

空气信息综合管理评估系统^①

王有帅, 陈梅, 陈艺丹

(兰州交通大学 电子与信息工程学院, 兰州 730070)

通信作者: 王有帅, E-mail: wyoahmx@163.com



摘要: 近年来, 重工业的快速发展加剧了空气质量下降, 环境治理变得尤为重要。然而, 在国内外现有的空气质量评估系统大多形式单一、准确度低、评估范围有限, 并不能较精确的以多样化的形式展示空气质量情况。本文设计实现了一套集数据收集、标准化处理及空气质量评估为一体的空气质量评估系统。该系统采用网络爬虫从空气质量监测网站获取数据, 并对数据进行抽取、清洗、单位转化、污染物分类及多种标准化方法处理, 提高空气质量的评估精确度。系统评估方式多样, 有按时、日、月、监测站位置及污染物种类等, 解决了大多数空气质量评估系统形式单一和评估不准确的问题。该系统可以为用户提供实时的空气质量情况, 还可以为大气污染相关科研人员提供准确的数据预处理结果。系统运行稳定可靠, 界面友好, 功能丰富, 能满足空气信息综合管理评估的需求。

关键词: 数据预处理; 标准化; 缺失值; 无效值; 空气质量评估; 数据挖掘

引用格式: 王有帅, 陈梅, 陈艺丹. 空气信息综合管理评估系统. 计算机系统应用, 2022, 31(4): 91-98. <http://www.c-s-a.org.cn/1003-3254/8434.html>

Air Information Integrated Management and Evaluation System

WANG You-Shuai, CHEN Mei, CHEN Yi-Dan

(School of Electronics and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China)

Abstract: In recent years, the rapid development of heavy industry has exacerbated the deterioration of air quality, and environmental governance has become increasingly important. However, most of the existing air quality assessment systems at home and abroad have single forms, low accuracy and limited assessment ranges. In other words, they cannot accurately display the air quality situation in a diversified way. This study develops an air quality assessment system which integrates data collection, standardized processing and air quality assessment. The system has the following advantages. It obtains data from air quality monitoring websites with Web crawlers and enables data extraction, data cleaning, data unit conversion, pollutant classification and data processing with a variety of standardization methods to ensure the accuracy of air quality assessment. It has numerous assessment modes, involving the assessment by hour, day, month, monitoring station location and pollutant type, which addresses the single forms and inaccurate assessment of most air quality assessment systems. It can provide real-time air quality information for users and accurate data preprocessing results for researchers. With stable and reliable operation, friendly interface and rich functions, the system can meet the need for comprehensive air information management and evaluation.

Key words: data preprocessing; standardization; missing values; invalid values; air quality assessment; data mining

空气是人类赖以生存的宝贵资源, 近年来遭到严重的破坏, 空气污染已成为一个亟待解决且严重威胁

人类健康生活的世界性问题。据科学报告^[1-3], 人类因呼吸疾病导致的死亡率与空气污染加剧具有明显的相

^① 基金项目: 国家自然科学基金 (61762057)

收稿时间: 2021-07-04; 修改时间: 2021-07-30; 采用时间: 2021-08-13; csa 在线出版时间: 2022-03-22

关性关系,尤其是可吸入颗粒物,如 $PM_{2.5}$ 、 PM_{10} 、 SO_2 、 NO_2 及 O_3 等,可渗透到肺部和心血管系统导致中风和肺癌等严重疾病^[4].因此,对空气质量进行评估可让人们了解自己所处的空气质量情况.在户外时做好防护准备,从而提高人们的环保意识,减少污染排放.

环境研究者们为解决空气污染问题曾提出了多种空气质量评估系统. Dionova 等提出利用模糊推理系统进行环境室内空气质量评价系统^[5],该系统采用4种室内空气污染物(IAP)和4种热舒适污染物(TCP),使用室内空气质量指数(EIAQI)对污染状况进行量化,采用聚类算法对模糊逻辑控制器的IAQI值和TCI值进行划分,从而有效区分室内空气和热舒适污染物. Jasim 等提出集成化智能算法和地理信息系统建模的城市空气质量评估系统^[6],使用机器学习和地理信息系统建立空气污染模型,并用主成分分析算法和支持向量回归对该模型进行优化,对空气质量进行有效的评估. Xu 等提出基于云模型的空气质量评估系统^[7],使用云模型空气质量评估方法对空气质量数据进行定性和定量分析,结合熵值和超熵值对分析结果的模糊性和随机性进行深入分析,得出大气污染等级. 张澍一等提出气象调整下的区域空气质量评估^[8],使用一种消除气象干扰的时空调整方法.通过控制气象条件调整污染物的时空平均浓度,捕捉到潜在排放量的变化,从而规避因气象条件造成的污染物浓度的变化. Xu 等提出的中国城市空气质量预警系统^[9],采用“分解与集成”理论,结合数据处理技术、支持向量机、仿生优化算法和留置一权决策策略提出了一种混合预测模型,该模型与模糊评估方法共同组成了空气质量的评价系统,在预测和评估方面具有一定的优势.上述提到的空气质量评估方式在某些特定的情况下,评估效果较好,但不能大规模使用,具有一定的局限性.

本文针对上述空气质量评估系统评估形式单一、准确度欠佳等缺点提出了空气信息管理评估系统,旨在应用多样化方法提高空气质量评估准确度和以多种方式展示评估结果.在文中介绍了空气质量数据收集、加工处理及空气质量评估展示,实现了空气质量数据一站式处理,评估结果多样化展示.文中对系统的架构设计、系统功能、主要技术及系统展示进行了详细的介绍.

1 系统设计

1.1 总体架构

本系统针对现有系统评估形式单一,准确度低,评估范围有限等缺点设计的一套集数据预处理、标准化及多样化评估为一体的空气质量评估系统.系统采用BS分层架构,具有开发及维护成本低,分布性和共享性强,访问简单等优势.如图1所示,该系统由3部分组成:数据预处理、数据标准化和空气质量评估.数据预处理是对原始数据进行抽取、清洗、单位转化等;数据标准化是将数据进行多种标准化处理;空气质量评估是通过按时、日、月、采集点和污染物等多种检索方式对空气质量进行评估展示.

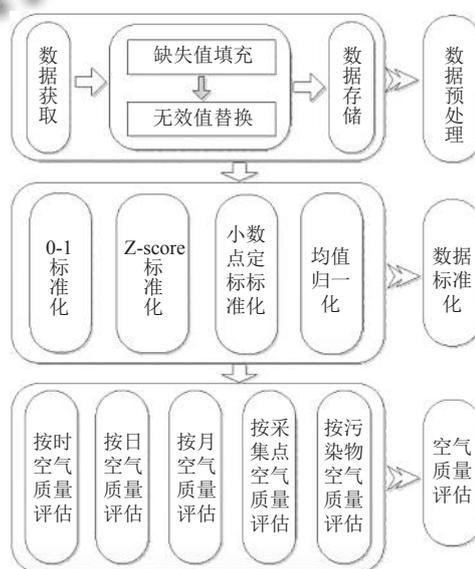


图1 系统架构图

1.2 系统流程图

系统流程图如图2所示,系统运行由数据导入开始.检测数据是否有缺失值和无效值,若有则根据图3进行缺失值填充或图4进行无效值替换.然后根据式(1)~式(4)对数据进行标准化处理.标准化处理后的数据可以通过检索进行空气质量评估展示,也可以保存至本地数据库或以文件的方式导出.

1.3 数据流图

空气质量智能评估系统的核心目标是实现各种空气质量数据的统一收集、处理、存储及对空气质量的评估展示.功能主要有数据导入、数据加工、数据存储、导出及空气质量评估展示,数据流图如图5所示.

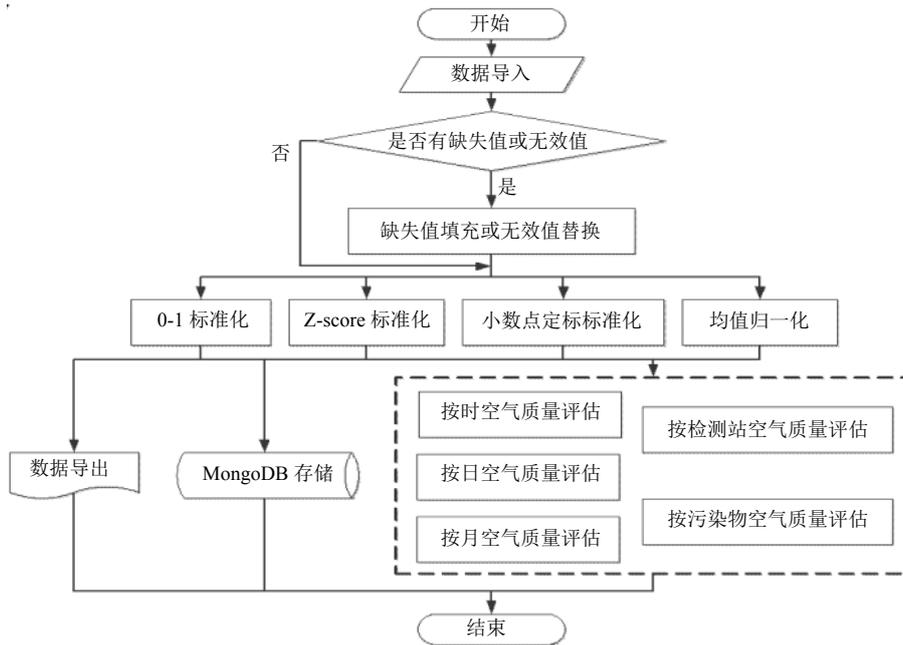


图2 系统流程图

SO ₂ mg/m ³
0.041
0.041
0.043
0.044
0.042
0.040
0.04125
0.037
0.037
0.036
0.04
0.06

0.21 (for 0.041, 0.043, 0.044, 0.042, 0.040)
 +
 0.11 (for 0.037, 0.037, 0.036)
 } 0.33/8=0.04125

图3 缺失值填充

PM _{2.5} mg/m ³	PM _{2.5} mg/m ³
0.125	0.125
0.128	0.128
0	0.01
0	0.01
0	0.01
0.107	0.107
0.066	0.066
0	0.01
0.059	0.059
0.056	0.056
0	0.01
0.104	0.104

图4 无效值替换

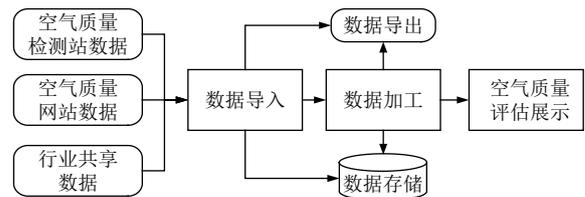


图5 数据流图

(1) 数据导入. 将从空气质量检测站、空气质量数据网站或行业共享等收集到的空气质量数据导入系统. 按业务需求将数据分发给数据加工模块, 也可以存储在数据库或以文件的形式导出.

(2) 数据加工. 按需求对接收到的数据进行抽取、清洗、标准化、单位转化、污染物分类等, 提交给空气质量评估模块, 也可以将处理后的数据存储于数据库或以文件的形式导出.

(3) 空气质量评估展示. 通过对数据加工后的数据进行计算空气质量指数, 然后按时、日、月、监测站位置及污染物种类等检索方式对空气质量进行评估展示.

1.4 关键技术及数据存储

本文采用 Flask 框架^[10]开发, 该框架为 Python 语言中轻量级框架, 具有核心简单、灵活、轻便高效和使用文档完整等优势. 空气质量数据存储的难点在于数据量大、维度高、内容复杂及存取频繁等. 传统的关系型数据库 MySQL^[11], Oracle^[12]等无法实现动态扩

x_{\min} 为总体数据中的最小值。

2.4 空气质量指数

根据中国环境部门2012年发布的《环境空气质量指数(AQI)技术规定(试行)》(HJ633-2012)^[19]和《环境空气质量标准》(GB3095-2012)^[20]规定,采用空气质量指数AQI衡量空气质量。通常情况下,空气污染物含有很多种类,如PM_{2.5}、PM₁₀、SO₂、NO₂及O₃等,首先用各污染物的实测浓度值,按式(5)分别计算空气质量分指数IAQI(individual air quality index)^[19]:

$$IAQI_p = \frac{IAQI_{Hi} - IAQI_{Lo}}{BP_{Hi} - BP_{Lo}}(C_p - BP_{Lo}) + IAQI_{Lo} \quad (5)$$

其中,IAQI_p为相应地区的污染物P的空气质量分指数;C_p为相应地区的污染物P的浓度值;BP_{Hi}为空气质量分指数对应的污染物浓度限值表中与C_p相近的污染物浓度限值的高位值;BP_{Lo}为空气质量分指数对应的

污染物浓度值限制表中与C_p相近的污染物浓度限值的低位值;IAQI_{Hi}为空气质量分指数对应的污染物浓度限值表中与BP_{Hi}对应的空气质量分指数;IAQI_{Lo}为空气质量分指数对应的污染物浓度限值表中与BP_{Lo}对应的空气质量分指数。在得到各个污染物空气质量分指数后,由式(6)计算空气质量指数^[19]。

$$AQI = \max\{IAQI_1, IAQI_2, IAQI_3, \dots, IAQI_n\} \quad (6)$$

其中,IAQI为空气污染物分指数,n为污染物。由式(6)可得,AQI取污染物分指数IAQI的最大值。当AQI大于50时,IAQI最大的污染物为首要污染物;若IAQI最大的污染物为两项或更多时,并列为首要污染物;当IAQI大于100时,该污染物为超标污染物。如表1所示,本文根据当地AQI值给出当前空气质量等级、类别、表示颜色、对健康生活影响及对各类人群应采取的措施等。

表1 空气质量指数对照表

AQI	级别	类别及表示颜色	对健康影响情况	建议应采取的措施
0-50	一级	优 绿色	空气质量令人满意,基本无污染。	各类人群可正常活动。
51-100	二级	良 黄色	空气质量可接受,但某些污染物可能对极少数异常敏感人群健康有较弱影响。	极少数异常敏感人群应减少户外活动。
101-150	三级	轻度污染 橙色	易感人群症状有轻度加剧,健康人群出现刺激症状。	儿童、老年人及心脏病和呼吸系统疾病患者应减少长时间、高强度的户外锻炼。
151-200	四级	中度污染 红色	进一步加剧易感人群症状,可能对正常人群心脏、呼吸系统有影响;	儿童、老年人及心脏病和呼吸系统疾病患者避免长时间、高强度的户外锻炼,一般人群适量减少户外运动。
201-300	五级	重度污染 紫色	心脏病和肺病患者症状显著加剧,运动耐受力降低,健康人群普遍出现症状。	儿童、老年人及心脏病和肺部患者应停留在室内,停止户外运动,一般人群减少户外运动。
>300	六级	严重污染 褐红色	健康人群运动耐受力降低,有明显强烈症状,提前出现某些疾病。	儿童、老年人和病人应当留在室内,避免体力消耗,一般人群应避免户外活动。

3 系统实现

3.1 系统设计

依据如图1所示的系统架构设计,本文系统功能如图7所示,空气信息综合管理评估系统由数据预处理、数据标准化处理、空气质量评估和系统管理4个模块构成。其中,数据预处理模块由数据导入、缺失值填充和无效值替换构成;数据标准化处理模块由0-1标准化、Z-score标准化、小数点定标标准化和均值归一化组成;空气质量评估由按时、日、月、监测站位置及污染物种类空气质量评估组成;系统管理模块由用户管理、日志管理和自定义管理组成。

3.2 系统实现

本文系统空气质量评估采用多样化检索方式,由按时、日、月、监测站位置及污染物种类评估组成。

下面对系统运行效果的部分功能进行展示。图8是在多个监测站按污染物种类以饼图、柱状图和折线图的检索方式展示以月为单位的环境污染物浓度情况,用户按需求选择污染物及展示方式。

图9是空气质量数据进行如图3、图4的无效值和缺失值处理后的展示;图10是对经过如图9处理之后的数据进行0-1标准化、Z-score标准化、小数定标标准化和均值归一法数据标准化展示;图11是对标准化后的数据进行按时空气质量评估展示;图12是对标准化后的数据进行按日空气质量评估展示。

在得到污染物的空气质量指数后,参考表1获取空气质量评估信息。如图13所示,计算空气质量分指数IAQI,获取IAQI中的最大值为空气质量指数AQI,经分析得首要污染物为PM_{2.5},无超标污染物,当天空

气质量指数级别为三级、属于轻度污染;易感人群症状有轻度加剧,健康人群出现刺激症状;建议儿童、老

年人及患有心脏病或呼吸系统疾病的人应减少长时间、高强度的户外锻炼。

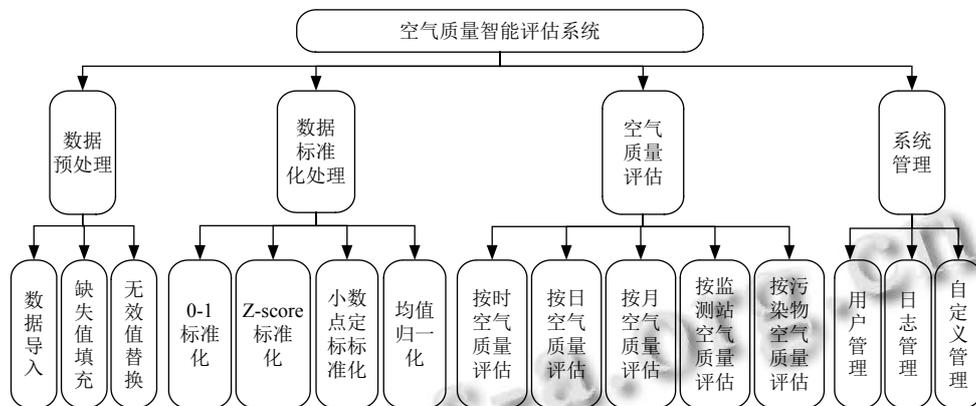


图7 系统功能

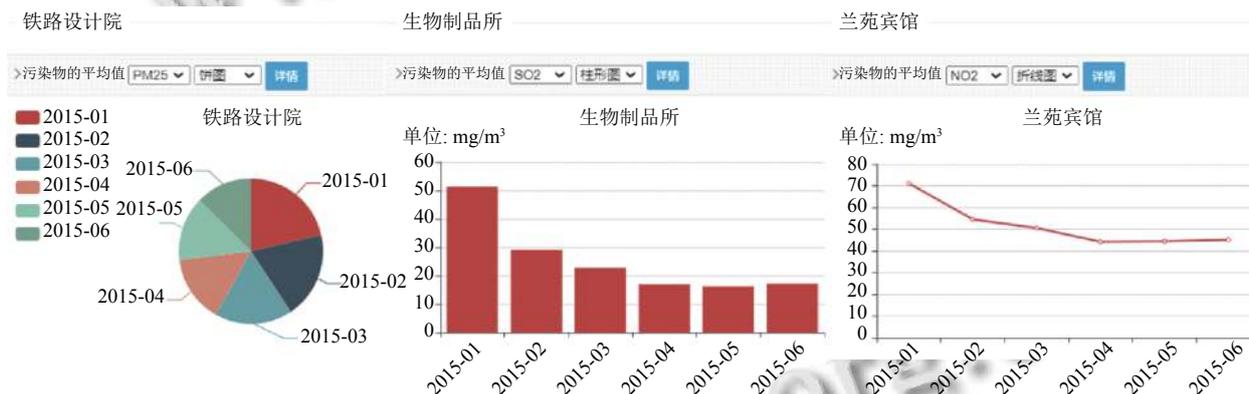


图8 部分监测站按需求展示结果

后台主页 数据列表 × 标准化处理 × 缺失值处理 ×

222222222222 铁路设计院 时间: 年 月 日 年 月 日 搜索 处理当前页

时间	铁路设计院	PM2.5 $\mu\text{g}/\text{m}^3$	SO2 $\mu\text{g}/\text{m}^3$	NO $\mu\text{g}/\text{m}^3$	NO2 $\mu\text{g}/\text{m}^3$	NOx $\mu\text{g}/\text{m}^3$	O3 $\mu\text{g}/\text{m}^3$	CO $\mu\text{g}/\text{m}^3$	PM10 $\mu\text{g}/\text{m}^3$
<input type="checkbox"/> 2015-03-01 00:00:00	生物制品所	true	true	true	true	true	true	true	true
<input type="checkbox"/> 2015-03-01 01:00:00	兰苑宾馆	true	true	true	true	true	true	true	true
<input type="checkbox"/> 2015-03-01 02:00:00	兰大附中校区	true	true	true	true	true	true	true	true
<input type="checkbox"/> 2015-03-01 03:00:00	职工医院	true	true	true	true	true	true	true	true
<input type="checkbox"/> 2015-03-01 04:00:00	舟曲中学	true	true	true	true	true	true	true	true
<input type="checkbox"/> 2015-03-01 05:00:00	西园环保局	true	true	true	true	true	true	true	true
<input type="checkbox"/> 2015-03-01 06:00:00	师范大学	true	true	true	true	true	true	true	true
<input type="checkbox"/> 2015-03-01 06:00:00	铁路设计院	true	true	true	true	true	true	true	true

图9 缺失值及无效值检测处理



图 10 数据标准化处理



图 11 按时计算空气质量指数

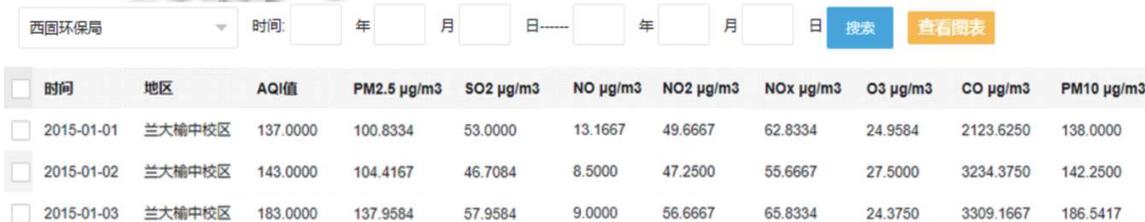


图 12 按日计算空气质量指数

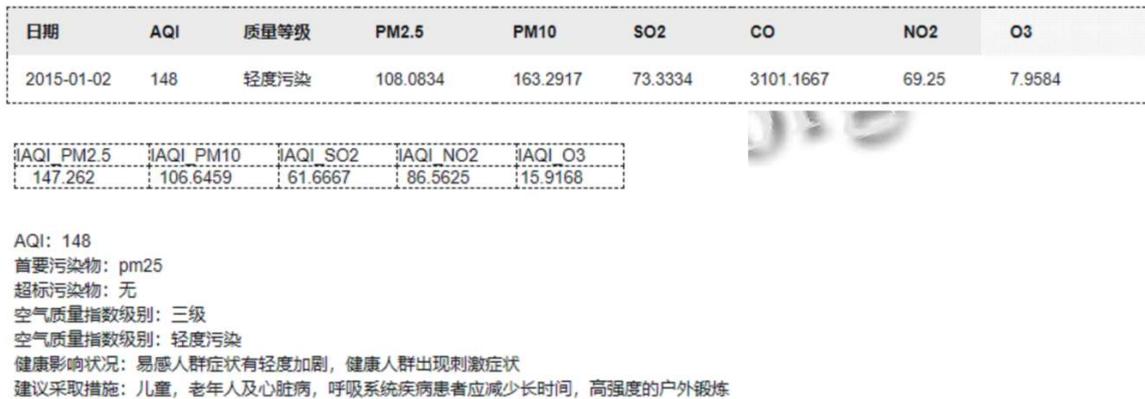


图 13 按日空气质量评估

在本文评估系统中, 按需求选择检测站, 按月评估为根据当月的空气质量数据, 计算空气质量等级为优、良、轻度污染、中度污染和重度污染的天数, 及其所占的概率. 如图 14 所示, 在铁路设计院空气监测站某年 1 月份的空气质量评估结果中, 空气质量等级为优和重度污染的天数为 0、良为 16 天, 轻度污染为

14 天及重度污染为 1 天. 以扇形图表示空气质量等级, 等级为优和重度污染占 0%, 等级为良占 51.62%, 等级为轻度污染占 45.17%, 等级为中度污染占 3.23%.

3.3 系统运行

本系统已开发完毕, 且已在 2021 年 6 月挂载在阿里云服务器, 访问地址为 <http://47.242.77.108:5001/>, 目

前使用于兰州市空气质量评估,使用效果较好.运行环境为1 vCPU,16 GB内存,40 GB高效云盘,1 Mb/s带宽及Linux操作系统.

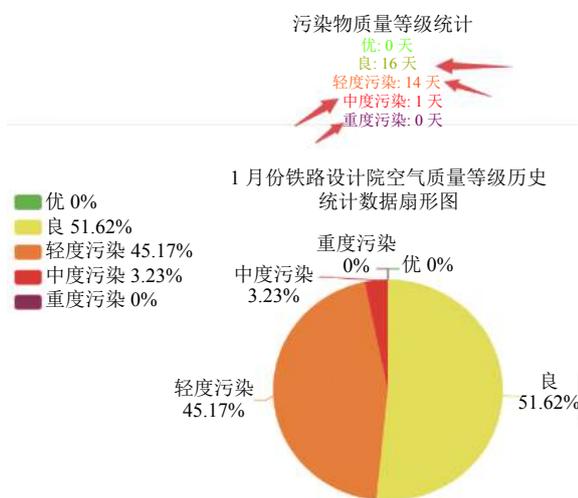


图14 按月空气质量评估

4 结论与展望

空气信息综合管理评估系统构建了集数据预处理、标准化及多样评估为一体的多功能评估平台.主要承担了各地的空气质量评估,为用户提供实时的空气质量情况及出行建议.并对系统进行优化处理,响应速度快、功能完善、界面友好,为空气污染数据处理、空气质量评估提供了良好的支撑,得到众多大气污染相关科研人员的一致肯定,接下来的研究工作是进一步扩充完善系统功能,提升数据支撑和评估能力.

参考文献

- Mo Z, Fu QL, Zhang LF, *et al.* Acute effects of air pollution on respiratory disease mortalities and outpatients in Southeastern China. *Scientific Reports*, 2018, 8(1): 3461. [doi: 10.1038/s41598-018-19939-1]
- Mokoena KK, Ethan CJ, Yu Y, *et al.* Ambient air pollution and respiratory mortality in Xi'an, China: A time-series analysis. *Respiratory Research*, 2019, 20(1): 139. [doi: 10.1186/s12931-019-1117-8]
- Zhao YJ, Wang SY, Lang LL, *et al.* Ambient fine and coarse particulate matter pollution and respiratory morbidity in Dongguan, China. *Environmental Pollution*, 2017, 222: 126–131. [doi: 10.1016/j.envpol.2016.12.070]
- 王鹏跃, 郭茂祖, 赵玲玲, 等. 城市空气质量感知方法综述.

计算机科学, 2019, 46(S1): 35–40, 51.

- Dionova BW, Mohammed MN, Al-Zubaidi S, *et al.* Environment indoor air quality assessment using fuzzy inference system. *ICT Express*, 2020, 6(3): 185–194. [doi: 10.1016/j.ict.2020.05.007]
- Jasim OZ, Hamed NH, Abid MA. Urban air quality assessment using integrated artificial intelligence algorithms and geographic information system modeling in a highly congested area, Iraq. *Journal of Southwest Jiaotong University*, 2020, 55(1): 1–15. [doi: 10.35741/issn.0258-2724.55.1.1]
- Xu QW, Xu KL. Assessment of air quality using a cloud model method. *Royal Society Open Science*, 2018, 5(9): 171580. [doi: 10.1098/rsos.171580]
- 张澍一, 陈松蹊, 郭斌, 等. 气象调整下的区域空气质量评估. *中国科学: 数学*, 2020, 50(4): 527–558.
- Xu YZ, Yang WD, Wang JZ. Air quality early-warning system for cities in China. *Atmospheric Environment*, 2017, 148: 239–257. [doi: 10.1016/j.atmosenv.2016.10.046]
- Grinberg M. *Flask Web Development: Developing Web Applications with Python*. Sebastopol: O'Reilly, 2018.
- DuBois P. *MySQL*. 4th ed. Pearson Education, 2008.
- Loney K. *Oracle Database 10g: The Complete Reference*. Emeryville: McGraw-Hill/Osborne, 2004.
- Parker Z, Poe S, Vrbsky SV. Comparing NoSQL MongoDB to an SQL DB. *Proceedings of the 51st ACM Southeast Conference*. Savannah: ACM, 2013. 1–6.
- Chodorow K. *MongoDB: The Definitive Guide: Powerful and Scalable Data Storage*. 2nd ed. O'Reilly Media, 2013.
- 李超. 基于多层网络的京津冀区域大气污染时空关联效应研究 [博士学位论文]. 北京: 中国地质大学 (北京), 2019.
- Bashar M, Cumanan K, Burr AG, *et al.* On the uplink max-min SINR of cell-free massive MIMO systems. *IEEE Transactions on Wireless Communications*, 2019, 18(4): 2021–2036. [doi: 10.1109/TWC.2019.2892463]
- Cheadle C, Vawter MP, Freed WJ, *et al.* Analysis of microarray data using Z score transformation. *The Journal of Molecular Diagnostics*, 2003, 5(2): 73–81. [doi: 10.1016/S1525-1578(10)60455-2]
- Olick JK. What does it mean to normalize the past?: Official memory in German politics since 1989. *Social Science History*, 1998, 22(4): 547–571.
- 环境保护部. HJ 633-2012 环境空气质量指数 (AQI) 技术规范 (试行). 北京: 中国环境科学出版社, 2016.
- 中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会. GB 3095-2012 环境空气质量标准. 北京: 中国环境科学出版社, 2016.