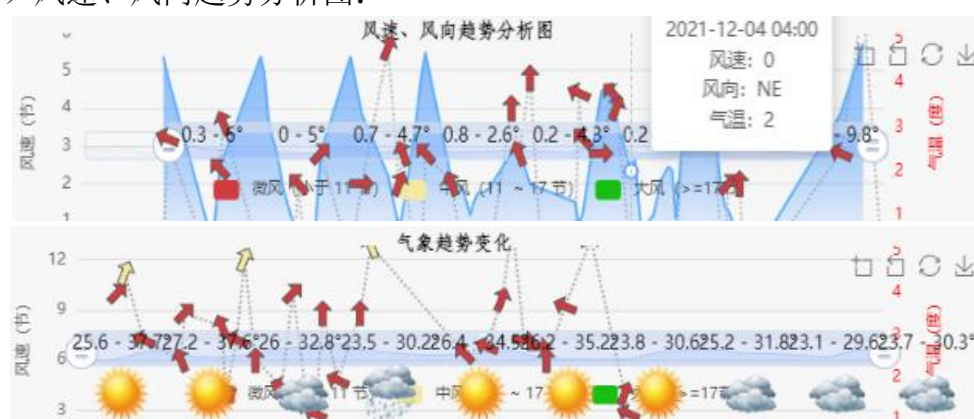


图 4-7 风速风向趋势分析图

### 视图设计：

图表采用折线面积图，箭头代表风向，虚线代表风速，蓝色折线代表气温，并分别用红色（微风，风速 $\leq 11$  节）、米黄色（中风， $11 \text{ 节} < \text{风速} < 17 \text{ 节}$ ）、绿色（大风，风速 $\geq 17$  节）代表不同风速等级，通过箭头颜色表现出来。

### 交互设计：

用户通过拖动时间轴或点击模块（1）中的时间选择组件来选择需要查看的数据，鼠标悬停在折线上时，可以查看具体时间、风速、风向和气温数据，方便用户结合其他视图分析污染物的影响情况。

### （4）AQI 质量等级日历图：

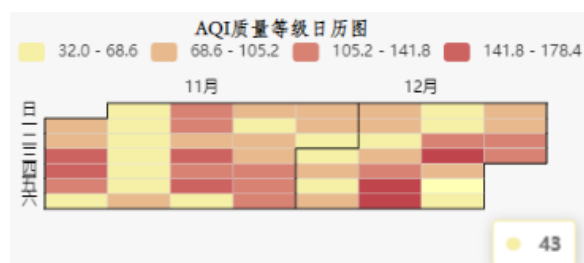


图 4-8 AQI 质量等级日历图



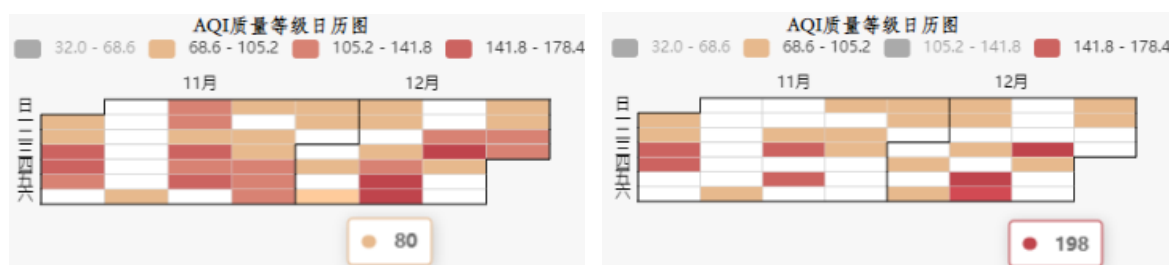


图 4-9 AQI 等级筛选操作

视图设计：

将日历图和热力图两者组合成时序日历热图，主要由时间变量和 AQI 等级变量组成，具体形式则是由小型色块有序且紧凑地以日历格式组成的图表，每个小型色块代表一天，色块颜色的深浅代表 AQI 数值大小，并将 AQI 分成四个等级。

交互设计：

用户将鼠标悬停在小色块上时可以查看当日的 AQI 指数（如图 4-8），并根据颜色判断 AQI 等级。另外用户也可以点击图表上方的等级图标进行数据筛选（如图 4-9），对当月或者当季的 AQI 发展态势进行分析。

#### （5）天气现象词云：

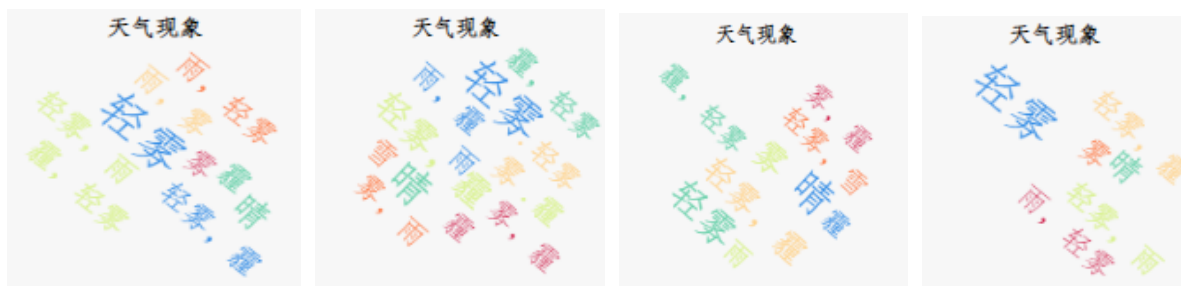


图 4-10 天气现象词云

视图设计：

各类天气现象名词通过使用不同颜色和文字样式大小来区分关键词的数据中的出现频率，以此判断该时间段内的天气状况。

交互设计：

用户可以通过点击选择时间跨度来查看这段时间内的气象关键词，鼠标停留在关键词上时可以查看到该日期对应的空气质量指数。

### 4.3.2 大气污染物与 AQI 数据对比分析模块

(1) 各类大气污染物与 AQI 对比分析图：

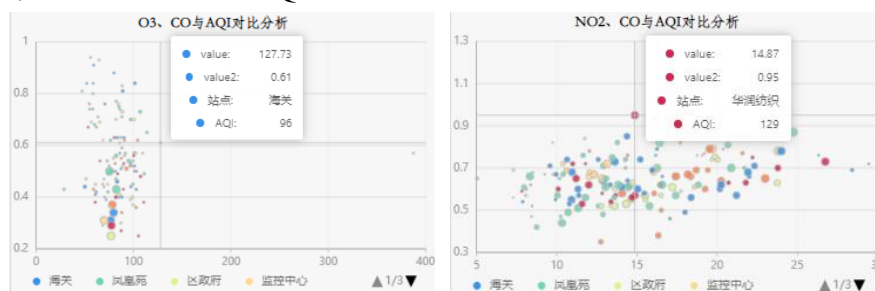


图 4-11 对比分析图

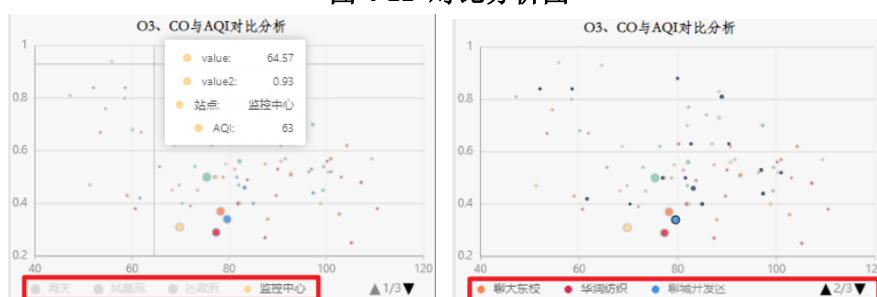


图 4-12 交互操作示意图

视图设计：

采用气泡散点图表示不同污染物与 AQI 之间的关联关系（如图 4-11），八种不同颜色代表不同污染物监测站点，气泡大小代表 AQI 指数大小，横轴与纵轴分别代表另外两种污染物浓度值。

交互设计：

用户可以将鼠标停留在气泡上方查看监测站点的具体污染物浓度与 AQI 值，并且可以点击图表下方的监测站点图标进行坐标筛选，点击图表右下方的三角形图标（▲）进行翻页查询，如图 4-12 所示。

(2) 污染物聚类分析模块：



图 4-13 污染物聚类分析图



图 4-14 交互操作示意图

视图设计：

城市污染类型聚类散点图（如图 4-13 所示）是基于污染类型指标的每日特征值数据进行聚类结果作为输入，将所有监测站点的污染物数据作为散点映射到坐标轴中，使用不同颜色映射不同的监测站点。

交互设计：

用户可以点击图表右侧是色块来筛选不同监测站点（如图 4-14 A），并且将鼠标悬停在色块上，图表内对应颜色也会弹出（如图 4-14 B），方便用户多种查看不同污染物对城市污染类型进行分类的结果的异同，寻找污染类型相似的站点，结合其他视图分析污染成因。

### （3）监测站点污染物波动趋势分析模块：

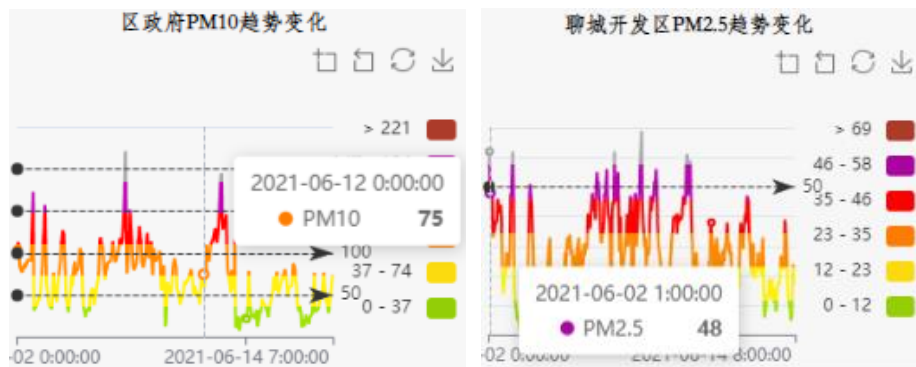


图 4-15 监测站点污染物波动趋势分析图

视图设计：

图表采用单轴折线图展示污染物在某个时间段内的走势，不同颜色区分污染物浓度大小，具体参考 AQI 等级规定标准颜色，颜色越深曲线越高表示污染物浓度越高。

交互设计：

用户可以在时间选择组件中自由设定时间段来查询各类污染物的变化趋势，也可以通过点击其他图表中的污染物信息或地图中的监测站点来分析污染物变化趋势，当鼠标悬停在图表内坐标点上方用户可以查看当前时间污染物的数据信息。

#### （4）监测站点污染物周期性变化分析模块：

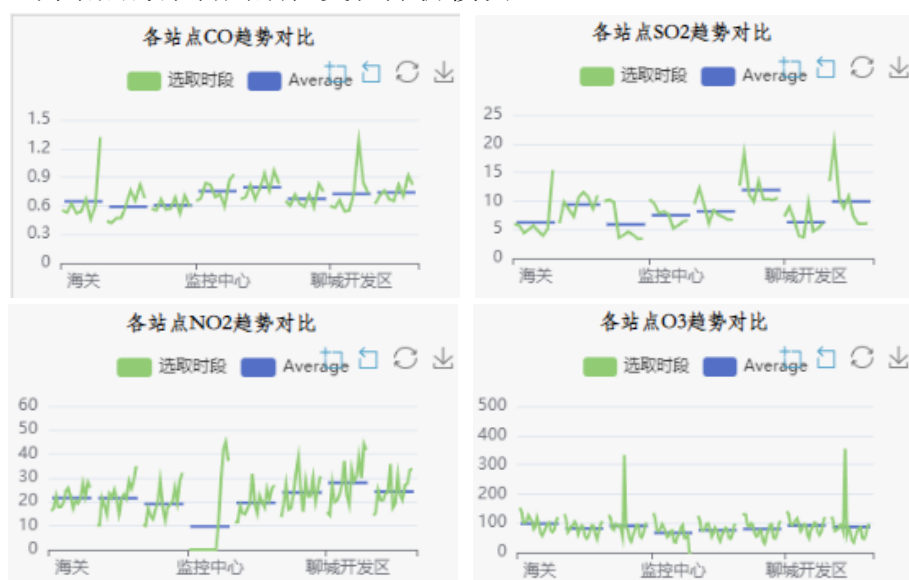


图 4-16 监测站点污染物周期性变化分析图



图 4-17 交互操作示意图

视图设计：

该模块是在基于周期图表(Cycle Plot)的基础上设计修改的，周期图表常用来分析商务金融数据，经测试发现此图表在大气污染数据分析的过程中也有较为良好的呈现效果，用户可以在视图中直观看出某时段内污染数据的最大、最小值（绿色折线）以及该时段内污染物浓度的均值（蓝色横线），辅助用户更准确的分析污染物的时空变化规律（如图 4-16 所示）。

交互设计：

用户将鼠标悬停在绿色折线上时，图表直接显示详细数值（图 4-17 A）；将鼠标悬停在蓝色区域时，会显示该段时间内污染物的均值（图 4-17 B）；用户点击图表右上方的区域进行选取并拖动鼠标选中需要进一步查看的区域，图表会自动放大到选中区域（图 4-17 C）；用户点击代表不同污染因子的图标可进一步筛选出相关数据信息，其余数据将灰度化（如图 4-17 D）。

#### （5）污染物与 AQI 变化趋势对比分析模块：

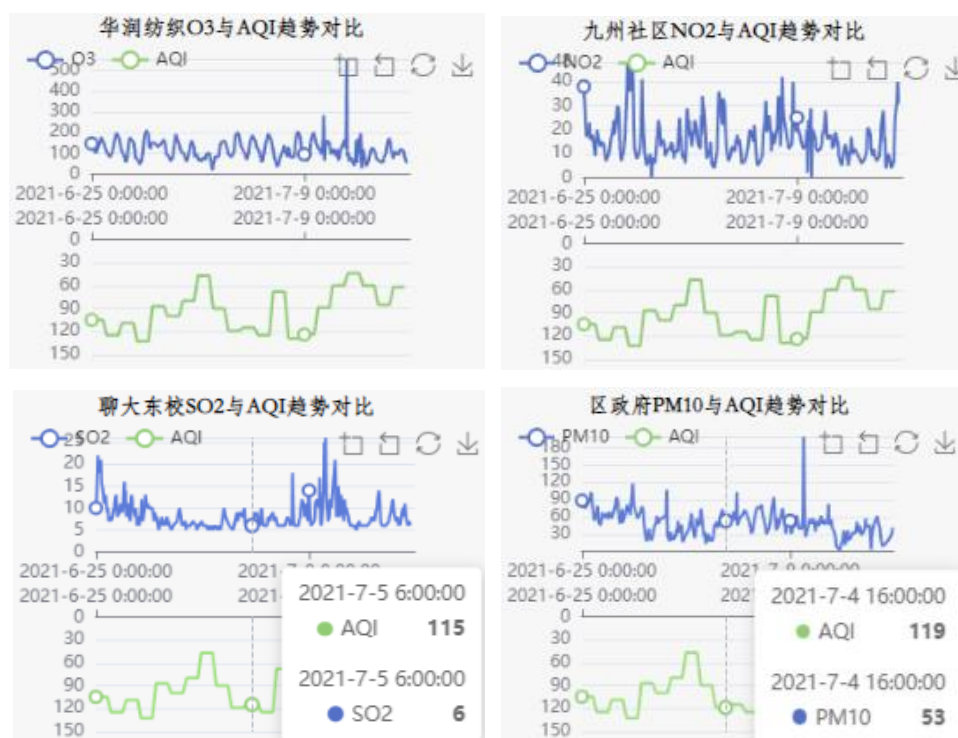


图 4-18 污染物与 AQI 变化趋势对比图

视图设计：



此图表基于雨量 Evaporation 关系图创建而成, 主要包括两个纵向数值轴和一个横向时间轴, 主要用来分析影响 AQI 指数变化的主要污染物, 其中蓝色折线代表污染物浓度变化, 绿色阶梯折线代表 AQI 指数的变化。

交互设计:

用户将鼠标悬停在任意折线上方, 可以同时查看当日的污染物浓度和 AQI 指数, 并且也可以通过鼠标拖动操作进行区域筛选自由查看某一时间段数据。

### 4.3.3 机动车流量与污染物相关性分析模块

(1) 机动车流量分析模块:



图 4-19 交通卡口机动车流量趋势分析图

视图设计:

采用面积折线图, 属于折线趋势分析图表的一种扩展形式, 数据起点基于前一数据集合, 用于表示每个数值所占大小随时间或机动车类别变化的变化趋势。用户可以任意查看每个交通卡口不同机动车型的过车情况。结合污染数据进行污染防治研究以及机动车污染物排放相关性分析。

交互设计:

用户可拖动图表下方时间轴自定义时间跨度查看车流量数据，帮助具体分析污染物成因及传播情况，同时也具有鼠标悬浮查看当日信息的交互操作。

## （2）机动车型与污染物相关性分析模块：

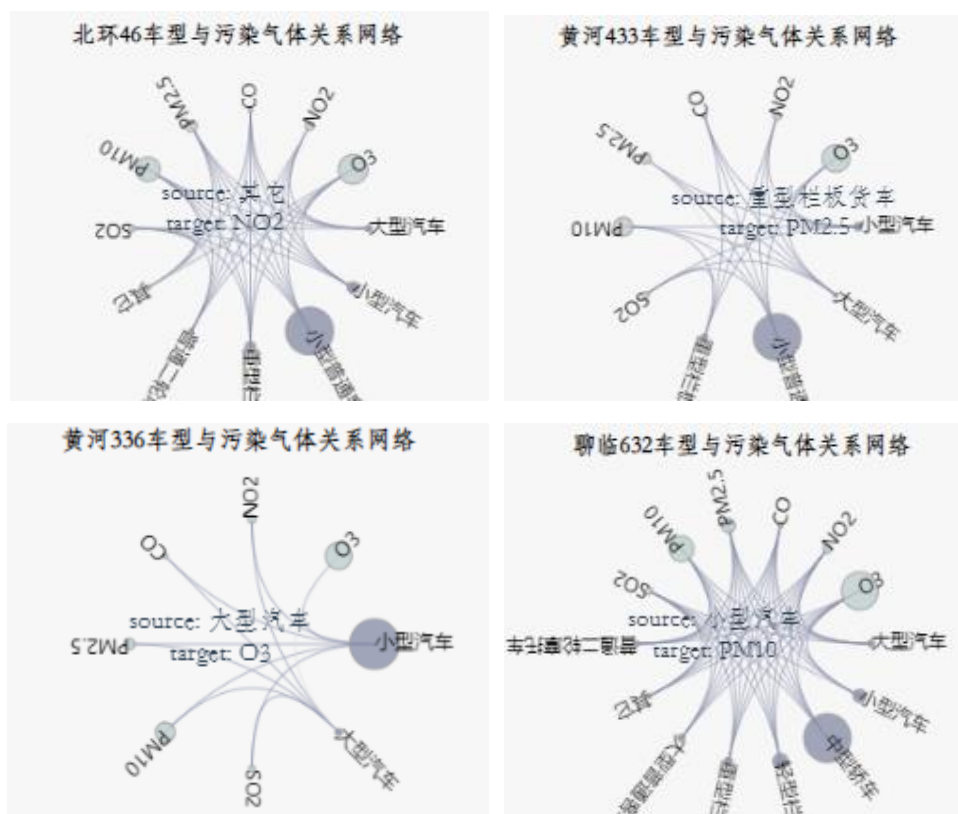


图 4-20 机动车型与污染物关系网络图

视图设计：

将机动车类型和各类污染物采用环形布局的关系网络图进行连接，主要由曲线，含有权重的气泡以及连接关系组成，用以研究数据之间的多对多关联关系。其中气泡大小随着权重值的变化而变化，用户可以以此对比分析各种类型的机动车流量对污染物排放的影响。

交互设计：

点击视图中的气泡会与其他模块视图产生联动，结合其他视图分析机动车对于污染物的影响情况。同时支持鼠标悬浮在气泡或曲线上方展示具体连接信息的操作。

## （3）卡口车流量配置建议模块：

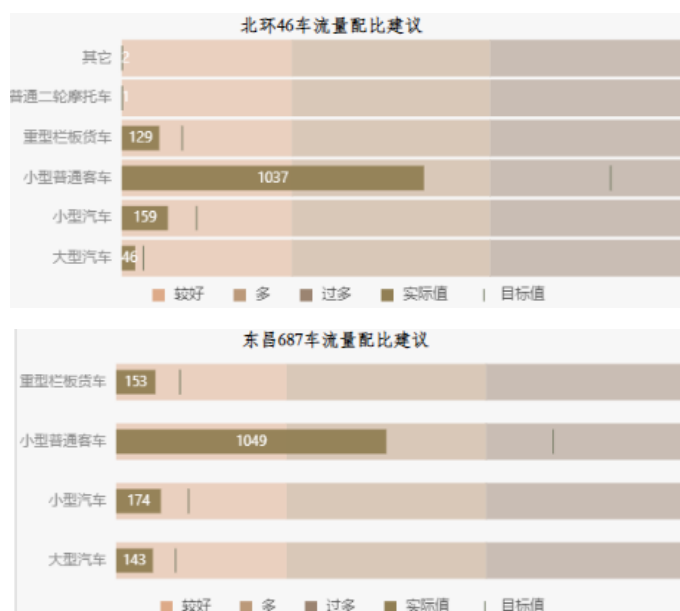


图 4-21 机动车流量配置建议图

视图设计：

在计算数据支持度和置信度得到强关联规则的基础上建立卡口流量阈值模型，设置机动车流量管控建议，采用横向条形图并在图表中设置三个等级（较好、多、过多）区分过车情况，并采用竖线表示车流量管控的阈值。

交互设计：

该视图通过点击机动车类型实现与其他视图的数据同步更新，并提供鼠标悬浮展示实际车流量和阈值车流量信息。

（4）机动车流量转移分析模块：

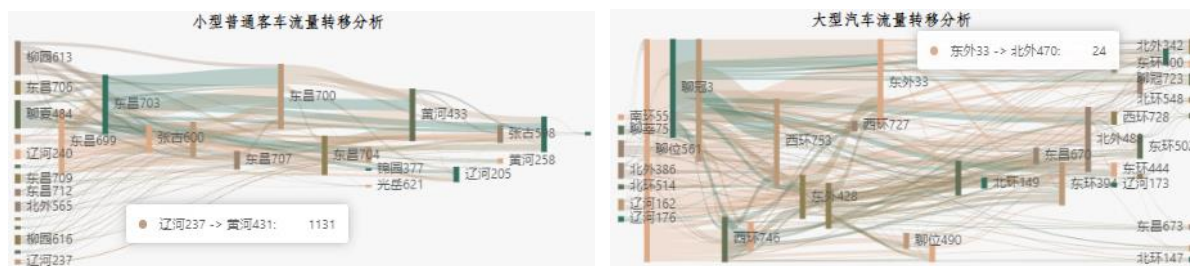


图 4-22 交通卡口间机动车流量转移分析图示例一



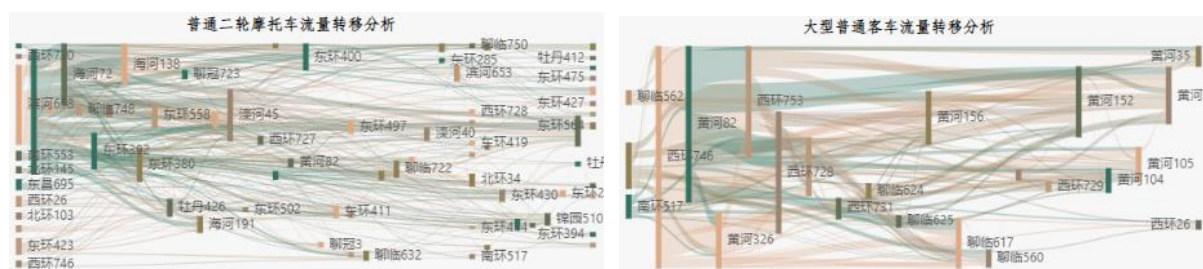


图 4-23 交通卡口间机动车流量转移分析图示例二

视图设计：

流量转移分析图主要基于桑基图的基础上设计的，主要由节点和边两种元素组成，每个节点代表一个卡口，每个边连接两个节点，边代表了流动的数据，边左侧为起点，右侧为终点，节点长度代表转移到其他节点的数量，边的长度代表卡口位置的远近，边的粗细代表车流量大小情况。

交互设计：

鼠标点击节点，可在地图视图中查看到站点坐标及流量转移情况，同时支持鼠标悬停展示节点间的车流量转移数据信息的查询操作。

#### 4.3.4 污染物之间相关性与数据预测综合分析模块

(1) 监测站点污染物相似度对比分析模块：

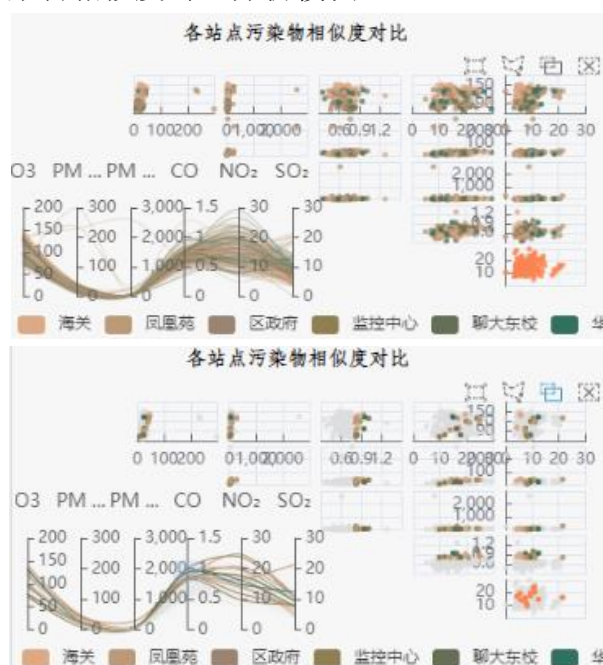


图 4-24 监测站点污染物相似度分析

视图设计：

图表采用散点矩阵与平行坐标相结合，形成多维度数据可视化编码方式。将各类污染物绘制在同一平行坐标图中，用户可以直观发现不同监测站点污染物的含量，分析不同监测站点之间污染物类型和含量的相似情况。

交互设计：

用户拖动鼠标实现不同污染物类别的筛选，同时支持点击图标筛选不同监测站点以及悬停展示污染物浓度信息的交互操作。

## （2）监测站点污染物分布规律分析模块：

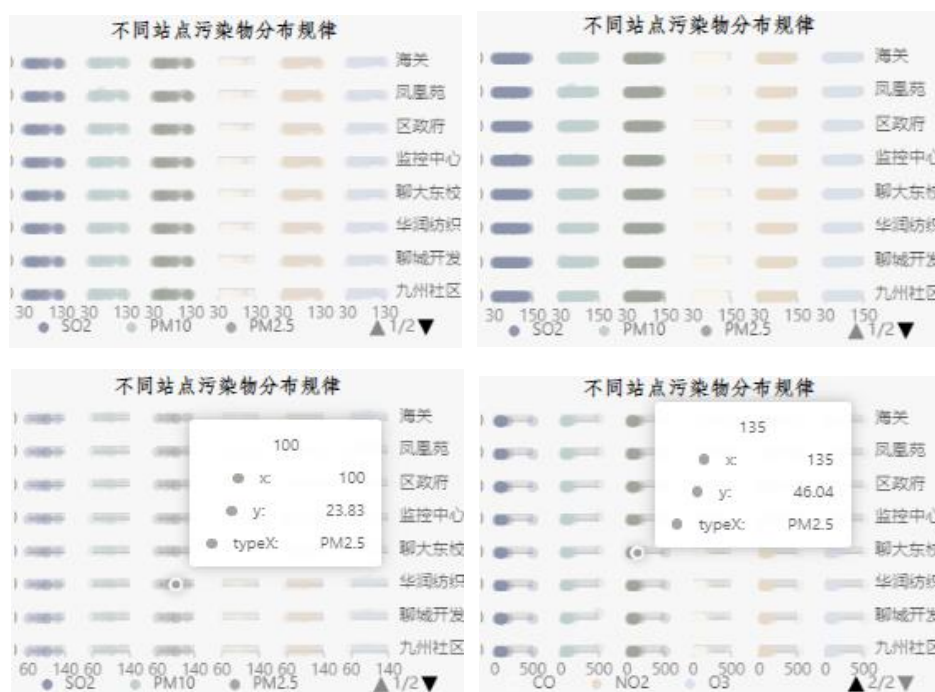


图 4-25 监测站点污染物分布规律图

视图设计：

采用高斯模糊后的散点图矩阵，主要用来挖掘污染物异常情况，用户可在图表中查看较为离散的异常点来挖掘污染物浓度的波动情况，结合其他视图分析出现污染物浓度异常情况的具体原因。

交互设计：

用户通过选择监测站点或时间，可以动态查询污染物浓度数据。鼠标悬停在散点上方可以查看具体坐标点信息。

(3) 监测站点污染物与卡口车流量拟合度分析模块：



图 4-26 监测站点污染物与卡口车流量拟合度分析图

视图设计：

主要采用对称层叠条形图，左侧表示某个监测站点中污染物浓度大小，右侧表示某交通卡口中某机动车类型的车流量数据。用户可以针对性的分析机动车流量对于污染物排放的影响程度，从而帮助相关工作人员更好的制定车辆管控策略。

交互设计：

用户将鼠标悬浮在条形图上时，可查看当日车流量与污染物浓度数据信息。

(4) 空气污染物及 AQI 预测分析模块：

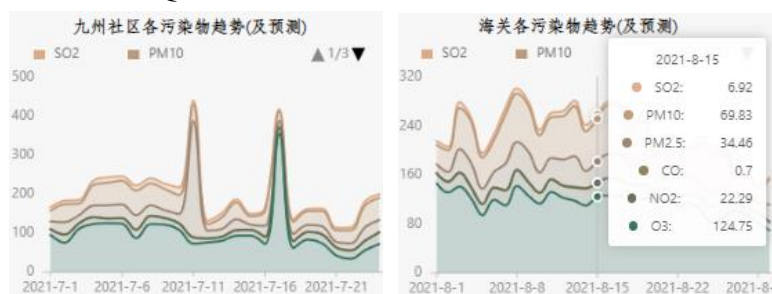


图 4-27 空气污染物预测分析图

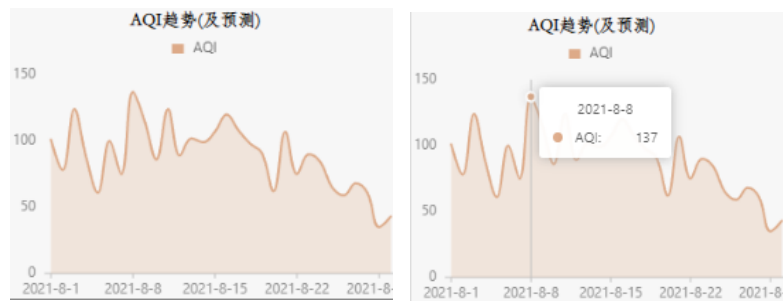


图 4-28 AQI 预测分析图

视图设计：

图表采用堆叠渐变面积图表示污染物的预测结果及变化情况，也能更直观的比较各个变量之间的大小关系和趋势以及异常情况。

交互设计：

用户可以通过拖动鼠标查询各类污染物和 AQI 指数的具体数值，辅助用户对相关污染物的分析研究。

#### 4.3.5 中心地图可视化绘制分析模块

（1）风场轨迹及车流量迁移模块：



图 4-29 中心地图模块

视图设计：

填充地图上侧 A、B 区域为交互操作控制台，用来在地图中展示相关数据信息；C 区域为各监测站点污染物排行；D 区域为各个监测站点污染物组成的雷达图，可直观分析出污染物占比情况；E 区域为主地图，主要用来展示监测站点的坐标以及污染物随风场转移轨迹和车流量迁移轨迹；F 区域为筛选操作控制台。

交互设计：

当用户点击 A 区域时，主地图（E 区域）会以柱状图的形式展示污染物数据在城区的分布情况；点击 B 区域时，地图中出现风场轨迹和车流量迁移轨迹；用户点击 E 地图中的柱状图时，视图中弹出 C、D 区域，对污染物进行排序操作；点击 F 区域时，可以在地图中进行区域选择操作，主要包括单击绘制节点，放行区域选择，圆形区域选择和自定义图案区域选择。

#### 4.4 本章小结

本章基于前述三章的研究内容对多源数据融合的可视化系统进行设计开发和实现，主要包括对系统运行环境、开发工具、系统主要特点和系统各模块进行了详细说明，为后续系统测试，应用分析等部分提供基础。

## 第 5 章 系统评估

本系统以山东省数据应用创新创业大赛聊城市创新应用赛题的基于主干道交通状况、车流量统计、天气状况等方面的脱敏数据，针对该数据集设计出适宜的可视化设计方案，进行全方位模型分析、可视化管理、预测评估，为大气环境保护、车辆管控方面的政策制定提供重要依据。

### 5.1 案例分析

#### 5.1.1 案例一：降水量对污染物浓度及 AQI 指数的影响分析

（1）降水量对 AQI 指数影响：

用户自主选择时间跨度（以 2021 年 6 月 1 日至 2021 年 7 月 30 日为例），通过 AQI 质量等级日历热图分析（如图 5-1 所示），6 月份每日 AQI 指数相比于 7 月份偏高，且通过天气属性面板查询得知影响 AQI 的因素可能和降水量有关。

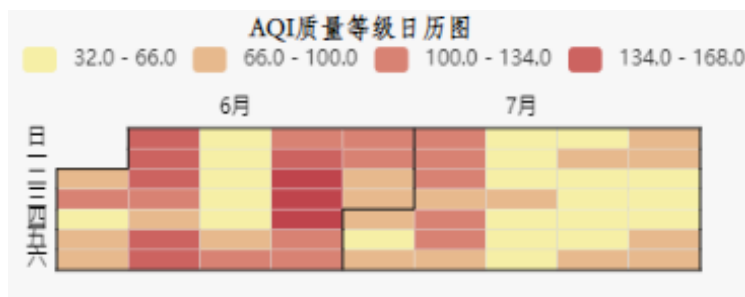


图 5-1 6、7 月份 AQI 质量等级日历图



图 5-2 天气属性面板查询

通过对降水量与 AQI 双轴折线图的分析，聊城市 7 月份中下旬的降水量明显多于 6 月份，且 6 月份存在降水的时段和日历图中 AQI 等级降低的时段相对应，更加证明了降



水量对 AQI 存在较大影响，空气质量变好的主要原因可能为强降水导致地面气温降低，大气垂直扩散条件相对增强，颗粒物沉降作用明显。



图 5-3 降水量与 AQI 关联分析

## （2）降雨量对其他大气污染物浓度影响分析：

分析图 5-3 可知，2021 年 6 月 28 日至 7 月 4 日、7 月 10 日至 7 月 19 日、7 月 27 日至 7 月 30 日三个时段存在较多降水，结合地图（图 5-4）查询污染物变化情况，以凤凰苑、聊城开发区、聊大东校为例（如图 5-5 所示）分析降水影响情况。



图 5-4 地图监测站点查询

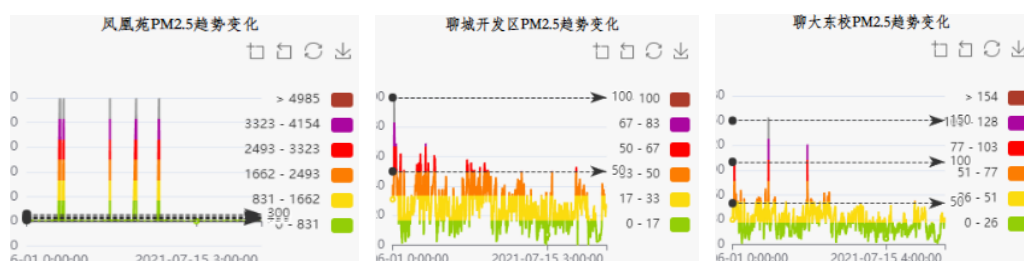
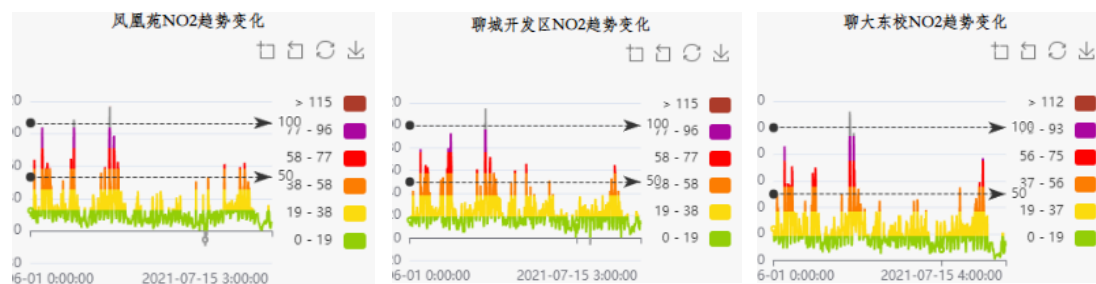
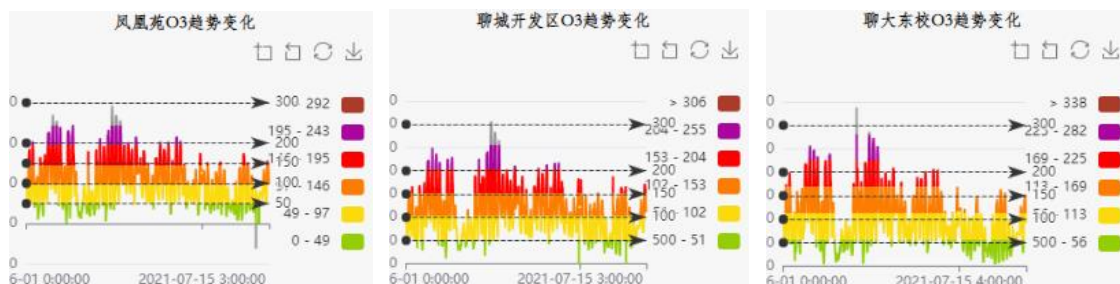


图 5-5 PM2.5 污染物浓度趋势分析

图 5-6 NO<sub>2</sub> 污染物浓度趋势分析图 5-7 O<sub>3</sub> 污染物浓度趋势分析

由上述图 5-5、5-6、5-7 中大气污染物浓度数值变化趋势分析可得,降雨量与 NO<sub>2</sub>、O<sub>3</sub>、PM<sub>2.5</sub> 浓度存在一定的反相关关系,其中 PM<sub>2.5</sub> 和 NO<sub>2</sub> 污染物受降水量的影响较大,主要原因为降水导致 PM<sub>2.5</sub> 可吸入颗粒沉降,空气中悬浮颗粒减少。7 月份时,白天气温较高,无降水时太阳辐射强烈,光化学反应较剧烈,故当存在降水时,NO<sub>2</sub> 与水 and 氧气发生反应,下降速率明显。O<sub>3</sub> 浓度也存在下降趋势,但不如其他污染物变化明显,能受夏季白天光照时间、气温等因素影响。

### 5.1.2 案例二：城市不同监测站点污染物对比及污染物与 AQI 关联分析

#### (1) 夏季时间污染物对比分析：

将时间选择为 2021 年 7 月 1 日至 2021 年 8 月 31 日,以 O<sub>3</sub> 污染物为例分析不同监测站点的相同特征。如图 5-6 所示,根据同期对比折线图发现区政府与九州社区在该段时间内的 O<sub>3</sub> 浓度具有高度相似性,结合其他图表对其进行分析。



图 5-8 污染物浓度趋势对比



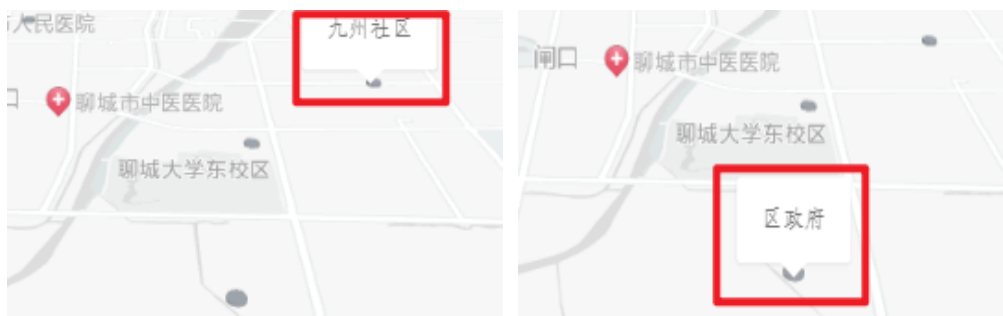


图 5-9 监测站点地理位置展示

首先根据地图（图 5-7）发现区政府和九州社区在地理空间位置上相近，都位于市区内，结合对比分析散点图（图 5-8），证明相关污染物在时空分布上相似，即  $O_3$  和其他大气污染指标与气象因素关联程度随区域不同变化较大。

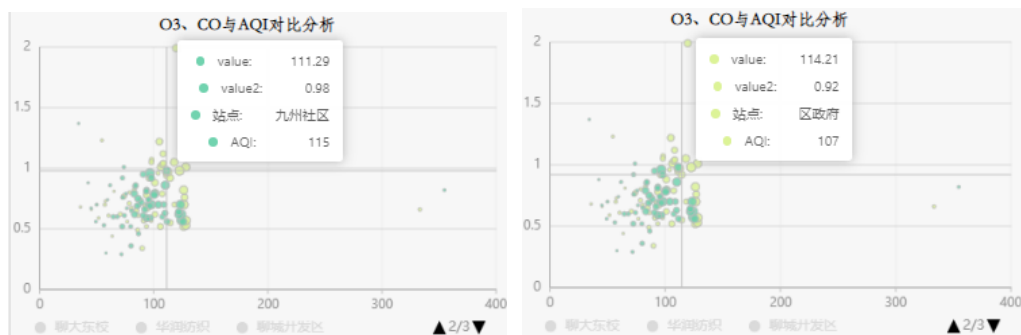


图 5-10 相关污染物对比分析

相关研究表明， $O_3$  是  $NO_2$ 、 $H_2O$  在太阳紫外线作用下，通过光化学反应生成的二次污染物，其影响范围较广、持续时间长、传输距离远，具有明显的区域性污染特征，因此对  $NO_2$  进行趋势对比分析（如图 5-8），发现两个监测站点间的  $NO_2$  浓度也具有相似的变化，验证了上述观点。

图 5-11  $NO_2$  变化趋势分析

## （2）污染物与 AQI 关联性分析：

对不同监测站点的各类污染物与对应时段的 AQI 数据进行比对，挖掘各类污染物对 AQI 指数的影响程度。通过分析发现，与 AQI 存在较强正相关性的污染物类型主要为：PM10、O3、PM2.5。

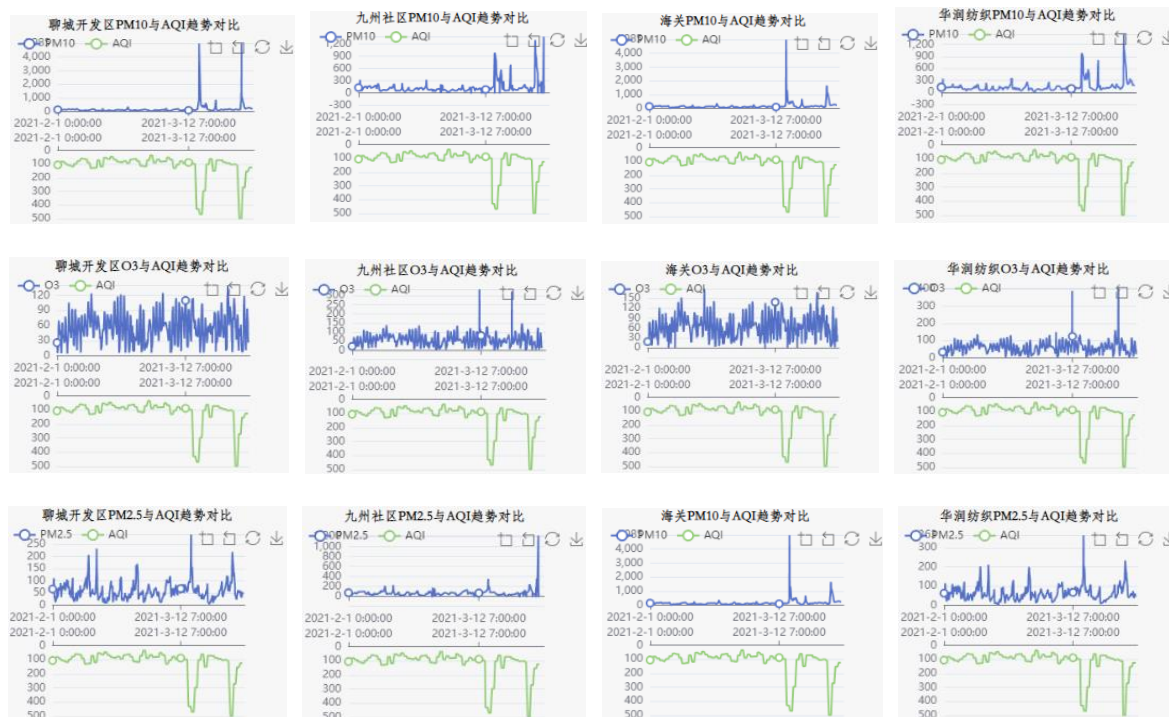


图 5-12 污染物与 AQI 关联分析图

### 5.1.3 案例三：交通数据对各类污染物的影响分析

选取 2021 年 6 月作为分析时段，通过 AQI 等级日历热图（图 5-10）可查看到该月 AQI 变化幅度较大，结合交通数据分析污染物变化规律较为明显。

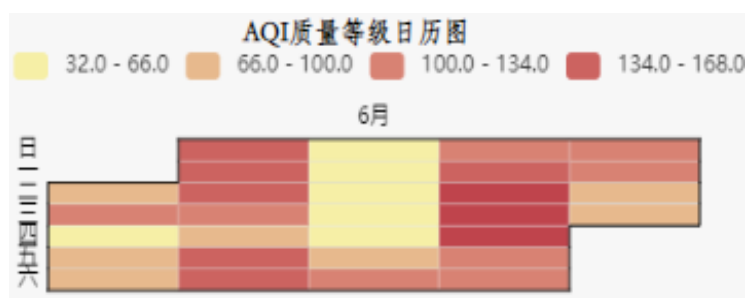


图 5-13 6 月份 AQI 指数等级分布

查看地图(图 5-11)中地理位置较为相近的污染物监测站点与机动车流量监测卡口，分析其相关性。案例选择凤凰苑监测站点与黄河 354 卡口作为研究对象。



图 5-14 污染物监测站点与卡口分布情况

通过分析机动车类型与污染物关系网络(图 5-12)发现,黄河 354 站点小型普通客车车流量较大, O<sub>3</sub>、PM<sub>10</sub> 污染物浓度较高, 因此可以猜测是小型普通客车的流量增加导致了汽车排放污染物增加发生化学反应生成的污染物浓度的增加。

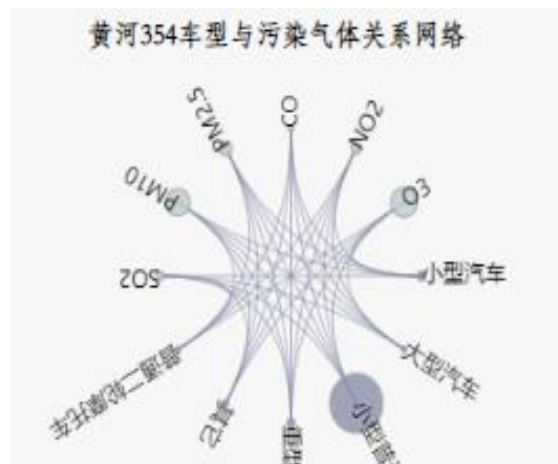


图 5-15 污染物监测站点与卡口分布情况

但是经过对大气污染物与卡口机动车流量拟合度分析之后, 发现 O<sub>3</sub> 浓度的变化和小型汽车的车流量关联较大(如图 5-13、5、14 所示), 与小型普通客车车流量关系较小。

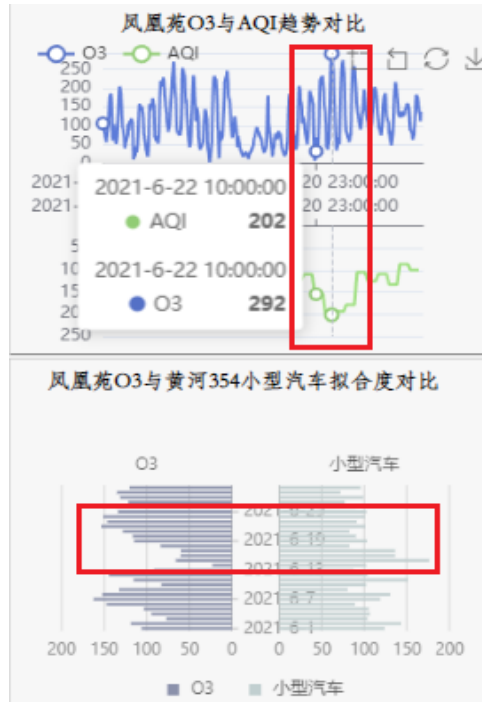


图 5-16 O<sub>3</sub> 与小型汽车流量分析

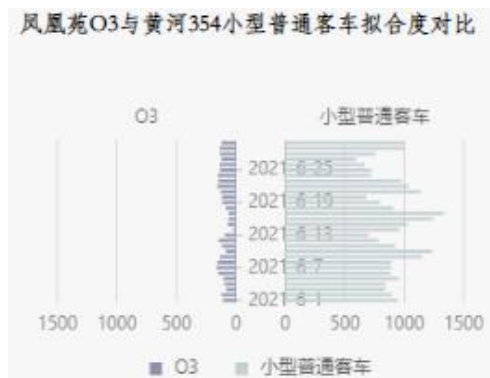


图 5-17 O<sub>3</sub> 与小型普通客车流量分析

PM<sub>10</sub> 浓度则受小型普通客车的影响较大，与 O<sub>3</sub> 变化相反，如图 5-15 所示。



图 5-18 PM<sub>10</sub> 与两类车流量分析

最后对机动车流量配置建议进行分析，图 5-16 所示，发现小型汽车流量相比于小型普通客车流量值更接近设定目标值，因此说明在日后夏季机动车管控中，凤凰苑附近的工作人员应重点控制小型汽车流量，以此控制污染物的排放。

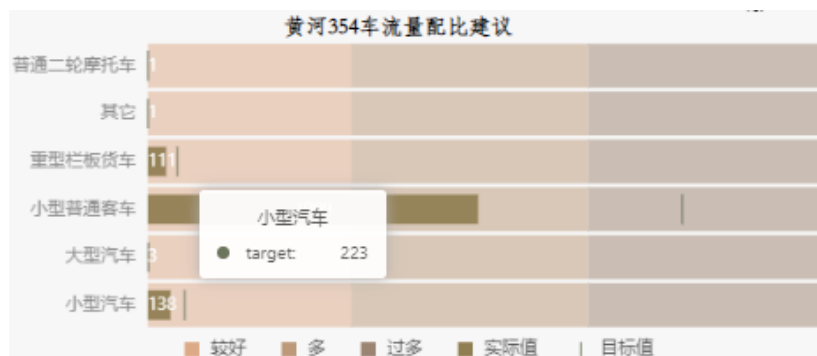


图 5-19 黄河 354 小型汽车配置比建议

#### 5.1.4 案例四：监测站点污染传播关系与风速风向对空气污染的影响分析

结合气象因素随时间序列对污染物浓度值的变化，进一步挖掘污染物扩散特点，探究影响大气污染时空分布模式的因素。

以  $\text{SO}_2$  污染物为例，对各监测站点污染物浓度进行排序，结合其他图表分析  $\text{SO}_2$  的传播路径以及形成原因：

选取 2021 年 3 月作为研究时间，通过对主地图分析发现海关、华润纺织以及聊大东校区的  $\text{SO}_2$  浓度排名靠前，而监控中心的  $\text{SO}_2$  浓度与其他监测站点相差较大。



图 5-20 市区内监测站点位置与风向展示





图 5-21 城郊监测站点展示

通过对风向（蓝色箭头）的分析，发现该时段内的风基本都由凤凰苑、区政府转移到海关、聊大东校等区域。风场较为密集的区域，例如聊城开发区、九州社区等  $\text{SO}_2$  浓度排名较为靠后，可能是因为空气流动大、地理位置空旷偏僻等原因。

针对区政府、聊大东校  $\text{SO}_2$  浓度较高的情况，通过对地图中机动车流量监测卡口（图 5-19）分析发现，区政府与张吉 600、东昌 639 以及光岳 315 卡口较近，故进一步分析卡口车流量信息挖掘污染物的传播情况。



图 5-22 机动车卡口及轨迹迁移展示

如上图所示，橙色飞线为机动车迁移流向情况，圆形图标为卡口流量信息，圆形半径越大车流量越大，颜色越深机动车类型越多。

通过图 5-20 机动车流向分析图分析可知机动车由张吉 600 卡口迁往黄河 433 卡口过程中，需要经过卡口东昌 639 和光岳 315，由此可得区政府和聊大东校污染物浓度较高可能与机动车流量迁移有关。

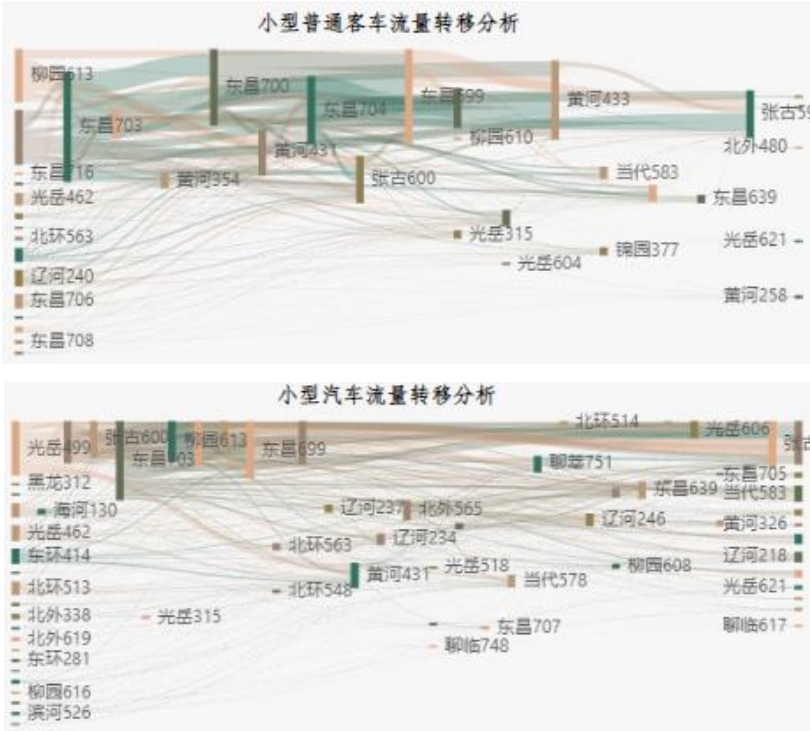


图 5-23 机动车卡口流量转移分析图

另外，通过分析距离海关较近的黄河 433 卡口机动车流量可知影响海关站点污染物的机动车类型为小型普通客车，具体如图 5-21 所示。

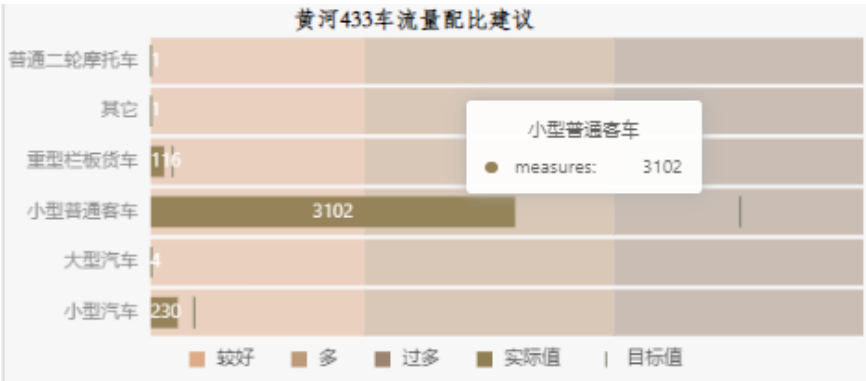


图 5-24 黄河 433 卡口车流量查询

根据图 5-18 中的城郊监测站点显示，华润纺织与监控中心站点 SO<sub>2</sub> 污染物浓度相差较大，下面将对其进行详细分析。



图 5-25 全局风场轨迹图

如图 5-21 所示，通过查看整体的风场轨迹发现，风轨迹整体向华润纺织方向迁移，且通过查询华润纺织站点的相关情况（如图 5-22），发现其所在位置为聊城市经济开发区，在聊城市经济发展中占据重要地位，故此地的其他生产活动也可能会影响污染物浓度。

山东聊城华润纺织有限公司

联系方式企业信用

公司介绍

公司名称	山东聊城华润纺织有限公司
经营模式	制造商、工厂
所在地区	中国山东 聊城 聊城市
主营产品/服务	紧密纺,环锭纺,赛络纺,单纱,股纱,烧毛纱,松式筒纱,S+Z捻,强捻纱,上蜡纱
采购产品	
详细地址	聊城市经济开发区辽河路12号

图 5-26 华润纺织站点详细信息



## 5.2 系统测试

### 5.2.1 系统视图测试

对基于多源数据融合的城市气象污染可视分析系统的功能测试，主要目标为测试不同视图的内容、图表、文字是否可以正常展示，各个功能点是否可以正常交互操作，具体如表 5-1 所示。

表 5-1 可视分析系统视图功能测试

序号	测试项目	测试步骤	预期结果	是否符合预期
1	地图主视图	运行系统	地图及操作台正常加载	是
2	气象属性面板	进入系统	各气象属性数据正常显示	是
3	AQI 与降水量关联关系视图	进入系统	AQI 与降水量数据正常加载，视图正常显示	是
4	风向、风速趋势分析视图	进入系统	风向、风速趋势分析视图正常显示	是
5	AQI 日历热图	进入系统	AQI 日历热图正常显示	是
6	天气现象词云视图	进入系统	词云正常显示	是
7	污染物与 AQI 对比分析视图	进入系统	污染物与 AQI 对比分析视图正常显示	是
8	污染物聚类模型视图	进入系统	污染物聚类模型视图正常显示	是
9	污染物趋势变化分析模块	进入系统	污染物浓度变化分析模块正常显示	是
10	各站点污染物周期变化对比分析模块	进入系统	各站点污染物周期变化分析对比模块正常显示	是

续表 5-1 可视分析系统视图功能测试

序号	测试项目	测试步骤	预期结果	是否符合预期
11	污染物与 AQI 对比分析模块	进入系统	污染物与 AQI 对比分析模块正常显示	是
12	交通卡口机动车流量监测分析模块	进入系统	交通卡口车流量监测分析模块正常显示	是
13	机动车型与大气污染物浓度关联关系视图	进入系统	机动车类型与污染物关联关系视图正常显示	是
14	交通卡口机动车流量配置建议视图	进入系统	交通卡口机动车流量配置建议视图正常显示	是
15	机动车流量转移分析视图	进入系统	机动车流量转移分析视图正常显示	是
16	污染物相似度对比分析模块	进入视图	污染物相似度对比分析模块正常显示	是
17	污染物预测分析视图	进入系统	污染物预测分析视图正常显示	是
19	AQI 预测分析模块	进入系统	AQI 预测分析模块正常显示	是
20	污染物浓度排序视图	进入系统	大气污染物浓度值排序视图正常显示	是

### 5.2.2 系统交互测试

对基于多源数据融合的城市气象污染可视分析系统的交互测试，主要为了测试验证系统中的人机交互操作是否可以正常进行。经多次测试，14 个交互操作均符合预期效果。

表 5-2 可视分析系统交互测试

序号	测试项目	测试步骤	预期结果	是否符合预期
1	地图界面缩放操作	进入系统，滑动鼠标滚轮	地图可随滚轮滑动进行缩放	是
2	地图界面拖动操作	进入系统，鼠标点击并拖动地图	地图随鼠标进行移动	是
3	地图区域选择操作	进入系统，鼠标点击选择自定义绘制节点并拖动鼠标划分区域	地图正常显示被划分的区域	是
4	时间选择组件	进入系统，鼠标左击时间选取组件自定义选取时间跨度	选定时间后正常显示	是
5	时间轴拖动操作	进入系统，鼠标左击并拖拽时间轴组件	视图正常显示所选择时间跨度	是
6	图例信息筛选操作	进入系统，鼠标左击视图中的图例	视图中被选择的信息单独展示	是
7	多维坐标刷选数据	进入系统，鼠标自定义选择维度坐标轴	未被选择的数据灰度化	是
8	地图与污染物趋势分析图交互	进入系统，鼠标左击地图模块中的监测站点	分析图中变换站点数据信息	是
9	机动车及污染物关系网络图与拟合度对比分析图交互	进入系统，鼠标左击关系网络图中的气泡	污染物与机动车流量拟合度对比分析模块数据更新	是
10	地图操作台与地图模块交互	进入系统，鼠标左击控制台按钮	地图模块显示对应信息	是

续表 5-2 可视分析系统交互测试

序号	测试项目	测试步骤	预期结果	是否符合预期
11	图表信息查询交互操作	进入系统，鼠标悬停在图表上方	图表正常显示其对应具体内容	是
12	机动车流量与地图模块交互操作	进入系统，鼠标左击车流量分析视图	地图模块交通卡口映射为选择区域内车流量通过情况	是
13	地图监测站点展示模块与污染物排序分析视图交互操作	进入系统，鼠标左击地图中的监测站点	地图左侧弹出站点内各类污染物的排序视图	是
14	地图监测站点展示模块与污染物聚类模型分析模块	进入系统，鼠标左击监测站点污染物柱状图	聚类模型视图刷新污染数据	是

### 5.3 本章小结

本章通过案例分析对系统进行评估，从多角度分析了大气污染时空分布模式、时序演变趋势、比较污染差异、评估大气污染状况以及结合交通数据及气象因素分析了污染成因等任务，对数据进行了多角度、多方面的分析研究，为本课题最重要的核心研究内容。另外对可视分析系统进行了相关测试工作，保证了系统运行的质量和可靠性。

## 总结与展望

本课题利用基于高纬度、时序特点的气象、空气污染以及交通流量相关数据，利用计算机科学与技术相关专业知识构建可视化分析系统。主要围绕分析大气污染时空分布与演变态势、气象属性与大气污染关联关系、分析大气污染成因、不同区域大气污染分布差异、道路机动车流量时空分布变化以及空气污染与卡口机动车流量及风场轨迹关联性等任务，对数据进行多方面处理，并最终通过案例分析对系统进行评估工作，主要包含以下四个部分。

（1）数据处理：系统使用 Text-Rank 算法进行气象文本数据关键词提取、利用 LSTM 长短期时序记忆神经网络构建大气污染预测模型、通过计算皮尔森相关系数（PCC）对机动车流量数据与污染物数据进行相关性度量分析以及分析计算 IAQI 与 AQI 指数对相关数据进行分析。

（2）可视化图表编码：主要采用仪表盘、折线趋势图、日历热图、热力散点图、折线走势图、气泡图、散点图、桑基图、堆叠面积图以及多维坐标等图表，用户可以通过上述图表对系统进行可视分析探索，从而辅助工作人员因地制宜地制定大气污染防治策略，在污染物海量数据分析、环境保护以及污染物浓度异常应急事件分析处理决策方面发挥重要作用。

（3）人机交互操作：系统交互采取了复选、单击、缩聚、拖动以及鼠标指针悬浮等多种方式，确保用户在分析任务目标时视图之间的协同交互，从而有助于用户在实际应用中探索挖掘更多的数据关联信息。

（4）系统测试评估：通过对系统的真实交互操作，验证了系统性能、可交互性与实际应用价值，为未来系统的优化改进提供了重要依据。

本课题开发的系统还存在较大改进空间，例如增加经济生产类型数据结合污染物数据进行分析，提升数据分析的丰富性，添加更多符合人机工程学的高效交互方式，实现更多的视图协同联动交互，使得分析过程更加顺滑流畅等。在未来的系统开发优化工作中，将考虑融合更大数据量的多源数据，提高城市气象污染数据分析的挖掘深度与广度。

## 致谢

行文至此，落笔为终，始于初秋终于夏。

饮水思其源，学成念吾师。首先感谢王松老师在我的毕业设计实现和论文撰写过程中给予的悉心指导。从大二时便一直跟随王松老师在实验室中进行相关科研训练，极大培养了我的科研能力。另外老师严谨细心的治学态度、勇于开拓的创新精神以及灵活敏锐的思维方式深深影响和激励着我。在此，再次向王松老师表示崇高的敬意和衷心的感谢。

同时也要感谢陈华容、覃仁超、王昆等任课老师，其亲切自然、幽默风趣又不失严谨的授课方式以及渴求知识、探索真理的热烈氛围激发了我对计算机专业知识的热爱与追求。感谢班主任陈永辉、蔡茂蓉老师在我大学学习生活期间的指导帮助。另外也要感谢刘艳春老师对于我不断追求理想的肯定与鼓励。愿各位恩师身体健康，生活幸福，工作顺利，桃李芬芳。

感谢 245 实验室同门的胡浩、吴潼、向英佳等师兄师姐的帮助，使我的科研之路虽困难重重，但仍感觉较为顺利，那段熬夜通宵准备比赛，并肩奋斗充满欢声笑语的时光至今历历在目。祝愿各位前程似锦。

纵使沧海变桑田，足下万里各有如画千秋。感谢明宇、鑫辰、大宝、阿伟、泽润等同窗好友给我的大学生活增添了许多色彩。也要感谢张丽丽等学长学姐对于我大学生活学习上的帮助，愿我们此去奔赴山海，再相逢依旧如故。

岁月清浅，感念相伴。感谢博洋、靖裕、琳昊、庆林、兴达、星宇、泽浩七位挚友，年少轻狂，幸福时光。孤身求学千里外，幸得诸位照拂。风雨同舟，相互分忧，倾囊相助，互为解难。

父母之爱子，则为之计深远。感谢父母予我暖衣饱食，除我后顾之忧，爱不挂口，常记于心。另外也要感谢其他亲朋好友在我十余年求学生涯中给予的精神物质上的支持。

再见少年拉满弓，不惧岁月不惧风。我在青龙大道 59 号龙山脚下的大学生活也即将结束。山鸟与鱼不同路，我也应该怀着对春日秋月，夏雨冬雪的期待，奋力向前走去。

士者，国之重器。为天地立心，为生民立命，为往圣继绝学，为万世开太平。我亦不会停止我的求学之路。感谢一直努力奋斗的自己，我深知自己阅历尚浅，步履所致不远，但心中仍然满怀对生活，对人生的热爱。

最后，非常衷心感谢各位评审专家教授可以在百忙之中评阅我的毕业论文，能得到您的指导与批评是我的荣幸！

文有所长，感激之情无尽矣。

2023.5.21 癸卯年四月初三

四川绵阳

## 参考文献

- [1]H. Qu, et al. Visual Analysis of the Air Pollution Problem in Hong Kong[J]. Ieee Transactions on Visualization & Computer Graphics, 2007, 13(6): 1408-1415.
- [2]蒋志方,孟祥旭,樊方芳,等. 基于圆环段技术的空气污染物浓度数据可视化[J]. 系统仿真学报, 2006, 18(10): 2859-2861.
- [3]马春燕,李璐池,薛文燕. 空气污染特征与分布规律的可视化技术研究[J]. 西北工业大学学报, 2017, 35(6): 1073-1078.
- [4]孙国道,胡亚娟,蒋莉,等. 基于城市群空气质量数据的可视分析方法研究[J]. 计算机辅助设计与图形学学报, 2017, 29(1): 17-26.
- [5]雷宝. 基于大数据的探索性空气质量数据分析[J]. 电子世界, 2017, (16): 5-6.
- [6]孙浩鑫. 基于时序的山东省环境数据可视分析研究[D]. 山东: 山东大学, 2016
- [7]G. Zhao, G. Huang, H. He, et al. Innovative Spatial-Temporal Network Modeling and Analysis Method of Air Quality[J]. Ieee Access, 2019(7): 1-1.
- [8]张涛. 空气质量分析系统的设计与实现[D]. 河北: 河北科技大学, 2016
- [9]Mengjie Zhou, Rui Wang, Shumin Mai, et al. Spatial and temporal patterns of air quality in the three economic zones of China[J]. Journal of Maps, 2016(12): 1-7.
- [10]王义祥,孙振海,张攀攀,等. 基于 GIS 空气质量模拟结果的表达方法研究[J]. 地理与地理信息科学, 2011, 27(2): 44-49.
- [11]赵桢. 城市空气质量可视化分析实现[D]. 黑龙江: 黑龙江大学, 2016
- [12]Yi Du, Cuixia Ma, Chao Wu, et al. A Visual Analytics Approach for Station-Based Air Quality Data[J]. Sensors, 2017, 17(1): 30-47.
- [13]N. Shaadan, L. N. Nazeri, Mfm Jalani, et al. Data visualization of temporal ozone pollution between urban and sub-urban locations in Selangor Malaysia[J]. African Journals Online (ajol), 2018, 9(6S): 490-507.
- [14]Ki Chul Choi, Jae Bum Lee, Jung Hun Woo, et al. Development of an Emissions Processing System for Climate Scenario Inventories to Support Global and Asian Air Quality Modeling Studies[J]. Asian Journal of Atmospheric Environment, 2018, 11(4): 330-343.
- [15]Li Huan, Fan Hong, Mao Feiyue. A Visualization Approach to Air Pollution Data Exploration—A Case Study of Air Quality Index (PM2.5) in Beijing, China[J]. Atmosphere, 2016, 7(3): 35-55.



- [16] Ningbo Jiang, Alan Betts, Matt Riley. Summarising climate and air quality (ozone) data on self-organising maps: a Sydney case study[J]. Environmental Monitoring & Assessment, 2016, 188(2): 103-119.
- [17] C. Zhang, M. Chen, R. Li, et al. A virtual geographic environment system for multiscale air quality analysis and decision making: A case study of SO<sub>2</sub> concentration simulation[J]. Applied Geography, 2015(63): 326-336.
- [18] S. Meinardi, P. Nissenson, B. Barletta, et al. Influence of the public transportation system on the air quality of a major urban center. A case study: Milan, Italy[J]. Atmospheric Environment, 2008, 42(34): 7915-7923.
- [19] B. Barry, P. Li. Overview of the Model and Observation Evaluation Toolkit (MONET) Version 1.0 for Evaluating Atmospheric Transport Models[J]. Atmosphere, 2017, 8(12): 210-222.
- [20] S. M. An, H. Y. Lee, B. Kim, et al. Geospatial spreadsheets with microscale air quality visualization and synchronization for supporting multiple-scenario visual collaboration[J]. International Journal of Geographical Information Science, 2014, 28(11): 2511-2532.
- [21] 苏金树, 张博锋, 徐昕. 基于机器学习的文本分类技术研究进展[J]. 软件学报, 2006, 17(9): 1848-1859.
- [22] 陈凯, 朱钰. 机器学习及其相关算法综述[J]. 统计与信息论坛, 2007, 22(5): 105-112.
- [23] 凌德森, 王晓凯, 朱涛. 基于自适应时序分解的空气污染物浓度预测[J]. 测控技术, 2023, 42(1): 83-91, 98.
- [24] 李钢, 宋海军, 蒋文静. 用数学方法对郑州城市主要空气污染物预测的机理研究[J]. 中原工学院学报, 2013, 24(5): 41-43.
- [25] 孟凡荣, 庄朋, 闫秋艳. 基于加权移动平均的数据流预测模型[J]. 计算机应用研究, 2009, 26(10): 3680-3682, 3686.
- [26] 马海良, 王若梅, 庞庆华. 工业废水污染与城镇化进程的动态关系分析[J]. 水利经济, 2015, 33(6): 50-54, 74+81-82.
- [27] 赵子杰, 王晨旻, 林沛辰, 等. 基于时间序列的上海市空气质量变化与污染物特征分析[J]. 统计学与应用, 2021, 10(1): 115-131.
- [28] M. Miljanovic. Comparative analysis of Recurrent and Finite Impulse Response Neural Networks in Time Series Prediction[J]. Indian Journal of Computer Science & Engineering, 2012, 3(1): 180-191.
- [29] Y. Lecun, Y. Bengio, G. Hinton. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [30] 罗敏刚. 探析大数据可视化技术与工具[J]. 科技视界, 2020, (9): 159-161.

- [31]Y. Bengio, Réjean Ducharme, Pascal Vincent, et al. A Neural Probabilistic Language Model[J]. Journal of Machine Learning Research, 2003(3): 1137-1155.
- [32]W. Li, J. Zhao. TextRank Algorithm by Exploiting Wikipedia for Short Text Keywords Extraction[C]//International Conference on Information Science & Control Engineering, 2016.
- [33]Tu,Shouzhong,Huang, et al. Mining microblog user interests based on TextRank with TF-IDF factor[J]. 中国邮电高校学报: 英文版, 2016, 23(5): 40-46.
- [34]C. Zhang, D. Zeng, J. Li, et al. Sentiment analysis of Chinese documents: From sentence to document level[J]. Journal of the Association for Information Science and Technology, 2009, 60(12): 2474-2487.
- [35]L. I. Ai-Ping, D. I. Peng, L. G. Duan. Document Sentiment Orientation Analysis Based on Sentence Weighted Algorithm[J]. Journal of Chinese Computer Systems, 2015(36): 2252-2256.
- [36]Rada Mihalcea, Paul Tarau. TextRank: Bringing Order into Texts[D]. Department of Computer Science University of North Texas. North Texas, 2004.
- [37]J. Schmidhuber. Deep Learning in Neural Networks: An Overview[J]. Neural Networks, 2014(61): 85-117.
- [38]S. Hochreiter, J. Schmidhuber. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [39]白盛楠,申晓留. 基于 LSTM 循环神经网络的 PMsub2.5/sub 预测[J]. 计算机应用与软件, 2019, (1): 67-70,104.