

基于置信规则库的空气质量预测模型

梁泽森, 付 伟

哈尔滨师范大学计算机科学与信息工程学院, 黑龙江 哈尔滨

收稿日期: 2024年11月1日; 录用日期: 2024年11月29日; 发布日期: 2024年12月9日

摘 要

空气质量一直是公众和社会最关注的环境问题之一, 其质量已经因为经济活动而受到损害。然而, 由于空气的复杂性以及人类活动对空气的影响, 预测空气质量是一项不小的挑战。传统上使用单一的机器学习模型预测空气质量时, 在各种AQI波动趋势下难以获得良好的预测结果。为了有效地解决这一问题, 本文提出了一种分层置信规则库(Hierarchy Belief Rule Base-RFI)预测模型, 为空气质量预测提供了一种有用的方法。首先, 为了提高模型的精准度, 采用了随机森林来对关键特征进行筛选, 同时对未被筛选出的特征进行ER融合。其次, 提出了空气质量预测方法的可解释性准则, 以规范整个建模过程的可解释性。然后给出了模型的推理和优化方法。最后通过实验验证了该模型的有效性。

关键词

空气质量指数预测, 可解释性, 置信规则库, 证据推理

Air Quality Prediction Model Based on Confidence Rule Base

Zesen Liang, Wei Fu

School of Computer Science and Information Engineering, Harbin Normal University, Harbin Heilongjiang

Received: Nov. 1st, 2024; accepted: Nov. 29th, 2024; published: Dec. 9th, 2024

Abstract

Air quality has always been one of the top environmental concerns of the public and society, and its quality has been compromised by economic activities. However, due to the complexity of the air and the impact of human activities on it, predicting air quality is not a small challenge. Traditionally, when a single machine learning model is used to predict air quality, it is difficult to obtain good prediction results under various AQI fluctuation trends. In order to solve this problem effectively, a prediction model based on Hierarchy Belief Rule Base-RFI is proposed in this paper, which provides

a useful method for air quality prediction. First of all, in order to improve the accuracy of the model, random forest is used to screen the key features, and ER fusion is carried out on the features that are not screened. Secondly, the interpretability criteria of air quality prediction methods are proposed to standardize the interpretability of the whole modeling process. Then the inference and optimization methods of the model are given. Finally, the validity of the model is verified by experiments.

Keywords

AQI Prediction, Interpretability, Confidence Rule Base, Evidence Reasoning

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

空气污染是 21 世纪普遍存在的环境问题。随着工业化和城市化的快速发展, 大气污染日益严重, 严重影响了人类的生存环境和健康[1]。Li 等人得出结论, 由于中国的环境空气污染, 户外体育运动会带来很大健康风险[2][3]。根据《中国环境空气质量标准》(GB30952012), 用于监测空气质量的常规空气污染物有六种: 二氧化硫(SO₂)、二氧化氮(NO₂)、粒径小于 10 微米的颗粒物(PM₁₀)、粒径小于 2.5 微米的颗粒物(PM_{2.5})、臭氧(O₃)和一氧化碳(CO) [4]-[6]。这些污染物对人体健康有不利影响。国际能源署估计, 空气污染每年造成 650 万人过早死亡, 而长期暴露于细颗粒(如 PM_{2.5})或交通相关污染物等污染物与肺癌、冠心病和其他疾病的较高发病率有关[7][8]。因此, 空气质量预测的研究尤为重要, 被认为是保护环境的关键因素。空气质量监测、建模和准确预测对于清楚地了解未来的污染水平及其相关的健康风险非常重要。

在目前的研究中, 空气质量预测模型大致有三类: 黑盒模型、白盒模型和灰盒模型。Raimondo 等人[9]、Garcia 等人[10]和 Park 等人[11]使用神经网络建立模型来预测单个污染物的流行情况, 例如, 测量小于 10 微米的颗粒物(PM₁₀)。Raimondo 等使用支持向量机(SVM)和人工神经网络(ANN)对模型进行训练。Yu 等[12]提出了一种随机森林方法, 命名为 RAQ, 用于 AQI 类别预测。随后, Yi 等[13]应用深度神经网络进行 AQI 类别预测。Veljanovska 和 Dimoski [14]采用不同的设置来优于 k-最近邻(k-NN)、决策树和支持向量机来预测 AQI 水平。他们的人工神经网络模型达到了 92.3%的准确率, 优于所有其他经过测试的算法。但是以神经网络为代表的黑盒模型的内部参数和结构难以得到验证, 这类预测模型缺乏可解释性, 使得用户难以理解预测背后的原因。给实际应用带来了潜在的风险。为了得到可靠的预测结果, 需要从可解释的角度出发构建出合理的模型从而获得可靠的预测结果。为了解决这一问题, 研究人员正在开发一种混合模型, 这一模型旨在提供准确透明且具有可解释的预测, 使用户能根据模型的输出结果从而做出明智的决策。

灰盒模型结合了黑盒模型和白盒模型的优点, 很好地在准确性和可解释性之间取得了平衡。它可以像黑盒模型一样捕捉数据中的复杂关系, 同时也可以像白盒模型一样提供对决策过程的洞察。模型的体系结构和参数可以根据问题的领域知识进行调整, 从而提高了模型的准确性和可解释性。

BRB 作为一种典型的灰盒模型, 它可以整合不同来源的证据, 做出准确的预测和决策。其次, BRB 模型允许将专家知识纳入决策过程。这些专业知识可以帮助提高模型的准确性和可解释性。

2. RFI-HBRB 模型的问题描述

空气质量预测对我们的生存环境和身体健康至关重要, 因此建立有效准确且具有可解释性的预测模型直接关系到我们能否采取正确的措施应对空气污染带来的危害。本节给出了基于 BRB 的空气质量指数的问题描述。然后, 构建了 RFI-HBRB 的基本模型框架。

本文在构建 RFI-HBRB 空气质量指数预测模型时考虑了以下三个问题:

问题 1. 在实际预测中, 影响环境质量的指标较多, 例如 PM2.5、PM10 等, 然而有些指标的数据同时包含定量信息和定性信息, 有些指标对于空气质量的影响程度几乎可以忽略不计。若将所有指标全部作为评估指标, 则会造成计算量较大, 并且对于模型的最终预测准确率也会有影响。因此, 如何选取有效的评估指标对于模型的建立极其重要。指标的选取可以描述为:

$$O = \{a_1, a_2, \dots, a_n\} \quad (1)$$

$$O^* = S(O) \quad (2)$$

$$O^* = \{a_1, a_2, \dots, a_n\} \quad (3)$$

其中, O 表示原始特征集, $S(\)$ 表示筛选方法, O^* 表示筛选后得到的特征集。

问题 2. 在筛选出有效特征指标后, 未被筛选出的指标对预测结果仍然有影响, 如何处理这些特征指标使其在模型预测中发挥积极的作用是本文要解决的第二个问题。未被筛选出的特征指标处理可以描述为:

$$P = \{a_1, a_2, \dots, a_m\} \quad (4)$$

$$T = ER(P) \quad (5)$$

其中, P 表示筛选后的特征集, $ER(\)$ 表示证据推理方法, T 表示进行融合后的结果。

问题 3. 如何在前人提出的 BRB 可解释性一般准则的基础上总结出适用于预测模型的 BRB 可解释性准则。Cao 等人对 BRB 可解释性进行了全面的综述, 并提出了八个通用的标准。来指导可解释性 BRB 的建立。因此, 本文在一般准则的基础上提出可解释性准则, 如式(6):

Interpretability criteria:

$$\{C|C_1, C_2, \dots, C_n\} \quad (6)$$

其中, C 表示可解释的准则集, n 表示准则数。

最终的模型推理描述为式(7):

$$I = g\{O^*, C, EZ\} \quad (7)$$

其中, O^* 表示空气质量预测系统的输入数据。EZ 代表专家知识。 I 表示空气质量指数预测结果集合。 g 为非线性函数, 表示系统特征与预测值之间的关系。 $g\{\ }$ 表示空气质量预测模型。

3. 基于 RFI-HBRB 的空气质量预测方法

新提出的基于 RFI-HBRB 的模型结构如图 1 所示, 该模型由四个部分组成。第一部分是特征筛选, 采用随机森林(Random Forest)进行特征重要性排序, 选取对于结果影响较大的指标。第二部分是使用证据推理(ER)来融合多个特征指标的信息。第三部分是基于可解释的分层 BRB 推理方法。第四部分是 BRB 的优化训练算法。

3.1. 基于 Random Forest 的空气质量预测特征筛选

随机森林(Random Forest, RF)作为一种集成机器学习方法, 它利用随机重采样技术 bootstrap 和节点

随机分裂技术构建多颗决策树, 通过投票得到最终结果。随机森林具有分析复杂相互作用分类特征的能力, 对于噪声数据和存在缺失值的数据具有很好的鲁棒性, 并且具有较快的学习速度, 其变量重要性度量可以作为高维数据的特征选择工具, 近年来已经被广泛应用于各种分类、预测、特征选择以及异常点检测问题中。

在这一部分中, 使用随机森林特征选择算法对特征重要性进行排序, 从而选出最具有有效性的特征。特征重要性计算是 Random Forest 内嵌的功能, 以袋外(Out-Of-Bag, OOB)数据分类准确率的变量为评价准则。

假设有 bootstrap 样本 $k=1, 2, \dots, K$, K 表示训练样本的个数, 每个样本有 N 个特征, 特征重要性排序的计算步骤如下:

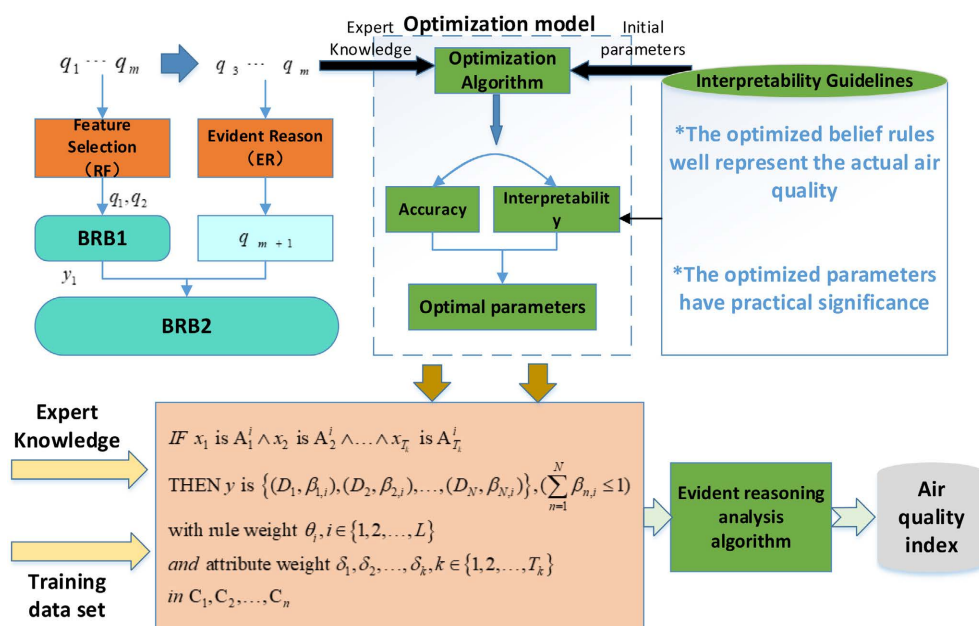


Figure 1. Air quality prediction method based on RFI-HBRB

图 1. 基于 RFI-HBRB 的空气质量预测方法

步骤 1: 初始化 $k=1$, 创建决策树 T_k ;

步骤 2: 对第 k 个 bootstrap 样本训练 T_k , 计算第 k 个 OOB 数据的分类准确率 L_k^{oob} 。

步骤 3: 随即对 OOB 数据集的特征 X_j , $j=1, 2, \dots, N$ 施加扰动, 再次计算 $L_{k,j}^{oob}$ 。

步骤 4: 对于 $k=2, 3, \dots, K$ 重复步骤 1~步骤 3。

步骤 5: 特征 X_j 的重要性度量 P_j 通过式(8)计算

$$P_j = \frac{1}{K} \sum_{k=1}^K (L_k^{oob} - L_{k,j}^{oob}) \quad (8)$$

步骤 6: 对 P_j 降序排列, 得到特征重要性排序。

3.2. 基于证据推理的空气质量指标评估模型

在空气质量预测中, 对于一些极端天气, 可能会大大提高数据获取的难度, 从而对空气质量预测带来很大的影响。另一方面, 如果只用随机森林筛选出的具有有效性指标进行预测, 而放弃未被筛选出的指标, 分析过程孤立且结论缺乏说服力。同时筛选指标会造成信息损失, 预测结果存在误差。因此, 本

文采用证据推理(ER)算法对筛选后余下的指标进行融合, 确保使用数据的完整性。其流程如图 2 所示。主要实现步骤如下:

步骤 1: 根据专家知识初始化指标中不同评估等级的置信度, 第 i 个指标可以描述为:

$$e_i = \{(\beta_r, \sigma_{r,i}), (\theta, \sigma_{\theta,i}), r=1 \cdots R, i=1 \cdots N\} \quad (9)$$

其中 β_r 表示第 r 个指标数据的评估级别, $\sigma_{r,i}$ 表示在证据 e_i 下被评估为等级 β_r 的置信度。 θ 表示全局无知, $\sigma_{\theta,i}$ 表示第 i 个指标中数据未分配的置信度。

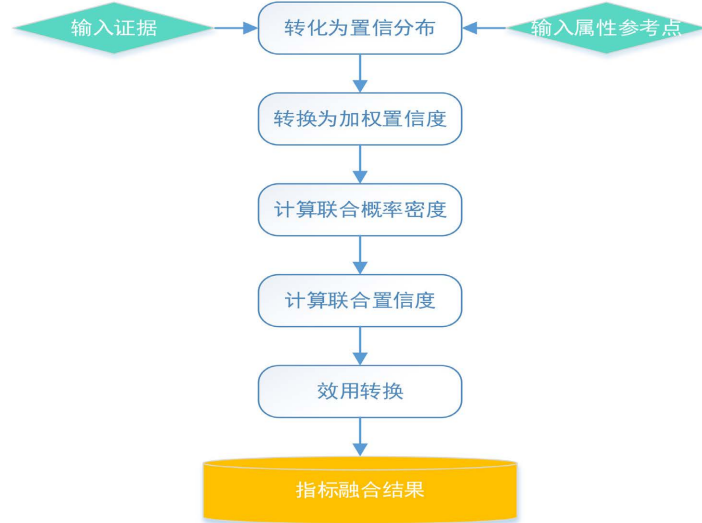


Figure 2. Flow chart of evidence reasoning
图 2. 证据推理流程图

步骤 2: 计算证据的基本概率质量, 如式(10~12)描述:

$$m_{r,i} = 1 - w_i \sigma_{r,i} \quad (10)$$

$$m_{\delta,i} = 1 - w_i \sum_{r=1}^R \sigma_{r,i} \quad (11)$$

$$m_{\epsilon,i} = 1 - w_i \quad (12)$$

其中, $m_{r,i}$ 表示的是第 r 个指标中第 i 个指标等级的基本概率质量, $m_{\delta,i}$ 代表未分配给数据等级集合的基本概率质量, $m_{\epsilon,i}$ 为缺少的第 i 个数据指标概率质量。

步骤 3: 基于 Dempster 规则求解组合概率质量:

$$m_{r,J(i+1)} = F_{J(i+1)} [m_{r,J(i)} m_{r,i+1} + m_{r,J(i)} m_{\theta,i+1} + m_{\theta,J(i)} m_{r,i+1}] \quad (13)$$

$$m_{\theta,J(i)} = m_{\theta,J(i)} + m_{u,J(i)} \quad (14)$$

$$m_{u,J(i+1)} = F_{J(i+1)} [m_{u,J(i)} m_{u,i+1} + m_{u,J(i)} m_{\theta,i+1} + m_{\theta,J(i)} m_{u,i+1}] \quad (15)$$

$$m_{L(\Theta),J(i+1)} = F_{J(i+1)} [m_{\theta,J(i)} m_{\theta,i+1}] \quad (16)$$

$$F_{J(i+1)} = \frac{1}{1 - \sum_{r=1}^R \sum_{s=1}^R m_{r,J(i)} m_{s,i+1}} \quad (17)$$

步骤 4: 根据上述结果计算联合置信度:

$$\sigma_r = \frac{m_{r,J(i)}}{1 - m_{L(\Theta)J(i)}} \quad (18)$$

步骤 5: 假设评估等级 β_r 的效用为 $u(\beta_r)$, 那么评估的期望效用计算如下:

$$x'_i = \sum_{r=1}^R u(\beta_r) \sigma_r \quad (19)$$

基于上述分析, 式(19)的结果能对影响空气质量的指标进行有效的评估, 某类指标的效用值越高, 代表该类属性对空气质量预测的影响越大, 验证了空气质量预测方案的有效性。

3.3. 模型构建

在实际问题中存在各种不确定因素, 特别是人类活动, 使得对于空气质量预测难度较大, 而 BRB 模型是一种具有专家系统和数据驱动模型特征的新型智能专家系统。BRB 是采用 Yang 等人提出的证据推理方法的规则库。在该方法中, 第条 IF-THEN 置信规则表示为式(20): 其中, $x_i (i=1 \cdots M)$ 是指空气质量预测评估指标; $A_i (i=1 \cdots M)$ 表示评估指标的参考值集合; $H_i (i=1 \cdots N)$ 是 RFI-HBRB 模型输出结果的参考等级; $\beta_i (i=1 \cdots N)$ 表示第 k 条置信规则下, 每个结果对应的置信度; $\theta_i (i=1, \cdots, K)$ 表示第 i 条置信规则的规则权重; K 表示置信规则的数量; $\delta_i (i=1 \cdots M)$ 表示第 i 个评估指标的属性权重; M 表示评估指标的数量。

$$\begin{aligned} R_k : & \text{if } x_1 \text{ is } A_1 \wedge x_2 \text{ is } A_2 \wedge \cdots \wedge x_M \text{ is } A_M \\ & \text{They result is } \{(H_1, \beta_1), (H_2, \beta_2), \cdots, (H_N, \beta_N)\} \\ & \text{with rule weight } \theta_1, \theta_2, \cdots, \theta_K \\ & \text{and attribute weight } \delta_1, \delta_2, \cdots, \delta_M \end{aligned} \quad (20)$$

可以看出, BRB 可以很好地处理定性和定量信息, 并且建模容易理解。但在 BRB 的应用中由于影响指标较多, 输入属性过多会导致规则爆炸。因此, 考虑到具有多重属性, 设计了分层结构, 这样的模型结构具有良好的可扩展性, 能够更好的应用于空气质量预测问题。

该方法将随机森林筛选出的属性作为输入, 并根据专家知识建立的 BRB 输出相应的结果, 然后继续作为下一层的输入值, 直到最后一层。然后考虑采用可解释准则的优化算法进行优化。最终在可解释模型下得到高精度的预测结果。其推理过程如下。

步骤 1: 初始化。通过专家知识构建初始化规则置信度。

步骤 2: 使用下列公式计算输入与参考值的匹配度:

$$a_i^s = \begin{cases} \frac{B_i^{n+1} - x_i}{B_i^{n+1} - B_i^n}, & s = n (A_i^n \leq x_i \leq A_i^{n+1}) \\ \frac{x_i - B_i^n}{B_i^{n+1} - B_i^n}, & s = n + 1 \\ 0, & s = 1, 2, \cdots, S (s \neq n, n + 1) \end{cases} \quad (21)$$

其中, a_i^s 表示第 i 个输入 x_i 对于参考值 A_i 的匹配度。

步骤 3: 计算规则激活权重, 公式如下:

$$\omega_k = \frac{\theta_k \prod_{i=1}^T (a_i^s)^{\delta_i}}{\sum_{l=1}^L \theta_l \prod_{i=1}^T (a_i^s)^{\delta_i}} \quad (22)$$

其中, ω_k 表示第 k 条规则的激活权重, θ_k 表示第 k 条规则的规则权重, δ_i 表示第 i 个属性权重, T 表示先行属性的数量。

步骤 4: 使用 ER 分析算法生成每个空气质量预测结果的置信度, 公式如下:

$$\beta = \frac{\mu \times \left[\prod_{k=1}^L (\omega_k \beta_{n,k}) + 1 - \omega_k \sum_{i=1}^N \beta_{i,k} - \prod_{k=1}^L 1 - \omega_k \sum_{i=1}^N \beta_{i,k} \right]}{1 - \mu \times \left[\prod_{k=1}^L 1 - \omega_k \right]} \quad (23)$$

$$\mu = \frac{1}{\left[\sum_{n=1}^N \prod_{k=1}^L \left(\omega_k \beta_{n,k} + 1 - \omega_k \sum_{i=1}^N \beta_{i,k} \right) - (N-1) \prod_{k=1}^L 1 - \omega_k \sum_{i=1}^N \beta_{i,k} \right]} \quad (24)$$

步骤 5: 最终输出为:

$$y_r = \sum_{n=1}^N D_n \beta_n \quad (25)$$

其中, y_r 表示模型的实际输出, β_n 表示相对于评价结果 D_n 的置信度。

3.4. 模型优化过程

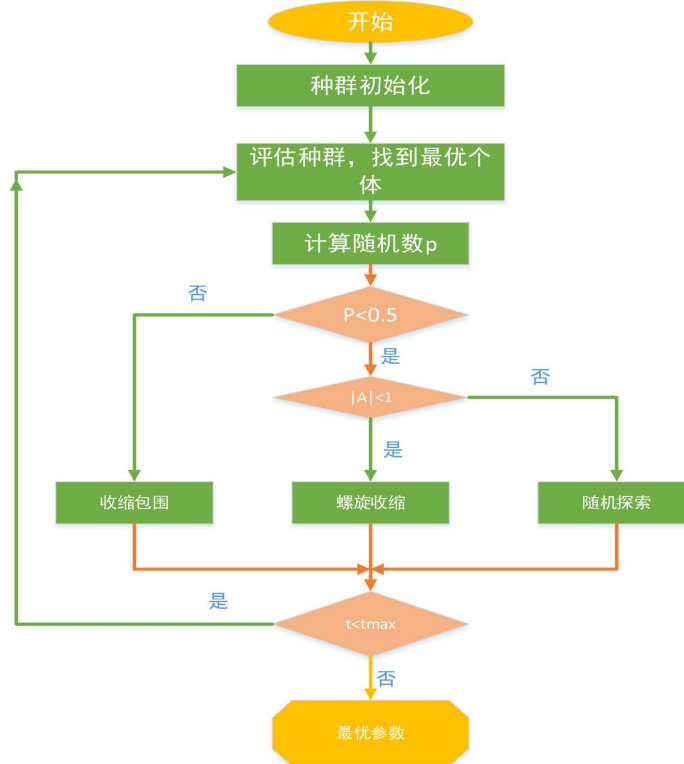


Figure 3. Flowchart of the whale optimization algorithm
图 3. 鲸鱼优化算法流程图

考虑到空气质量受到环境和人类活动的干扰以及专家知识的有限性影响, 模型难以准确预测空气质量
量的真实情况。因此, RFI-HBRB 的优化对提高模型的精度是至关重要的。鲸鱼优化算法(Whale Optimi-
zation Algorithm, WOA)是由 MIRIALILI 提出的一种模拟座头鲸捕食猎物行为的新型智能优化算法。该算
法将猎物的位置抽象为一个对应的最优解, 鲸鱼通过种群个体搜索, 收缩包围猎物来实现最终达到猎物
位置即最优解位置。由于鲸鱼算法原理简单, 参数设置少和具有较强的全局搜索能力的优点。算法的简
单步骤如图 3 所示。

4. 案例研究

4.1. 实验过程

实验主要包括数据下载、数据处理、初始模型建立、模型训练、结果对比分析等部分。空气质量预
测的实验过程如图 4 所示。

Table 1. Initial reference values
表 1. 初始参考值

	参考点	VL	L	M	H
BRB1	X1	2	56	186	380
	X2	6	49	129	560
BRB2	X3	0.22	0.65	0.83	0.98
	X4	15	49	129	440

Table 2. Initial rules for BRB
表 2. BRB 的初始规则

Rule No.	Indicators (x_1, x_2)	Rule Weight Constraint $\zeta_k \sim G_1$	Rule Weight ζ_k	Initial Belief $\{\alpha_1^k, \alpha_2^k, \alpha_3^k, \alpha_4^k\}$	The Belief Constraint $\{\alpha_1^k, \alpha_2^k, \alpha_3^k, \alpha_4^k\}$
1	VL ^{VL}	0.5-1	1	{0.9, 0.05, 0.05, 0}	{0.8-0.9, 0-0.1, 0-0.1, 0-0.1}
2	VL ^L	0.5-1	1	{0.75, 0.2, 0.05, 0}	{0.7-0.8, 0.2-0.3, 0-0.1, 0-0.1}
3	VL ^M	0.5-1	1	{0.3, 0.4, 0.2, 0.1}	{0.2-0.4, 0.4-0.5, 0.1-0.3, 0-0.1}
4	VL ^H	0.5-1	1	{0.05, 0.35, 0.3, 0.3}	{0-0.1, 0.1-0.2, 0.2-0.4, 0.3-0.5}
5	L ^{VL}	0.5-1	1	{0.9, 0.1, 0, 0}	{0.9-1, 0-0.1, 0-0.1, 0-0.1}
6	L ^L	0.5-1	1	{0.75, 0.15, 0.1, 0}	{0.7-0.8, 0.1-0.2, 0-0.1, 0-0.1}
7	L ^M	0.5-1	1	{0.25, 0.35, 0.3, 0.1}	{0.2-0.3, 0.3-0.4, 0.3-0.4, 0-0.1}
8	L ^H	0.5-1	1	{0.2, 0.3, 0.25, 0.25}	{0.1-0.3, 0.3-0.4, 0.2-0.3, 0.2-0.3}
9	M ^{VL}	0.5-1	1	{0.25, 0.35, 0.3, 0.1}	{0.2-0.3, 0.3-0.4, 0.2-0.4, 0-0.1}
10	M ^L	0.5-1	1	{0.55, 0.2, 0.15, 0.1}	{0.5-0.6, 0.1-0.3, 0.1-0.2, 0-0.1}
11	M ^M	0.5-1	1	{0.2, 0.25, 0.4, 0.15}	{0.1-0.2, 0.2-0.3, 0.4-0.5, 0.1-0.2}
12	M ^H	0.5-1	1	{0.05, 0.1, 0.3, 0.55}	{0-0.1, 0.1-0.2, 0.2-0.3, 0.5-0.6}
13	H ^{VL}	0.5-1	1	{0.05, 0.5, 0.25, 0.2}	{0-0.1, 0.4-0.6, 0.2-0.3, 0-0.2}
14	H ^L	0.5-1	1	{0.25, 0.5, 0.15, 0.1}	{0.2-0.3, 0.4-0.6, 0.1-0.2, 0.1-0.2}
15	H ^M	0.5-1	1	{0.05, 0.1, 0.2, 0.65}	{0-0.1, 0-0.1, 0.1-0.3, 0.6-0.7}
16	H ^H	0.5-1	1	{0, 0.1, 0.15, 0.75}	{0-0.1, 0-0.1, 0.1-0.2, 0.7-0.8}

根据专家知识的判断, 选择四个语义值来描述系统状态, 即“非常低”(VL)、“低”(L)、“中等”(M)和“高”(H)。BRB1 和 BRB2 的属性权重及初始参考值如表 1 所示。初始规则和初始规则权重如表 2 所示。



Figure 4. Experimental process of air quality prediction
图 4. 空气质量预测实验过程

4.2. 模型的实验步骤

Table 3. Optimized rules
表 3. 优化后的规则

Rule No.	Indicators (x_1, x_2)	Rule Weight ζ_k	The Optimized Belief $\{\alpha_1^k, \alpha_2^k, \alpha_3^k, \alpha_4^k\}$
1	VL ^{VL}	0.9522	{0.95, 0.05, 0, 0}
2	VL ^L	0.5621	{0.75, 0.25, 0, 0}
3	VL ^M	1	{0.3, 0.4, 0.2, 0.1}
4	VL ^H	1	{0.05, 0.35, 0.3, 0.3}
5	L ^{VL}	0.6427	{0.85, 0.15, 0, 0}
6	L ^L	0.9452	{0.7, 0.15, 0.1, 0.05}
7	L ^M	0.9022	{0.3, 0.35, 0.3, 0.05}
8	L ^H	1	{0.2, 0.3, 0.25, 0.25}
9	M ^{VL}	0.8385	{0.3, 0.35, 0.3, 0.05}
10	M ^L	0.6173	{0.6, 0.2, 0.15, 0.05}
11	M ^M	0.5509	{0.15, 0.25, 0.4, 0.2}
12	M ^H	0.9633	{0, 0.05, 0.3, 0.65}
13	H ^{VL}	1	{0.05, 0.5, 0.25, 0.2}
14	H ^L	0.8777	{0.2, 0.5, 0.2, 0.1}
15	H ^M	0.8273	{0., 0.1, 0.1, 0.8}
16	H ^H	0.9978	{0, 0, 0.15, 0.85}

首先分析数据集的激活权重，并标记不激活权重。图 5 显示 RFI-HBRB 各规则的激活权重。图中 3、4、8、13 未激活。这表明在有限的数据集中，这些规则对结果没有影响。因此，在优化过程中，如果需要调整相关参数，为避免破坏模型的可解释性，需要对这些规则进行标记和修正。

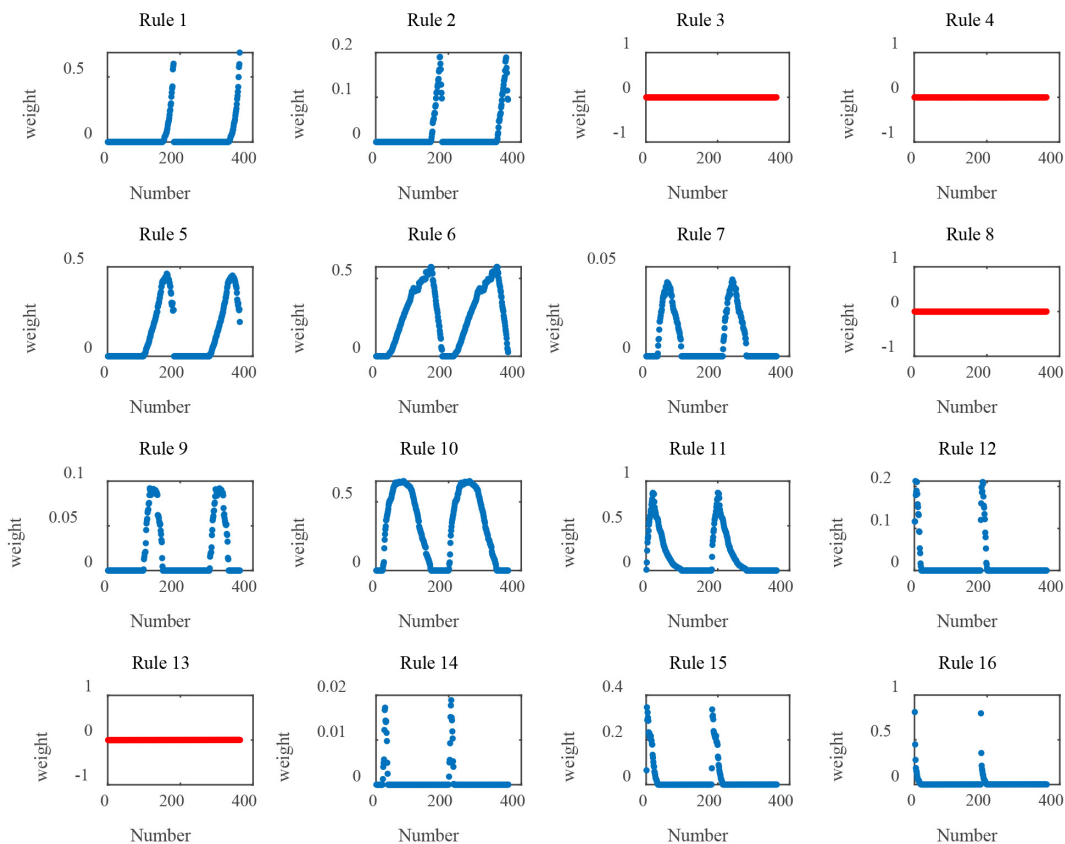


Figure 5. RFI-HBRB rule weight activation analysis
图 5. RFI-HBRB 规则权重激活分析

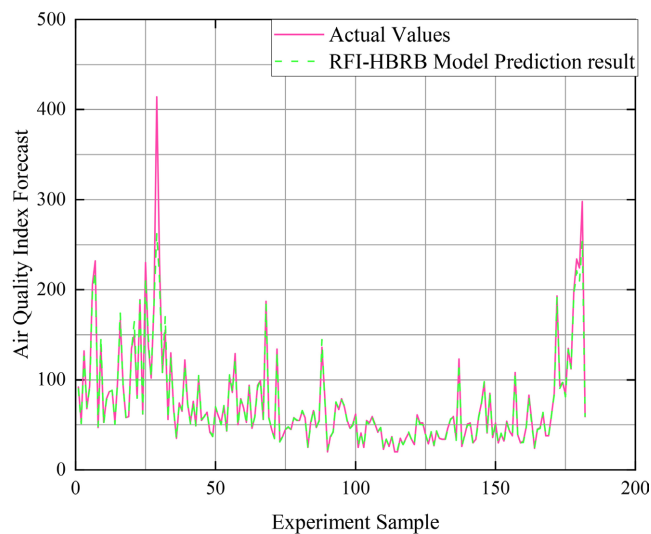


Figure 6. Predicted and true values of RFI-HBRB
图 6. RFI-HBRB 的预测值和真实值

初始参考值和最佳参考值的可行区间如表 1 和表 2 所示, 表 3 所示的优化规则显示了调整后优化算法的有效性。预测结果如图 6 所示。

4.3. 分析实验结果

为了更好地评估模型性能, 我们选择了 BP 神经网络(BP Neural Networks)、KNN、ELM、SVM、RF (Random Forest)和 K-means 六种算法进行对比试验。此外还比较了专家知识构建的 BRB0、原始 BRB1 和 RFI-HBRB。本文选择 MAE、RMSE 和 R^2 进行综合评价和模型比较。

4.4. 精度分析

Table 4. Comparison of RFI-HBRB and other model predictions
表 4. RFI-HBRB 和其他模型预测的比较

No.	Method	MAE	RMSE	R^2
1	RFI-HBRB	3.6229	11.4485	0.9854
2	BP neural works	9.9683	13.03	0.95271
3	KNN	10.72	16.63	0.9095
4	ELM	12.22	21.84	0.844
5	SVM	15.8	25.1	0.75
6	RF	12.5	19.9	0.87
7	K-means	12.54	16.56	0.938

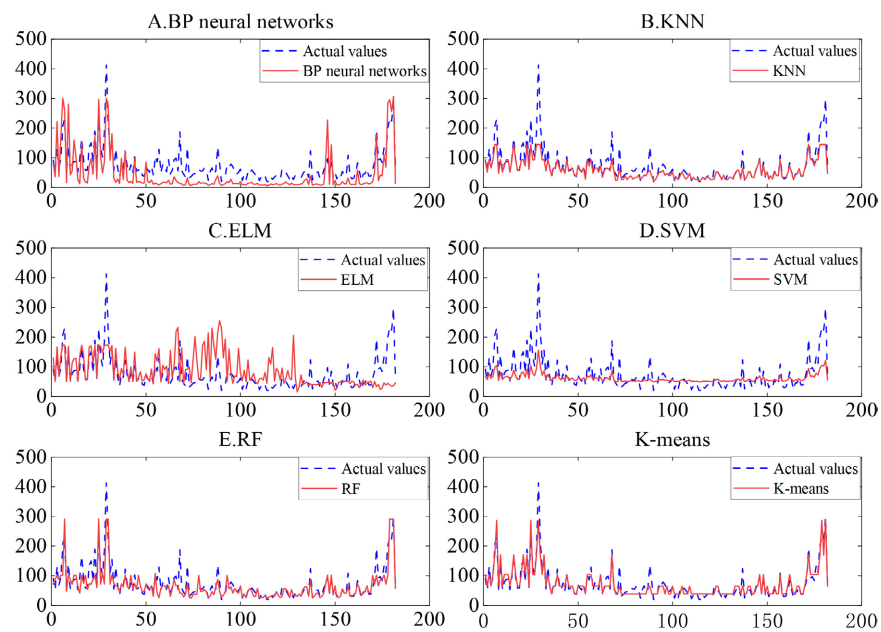


Figure 7. Comparison between other algorithms and real values
图 7. 其他算法和真实值比较

在精度分析中, 实验结果分为两部分进行描述。为了有效地对模型进行综合评价, 本节使用了 MAE、RMSE 和 R^2 作为评价指标。计算公式如式(26)。RFI-HBRB 结合 BP Neural Networks、KNN、ELM、SVM、RF 和 K-means。结果如表 4 所示。

$$MAE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i) \quad RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (\bar{y}_i - y_i)^2} \quad (26)$$

在第一部分, RFI-HBRB 的输出值和真实值如图 6 所示。如图中所示, 预测值和真实值拟合程度较好, 说明该模型对于空气质量预测的有效性。第二部分是 BP neural networks、KNN、ELM、SVM、RF 和 K-means 与真实值的比较。如图 7。

由此可以得出, 本文所使用的模型相较于其他模型拥有很高的预测准确率, 适用于空气质量预测领域。

5. 结论

考虑到空气质量在时间和空间上复杂和不断变化的性质, 准确预测空气质量是一项艰巨的任务。本文在 BRB 的基础上构建了 RFI-HBRB 模型, 为空气质量预测提供了一种高度准确和可信的方法。研究人员已经证明了 BRB 专家系统预测未来空气质量的能力。RFI-HBRB 模型增强了 BRB 的可解释性, 显示了 RFI-HBRB 在空气质量预测中的优势。最终结果表明, RFI-HBRB 模型在提高可解释性的同时, MAE 稳定在 2~4 之间, 基本达到了初始 BRB 的精度。这一成就不仅强调了置信规则库和专家知识在预测空气质量方面的有效性, 而且强调了通过创新方法解决环境问题的潜力。

总之, 我们的研究突出 RFI-HBRB 模型在空气质量预测方面的意义, 并证明了它优于传统的深度学习算法。从我们的研究获得的见解具有普遍适用性, 为解决全球范围内的空气质量问题提供了机会。通过置信规则库和专家系统的协同作用, 我们为可持续和更健康的未来铺平了道路。

参考文献

- [1] Li, X., Jin, L. and Kan, H. (2019) Air Pollution: A Global Problem Needs Local Fixes. *Nature*, **570**, 437-439. <https://doi.org/10.1038/d41586-019-01960-7>
- [2] Han, Y., Lam, J.C.K. and Li, V.O.K. (2018) A Bayesian LSTM Model to Evaluate the Effects of Air Pollution Control Regulations in China. 2018 *IEEE International Conference on Big Data (Big Data)*, Seattle, 10-13 December 2018, 4465-4468. <https://doi.org/10.1109/bigdata.2018.8622417>
- [3] Bai, L., Wang, J., Ma, X. and Lu, H. (2018) Air Pollution Forecasts: An Overview. *International Journal of Environmental Research and Public Health*, **15**, Article 780. <https://doi.org/10.3390/ijerph15040780>
- [4] Ding, Y. and Xue, Y. (2019) A Deep Learning Approach to Writer Identification Using Inertial Sensor Data of Air-handwriting. *IEICE Transactions on Information and Systems*, **102**, 2059-2063. <https://doi.org/10.1587/transinf.2019edl8070>
- [5] Dotse, S., Petra, M.I., Dagar, L. and De Silva, L.C. (2018) Application of Computational Intelligence Techniques to Forecast Daily PM 10 Exceedances in Brunei Darussalam. *Atmospheric Pollution Research*, **9**, 358-368. <https://doi.org/10.1016/j.apr.2017.11.004>
- [6] Jia, M., Komeily, A., Wang, Y. and Srinivasan, R.S. (2019) Adopting Internet of Things for the Development of Smart Buildings: A Review of Enabling Technologies and Applications. *Automation in Construction*, **101**, 111-126. <https://doi.org/10.1016/j.autcon.2019.01.023>
- [7] Abirami, S., Chitra, P., Madhumitha, R. and Ragul Kesavan, S. (2020) Hybrid Spatio-Temporal Deep Learning Framework for Particulate Matter (PM_{2.5}) Concentration Forecasting. 2020 *International Conference on Innovative Trends in Information Technology (ICITIIT)*, Kottayam, 13-14 February 2020, 1-6. <https://doi.org/10.1109/icitit49094.2020.9071548>
- [8] Cheng, Y., Zhang, S., Huan, C., Oladokun, M.O. and Lin, Z. (2019) Optimization on Fresh Outdoor Air Ratio of Air Conditioning System with Stratum Ventilation for Both Targeted Indoor Air Quality and Maximal Energy Saving. *Building and Environment*, **147**, 11-22. <https://doi.org/10.1016/j.buildenv.2018.10.009>
- [9] Raimondo, G., Montuori, A., Moniaci, W., Pasero, E. and Almkvist, E. (2007) A Machine Learning Tool to Forecast PM10 Level. *Proceedings of the Fifth Conference on Artificial Intelligence Applications to Environmental Science*, San Antonio, 14-18 January 2007, 1-9.
- [10] Garcia, J.M., Teodoro, F., Cerdeira, R., Coelho, L.M.R., Kumar, P. and Carvalho, M.G. (2016) Developing a

- Methodology to Predict PM₁₀ Concentrations in Urban Areas Using Generalized Linear Models. *Environmental Technology*, **37**, 2316-2325. <https://doi.org/10.1080/09593330.2016.1149228>
- [11] Park, S., Kim, M., Kim, M., Namgung, H., Kim, K., Cho, K.H., *et al.* (2018) Predicting PM10 Concentration in Seoul Metropolitan Subway Stations Using Artificial Neural Network (ANN). *Journal of Hazardous Materials*, **341**, 75-82. <https://doi.org/10.1016/j.jhazmat.2017.07.050>
 - [12] Yu, R., Yang, Y., Yang, L., Han, G. and Move, O. (2016) RAQ—A Random Forest Approach for Predicting Air Quality in Urban Sensing Systems. *Sensors*, **16**, Article 86. <https://doi.org/10.3390/s16010086>
 - [13] Yi, X., Zhang, J., Wang, Z., Li, T. and Zheng, Y. (2018) Deep Distributed Fusion Network for Air Quality Prediction. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, London, 19-23 August 2018, 965-973. <https://doi.org/10.1145/3219819.3219822>
 - [14] Veljanovska, K. and Dimoski, A. (2018) Air Quality Index Prediction Using Simple Machine Learning Algorithms. *International Journal of Emerging Trends & Technology in Computer Science*, **7**, 25-30.