

基于 Gradient Boosting 算法的 ERMS 辐射数据预测^①



朱武峰¹, 王廷银¹, 林明贵², 苏伟达¹, 李汪彪¹, 吴允平^{1,3,4}

¹(福建师范大学 光电与信息工程学院, 福州 350007)

²(福建省辐射环境监督站, 福州 350013)

³(数字福建环境监测物联网实验室, 福州 350117)

⁴(福建省光电传感应用工程技术研究中心, 福州 350000)

通讯作者: 吴允平, E-mail: wyp@fjnu.edu.cn

摘要: 影响核辐射监测站点辐射监测 HPIC 剂量率实时数据准确性的组成因素多且复杂, 如自然因素的降雨、温湿度、风向及太阳辐射等, 客观因素的设备异常及放射性状况等; 以致在实际应用中发现辐射监测状态异常时, 很难分析出是什么原因导致的监测数据偏离. 结合 ERMS 海量历史辐射序列监测数据, 深入挖掘降雨、温湿度、气压、风向、太阳辐射天顶方向电子量及周边各站点辐射数值等特征因子集, 基于 Gradient Boosting 算法 (简称 GB 算法) 建立起 HPIC 剂量率辐射数据的在线预测模型, 有效融合自然特征因子, 降低了自然因子对 HPIC 剂量率辐射监测数值异常的分析及判读的干扰作用, 提高了对 ERMS 辐射异常发现的辅助判断能力及维保效率.

关键词: HPIC 剂量率; 异常; ERMS; GB 算法; 预测

引用格式: 朱武峰, 王廷银, 林明贵, 苏伟达, 李汪彪, 吴允平. 基于 Gradient Boosting 算法的 ERMS 辐射数据预测. 计算机系统应用, 2019, 28(11): 37-44. <http://www.c-s-a.org.cn/1003-3254/7146.html>

Prediction of ERMS Radiation Data Based on Gradient Boosting Algorithm

ZHU Wu-Feng¹, WANG Ting-Yin¹, LIN Ming-Gui², SU Wei-Da¹, LI Wang-Biao¹, WU Yun-Ping^{1,3,4}

¹(College of Photonic and Electronic Engineering, Fujian Normal University, Fuzhou 350007, China)

²(Fujian Radiation Environment Supervision Station, Fuzhou 350013, China)

³(Digital Fujian Environmental Monitoring IoT Laboratory, Fuzhou 350117, China)

⁴(Fujian Provincial Engineering Technology Research Center of Photoelectric Sensing Application, Fuzhou 350007, China)

Abstract: The factors affecting the accuracy of real-time data of HPIC dose rate in nuclear radiation monitoring stations are complex, such as natural factors of rainfall, temperature and humidity, wind direction and solar radiation, objective factors of equipment anomalies and radioactivity, etc. When it is found that the radiation monitoring state is abnormal, it is difficult to analyze the cause of the deviation of the monitoring data. Combined with the monitoring data of massive historical radiation series of ERMS, the characteristics of rainfall, temperature and humidity, air pressure, wind direction, electrons in the zenith direction of solar radiation and the radiation values of surrounding sites are deeply explored. HPIC is established based on the Gradient Boosting algorithm (referred to as GB algorithm). The online prediction model of dose rate radiation data effectively combines the natural characteristic factors, reduces the natural factor's analysis of the HPIC dose rate radiation monitoring numerical anomaly and the interference effect of interpretation, and improves the auxiliary judgment ability and maintenance efficiency of ERMS radiation abnormality discovery.

Key words: HPIC dose rate; abnormal; ERMS; GB algorithm; prediction

① 基金项目: 国家自然科学基金海峡联合基金 (U1805263); 福建省自然科学基金 (2019J01427); 福建省引导性项目 (2019H0009)

Foundation item: Joint Fund for the Promotion of Scientific Cooperation across the Taiwan Straits of National Natural Science Foundation of China (U1805263); Natural Science Foundation of Fujian Province (2019J01427); Guiding Project of Fujian Province (2019H0009)

收稿时间: 2019-04-12; 修改时间: 2019-05-08; 采用时间: 2019-05-16; csa 在线出版时间: 2019-11-06

引言

核能发电不仅经济高效、排放无污染,而且原料来源广泛,是当今社会重要的电力能源之一^[1],通常,核电站会围绕自身建立辐射监测系统(Radiation Monitoring System, RMS),以保障运行安全^[2];其中,环境辐射监测系统(ERMS)是RMS的一个重要组成部分;所谓ERMS,是为了保证核电站外围环境安全,在核电站四周若干公里范围内设置若干个监测站点,每个监测站点都会配备监测仪器和通信装置,以便通过网络把数据传到计算机.其系统数据采集框架如图1所示,监测站采集设备主要分为辐射数据采集设备(如NaI能谱探测仪和高压电离室)和气象数据采集设备(如雨量计、温度传感器、风速传感器等),采集的数据种类主要有HPIC剂量率、雨量、气温、湿度及风速风向等;其中,HPIC剂量率是指示辐射监测站点实时 γ 辐射空气吸收剂量率的重要指标.

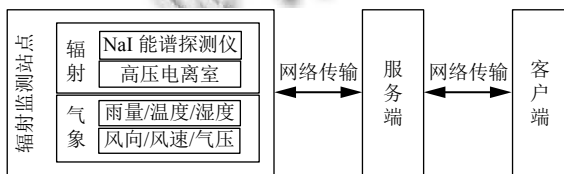


图1 ERMS系统数据采集系统框架图

这些数据蕴含着核电站运行安全性的重要价值指示信息,一直以来都是各国保障核电安全的研究热点,具有重大的研究意义.

国外,2015年,Chen等人^[3]针对 γ 辐射剂量率的仿真模拟问题,以蒙特卡罗方法为基础提出一种可用于进行曝光剂量估计的光谱测定G(E)函数方法,用于进行 γ 剂量率的仿真模拟,获得的仿真值与电离室测量的 γ 剂量率实际值之间最大偏差仅6.31%,对确保 γ 剂量率的可靠性具有很大应用价值.2017年,印尼国家核安全局的Susila等人^[4]则对塞彭核设施周围5年的序列辐射数据应用关联分析方法进行分析,发现 γ 辐射剂量率的数值与空气中碘和氙的放射性同位素含量具有较强的正相关关系;同年,Bossew等人^[5]指出空气中的天然放射性元素氡与 γ 辐射剂量率也有较强关系,采用统计关联方法对欧洲地区两者之间进行相关性研究,得出大部分地区氡对 γ 辐射剂量率的监测结果的贡献值低于5 nSv/h,部分地区则达到30 nSv/h,这对加强核辐射环境监测的预警能力具有很好的参考意义.

国内,2015年,朱耀明、林明贵等人^[6]提出要加强对

ERMS数据应用能力和人员配置管理问题,指出对ERMS管理维护上要安排专业工作人员,以便可以立时应对不同紧急情况,如当出现辐射数据异常或多个站点 γ 辐射剂量率数值超高,触发阈值报警时,确保工作人员可以及时收到预警信息并迅速进行处理,查找问题根源以寻求解决.2017年,高泽泉等人^[7]则应用线性统计方法对序列辐射数据中降雨与 γ 辐射剂量率的相关性进行分析发现,降雨天气时,降雨会导致 γ 辐射剂量率的升高,造成 γ 辐射剂量率的实时数值的不准确性,这对我们在日常监测中进行 γ 辐射剂量率数值的判断是一个很好的参考指标.2018年,罗敦焯等人^[8]也基于线性统计方法、关联分析及可视化技术对日常监测中 γ 辐射剂量率的特征影响因子进行了更全面的挖掘分析,总结了众多与 γ 辐射剂量率数据相关的特征影响因子,如宇宙射线、自然放射性物质、温湿度、风向及气压等气象因素,这对建立和完善 γ 辐射剂量率数值可靠性评价指标具有很大意义.

总的来说,在ERMS日常监测过程中, γ 辐射监测数值的影响因素较多,如上述的源相关的放射性物质、降雨、温湿度、风向及气压等自然因素,还有设备老化故障等都会导致 γ 辐射监测数据的不准确性.近年,我国物联网技术愈发成熟,数据资源获取能力逐步增强,围绕ERMS的设备可靠性、数据可信度以及源相关性等方面取了一些进展,但ERMS辐射数据分析却仍以实时数据判定、事后报警为主;围绕 γ 剂量率监测也进行相关影响因子的定性分析,提供了对实时 γ 剂量率监测准确性的辅助判断依据,但是在一定程度上对各影响因子融合进行数据价值挖掘未作出过多研究,如在自然因子影响下,如何有效识别和降低自然因素干扰,提高对HPIC剂量率的可靠性评估能力.而当今数据挖掘在医疗卫生^[9]、网络安全^[10]、企业管理^[11]、城市交通^[12]及工业生产^[13]等诸多领域都取得了显著的应用成果,为我们提升智能化监管效率和实现新的监管技术创新指出了新思路.其中以GB回归为代表的的人工智能算法,在解决回归问题上发挥了巨大的优势.2013年,山东大学陈爽爽等^[14]人应用GB算法对癫痫及复发概率数据进行建模,取得很好的检测效果,达到了初步临床实验的标准;2015年,瑞典艾滋病研究团队为预测戒烟成功率和艾滋病复发率,采用GB算法模型进行回归分析,取得很好的效果,具有很强的实用性^[15];而纵观我们的辐射监测数据、太阳活

动数据及气象数据,整体为时间序列的离散值,以HPIC剂量率值作为 γ 剂量率监测的重要指标,特征数据有类别特征(如风向)、离散值(如雨量、各监测站点HPIC剂量率历史数据、温湿度及天顶方向电子量VETC等),其总体数据特征完全符合GB算法模型,而且对模型误差,我们不排除其他因素干扰,充分考虑作为回归模型输入特征,可以一定程度排除未知因素干扰.GB算法对帮助我们解决ERMS中HPIC剂量率在线预测的问题是一个较佳选择.

1 GB 算法理论

GB算法^[16]是一种机器学习方法,其算法的核心思想在于:将损失函数看作模型的“靠谱程度”,当损失函数数值较大时,说明模型的可信度较低,预测结果的准确率较差.因此,我们通常会根据起始的损失函数,进行损失函数的优化工作,通常做法是根据梯度下降法来实现损失函数在梯度方向上的不断迭代减小直至收敛,此时模型最优,残差也达到最小值,残差通常认为就是目标实际值与模型预测值的误差.基于GB算法进行回归预测大致流程如图2所示,首先基于训练集建立一个基模型,然后将这个模型的残差作为下一个模型的优化学习目标输入,得到新的基模型,不断重复此迭代过程,直到模型的残差达到理想数值范围内.

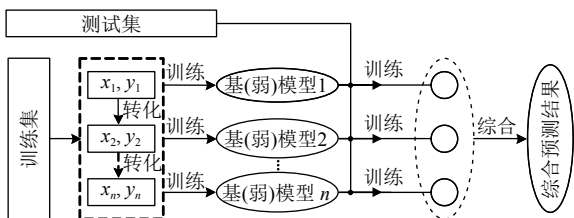


图2 Gradient Boosting 回归模型预测流程图

算法基本形式可表示如下:

算法的输入数据集是一组属性值 $x = \{x_1, x_2, \dots, x_m\}$ 及实际值 $y = \{y_1, y_2, \dots, y_m\}$, 设定与之间具有某种回归关系:

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_mx_m \quad (1)$$

设定损失函数为 $L(y, f(x))$, 基函数是 $\{a(x; \gamma)\}$. 首先, 设定初始化预测函数为:

$$f_0(x) = 0 \quad (2)$$

其次, 设置算法最大迭代次数 N , 在设定的迭代次数基础上:

(1) 极小化算法的损失函数, 使用如下方式求得使损失函数最小化的最佳参数 b_n, γ_n :

$$(b_n, \gamma_n) = \arg \min \sum_{i=1}^n L(y_i, f_{n-1}(x_i) + ba(x_i; \gamma)) \quad (3)$$

(2) 迭代更新预测函数, 在旧的弱学习器基础上不断学习新的弱学习器(决策树)模型 $f_n(x)$ 来优化模型损失函数, 如下式:

$$f_n(x) = f_{n-1}(x) + b_n a(x; \gamma_n) \quad (4)$$

(3) 累加预测函数, 求得最终强学习器预测函数公式如下:

$$f(x) = \sum_{n=1}^N b_n a(x, \gamma_n) \quad (5)$$

GB算法是一种集成学习算法. 该算法基于多个弱学习器组成强学习器的思想, 可以很好地克服算法建模中存在的诸如过拟合问题等, 在解决分类回归问题上具有很好的适用性.

2 GB 算法在辐射监测数据挖掘的应用

2.1 环境辐射监测系统概况

如图3所示为福建省某S核电站ERMS辐射监测站点空间分布图, 以此核电站为中心, 在周围设置了11个监测子站; 从图3中可以看出, S_1-S_{11} 号自动站围绕核电站整体呈现放射状分布, 保证充分检测四周环境放射性核素剂量, 同时实施多点连续监测, 在一定程度上可以更好地防止突发情况发生, 提高安全效率, 这样就形成了区域核辐射外围网络分布.

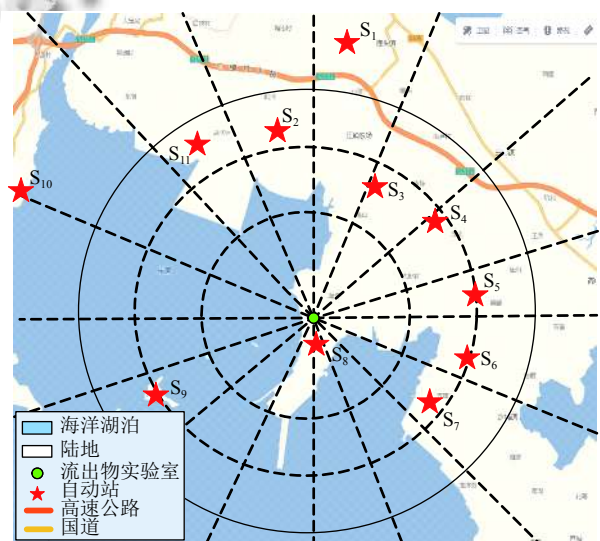


图3 某核电站 ERMS 辐射监测站空间分布图

2.2 数据来源

数据资料主要分 3 部分: (1) 辐射监测子站环境日常监测数据可以从某核电站辐射监测中心平台数据库直接导出 2015–2017 年文本格式历史辐射监测数据;

(2) 所需太阳日常活动历史数据从空间环境预报中心直接获取, 以对方提供的 2015–2017 年文本格式天顶方向电子量 VTEC 数据为主; (3) 气象数据指标可从气象局网站通过网页记录获取, 相关数据指标如表 1.

表 1 数据来源

种类	辐射数据				气象数据		太阳活动数据	
指标名称	气温	气压	雨量	湿度	HPIC 剂量率	风速	风向	天顶方向电子量 VTEC
数据来源	平台数据库					气象局网站		空间环境预报中心

2.3 实验方案

2.3.1 实验环境

硬件方面, 本实验运用 8 核、16 GB 内存 PC 机进行.

软件方面, 在 Windows10 系统环境下配置 JDK1.8、Anaconda3.4 及 MySql 等, 以 Mysql 作为文件存储数据库进行数据存储, 以 Anaconda3.4 自带软件 Spyder 作为实验编程平台, 以 Python 作为程序编程语言, 借助 Python 第三方科学计算相关类库 (如 numpy、pandas、matplotlib 及 sklearn 等) 进行整体数据挖掘分析工作; 具体实验环境配置如表 2 所示.

表 2 实验环境配置信息

项目	版本配置	项目	版本配置
操作系统版本	Windows10	硬件配置	8 核、16 GB 内存
Anaconda 版本	Anaconda3.4	scikit-learn 版本	scikit-learn0.19.1
JDK 版本	JDK1.8.102	pandas 版本	pandas0.23.1
Mysql 版本	Mysql5.7	numpy 版本	numpy1.14.5
SQLyog	SQLyog12.3.3	matplotlib 版本	matplotlib2.2.2

2.3.2 模型设计方案

如图 4 所示为本实验模型构建整体实验过程方案, 主要包括数据存储、数据预处理、数据抽取、数据建模及模型应用五部分. 首先, 实验数据主要为 3.2 中阐述的历史离线数据集, 数据特征为时间序列的离散型数据, 所以采用 MySQL 数据库进行数据持久存储; 其次, 数据存在来源不一, 有如风向等类别特征, 各属性特征之间量纲也存在不统一, 数据缺失等问题, 因此为获得较好质量的建模数据, 必须进行一定的数据集成、清洗、规整及变换等预处理等工作; 然后, 对处理后的建模数据按一定比例进行拆分以获得训练集和测试集两部分; 接着, 使用训练数据对 GB 模型进行训练工作, 同时结合交叉验证法和网格搜索法进行模型的评估优化工作, 以获得最优的模型; 最后, 将构建好的 HPIC 剂量率预测模型在测试集上进行测试, 测试模型的预测效果.

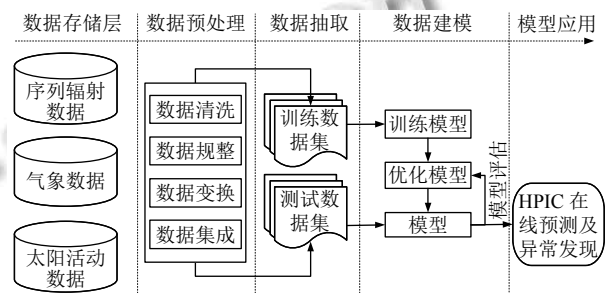


图 4 HPIC 剂量率在线预测流程方案

2.4 相关特征分析

我们以 S₁ 辐射监测站为目标研究对象. 如图 5 所示是采用可视化技术展现出的某降雨时段前后, 其 HPIC 剂量率监测数据数值随时间的变化信息图; 从图中可以看出 HPIC 剂量率在降雨时段会有明显的上升, 随着降雨停后, 其数值又缓慢降低回至正常水平.

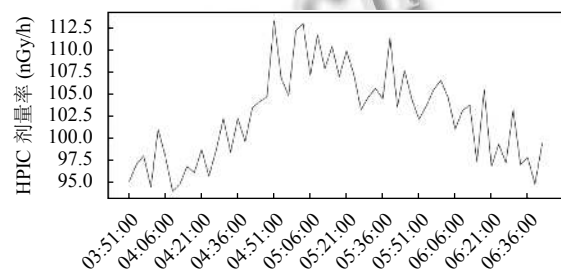


图 5 降雨时段 HPIC 波动图

根据该核电站区域核辐射各站点 2015–2017 年三年辐射序列数据进行异常值处理及数据标准化后, 对其温湿度、气压与 HPIC 剂量率进行皮尔逊相关系数分析, 分析结果如表 3 所示: 温度整体与 HPIC 剂量率数值之间的相关性较弱, 但是在 S₇ 站点相关性较强; 各辐射监测站点湿度与 HPIC 剂量率之间存在很强的正相关; 对气压相关性来说, 除了 S₂、S₁₀ 与 S₁₁ 3 个站点与 HPIC 剂量率相关性较弱, 整体上与 HPIC 剂量率数值之间存在着很强的负相关关系.

表3 温度、湿度及气压与 HPIC 剂量率皮尔逊相关系数表

站点	S ₁	S ₂	S ₃	S ₄	S ₅	S ₆	S ₇	S ₈	S ₉	S ₁₀	S ₁₁
温度	0.32	0.34	0.21	0.23	0.29	0.25	0.46	0.31	0.30	0.17	0.27
湿度	0.74	0.65	0.75	0.67	0.82	0.72	0.59	0.69	0.76	0.79	0.70
气压	-0.54	-0.34	-0.47	-0.63	-0.56	-0.72	-0.57	-0.73	-0.61	-0.23	-0.37

进一步具体分析 S₁ 站点温度与 HPIC 剂量率之间存在的相关关系, 将温度等宽分为 3 个区间: 偏低温、偏中温及偏高温, 如图 6 所示, 对每个区间 HPIC 剂量率标准化后数据采用箱线图统计分析可以得知, HPIC 剂量率在偏高温时集中偏大, 低温时偏小, 两者整体特点表现为温度越高, HPIC 剂量率越高; 因此我们可以认为温度也是 HPIC 剂量率的特征影响因子之一, 两者之间存在着正相关关系。

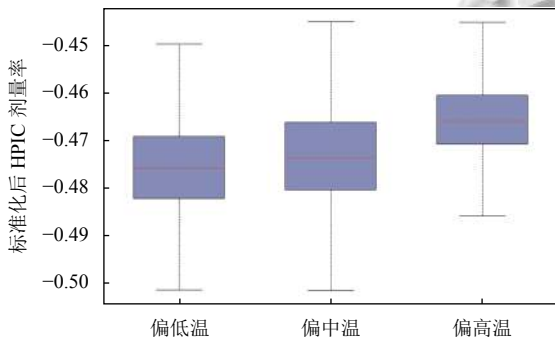


图6 2015年度 S₁点 HPIC 剂量率在温度区间数据分布箱线图

如图 7 所示, 先通过标准化方法消除量纲问题, 采用线性分析方法对 VTEC 与 HPIC 剂量率数值进行相关性分析得出, 在 VTEC 电子含量与 HPIC 剂量率呈现明显的同升同降变化规律, 采用皮尔逊算法计算得出

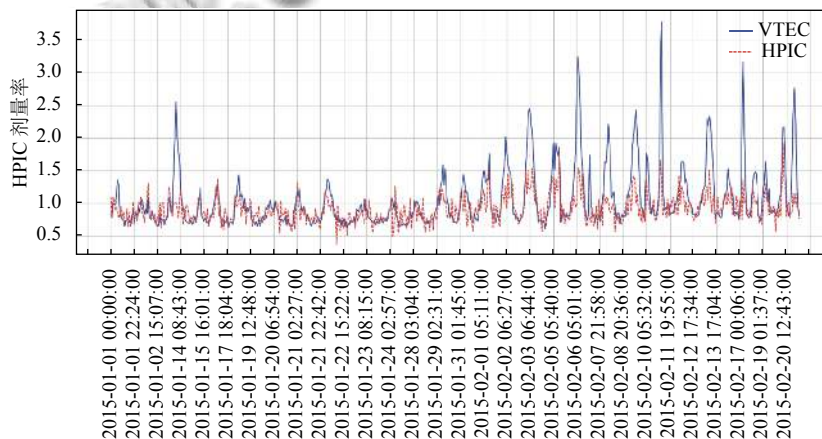


图7 VTEC 与剂量率同一时段波动对比图

两者相关系数值 r 高达 0.669, 具有显著的相关性。

如图 8 所示, 以箱线图方法统计出 8 个不同风向时 HPIC 剂量率数据信息, 可以看出, HPIC 剂量率数据在不同风向上, 其数据集中程度有稍许差异, 从中位数可以看出, 在风向偏南风及西南风时, 其 HPIC 剂量率数值中位数明显较大, 且数据整体数值较偏北风及东北风时明显偏大; 在风向为东风和东北风时, 其 HPIC 剂量率数值中位数是最低, 整体数值也是集中偏低, 不同风向上 HPIC 剂量率数值差值约在 (0~4) nGy/h 之间. 考虑风向为类别特征属性, 所以实验中采用 one-hot 对属性值进行了编码转换为八维的 0 或 1 的数值型。

再如表 4, 通过计算 S₁ 目标监测站点与其它站点之间 HPIC 剂量率皮尔逊系数得知, 除了 S₅、S₁₀ 站点, S₁ 站点与其余站点具有较强的相关性; 所以亦可以将其他监测站点监测数据作为当前监测数据的特征因子进行定性的预测; 我们考虑相关性较强的几个站点, 将与目标站点之间皮尔逊系数在 0.35 以上的 6 个站点作为模型特征参数输入, 进行模型构建。

同时, 考虑目标属性 HPIC 剂量率的时间序列离散特征, 其自身在时间前后也是有着较强的关联性; 所以为实现模型对当前时 HPIC 剂量率目标属性值预测能力, 本实验也考虑加入上一时刻的 HPIC 剂量率目标属性值作为特征输入, 以此作为基准。

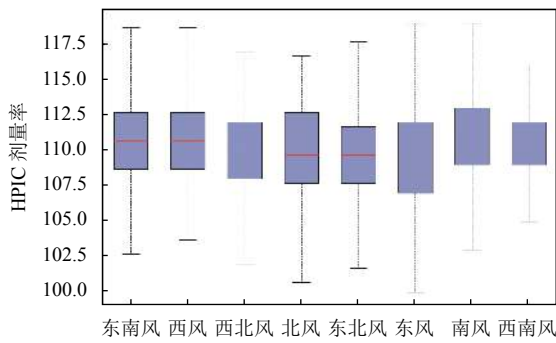


图8 风向与HPIC剂量率关系箱线图

综合看来,上述降雨、温湿度、气压、太阳辐射的VTEC及风向都与HPIC剂量率之间有着紧密的强关联性。这些自然因素对HPIC剂量率的日常监测会产生较强的干扰作用,不利于异常情况发生时,对异常成因的即时分析及准确定位,如当监测过程中发现某些站点HPIC剂量率骤然上升时,而此时区域内同时伴随上述不利于异常判断的自然因素发生,这时要及时得出异常成因就比较困难,可能需要提前进行预警,实行人工检测,比较费时耗力。因而,我们在此基础上设计如下在线预测模型,在一定程度上降低自然因素的综

合干扰,帮助进行异常情况的快速检出工作。

2.5 模型构建及验证

2.5.1 模型构建

如图9所示为我们所提出的基于GB算法的HPIC剂量率在线预测模型,输入采用第二章中提出的各种相关特征属性参数,如温度、湿度及风向等气象参数,天顶方向电子总量VTEC,与目标站点具有时空关联性的其它各站点HPIC剂量率数值,最后加上自身在时间前后存在时间关联性的目标站点上一时刻HPIC剂量率数值,输出采用当前时刻目标HPIC剂量率数值,这样就形成20个维度的特征输入,1个HPIC剂量率数值的输出。经过类别特征one-hot编码、零-均值标准化、缺失值填补、数据去重及异常值处理等数据预处理工作后,获得了质量较好的数据样本,然后将数据按照2:8的比例进行测试集和训练集的划分,再采用交叉验证方法将训练集等量划分为10等份,进行10折交叉验证,同时结合GridSearch网格寻优算法进行超参空间的构建,进行模型参数优化和选择,以此得出最优的GB预测模型。

表4 目标监测站点与其他监测子站HPIC监测值显著相关系数表

监测站点	S ₂	S ₃	S ₄	S ₅	S ₆	S ₇	S ₈	S ₉	S ₁₀	S ₁₁
皮尔逊系数	0.561	0.355	0.341	0.209	0.285	-0.549	-0.567	-0.428	-0.193	-0.561

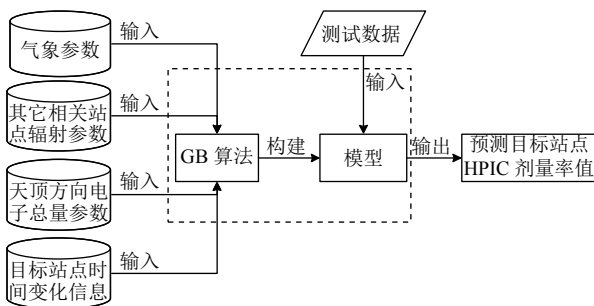


图9 基于GB算法的HPIC剂量率预测模型

实验过程中,对于GB算法模型的性能参数最主要有两个:模型中弱学习器的数量以及寻找最佳分割点要考虑的特征数量;这两个性能参数都是需要我们去设定,弱学器数量从10到140每隔20个进行数量的选取设定,最佳分割点的特征数量从16到最大特征数量20个每隔1个进行选取设定,总共取得5个特征数量取值;其模型在每种情况下的决定系数R-square和平均绝对误差MAE结果如图10和图11所示。

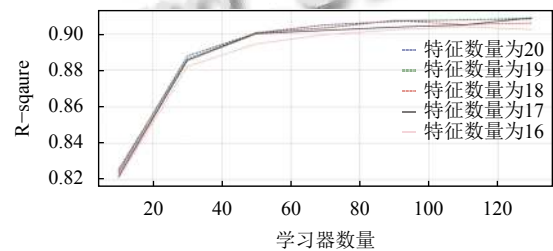


图10 GB模型参数的选取对评价指标R-square决定系数的影响

从图10和图11中可以看出,GB模型的R-square决定系数值随弱学习器数量的增加及特征数量的增加而增大;平均绝对误差MAE则随弱学习器数量的增加及特征数量的增加而减小;但是当弱学习器的数量达到120以上,其性能就开始趋于平缓;然而弱学器数量和特征数量越多,整体模型构建时间效率偏低;综合图中显示情况来看,设定弱学器数量为120,特征数量为19时,模型效果最好。

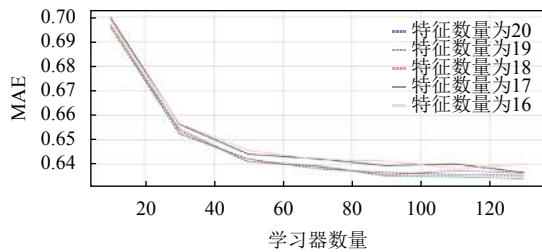


图 11 GB 模型参数的选取对评价指标平均绝对误差 MAE 的影响

2.5.2 模型测试

将训练好的模型对测试集进行预测, 预测结果可以用线性拟合图和散点图直观展示, 如图 12 和图 13 所示; 从图 12 中可以看出预测曲线与实际值曲线拟合性较好, 计算得出决定系数 R-Square 约 0.91, MAE 约 0.635, 但是在一些这些极值点处的预测效果会有所偏差. 监测站 HPIC 剂量率监测仪器高压电离室测量阈值误差一般在 ± 5 nGy 左右, 所以这个预测差额幅度在可接受范围内; 图 13 中 HPIC 剂量率预测值与实际值基本分布在图形 45°斜对角附近, 所以认为模型预测精度还是较好的.

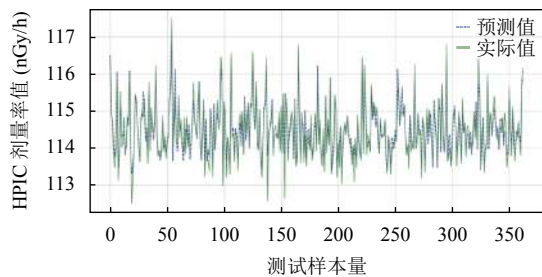


图 12 目标站点 HPIC 剂量率预测值与实际值拟合对比

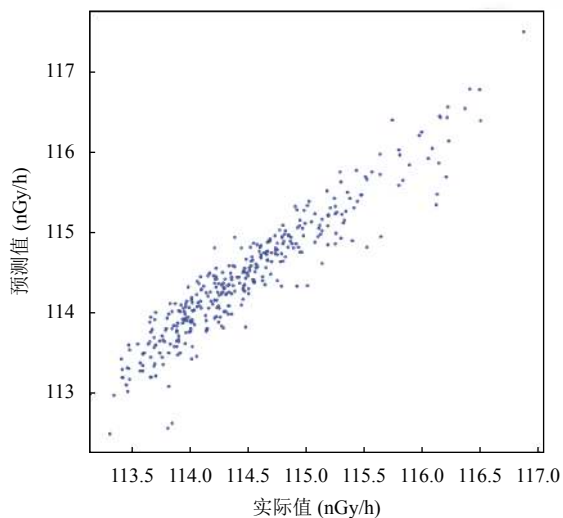


图 13 HPIC 预测结果和实际结果散点图

综上, 此模型较好地融合了各种自然特征影响因子, 并考虑目标站点与自身、各站点在 HPIC 剂量率值之间关联性构建起来的 HPIC 实时预测模型, 模型实验结果较好, 若能有效结合预测值与实际值误差阈值方法, 在一定程度上是可以帮助降低自然因素干扰辐射数据异常分析判断的影响, 实现对设备故障、放射性状况导致的辐射数据异常的更准确的定位; 比如, 设置误差阈值为 5 nGy, 可以假定当预测值与实际值的绝对误差低于 5 nGy, 辐射监测数据无异常情况发生; 当预测值与实际值的绝对误差高于 5 nGy, 辐射监测数据可能就出现了异常, 并即时向工作人员发出预警, 立时进行处理. 这对提高 ERMS 的异常发现能力和维保效率具有很大应用价值.

3 结论

首先, 感谢省级辐射中心、国家空间环境预报中心提供的数据支撑. 我们以机器学习算法进行数据的挖掘工作, 基于核电站积累的海量历史数据, 并引入气象数据及太阳活动数据, 充分考虑影响辐射监测中 HPIC 剂量率数值的重要特征因子, 并结合与目标监测站 HPIC 剂量率数值相关的上一时刻 HPIC 剂量率及其它相关站点的 HPIC 剂量率数值作特征输入, 进行大量的数据规整工作, 以 GB 回归模型建立起 HPIC 剂量率数值在线预测模型, 实现对当前时刻 HPIC 剂量率值的精准预测. 这对提高核电站对偏远外围监测站的环境辐射监测异常检测效率、管理水平及 ERMS 维保工作效率具有很大的现实意义和理论价值.

HPIC 剂量率数值除了受文中分析的多个自然因素的影响外, 还与空气中微量元素含量、维保日志、雷电、潮汐等有关, 但是囿于资源有限, 还未获取到这些数据, 如果增加这些数据进行挖掘分析, 可能得到更多有用的信息用于 HPIC 剂量率的预测和异常发现, 也是我们下一步将要进行的研究工作所在.

参考文献

- 1 汤旸. 核能发电的优势与发展前景. 科技展望, 2016, 26(28): 113. [doi: 10.3969/j.issn.1672-8289.2016.28.104]
- 2 孙力平, 孙鸣. 核电站保护屏障的监护神——核电站辐射监测系统 (RMS) 概述. 中国核电, 2008, 1(3): 252-257.
- 3 Chen W, Feng TC, Liu J, *et al.* A method based on Monte Carlo simulation for the determination of the G(E) function. Radiation Protection Dosimetry, 2015, 163(2): 217-221.

- [doi: [10.1093/rpd/ncu145](https://doi.org/10.1093/rpd/ncu145)]
- 4 Susila IP, Yuniarto A, Cahyana C. Monitoring and analysis of environmental gamma dose rate around serpong nuclear complex. *Atom Indonesia*, 2017, 43(2): 87. [doi: [10.17146/aij.2017.681](https://doi.org/10.17146/aij.2017.681)]
 - 5 Bossew P, Cinelli G, Hernández-Ceballos M, *et al.* Estimating the terrestrial gamma dose rate by decomposition of the ambient dose equivalent rate. *Journal of Environmental Radioactivity*, 2017, 166: 296–308. [doi: [10.1016/j.jenvrad.2016.02.013](https://doi.org/10.1016/j.jenvrad.2016.02.013)]
 - 6 朱耀明, 林明贵. 宁德核电厂外围环境 γ 辐射连续监测系统. *海峡科学*, 2015, (6): 75–78. [doi: [10.3969/j.issn.1673-8683.2015.06.026](https://doi.org/10.3969/j.issn.1673-8683.2015.06.026)]
 - 7 高泽全, 王瑞俊, 保莉, 等. 环境 γ 辐射剂量率受降水影响规律研究. *核电子学与探测技术*, 2017, 37(1): 4–7. [doi: [10.3969/j.issn.0258-0934.2017.01.002](https://doi.org/10.3969/j.issn.0258-0934.2017.01.002)]
 - 8 罗敦焯, 沙向东, 上官志洪, 等. 环境 γ 辐射剂量率连续监测数据影响因素和特征分析. *辐射防护*, 2018, 38(4): 308–317.
 - 9 张晓航, 石清磊, 王斌, 等. 机器学习算法在中医诊疗中的研究综述. *计算机科学*, 2018, 45(S2): 32–36.
 - 10 黄宁, 李玉龙, 陕永飞, 等. 数据挖掘在网络入侵检测系统中的应用. *计算机安全*, 2010, (10): 40–42. [doi: [10.3969/j.issn.1671-0428.2010.10.014](https://doi.org/10.3969/j.issn.1671-0428.2010.10.014)]
 - 11 张承畅, 张华誉, 罗建昌, 等. 基于云计算和改进 K-means 算法的海量用电数据分析方法. *计算机应用*, 2018, 38(1): 159–164. [doi: [10.11772/j.issn.1001-9081.2017071660](https://doi.org/10.11772/j.issn.1001-9081.2017071660)]
 - 12 丁立言. 数据挖掘在高速公路逃费分析中的应用. *交通世界*, 2019, (1): 18–19. [doi: [10.3969/j.issn.1006-8872.2019.01.007](https://doi.org/10.3969/j.issn.1006-8872.2019.01.007)]
 - 13 宋彦坡, 彭小奇. 数据挖掘技术及其在工业生产中的应用. *计算机测量与控制*, 2004, 12(10): 944–947. [doi: [10.3321/j.issn:1671-4598.2004.10.015](https://doi.org/10.3321/j.issn:1671-4598.2004.10.015)]
 - 14 陈爽爽. 基于 Gradient Boosting 算法的癫痫检测[硕士学位论文]. 济南: 山东大学, 2013.
 - 15 Schäfer J, Young J, Bernasconi E, *et al.* Predicting smoking cessation and its relapse in HIV-infected patients: The Swiss HIV cohort study. *HIV Medicine*, 2015, 16(1): 3–14. [doi: [10.1111/hiv.12165](https://doi.org/10.1111/hiv.12165)]
 - 16 Friedman JH. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 2001, 29(5): 1189–1232.