

基于敏感参数发现的区域重点污染物浓度预测^①



潘欣玉, 郑 亮

(中国石油大学(华东) 计算机科学与技术学院, 青岛 266580)

通讯作者: 潘欣玉, E-mail: pxy15653289289@163.com

摘 要: 污染物浓度变化趋势对于环境监测工作意义重大. 现今各种前馈神经网络预测模型的输出结果仅与当前输入有关, 无法研究污染物数据前后依赖关系. 且多种污染物具有相同排放源, 污染物间往往存在潜在关联关系, 一种污染物的变化可能反映另一种污染物变化, 所以在预测中需考虑其他敏感参数的影响. 针对上述两个问题, 提出一种基于敏感参数发现的区域重点污染物浓度预测方法. 应用关联规则算法及多元回归分析挖掘出各污染物的敏感参数, 构建多变量 LSTM 预测模型, 将待预测污染物及其敏感参数作为预测模型特征变量, 进行污染物的浓度预测. 实验结果表明本文方法可有效预测污染物浓度变化趋势, 预测效果优于未经关系发现的 LSTM 模型.

关键词: 污染物浓度预测; 关联规则算法; 多元回归分析; LSTM 预测模型

引用格式: 潘欣玉, 郑亮. 基于敏感参数发现的区域重点污染物浓度预测. 计算机系统应用, 2021, 30(10):202-209. <http://www.c-s-a.org.cn/1003-3254/8088.html>

Regional Key Pollutant Concentration Prediction Based on Sensitive Parameter Discovery

PAN Xin-Yu, ZHENG Liang

(College of Computer Science and Technology, China University of Petroleum, Qingdao 266580, China)

Abstract: The variation trend in pollutant concentration is of great significance for environmental monitoring. At present, the output of various feedforward neural network prediction models is only related to the current input, so it is impossible to study the dependence of pollutant data before and after. Multiple pollutants have the same emission source, and there is a potential correlation between pollutants. The change in one pollutant may reflect that in another pollutant, so the influence of other sensitive parameters should be considered in the prediction. To solve the above two problems, this study proposes a regional key pollutant concentration prediction method based on sensitive parameter discovery. The association rule algorithm and multiple regression analysis are used to mine the sensitive parameters of each pollutant, and a multivariable LSTM prediction model is constructed. The pollutants to be predicted and their sensitive parameters are taken as the characteristic variables of the model to predict the pollutant concentration. The experimental results show that the proposed method can reliably predict the variation trend in pollutant concentration and it performs better than the LSTM model without relation discovery.

Key words: pollutant concentration prediction; association rule algorithm; multiple regression analysis; LSTM prediction model

近年来, 空气质量成为人们越来越关心的问题, 大气污染物浓度预测对于预防地区污染具有重要的意义. 区域内的污染物浓度预测, 可以对区域空气质量的微变

化及时做出响应. 污染物的预警对于区域可持续发展具有十分重要的作用, 可为政府决策制定、环境监管^[1-3] 执行提供技术支持, 因此研究污染物预测工作具

① 收稿时间: 2020-12-26; 修改时间: 2021-01-25; 采用时间: 2021-01-27

有重要意义。

现阶段对于污染物预测的研究包含数值模拟及统计计算两部分。数值模拟^[4,5]方法充分研究污染物在大气中的迁移变化,预测的精确度较高。但在预测过程中需要收集复杂的物化数据,计算过程冗杂,且需应用气象、物化等多学科知识,预测工作较为繁琐;统计计算为污染物预测工作提供了广泛思路,可应用各种神经网络模型进行污染物预测工作^[6,7],通过对神经网络模型的构建及统计计算可以实现简易精确的预测。然而当前神经网络预测模型多为前馈型神经网络,其输出结果仅与当前输入有关,无法研究污染物数据前后依赖关系,因而需构建一种可以对前后依赖序列数据进行预测的模型。且由于多种污染物具有相同排放源,所以多个污染物间往往存在潜在关联关系,使得一种污染物浓度的变化可反映出其他污染物浓度变化情况,以往预测方法中,无论数值模拟还是统计计算均未考虑应用污染物间关联关系进行预测工作。

针对上述两个问题,本文充分考虑 LSTM 神经网络在处理时间序列数据前后依赖关系的优势,应用待预测污染物及其敏感参数之间的关联关系,实现区域污染物的浓度预测工作。本文将关联规则算法^[8,9]与多元回归分析^[10,11]结合应用,发现待预测污染物的敏感参数,在此基础上构建 LSTM 模型,将待预测污染物及其敏感参数,全部作为 LSTM 预测模型的输入变量,从而更加准确的实现对各种重点污染物浓度的预测。

1 相关工作

当前污染物浓度分析预测方法主要有机理类与非机理类两个方向。机理类方法中数值模拟^[1,2]涉及复杂物化流程,需要建立完整的污染物扩散过程相关物化模型^[12],将污染物的排放、扩散、迁移、转化过程使用复杂的偏微分方程来解析,通过展示模式得到空气污染物浓度的空间分布及变化趋势。

非机理类方法对污染物排放的复杂机理过程不做过多要求,主要是基于历史数据应用统计学方法构建神经网络预测模型^[3,4]。孙宝磊等^[13,14]建立 BP 预测模型对污染物的日均浓度进行预测,通过对模拟输出值和期望输出值误差的迭代计算调整,使 BP 神经网络预测输出不断逼近期望输出,直至误差满足要求。方彦^[15]构建 RBF 神经网络模型,基于历史数据作为训练样本构建模型并以同期数据作为测试样本,并选取 2018 年

同期数据作为测试样本对区域污染物进行预测。郭晓君等^[16,17]基于 GM(1,1) 模型、Verhulst 模型和 SCGM(1,1) 模型建立组合灰色预测模型,运用预测有效度方法确定组合预测模型的权重系数,利用模型预测污染物排放量,通过灰色系统 GM(1,1) 残差修正模型对大气污染物浓度进行预测。支持向量机在大气污染物浓度预测方面也有应用^[18-21],在对各类影响因子进行分析的基础上进行建模预测,利用 PCA 方法对输入因子降维形成支持向量机的训练样本集,建立大气污染预测模型。Haewon 等通过构建聚类回归模型^[22-24]进行污染物浓度预测,根据气象要素与污染物浓度数据资料,分析天气变化规律,挖掘若干种天气类型,并分析各种天气类型的典型参数,将这些参数与相应的环境质量实测数据通过回归分析^[25-27]建立定量关系,根据定量关系进行污染物浓度预测。杜续等^[28]针对神经网络算法在当前污染物浓度预测领域存在的易过拟合、网络结构复杂、学习效率低等问题,引入随机森林回归算法,分析气象条件、大气污染物浓度和季节所包含的多项特征因素,通过调整参数的最优组合,设计出一种新的污染物浓度预测模型—RFRP 模型。马天成等^[29]采用一种改进型 PSO 优化的模糊神经网络,将粒子群算法与模糊神经网络进行融合,发挥 PSO 算法全局寻优的特点,预测污染物浓度的变化规律。

在污染物预测领域,机理类方法预测过程较为复杂,需要综合污染物迁移过程中各种物化因素,因此机理类方法通用性较低。在以往非机理类模型中,前馈型神经网络未考虑时间序列依赖关系。以上两种方法均未考虑污染物间关联关系的影响,准确性有待提高。因此,本文充分考虑污染物间的具有关联关系的特点,以此为基础建立基于敏感参数发现的 LSTM 预测模型^[30-33]的污染物预测模型,应用待预测污染物及其敏感参数的关联关系来提高模型的预测精确度。

2 区域重点污染物浓度预测方法

在现实情况中,由于污染源并非排放单一类型的污染物,多种污染物具有相同排放源,污染物间往往存在潜在关系,一种污染物的变化可能反映另一种污染物变化,所以在污染物浓度预测工作中可应用污染物间的关系进行预测,即需考虑待预测污染物敏感参数这一影响因素。基于上述问题,本文将污染物预测工作分为两阶段进行。

(1) 基于关联规则与多元回归分析的敏感参数发现方法. 在该方法中, 首先应用关联规则算法对污染物数据集进行关联关系发现, 初步挖掘出各污染物的敏感参数. 再应用多元回归分析实现各污染物间相关性分析, 将各污染物的相关性分析结果与关联性分析结果取交集, 保留二者共同部分, 保证敏感参数发现的完备性与准确性, 并通过对比试验验证了敏感参数的有效性. 最终实现各污染物敏感参数的发现, 为后续污染物预测工作的开展奠定基础.

(2) 基于敏感参数发现的污染物预测方法. 由于污染物数据存在时间序列上的前后依赖关系, 将待预测污染物及其敏感参数作为预测模型特征变量, 实现污染物浓度预测工作. 在该部分工作中, 通过对现有的 LSTM 预测模型的改进完成污染物浓度预测工作.

2.1 基于关联规则的重点污染物关系发现

本文采用关联规则算法进行污染物间关联性关系发现. 针对污染物参数数据, 首先对污染物参数值分区间处理, 依据污染物数值最小值及最大值对参数数据三等分. 例如 SO_2 用 S 表示, 分为 S_1 、 S_2 、 S_3 . S_1 表示 SO_2 浓度很小, S_2 表示 SO_2 浓度中等, S_3 表示 SO_2 浓度大. O_3 用 O 表示, NO_2 用 N 表示, CO 用 C 表示, $\text{PM}_{2.5}$ 用 P 表示, PM_{10} 用 M 表示. 其中, 参数程度词的等级越高代表的值越大.

(1) 首先基于式 (1) 计算每个参数程度词集的支持度.

$$S(l) = \frac{N_l}{N} \times 100\% \quad (1)$$

其中, $S(l)$ 表示参数程度词 l 的支持度, N_l 表示参数程度词 l 在所有污染物参数数据中出现的总次数, N 表示污染物参数数据集的数据量, 支持度反映了污染物程度词 l 在数据集中出现的频次, 支持度越大表示程度词 l 出现得越频繁. 此处设置一个支持度阈值 S_T , 支持度大于支持度阈值 S_T 的程度词 l 集合记为一项频繁项集 L_1 .

(2) 不断迭代使用上一次得到的 $(k-1)$ 项频繁项集, 计算频繁项集中参数程度词的支持度, 保留满足支持度阈值 S_T 的参数程度集, 产生新的候选 k 项频繁项集 L_k , 直到没有新的频繁项集产生.

(3) 获取 k 项频繁项集 L_k 的关联规则 $L_s \rightarrow L_{k-s}$, 其中, L_s 表示由 s 个参数程度词组成的程度词集, L_{k-s} 表示去掉 L_k 中 s 个参数程度词剩余的程度词组成的程度词集.

(4) 计算每个候选关联规则的置信度, 置信度的计

算公式如式 (2).

$$C(X \rightarrow Y) = \frac{N_{XY}}{N_X} \times 100\% \quad (2)$$

其中, X 与 Y 表示两个程度词集, $C(X \rightarrow Y)$ 表示程度词集 X 到程度词集 Y 的置信度, N_{XY} 表示程度词集 X 与程度词集 Y 同时在全体数据集中出现的次数, N_X 表示程度词集 X 在全体数据集中出现的次数. 置信度以条件概率的方式反映了两个程度词集同时出现的概率, 置信度越高表示两个程度词集同时出现的概率越高.

(5) 设置一个置信度阈值 C_T , 当一个项集内所有子集之间的置信度都满足置信度阈值 C_T 时, 则认为该项集内的所有程度词之间具有较强的关联性, 并留下这样的项集, 保留到关联规则集中.

2.2 基于多元回归分析的污染物相关性分析

针对污染物指标数据, 应用回归分析中的相关系数法获取各个污染物指标之间的相关系数, 发现各个污染物指标之间的相关性程度, 综合污染物间关联规则进一步发现污染物间关系.

回归分析使用相关系数来度量污染物指标之间线性相关关系. 其定义如式 (3) 所示.

$$V = \frac{n \sum xy - \sum x \sum y}{n \left(\sum x^2 - (\sum x)^2 \right) \left(\sum y^2 - (\sum y)^2 \right)} \quad (3)$$

求得两污染物指标相关系数后, 应用 T 检验判断自变量 X 中影响因子对因变量 Y 的影响显著性程度, T 统计量计算公式如下:

$$t = \frac{\bar{x}_{l1} - \bar{x}_{l2}}{\sqrt{\frac{(\sigma_{xi})^2 + (\sigma_{yi})^2 - 2\gamma\sigma_{xi}\sigma_{yi}}{n-1}}} \quad (4)$$

其中, \bar{x}_{l1} 、 \bar{x}_{l2} 分别为两污染物指标样本的平均数, σ_{xi} 、 σ_{yi} 分别为两污染物指标样本的方差, γ 为相关指标样本的相关系数.

根据 t 检验表, 查得显著水平 α 的临界值 t_α . 若 $t > t_\alpha$, 则拒绝原假设, 说明两指标间具有较强相关性; 否则, 说明两指标间相关性不强.

2.3 基于 LSTM 的重点污染物预测模型

本文在进行污染物预测研究时充分考虑污染物敏感参数及时间序列关系, 通过对 LSTM 预测模型进行改进, 实现污染物预测工作. 本文预测工作的改进主要有以下两点:

(1) 模型预测参数及输入特征的选择. 在模型构建之前引入参数 *choose_target* 来进行待预测污染物的选择, *choose_target* 取值为 0-5, 分别代表 SO₂、NO₂、PM₁₀、PM_{2.5}、O₃、CO 六种污染物, 选择不同参数值即对不同污染物预测, 例如 *choose_target*=3 则代表此时的待预测污染物为 PM_{2.5}. 设置参数 *needed_target* 来进行待预测污染物输入特征的选择, *needed_target* 取值同样为 0-5. 例如 *needed_target* = [0, 1, 2] 则代表此时预测模型输入特征为 SO₂、NO₂、PM₁₀.

(2) 本模型在预测过程中采用了滚动预测的方式. 采用“早停止”策略判断模型收敛性, 首先将污染物输入特征数据输入到模型进行浓度预测, 得出未来一天的污染浓度数据并记录; 然后将实际输出结果加入下一条样本的污染物浓度特征中, 与其他敏感参数特征共同作为训练样本再次预测未来一天的污染浓度数据, 依次类推, 得到预测结果.

基于 LSTM 重点污染物预测模型框架如图 1 所示.

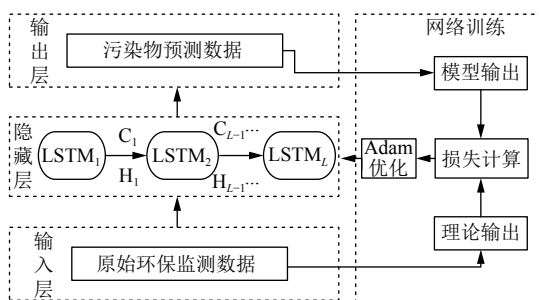


图 1 基于 LSTM 主要污染物预测模型框架

本文使用 Keras 深度框架构建 LSTM 模型. 预测模型的构建流程包括以下 4 个步骤.

(1) 构建模型

本预测模型包含一个 LSTM 神经网络层和一个全连接层, 两层之间顺序连接. 首先设置 LSTM 神经网络层: 一种污染物具有 n 个敏感参数, 则训练集特征数为 $n+1$, 因此模型的输入层特征维度设置为 $n+1$, 并设置时间步长及循环神经网络结构内部 Sigmoid 层和 tanh 层的神经元数量, 最后设置预测模型输出层, 用于输出污染物预测结果.

(2) 模型参数设置

本模型采用均方误差作为模型的损失函数, 采用 Adam 优化算法作为模型优化器, 在模型训练过程中设置每次迭代训练的样本数据量、训练轮数、模型学习率.

(3) 模型预测过程

模型采用滚动预测的方式, 首先将污染物输入特征数据输入到模型进行浓度预测, 得出未来一天的污染浓度数据并记录; 然后将实际输出结果加入下一条样本的污染物浓度特征中, 与其他敏感参数特征共同作为训练样本再次预测未来一天的污染浓度数据, 依次类推, 得到预测结果.

(4) 模型评价

用式 (5) 计算各污染物指标预测数据与真实数据间均方根误差 RMSE.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \hat{x}_i}{x_i} \right)^2} \quad (5)$$

其中, x_i 为第 i 个数据点污染物指标浓度的实测值, \hat{x}_i 为第 i 个数据点污染物指标浓度的预测值, n 为数据长度.

2.4 基于敏感参数发现的污染物浓度预测方法描述

(1) 对原始环保监测数据进行数据清洗、去除无用数据、数据标准化等操作得到预处理后的数据集.

(2) 扫描预处理后环保监测数据集, 得到满足支持度阈值 S_T 的 k 项频繁项集, 基于 k 项频繁项集组合得到候选关联规则集, 将所有满足置信度阈值 C_T 的污染物间关联规则加入关联规则集.

(3) 再次扫描环保监测数据集, 得到污染物两两之间的相关系数, 保留相关系数大于 0.5 的污染物关系, 将各污染物的相关性分析结果与关联性分析结果取交集, 保留二者共同部分, 实现污染物的敏感参数发现工作.

(4) 综合 (2)(3) 所发现的各污染物的敏感参数, 与待预测污染物及其敏感参数作为预测模型输入变量, 实现对污染物的浓度预测.

3 实验结果与分析

3.1 实验数据

本实验使用区域内 8 个环保监测点所得的环保监测数据, 实验数据共有 209 340 条数据.

3.2 数据预处理

本实验数据预处理操作包含数据清洗及数据标准化. 首先对数据进行数据清洗, 去除异常数据点及空白数据、重复数据. 其次, 进行数据标准化操作, 针对清洗后数据采用式 (6) 进行标准化操作.

$$z = \frac{x - \mu}{\sigma} \quad (6)$$

其中, x 为数据清洗后的环保监测数据, μ 为所有样本数据的均值, σ 为所有样本数据的标准差。

3.3 重点污染物关系发现

3.3.1 污染物间关联分析

设置支持度 5%, 置信度 60%, 通过污染物指标间关联规则挖掘, 得到污染物指标间的关联规则集, 可初步挖掘出污染物间的关联关系, 结果如表 1 所示。

3.3.2 污染物间相关性分析

为进一步发现污染物指标间的相关性, 对污染物指标进行相关性分析, 结果如表 2 所示。

当相关系数大于 0.5 则两种指标相关性较强, 综合表 1 与表 2 结果, 污染物间关系发现结果如表 3 所示。

表 1 重点污染物关联规则结果表

| 关联规则结果 | 出现次数及支持度 |
|--------------------------|------------------------------|
| ('O1', 'N2') | (8478, 0.1280065226253567) |
| ('C2', 'P2') | (8046, 0.12148389726865064) |
| ('P1', 'M2') | (8219, 0.1240959671452945) |
| ('C2', 'N3', 'O2', 'P3') | (8002, 0.12081955579713427) |
| ('C1', 'N2', 'P2') | (6693, 0.10105539701952258) |
| ('C2', 'M2') | (12766, 0.10547930727302925) |
| ('C3', 'N2') | (7047, 0.10640032613126783) |
| ('C1', 'N2', 'O2') | (7608, 0.11487067989310142) |
| ('O1', 'M2') | (7224, 0.10907279068714046) |

表 2 重点污染物相关系数表

| 污染物 | SO ₂ | NO ₂ | CO | O ₃ | PM ₁₀ | PM _{2.5} |
|-------------------|-----------------|-----------------|-----------|----------------|------------------|-------------------|
| SO ₂ | 1.000000 | 0.410310 | 0.490976 | -0.133552 | 0.408360 | 0.443194 |
| NO ₂ | 0.410310 | 1.000000 | 0.578502 | -0.518942 | 0.489390 | 0.547488 |
| CO | 0.490976 | 0.578502 | 1.000000 | -0.27499 | 0.599933 | 0.731420 |
| O ₃ | -0.133552 | -0.518942 | -0.274990 | 1.000000 | -0.133187 | -0.199924 |
| PM ₁₀ | 0.408360 | 0.42548 | 0.599933 | -0.133187 | 1.000000 | 0.829098 |
| PM _{2.5} | 0.443194 | 0.423837 | 0.731420 | -0.199924 | 0.829098 | 1.000000 |

表 3 重点污染物关系表

| 污染物 | 强相关污染物 |
|-------------------|--|
| NO ₂ | CO、O ₃ |
| CO | NO ₂ 、PM ₁₀ 、PM _{2.5} |
| O ₃ | NO ₂ |
| PM ₁₀ | CO、PM _{2.5} |
| PM _{2.5} | CO、PM ₁₀ |

3.4 重点污染物预测结果及评估

3.4.1 基于敏感参数发现的 LSTM 污染物预测

针对关联规则算法与多元回归分析挖掘的污染物关系, 预测模型构建时, 对 SO₂ 的预测输入变量设置为 SO₂; 对 NO₂ 的预测输入变量为 NO₂、CO、O₃; 对 CO 的预测输入变量为 CO、NO₂、PM₁₀、PM_{2.5}; 对 O₃ 的预测输入变量为 O₃、NO₂; 对 PM₁₀ 的预测输入变量为 PM₁₀、CO、PM_{2.5}; 对 PM_{2.5} 的预测输入变量为 PM_{2.5}、CO、PM₁₀。将数据预处理后的数据集中的后 600 条数据作为测试集数据, 其余作为训练集数据, 并将训练集中 10% 的数据作为验证集数据。采用前 24 个时刻的污染物浓度数据对下一时刻污染物浓度数据进行预测, 时间步长设置为 24, 每次迭代训练的样本数据量设置为 96, 将学习率设置为 0.0001, 训练轮数次

数设置为 50, 隐藏层神经元数量设置为 256。

如果模型两次迭代之间损失值变化已经很小, 那么说明模型收敛, 可以结束训练。因此, 为验证本文模型的收敛性, 引入“早停止”策略。当验证集与训练集上模型损失已不再变化时, 无论是否达到所设置的迭代次数, 均停止模型训练。

本文模型构建之初迭代次数设置为 50 次, 但是由图 2 所知, 当模型迭代次数为 16 次时, 模型已达收敛, 所以此时停止模型训练, 以减少模型损耗提高训练性能。

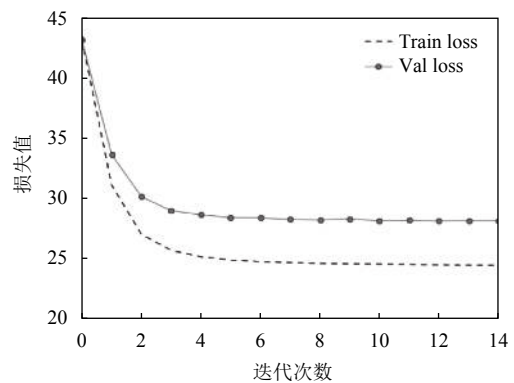


图 2 模型收敛性判别图

为检验和说明本文提出模型的预测性能,将本文污染物预测模型与未经敏感参数发现的 LSTM 模型进行对比预测.对比结果如图 3-图 12 所示.

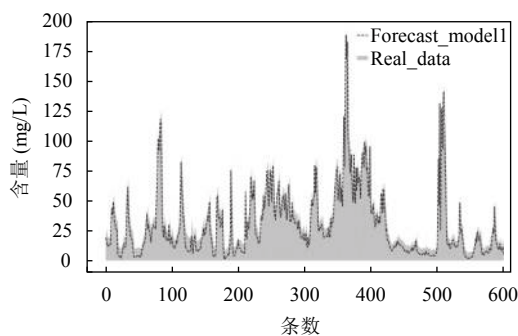


图 3 本文 SO₂ 预测结果拟合图

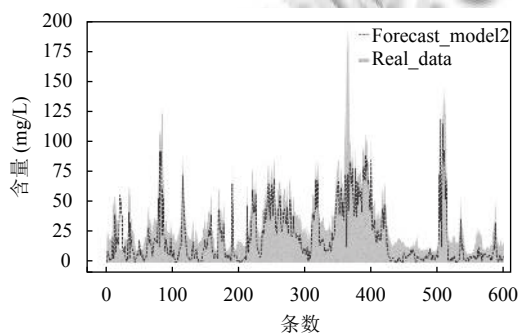


图 4 未经敏感参数发现 SO₂ 预测结果拟合图

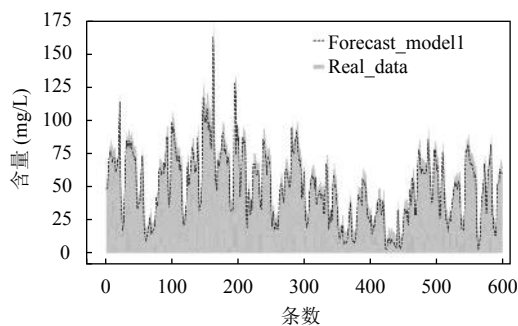


图 5 本文 NO₂ 预测结果拟合图

图 3-图 12 中, Forecast-model1 为经敏感参数发现的 LSTM 预测模型结果; Forecast-model2 为未经敏感参数发现的预测模型结果; Real_data 为污染物真实值.可以看出,本文模型预测效果优于未经敏感参数发现的对比模型,能够较好地跟踪各种污染物浓度变化趋势,并响应其波动变化,实现相对精确的预测.

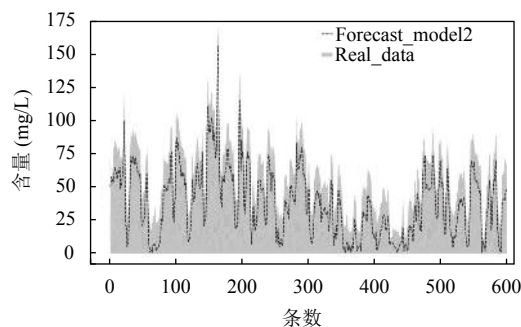


图 6 未经敏感参数发现 NO₂ 预测结果拟合图

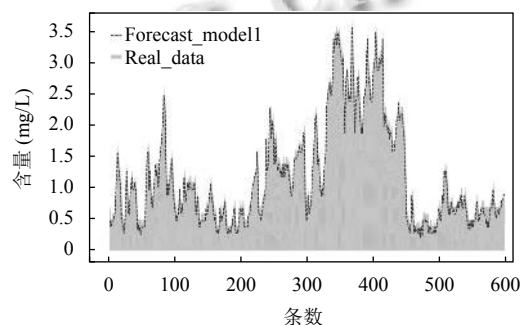


图 7 本文 CO 预测结果拟合图

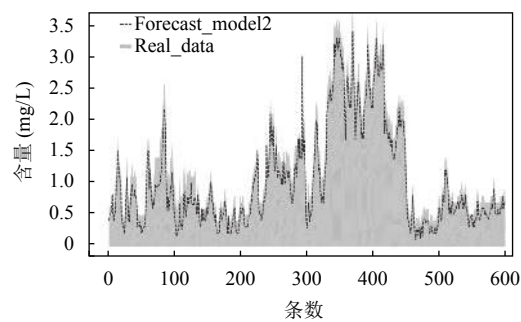


图 8 未经敏感参数发现 CO 预测结果拟合图

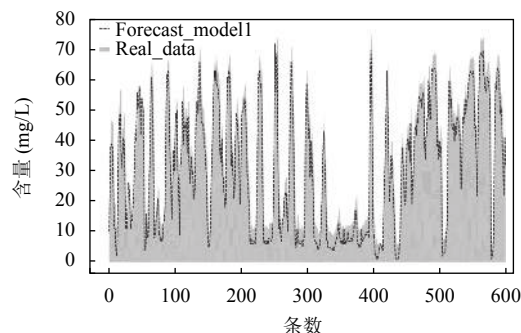


图 9 本文 O₃ 预测结果拟合图

3.4.2 污染物预测对比评估结果

根据 SO₂、NO₂、CO、O₃、PM₁₀、PM_{2.5} 六种污

染物预测值与真实值情况,两种污染物预测模型误差对比情况如表4所示。

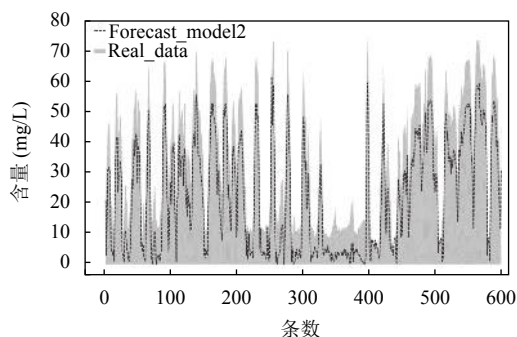


图10 未经敏感参数发现 O₃ 预测结果拟合图

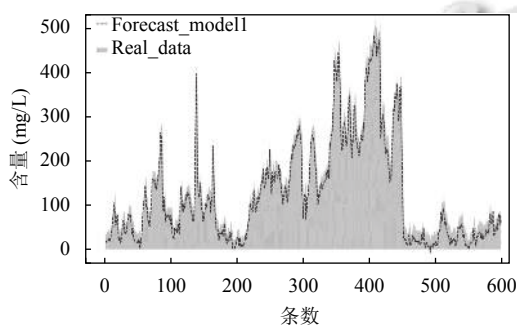


图11 本文 PM₁₀ 预测结果拟合图

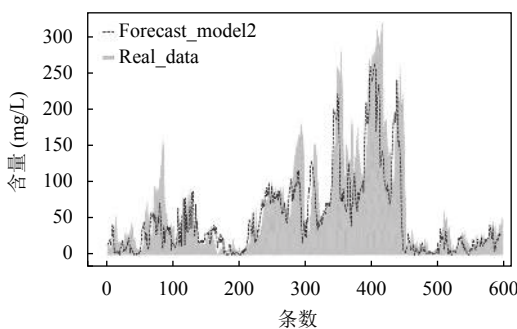


图12 未经敏感参数发现 PM₁₀ 预测结果拟合图

由图3-图14及表4所知,经污染物敏感参数发现的预测模型拟合效果与预测误差均优于对比模型.由此,证实了本文预测模型预测效果优于未经敏感参数发现的 LSTM 模型。

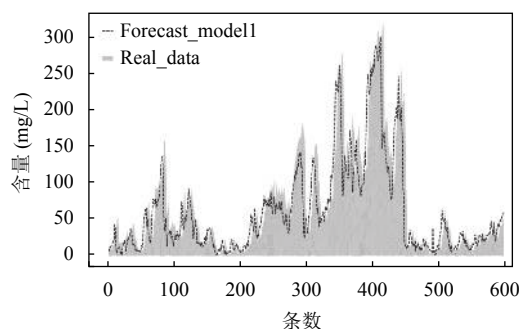


图13 本文 PM_{2.5} 预测结果拟合图

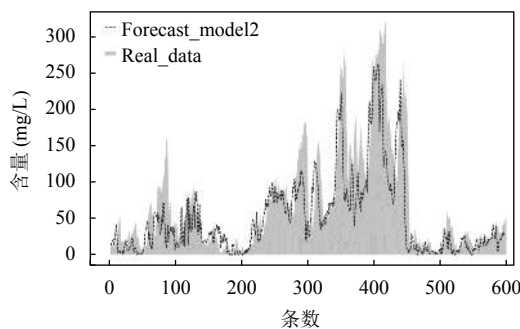


图14 未经敏感参数发现 PM_{2.5} 预测结果拟合图

4 结语

本文提出了一种基于敏感参数发现的区域重点污染物浓度预测方法.该方法主要包含3个模块,即数据预处理模块、重点污染物敏感参数发现模块、污染物预测模块.其中,在数据预处理模块主要是针对将原始数据进行数据清洗、数据标准化操作.在重点污染物敏感参数发现模块对环保监测数据应用关联规则算法、多元回归分析挖掘出各污染物的强相关污染物,为预测模型构建时输入变量的设置提供依据.在污染物预测模块构建 LSTM 污染物预测模型,基于挖掘所得各污染物间的敏感参数设置输入变量,使用均方根误差评估模型性能,证实本文模型较之以往未经污染物间关系发现的 LSTM 模型精确度较高.可有效实现环境污染预测完成由“滞后”到“超前”的转变.本文虽已实现重点污染物的预测,但是重点污染物的追踪工作同样是环保监测的重点工作,因此,下一步将研究重点污染物的溯源追踪工作,即依据污染物历史数据及各行业企业排污数据查清重点污染物的排放源头,从而

表4 模型均方根误差评估表

| 预测变量 | 平均绝对百分比误差 | | 均方根误差 | |
|-------------------|-----------|------------|------------|------------|
| | 本文模型 | 未经敏感参数发现模型 | 本文模型 | 未经敏感参数发现模型 |
| SO ₂ | 14.264783 | 31.846392 | 11.2338478 | 28.3729383 |
| NO ₂ | 10.483285 | 29.937285 | 14.2832338 | 33.3932837 |
| CO | 9.4829458 | 19.938564 | 16.2182782 | 54.2377678 |
| O ₃ | 16.947372 | 56.937562 | 9.2367262 | 20.1278731 |
| PM ₁₀ | 14.583756 | 33.185630 | 14.9237627 | 33.2972382 |
| PM _{2.5} | 18.563028 | 55.937583 | 10.3728371 | 26.7872362 |

与污染物预测共同为环保监测工作提供技术支撑。

参考文献

- 1 马莉娟, 付强, 姚雅伟. 我国环境监测方法标准体系的现状与发展构想. 中国环境监测, 2018, 34(5): 30-35.
- 2 詹志明, 尹文君. 环保大数据及其在环境污染防治管理创新中的应用. 环境保护, 2016, 44(6): 44-48.
- 3 徐敏, 孙海林. 从“数字环保”到“智慧环保”. 环境监测管理和技术, 2011, 23(4): 5-7, 26. [doi: 10.3969/j.issn.1006-2009.2011.04.002]
- 4 熊亚军, 徐敬, 孙兆彬, 等. 基于数据挖掘算法和数值模拟技术的大气污染减排效果评估. 环境科学学报, 2019, 39(1): 116-125.
- 5 卢晓华. 基于数值模拟的企业地下水重金属污染的环境影响预测评价. 安全与环境工程, 2014, 21(1): 93-97. [doi: 10.3969/j.issn.1671-1556.2014.01.018]
- 6 Ramyar S, Kianfar F. Forecasting crude oil prices: A comparison between artificial neural networks and vector autoregressive models. Computational Economics, 2019, 53(2): 743-761. [doi: 10.1007/s10614-017-9764-7]
- 7 Gao Y, Liu XD, Li XL, et al. A prediction approach on energy consumption for public buildings using mind evolutionary algorithm and BP neural network. Proceedings of 2018 IEEE 7th Data Driven Control and Learning Systems Conference. Enshi: IEEE, 2018. 385-389.
- 8 肖文, 胡娟, 周晓峰. 基于 MapReduce 计算模型的并行关联规则挖掘算法研究综述. 计算机应用研究, 2018, 35(1): 13-23. [doi: 10.3969/j.issn.1001-3695.2018.01.003]
- 9 刘莉萍, 章新友, 牛晓录, 等. 基于 Spark 的并行关联规则挖掘算法研究综述. 计算机工程与应用, 2019, 55(9): 1-9. [doi: 10.3778/j.issn.1002-8331.1811-0425]
- 10 Han XJ, Liu YL, Yang HY. A stereo matching algorithm guided by multiple linear regression. Journal of Computer-Aided Design & Computer Graphics, 2019, 31(1): 84-93.
- 11 刘畅, 马韵洁, 黄翔. 一种基于多元线性回归算法的车流量预测模型研究. 电子技术与软件工程, 2016, (21): 175.
- 12 王永权, 徐方勤, 燕彩蓉, 等. 融合气象参数及污染物浓度的空气质量预测方法. 计算机应用与软件, 2018, 35(8): 113-117. [doi: 10.3969/j.issn.1000-386x.2018.08.020]
- 13 孙宝磊, 孙嵩, 张朝能, 等. 基于 BP 神经网络的大气污染物浓度预测. 环境科学学报, 2017, 37(5): 1864-1871.
- 14 夏杰, 吴文青, 许海洋. 基于主成分分析的 BP 神经网络在水华预测中的应用. 世界生态学, 2018, 7(2): 53-60.
- 15 方彦. 基于灰色 RBF 神经网络的空气质量预测. 中国科技信息, 2018, (22): 100-102. [doi: 10.3969/j.issn.1001-8972.2018.22.038]
- 16 郭晓君, 刘思峰, 吴利丰. 污染物减排预测的灰色 Markov 组合模型与算法. 计算机应用研究, 2013, 30(12): 3670-3673. [doi: 10.3969/j.issn.1001-3695.2013.12.039]
- 17 张峰, 殷秀清, 董会忠. 组合灰色预测模型应用于山东省碳排放预测. 环境工程, 2015, 33(2): 147-152.
- 18 Shaban KB, Kadri A, Rezk E. Urban air pollution monitoring system with forecasting models. IEEE Sensors Journal, 2016, 16(8): 2598-2606. [doi: 10.1109/JSEN.2016.2514378]
- 19 Moazami S, Noori R, Amiri BJ, et al. Reliable prediction of carbon monoxide using developed support vector machine. Atmospheric Pollution Research, 2016, 7(3): 412-418. [doi: 10.1016/j.apr.2015.10.022]
- 20 付学良, 杨洋, 李纯子. 基于混沌—支持向量机的大气污染物浓度预测模型. 电子世界, 2013, (4): 78-79.
- 21 Liu GQ, Gao XD, You DY, et al. Prediction of high power laser welding status based on PCA and SVM classification of multiple sensors. Journal of Intelligent Manufacturing, 2019, 30(2): 821-832. [doi: 10.1007/s10845-016-1286-y]
- 22 Byeon H. A laryngeal disorders prediction model based on cluster analysis and regression analysis. Medicine, 2019, 98(31): e16686. [doi: 10.1097/MD.00000000000016686]
- 23 Punzo A, McNicholas PD. Robust clustering in regression analysis via the contaminated Gaussian cluster-weighted model. Journal of Classification, 2017, 34(2): 349-293.
- 24 沈劲, 钟流举, 何芳芳, 等. 基于聚类与多元回归的空气质量预报模型开发. 环境科学与技术, 2015, 38(2): 63-66.
- 25 付倩妮. 基于多元线性回归的雾霾预测方法研究. 计算机科学, 2016, 43(S1): 526-528.
- 26 白婷, 文继荣, 赵鑫, 等. 基于迭代回归树模型的跨平台长尾商品购买行为预测. 中文信息学报, 2017, 31(5): 185-193. [doi: 10.3969/j.issn.1003-0077.2017.05.026]
- 27 姚宏亮, 马晓琴, 王浩, 等. 基于形态特征与因果岭回归的股市态势预测算法. 计算机工程, 2016, 42(2): 175-183. [doi: 10.3969/j.issn.1000-3428.2016.02.032]
- 28 杜续, 冯景瑜, 吕少卿, 等. 基于随机森林回归分析的 PM_{2.5} 浓度预测模型. 电信科学, 2017, 33(7): 66-75.
- 29 马天成, 刘大铭, 李雪洁, 等. 基于改进型 PSO 的模糊神经网络 PM_{2.5} 浓度预测. 计算机工程与设计, 2014, 35(9): 3258-3262. [doi: 10.3969/j.issn.1000-7024.2014.09.054]
- 30 Selvin S, Vinayakumar R, Gopalakrishnan EA, et al. Stock price prediction using LSTM, RNN and CNN-sliding window model. Proceedings of 2017 International Conference on Advances in Computing, Communications and Informatics. Udipi: IEEE, 2017. 1643-1647.
- 31 Liu J, Wang G, Duan LY, et al. Skeleton-based human action recognition with global context-aware attention LSTM networks. IEEE Transactions on Image Processing, 2018, 27(4): 1586-1599. [doi: 10.1109/TIP.2017.2785279]
- 32 Qing X, Niu YG. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. Energy, 2018, 148: 461-468. [doi: 10.1016/j.energy.2018.01.177]
- 33 Karim F, Majumdar S, Darabi H, et al. Multivariate LSTM-FCNs for time series classification. Neural Networks, 2019, 116: 237-245. [doi: 10.1016/j.neunet.2019.04.014]