

基于大气环境监测数据的关联规则方法应用研究*

卢静 周围 李皓维 周运宝 于昊楠

北京交通大学交通大数据与人工智能教育部重点实验室, 计算机与信息技术学院
北京 100044

摘要 关联规则算法是挖掘数据隐藏价值的重要方法之一。本文对关联规则在大气环境领域的应用情况进行调研, 首先通过全球论文索引数据库来分析关联规则在大气环境中的研究趋势, 进而对关联规则两大类别: 基于兴趣度指标的穷举方法和基于启发式的方法进行工作原理、评价指标的对比分析, 最后, 梳理了大气污染物监测数据来源、关联规则在大气污染因素分析、预测方面的具体研究实例。希望本文为大气环境与计算机交叉领域的研究提供参考。

关键字 环境大数据, 关联规则, 关联分析, 群体智能, 数据挖掘

Research on the Application of Association Rule Method Based on Atmospheric Environment Monitoring Data

LU Jing ZHOU Wei LI Haowei ZHOU Yunbao YU Haonan

Key Laboratory of Big Data & Artificial Intelligence in Transportation, Ministry of Education, School of
Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China
wzhou@bjtu.edu.cn

Abstract—Association rule algorithm is one of the important methods for mining the hidden value of data. This paper investigates the application of association rules in the field of atmospheric environment. Firstly, web of sciences related database are used to analyze the research trends of association rules in the atmospheric environment. Then, the working principles and evaluation indicators of two major categories of association rules: interest based exhaustive methods and heuristic based methods are compared and analyzed. Finally, the sources of atmospheric pollutant monitoring data are listed, specific research examples of association rules in the analysis and prediction of air pollution factors are introduced. We hope this paper can provide references for the research on the intersection of atmospheric environment and computer science.

Keywords—Environmental Big data, Association Rules, Association Analysis, Swarm Intelligence, Data Mining

1 引言

随着信息技术的飞速发展, 各个领域积累了海量的数据。在大气环境监测领域, 大量的监测站点对不同的监测项目进行 24 小时连续在线监测, 使得大气环境监测数据呈爆炸式增长。1998 年, 我国开始启动对大气污染的监测与治理。2012 年, 国家环保部门开始逐步向社会公开一氧化碳 (CO)、二氧化氮 (NO₂)、臭氧 (O₃)、可吸入颗粒物 (PM₁₀)、细颗粒物 (PM_{2.5}) 和二氧化硫 (SO₂) 等基本大气污染物的小时浓度^[1]。截至 2017 年, 国控监测点位已建成 1436 个^[2], 同时各地的网格化监测网络也逐步推广, 环境大数据规模快速增长。

在碳监测方面, 十四五时期, 生态环境部先后印发了《碳监测评估试点工作方案》和《“十四五”生

态环境监测规划》^[3], 指出要推进碳监测评估试点, 组织重点行业企业温室气体排放监测试点, 组建重点城市温室气体监测网络, 合作共建全国碳监测网络。目前已建成一定数量的碳监测站点, 阶段性成果初步证实 CO₂ 在线监测具有较好应用前景。

在大气环境监测大数据的背景下, 如何识别和筛选大气环境监测数据中存在的丰富关联关系, 是大数据复杂关联关系挖掘的重要任务之一。本文的组织方式如下, 第 2 章, 通过全球论文索引数据库的引用数据, 来分析关联规则在大气环境中的研究趋势, 第 3 章对关联规则两大类别: 基于兴趣度指标的穷举方法和基于启发式的方法进行工作原理、评价指标的对比分析, 第 4 章梳理了大气污染物监测数据来源、关联规则在大气污染因素分析以及预测方面的具体应用实例。最后我们对本文进行了总结并展望了进一步的研究方向。

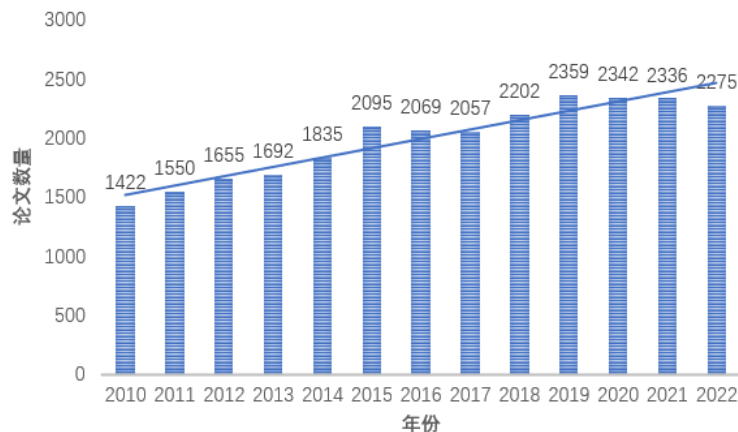
*基金资助: 本文得到河北省重点研发计划“碳排放监测系列装备研发与应用”(21373902D)资助。

2 关联规则的研究趋势

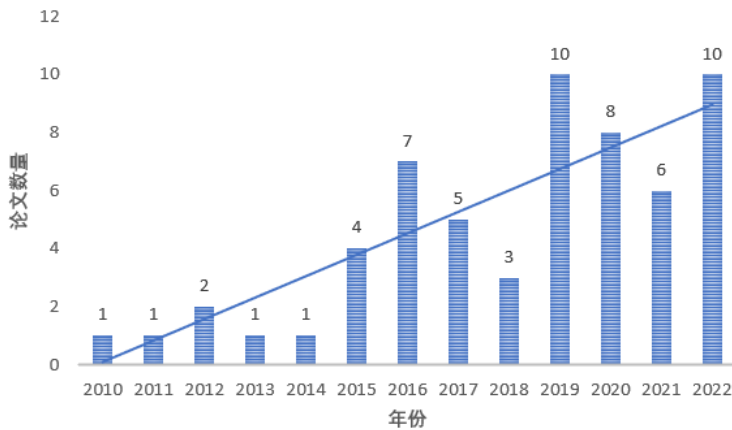
关联关系挖掘任务旨在探索变量间潜在的同现规律和关联模式。由于数据存在分布不确定、数据关联关系复杂多样、多种关联关系并存等特点,使得统计意义上的关联关系挖掘面临着诸多挑战^[4]。自1993年Agrawal等人^[5-6]提出了Apriori关联规则挖掘算法以来,关联规则的研究始终是学术界的热点,作为机器学习的经典算法之一,它可以用于探究多变量之间的关联关系。

在web of science中检索“association rules”关键字,时间范围选择“2010-01-01~2022-12-31”,共计检索出25889条结果,如图1(a)发现,自2010

年以来,学术界对关联规则的研究虽局部有较小的波动,但整体上呈上升趋势,充分说明关联规则挖掘算法是数据挖掘领域最重要的方法之一。大气环境监测领域中的关联规则挖掘试图在大气污染物、气象因素和温室气体等监测数据中提取价值。通过在web of science中检索“association rules”或“pattern mining”和“air pollution”关键字,时间范围选择“2010-01-01~2022-12-31”,共计检索出59条结果,如图1(b)可以看出,大气环境监测领域关联规则的研究整体上呈上升趋势,充分说明关联规则在大气环境监测领域的应用具有一定的意义,可以为相关交叉领域的学者和科技工作者开展相关研究提供借鉴参考。



(a) 关联规则算法研究整体趋势图



(b) 大气环境监测领域关联规则算法研究趋势图

图 1 关联规则算法研究趋势图

3 关联规则的算法分类与比较

本文在数据库 IEEE Xplore、Google Scholar、ScienceDirect 和 Web of Science 里查阅不同领域的

论文,梳理出如下表1所示的两大类关联规则算法,它们的基本思想、优缺点及经典算法列出。

关联规则最常用的指标是支持度(Support)和置信度(Confidence)。此外,为了挖掘出准确、可靠、

有趣且容易理解的规则，提升度^[7](Lift)、可理解度(Comprehensibility)以及属性振幅^[8](Amplitude of the attributes) 被相继提出。

表 1 关联规则算法分类与优缺点对比

类别	基于兴趣度指标的穷举方法	基于启发式的方法
基本思想	首先遍历数据库挖掘频繁项集, 然后从频繁项集中按照特定兴趣度指标提取关联规则	首先对关联规则进行编码, 然后通过启发式搜索找到关联规则的最优解集
优点	简单、容易理解	搜索效率高
缺点	计算效率低, 产生误导规则和冗余规则	遗漏潜在的关联规则
经典算法	Apriori算法	遗传算法

3.1 基于兴趣度指标的穷举方法

基于兴趣度指标的穷举方法是挖掘关联规则的主流方法。该方法分为两个步骤：第一步是遍历数据库寻找满足阈值要求的频繁项集；第二步是从频繁项集中生成满足兴趣度指标的关联规则。其具有简单、容易理解等优点，但也存在易产生误导性规则、冗余性规则和对大规模数据集缺乏可扩展性等问题。此类方法的代表主要包括 Apriori 算法^[5-6]、FP-growth 算法^[9]以及 Eclat 算法^[10]。其中 Apriori 算法采用广度优先搜索策略，它通过设置最小支持度阈值来挖掘频繁项集，然后通过设置最小置信度阈值在该频繁项集中生成规则。FP-growth 算法采用深度优先搜索策略，它通过构建 FP-tree，从 FP-tree 中挖掘频繁项集。Eclat 算法采用深度优先搜索策略。它通过扩展项集的前缀来生成候选项，直到找到一个不常见的前缀为止。

基于穷举的方法虽然较为经典且应用广泛，但是对于数值型数据，此类方法必须先将数据离散化，离散化的过程会损失原始数据的部分信息，且无法挖掘离散区间之外的有趣规则，因此获得的规则实际上是离散数据后的规则。

3.2 基于启发式的方法

基于启发式的方法是关联规则领域的另一大方向，常用于处理数值型数据。此方法不需要任何的预处理操作，简化了关联规则挖掘步骤。在关联规则挖掘过程中既可以自动找到合适的属性范围，又可以充分考虑属性之间的交互。常见的启发式算法如黏菌算法^[12]、遗传算法^[13]、粒子群算法^[14]以及混合算法^[15]等均已经被广泛应用于关联规则挖掘任务。通常来说，使用启发式搜索的方法挖掘关联规则首先需要将每个个体按照一定的编码方式编码成向量形式。编码方法主要分

为两种：密歇根方法和匹兹堡方法。密歇根方法将一个个体呈现为一条规则，该方法适合目标是找到一个高质量的帕累托前沿的问题。匹兹堡方法将一个个体呈现为一组规则，该方法适用于分类或聚类应用。算法产生的最终规则可由向量进行规则解码得到。

使用启发式搜索的方法一般需要同时考虑多个度量指标，因此也可以称之为一个多目标优化问题。在关联规则挖掘过程中，每条规则的产生过程可以看作是在 k 个属性中选取 d ($d \leq k$) 个属性来构成一条完整的规则。每个属性均存在“被选择为规则前件”、“被选择为规则后件”以及“不被选择”三种情况。算法的目的是选择最优的属性集并根据其前后件分类组成一条规则，所获得的规则尽可能准确、可靠、有趣且易于理解。常见的目标函数分别是支持度、置信度、可理解度及属性振幅等。

代表关联规则的候选解可以在目标值上优于、差于或等于其他候选解，也可以与其他候选解无关。一个规则A在任何目标上都不比另一个规则B差，且至少在一个目标上比规则B更好，则称规则A优于或支配规则B。因此，最优解所代表的规则不被任何其他规则支配，最优解集代表了高质量的非支配关联规则集合。当非支配解的个数超过阈值时，可以根据拥挤度距离^[16]等方式来删除多余的非支配解”。

使用群体智能优化算法实现关联规则挖掘，一般将数据集视为搜索空间，群体智能优化算法视为搜索策略来实现规则挖掘任务。主要步骤如下：群体智能优化算法首先考虑数据集中属性的下界和上界来初始化候选解，根据支配关系被支配候选解将被删除，非支配的候选解将被保留，使用拥挤度距离来选取每次迭代的最优解。其后，如果非支配候选解的数量超过确定的阈值，则使用拥挤度距离来删除彼此接近的规则。最后，使用群体智能优化算法更新个体的位置。当算法达到最大迭代次数或其他终止条件时，算法终止。

启发式搜索算法有着较低的执行时间，因此它可以在合理的时间内挖掘出有趣规则集，但由于其本身的搜索机制，该方法也可能遗漏某些有价值的规则。

3.3 小结

综上所述，基于兴趣度指标的穷举方法虽然较为经典且应用广泛，但是此类方法应用于大气环境监测数据时必须先将数据离散化，离散化的过程会损失原始数据的部分信息，且无法挖掘离散区间之外的有趣规则，因此获得的规则实际上是离散数据后的规则。基于启发式的方法可以作为分析数值型数据任务的具有前景的一种方法，它能简化关联规则的挖掘步骤，自动调整属性间隔，并找到一些高质量、可理解、可

解释且低冗余的规则, 但如何平衡算法的搜索能力, 避免规则遗漏仍是一个待解决的问题。

4 关联规则在大气环境监测领域的应用

4.1 大气环境监测数据源

在大气环境监测领域, 空气质量和气象条件的监测数据较为丰富。全国空气质量数据来源主要是中国环境监测总站的全国城市空气质量实时发布平台, 各省空气质量数据来源主要是各省生态环境厅首页下面的空气质量监测平台。以河北省为例, 在“河北生态环境厅”首页的“数据中心”, 点击“空气质量实时自动发布系统”即可跳转到监测数据页面。

此外, 随着遥感监测的不断完善以及各学者对空气污染的广泛研究, 陆续出现了高分辨率高质量空气污染数据集。例如: 由美国马里兰大学团队研究并由国家地球系统科学数据中心长江三角洲分中心(<http://nnu.geodata.cn/>)协助整合提供的中国高分辨率高质量近地表空气污染物数据集(CHAP)^[17], 该数据集基于多源卫星遥感和人工智能技术, 综合考虑了丰富地基观测、卫星遥感产品、大气再分析和模式排放清单等数据^[17]。目前应用较为广泛, 截至 2023 年 2 月, 已发表相关应用文章 130 余篇。

气象数据来源包括中国气象数据网-国家气象科学数据中心(<https://data.cma.cn>), 提供中国国家级地面站小时值数据, 包括气温、气压、相对湿度、风、降水量等要素小时观测值。美国国家气候数据中心(NCDC)(<https://www.ncei.noaa.gov>), 提供中国区域 1951 年至今的 900 多个站点的逐日数据。

4.2 大气污染物之间的关联关系挖掘

通过大气环境监测数据, 可以发现大气污染物之间的关联关系, 如二氧化硫和氮氧化物等污染物的相关性, 进一步可以帮助预测和防治大气污染。Chen 等人^[19]利用考虑时滞效应的动态 Apriori 算法, 挖掘不同城市 PM_{2.5} 极端污染事件的时空关联。Aggarwal 等人^[20]利用 HBST 算法从时间和位置感知的空气质量数据中识别和提取频繁模式。Borah 等人^[21]提出了一种高效的稀有模式挖掘技术, 从真实环境数据集中获取完整的稀有模式集。然后从生成的稀有模式中进一步提取代表空气污染物与其他环境参数之间相关性的重要稀有关联规则, 以确定空气污染的实质原因。钱怡欣^[11]结合 Apriori 算法和证据推理解决了多个监测站点数据信息挖掘过程中存在的数据差异化问题, 适用于多元化规则挖掘需求和规则集不相同情况下的融合需求。

4.3 大气污染物与其他因素之间的关联关系挖掘

通过关联规则挖掘方法研究大气污染物与气象条件, 经济发展, 医疗健康等相关领域之间的关系, 从而进一步深化对相关领域的理解。大量研究表明, 污染物与气象条件之间存在不同程度的关联关系。Godoy 等人^[22]利用 Apriori 算法识别污染物浓度较高时的主要气象条件。Martínez-Ballesteros 等人^[23]于 2009 年提出了基于遗传算法的关联规则挖掘算法, 挖掘温度、风和臭氧之间的关联关系, 2010 年提出了基于进化计算的定量关联规则挖掘算法挖掘大气污染与气候条件之间的关联规则^[24]。随后, 于 2011 年提出了一种基于遗传算法的定量关联规则挖掘算法, 并将其应用于由多个气候变量组成的多维时间序列, 结果具有良好的性能^[25]。Ma 等人^[26]运用关联规则算法, 从时间和空间的角度寻找影响因素, 并分析其对碳排放的共同影响。

“高能耗、高排放”的经济发展往往伴随着环境污染。通过了解二者的关系有利于在保持经济增长的基础上减少环境污染。Zeng 等人^[27]采用 Apriori 关联规则挖掘算法分析了 110 个地级市的环境、经济和社会特征。Li 等人^[28]利用 FP-growth 算法挖掘工业化水平与空气质量之间的关联规律。Li 等人^[29]利用 Apriori 关联规则挖掘技术从 44 个社会经济因素中识别出关键因素, 提取出它们与 PM_{2.5} 污染的潜在关联关系。并结合 160 个不同 PM_{2.5} 水平城市的年、季、周情景, 进一步揭示了不同类型城市中不同因子贡献差异的隐藏模式。Yang 等人^[30]利用 Apriori 关联规则挖掘技术, 在中国 297 个城市的宏观、中尺度和微观情景中, 揭示了臭氧变化与社会经济因素之间的隐藏模式。

研究表明, 即使空气污染水平很低, 空气污染也会影响身体的所有部位。因此, 探究污染物与健康之间的关系至关重要。Kanani 等人^[31]利用空间关联规则挖掘技术, 研究了空气污染对哮喘性过敏的影响, 并研究了哮喘性过敏的患病率与可能影响空气污染的环境特征之间的关系。同样, Karimipour 等人^[32]利用模糊关联规则挖掘研究空气污染对哮喘性过敏的影响。次年, Karimipour 等人^[33]利用模糊空间关联规则挖掘算法挖掘过敏患病率的空间分布与空气污染六参数以及与公园、道路的居住距离之间的关联规则。Lee 等人^[34]提出了基于模式的决策树(PBDT)和基于模式的类关联规则(PBCAR)两种数据挖掘方法, 挖掘儿童哮喘过敏与天气、空气污染等环境因素的关系。

4.4 大气污染预测

通过对历史监测数据进行关联规则挖掘,可以发现大气污染的规律和趋势,从而为大气污染的预测提供依据。Zhu 等人^[35]利用关系矩阵和相关运算,基于关系代数理论,利用优化关系关联规则求出所有频繁项集。挖掘空气污染物对下呼吸道疾病影响的关联规则,进一步采用数据处理成组方法预测下呼吸道疾病患者人数。Qin 等人^[36]利用 Apriori 算法对极端颗粒物数据进行分析,挖掘颗粒物时空变化及城际影响的定量关联规律。邢胜等人^[37]利用关联规则算法挖掘周边大城市污染对沧州空气环境的影响程度,进一步建立神经网络模型对大气污染进行预测。Jiang P 等人^[38]通过高维关联规则集成不同物体的时空链接,提出了一种混合策略用于 PM_{2.5} 浓度的预测。Yadav 等人^[39]改进连续目标顺序模式挖掘算法,该算法是一种基于 Apriori 的关联规则挖掘算法,生成一组有助于预测空气污染物浓度的关联规则。Varde 等人^[40]结合 Apriori 关联规则及聚类算法对空气质量进行预测。

此外,通过分析不同区域和时间段内的大气污染物浓度数据,挖掘它们之间的关系,可以应用于污染源的定位、制定合理的污染物排放标准、制定相应的污染物削减策略、以及帮助政府采取相应的措施来控制大气污染。总之,大气污染物之间的关联关系挖掘可以帮助我们更好地理解大气污染机理,从而制定更加科学的大气污染控制措施,保护人类身体健康和生存环境。

4.5 小结

综上所述,关联规则应用于大气环境监测领域主要集中在分析大气污染与其他因素之间的关联关系。研究方法大多基于经典的 Apriori 算法,基于群体智能的算法在近几年逐渐增多。研究人员多从事环境相关专业,计算机领域与环境领域的跨学科交叉研究不够深入。未来大气污染领域的研究将更加多元化和复杂化,需要更多跨学科的方法和技术,从多个角度对大气污染问题进行综合研究和探索。

5 结束语

本文展示了关联规则的特点及其在大气环境数据挖掘方面的巨大潜力。关联规则的优势之一在于它的输出很容易解释,便于使环境领域的学者理解并使用。使用关联规则不仅可以发掘空气污染物内部之间的关联关系,也可以发现空气污染物与其他难以明确发现的因素之间的关联关系。这些关系可以使我们能够更好地了解大气污染领域。此外,关联规则也可以进行数据特征选择,与其他方法结合,为下一步预测或者分类等任务提供依据。

未来研究方向有以下几个方面:进一步采用更加灵活和高效的关联规则挖掘算法深入研究不同污染物之间的关联规则,从而发现更加微小和隐蔽的关联关系,更加准确地描述不同污染物之间的关联关系;与时空序列预测模型进行融合,对大气污染趋势进行更加准确和稳定的预测;将关联规则用于源解析中,深入研究空气污染物的来源解析问题,以便制定更加精准和针对性的污染物控制策略。总之,未来可以通过深入研究关联规则在大气污染领域中的应用,开发更加准确和有效的数据挖掘算法和模型,以解决大气污染领域中的关键问题和挑战。

参考文献

- [1] 胡冬梅. 基于时空数据挖掘的城市空气污染分析与预测[D].清华大学,2018.
- [2] 中国环境监测总站. 国家环境空气质量监测网[EB/OL].2017.
http://www.cnemc.cn/zjj/jcwl/dqjcw/201711/t20171108_645109.shtml.
- [3] 中华人民共和国生态环境部. 生态环境部召开 1 月例行新闻发布会 [EB/OL]. 2023.
https://www.mee.gov.cn/ywdt/zbft/202301/t20230117_1013623.shtml.
- [4] 成红红. 基于粒计算的关联关系挖掘研究[D].山西大学,2020.
- [5] Agrawal R, Imielinski T, Swami A. Mining Association Rules Between Sets of Items in Large Databases[C]. //Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, United States, 1993:207-216.
- [6] Agrawal R, Srikant R. Fast Algorithms for Mining Association Rules[C]. //Proceedings of the 20th International Conference on Very Large Data Bases, Santiago de Chile, 1994:487-499.
- [7] Ghafari S M, Tjortjis C. A Survey on Association Rules Mining Using Heuristics[J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2019, 9(4): e1307.
- [8] Altay E V, Alatas B. Differential Evolution and Sine Cosine Algorithm Based Novel Hybrid Multi-Objective Approaches for Numerical Association Rule Mining[J]. Information Sciences, 2021, 554: 198-221.
- [9] Han J, Pei J, Yin Y, et al. Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach[J]. Data Mining and Knowledge Discovery, 2004, 8: 53-87.
- [10] Zaki M J. Scalable Algorithms for Association Mining[J]. IEEE Transactions on Knowledge and Data Engineering, 2000, 12(3): 372-390.
- [11] 钱怡欣. 基于 Apriori 算法和证据推理的大气环境关键规则挖掘研究[D].北京交通大学,2019.
- [12] Yacoubi S, Manita G, Amdouni H, et al. A Modified Multi-Objective Slime Mould Algorithm with Orthogonal Learning for Numerical Association Rules Mining[J]. Neural Computing and Applications, 2022: 1-27.
- [13] Kabir M M J, Xu S, Kang B H, et al. A New Multiple Seeds Based Genetic Algorithm for Discovering a Set of

- Interesting Boolean Association Rules[J]. *Expert Systems with Applications*, 2017, 74: 55-69.
- [14] Su T, Xu H, Zhou X. Particle Swarm Optimization-Based Association Rule Mining in Big Data Environment[J]. *IEEE Access*, 2019, 7: 161008-161016.
- [15] Altay E V, Alatas B. Differential Evolution and Sine Cosine Algorithm Based Novel Hybrid Multi-Objective Approaches for Numerical Association Rule Mining[J]. *Information Sciences*, 2021, 554: 198-221.
- [16] Deb K, Pratap A, Agarwal S, et al. A Fast and Elitist Multi-Objective Genetic Algorithm: NSGA-II[J]. *IEEE Transactions on Evolutionary Computation*, 2002, 6(2): 182-197.
- [17] Wei J, Li Z Q. (2023). China High resolution and high quality PM2.5 dataset (2000-2021). National Qinghai-Tibet Plateau Scientific Data Center, <https://data.tpdc.ac.cn/zh-hans/data/6168e75d-93ab-4e4a-b7ff-33152e49d0bf>.
- [18] Wei, J., Li, Z., Lyapustin, A., Sun, L., Peng, Y., Xue, W., Su, T., & Cribb, M. (2021). Reconstructing 1-km-resolution high-quality PM2.5 data records from 2000 to 2018 in China: spatiotemporal variations and policy implications. *Remote Sensing of Environment*, 252, 112136. <https://doi.org/10.1016/j.rse.2020.112136>
- [19] Chen J, Song X, Zang L, et al. Spatio-temporal association mining of intercity PM2. 5 pollution: hubei province in China as an example[J]. *Environmental Science and Pollution Research*, 2022: 1-14.
- [20] Aggarwal A, Toshniwal D. Frequent pattern mining on time and location aware air quality data[J]. *IEEE Access*, 2019, 7: 98921-98933.
- [21] Borah A, Nath B. An efficient method for mining rare association rules: a case study on air pollution[J]. *International Journal on Artificial Intelligence Tools*, 2021, 30(04): 2150018.
- [22] Godoy A R L, da Silva A E A. Spatial patterns and temporal variations of pollutants at 56 air quality monitoring stations in the state of São Paulo, Brazil[J]. *Environmental Monitoring and Assessment*, 2022, 194(12): 1-22.
- [23] Martínez-Ballesteros M, Martínez-Álvarez F, Troncoso A, et al. Quantitative association rules applied to climatological time series forecasting[C]//*Intelligent Data Engineering and Automated Learning-IDEAL 2009: 10th International Conference*, Burgos, Spain, September 23-26, 2009. *Proceedings 10*. Springer Berlin Heidelberg, 2009: 284-291.
- [24] Martínez-Ballesteros M, Troncoso A, Martínez-Álvarez F, et al. Mining quantitative association rules based on evolutionary computation and its application to atmospheric pollution[J]. *Integrated Computer-Aided Engineering*, 2010, 17(3): 227-242.
- [25] Martínez-Ballesteros M, Martínez-Álvarez F, Troncoso A, et al. An evolutionary algorithm to discover quantitative association rules in multidimensional time series[J]. *Soft Computing*, 2011, 15: 2065-2084.
- [26] Ma X, Jiang P, Jiang Q. Research and application of association rule algorithm and an optimized grey model in carbon emissions forecasting[J]. *Technological Forecasting and Social Change*, 2020, 158: 120159.
- [27] Zeng L, Wang B, Fan L, et al. Analyzing sustainability of chinese mining cities using an association rule mining approach[J]. *Resources Policy*, 2016, 49: 394-404.
- [28] Li T, Li Y, An D, et al. Mining of the association rules between industrialization level and air quality to inform high-quality development in China[J]. *Journal of environmental management*, 2019, 246: 564-574.
- [29] Li W, Yang G, Qian X. The socioeconomic factors influencing the PM2. 5 levels of 160 cities in China[J]. *Sustainable Cities and Society*, 2022, 84: 104023.
- [30] Yang G, Liu Y, Li W, et al. Association analysis between socioeconomic factors and urban ozone pollution in China[J]. *Environmental Science and Pollution Research*, 2022: 1-15.
- [31] Kanani Sadat Y, Karimipour F, Kanani Sadat A. Investigating the relation between prevalence of asthmatic allergy with the characteristics of the environment using association rule mining[J]. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2014, 40: 169-174.
- [32] Karimipour F, KANANI S Y. Investigating the relation between prevalence of asthmatic allergy with the characteristics of the environment using fuzzy association rule mining[J]. 2015.
- [33] Karimipour F, Kanani-Sadat Y. Mapping the vulnerability of asthmatic allergy prevalence based on environmental characteristics through fuzzy spatial association rule mining[J]. *Journal of Environmental Informatics*, 2016, 28(1): 1-10.
- [34] Lee C H, Chen J C Y, Tseng V S. A novel data mining mechanism considering bio-signal and environmental data with applications on asthma monitoring[J]. *Computer methods and programs in biomedicine*, 2011, 101(1): 44-61.
- [35] Zhu W, Wang J, Zhang W, et al. Short-term effects of air pollution on lower respiratory diseases and forecasting by the group method of data handling[J]. *Atmospheric environment*, 2012, 51: 29-38.
- [36] Qin S, Liu F, Wang C, et al. Spatial-temporal analysis and projection of extreme particulate matter (PM10 and PM2. 5) levels using association rules: A case study of the Jing-Jin-Ji region, China[J]. *Atmospheric Environment*, 2015, 120: 339-350.
- [37] 邢胜,王晓兰,金会赏.基于数据挖掘的大气污染预测分析研究[J]. *决策探索(中)*, 2019, No.630(10):84.
- [38] Jiang P, Dong Q, Li P. A novel hybrid strategy for PM2. 5 concentration analysis and prediction[J]. *Journal of environmental management*, 2017, 196: 443-457.
- [39] Yadav M, Jain S, Seeja K R. Prediction of air quality using time series data mining[C]//*International Conference on Innovative Computing and Communications*. Springer, Singapore, 2019: 13-20.
- [40] Varde A S, Pandey A, Du X. Prediction tool on fine particle pollutants and air quality for environmental engineering[J]. *SN Computer Science*, 2022, 3(3): 1-31.