

基于 KPCA-GA-ELM 的海底管道 外腐蚀速率预测技术

高 帅

(中海石油(中国)有限公司深圳分公司, 广东 深圳 518000)

摘 要: 为了掌握海底管道外腐蚀情况, 保证管道的安全运行, 对海洋环境腐蚀因素进行了梳理, 确定了8种影响海底管道外腐蚀速率的主要因素, 通过核主成分分析(kernel principal components analysis, KPCA)对影响因素进行优选, 并将优选后的样本放入极限学习机(extreme learning machine, ELM)进行训练, 以影响因素为输入, 腐蚀速率为输出, 采用改进的遗传算法(genetic algorithm, GA)对ELM的输入权值和隐含层偏差进行优化。结果表明, KPCA-GA-ELM模型最大相对误差2.43%, MSE和MAPE分别为0.148 33和1.15, 与其余三种模型相比, 预测精度最高。研究结果可为提高海底管道完整性管理水平提供技术支持。

关键词: 海底管道; 外腐蚀; 预测; KPCA; ELM; GA

中图分类号: TG174

文献标识码: B

DOI: 10.19291/j.cnki.1001-3938.2021.11.004

Prediction of Corrosion Rate of Submarine Pipeline based on KPCA-GA-ELM

GAO Shuai

(CNOOC China Limited, Shenzhen Branch, Shenzhen 518000, Guangdong, China)

Abstract: In order to master the corrosion situation of submarine pipeline and ensure the safe operation of pipeline, by sorting out the corrosion factors in the marine environment, eight main factors affecting the corrosion rate of submarine pipelines were determined. The influencing factors were optimized by KPCA, and the optimized samples were put into the extreme learning machine (ELM) for training. The influencing factors were taken as input and the corrosion rate as output. An improved genetic algorithm (GA) was used to optimize the input weight and hidden layer bias of ELM. The results show that the maximum relative error of KPCA-GA-ELM model is 2.43%, MSE and MAPE are 0.148 33 and 1.15, respectively. Compared with the other three models, the prediction accuracy is the highest. The research results can provide technical support for improving the integrity management level of submarine pipeline.

Key words: submarine pipeline; external corrosion; prediction; KPCA; ELM; GA

海底油气输送管道作为海上油气田开发生产系统与处理系统的重要连接部分, 是海上油气运输的重要通道。海水中含有大量腐蚀性较强的天然电解质, 为电化学反应创造了条件, 海底管道一旦泄漏, 会造成重大的经济损失和环境污染^[1-2]。目前, 对于海底管道完整性的研究

多偏向内腐蚀, 但在复杂多变的海洋环境中, 外腐蚀也是造成海底管道失效的原因之一^[3]。因此, 梳理海水腐蚀因素, 建立可靠的外腐蚀速率预测模型对于提高管道完整性管理水平有着重要意义。

目前, 国内外学者针对海底管道腐蚀问题

开展了一系列研究。张新生等^[4]在考虑随机效应的前提下,应用贝叶斯网络对腐蚀深度进行预测;毕傲睿等^[5]采用主成分分析对影响管道腐蚀的土壤因素进行了筛选,并结合维纳退化过程对管道剩余寿命进行了预测;王奇等^[6]采用 BP 神经网络对外腐蚀速率进行了预测;者娜等^[7]采用支持向量机模型对管道腐蚀速率进行了预测。以上成果对于海底管道腐蚀的研究具有一定意义,但贝叶斯网络、BP 神经网络及支持向量机模型结构复杂,在训练过程中容易陷入局部最优解;主成分分析主要解决线性问题,对于海洋环境而言,海底管道受多种腐蚀因素影响,主成分分析无法适应非线性、小样本的数据降维和映射问题。基于此,采用核主成分分析(kernel principal components analysis, KPCA)对影响海底管道腐蚀的因素进行筛选,并将数据代入极限学习机(extreme learning machine, ELM)进行训练^[8],采用遗传算法(genetic algorithm, GA)对 ELM 的输入权值和隐含层偏差进行优化,建立外腐蚀速率预测模型,以期海底管道的安全运行提供理论依据。

1 KPCA 原理

KPCA 在主成分分析基础上引入核函数的概念,对样本 x_k ($k=1,2,3,\dots,m$) 采用非线性函数 $\varphi(x_k)$ 进行变换,将其从低维空间映射到高维空间 F ,此时特征空间的协方差矩阵为

$$C = \frac{1}{m} \sum_{j=1}^m \varphi(x_k) \varphi(x_k)^T \quad (1)$$

定义协方差矩阵的特征向量 V 和特征值 λ , V 可用 $\varphi(x_k)$ 线性表示为

$$V = \sum_{j=1}^m a_j \varphi(x_k) \quad (2)$$

引入核矩阵 $K=K(x_i, x_j)=\varphi(x_i)\varphi(x_j)$, 对于任意样本,原空间 x 中的向量在特征空间中的主成分为 $\varphi(x)$ 在特征向量 V 上的投影,即

$$V \cdot \varphi(x) = \sum_{j=1}^m a_j \varphi(x_i) \varphi(x) = \sum_{j=1}^m a_j K(x_i, x) \quad (3)$$

式中: a ——核矩阵 K 的特征向量。

由于径向基函数可将数据映射到无限高维空间,故选用该函数作为核函数。经计算,提取累积方差百分比 $\geq 90\%$ 的主成分作为影响海底管道外腐蚀速率的主要因素。

2 ELM 算法

ELM 算法^[9]由黄广斌教授提出,用于训练单隐含层的前馈神经网络,可用于分类、预测和回归分析。与 BP 神经网络不同,ELM 算法的输入权值和隐含层偏差均随机产生,输出权值依据广义逆矩阵理论解析求出,无需误差反向传播训练权值,具有训练参数少、学习速度快、鲁棒性强的特点。

在 KPCA 数据降维的基础上,利用 ELM 算法挖掘腐蚀因素与腐蚀速率之间的内在联系,从而预测腐蚀速率。模型如下:

$$y_i = \sum_{j=1}^m g(w_{ij} \cdot x_i + b_j) \beta_{jk} \quad (4)$$

式中: y_i ——样本输出值;

x_i ——样本输入值;

w_{ij} ——第 i 个输入层节点与第 j 个隐含层节点之间的权值,即输入权值;

b_j ——隐含层神经元偏差;

β_{jk} ——第 j 个隐含层节点与第 k 个输出层节点之间的权值,即输出权值;

$g(x)$ ——隐含层激活函数;

m ——隐含层节点数,当给定隐含层的节点数后,可依据广义逆矩阵理论计算输出权值 β_{jk} ,使预测输出值 y_i 的误差最小。

为动态求解 ELM 算法中的输入权值和隐含层偏差,采用 GA 算法对其进行优化处理。GA 算法是模拟达尔文的生物进化自然选择和孟德尔的遗传进化过程来搜索最优解,属于并行、全局搜索方式,可在搜索过程中自动累积先验知识,并自适应控制搜索过程以获得全局最优解^[10-11]。

为进一步提高遗传算法在不同阶段的进化适应性,对 GA 算法进行改进。引入反余弦函数改变交叉、变异因子的计算方法,在保证种群多样性和稳定性的前提下,避免出现局部最优解。改进公式如下:

$$P_c = \begin{cases} P_{cmax} - \frac{P_{cmax} - P_{cmin}}{\cos(\frac{S_{avg} - S_i}{S_{avg} - S_{min}} \pi)}, & S_i \geq S_{avg} \\ 1 + e, & S_i < S_{avg} \end{cases} \quad (5)$$

$$P_m = \begin{cases} P_{mmax} - \frac{P_{mmax} - P_{mmin}}{\cos(\frac{S_{avg} - S_i}{S_{avg} - S_{min}} \pi)}, & S_i \geq S_{avg} \\ P_{mmin}, & S_i < S_{avg} \end{cases} \quad (6)$$

式中: P_c 和 P_m ——交叉、变异因子;

P_{cmax} 、 P_{cmin} ——交叉因子的最大值和最小值；
 P_{nmax} 和 P_{nmin} ——变异因子的最大值和最小值；

S_i ——当前适应度；

S_{avg} ——适应度平均值。

3 基于 KPCA-GA-ELM 算法的腐蚀预测模型

本研究提出的基于 KPCA-GA-ELM 算法的腐蚀预测模型详细流程如下：

(1) 数据预处理，构建海底管道外腐蚀评价指标体系，对数据进行归一化处理。假设有 N 个腐蚀样本，影响因素为 m 个， (x_i, y_i) 为第 i 个样本数据，其中 $x_i = [x_{i1}, x_{i2}, \dots, x_{im}]$ 作为输入， y_i 作为输出，采用极值法进行归一化处理，

$$x'_{ie} = \frac{x_{ie} - x_{emin}}{x_{emax} - x_{emin}} \quad (7)$$

式中： x'_{ie} ——第 i 个样本的第 e 个影响因素归一化后的数据；

x_{ie} ——归一化前的原始数据；

x_{emax} 和 x_{emin} ——样本范围内第 e 个影响因素的最大值和最小值。

(2) 将处理后的数据代入 KPCA 模型中进行数据优选。

(3) 将步骤 (2) 中数据分为训练集和验证集，随机抽取前者为 44 组，后者为 6 组，将训练集代入 ELM 模型，以预测值和真实值之间的平均相对变动率 (ARV) 作为适应度函数，通过 GA 模型的选择、交叉和变异，经多次迭代后不断优化 ELM 算法中的输入权值和隐含层偏差，得到最优的 GA-ELM 模型。

(4) 将验证集数据代入训练好的 GA-ELM 模型，并采用均方误差 (MSE)、平均相对误差 (MAPE) 和相关系数 R^2 评价预测结果，其中 MSE、MAPE 的数值越小，模型偏差越小； R^2 的值越接近 1，预测值与实际值之间的相关性越大，拟合优度越好。公式如下，

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - y_i^*)^2 \quad (8)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{(y_i - y_i^*)}{y_i^*} \right| \times 100\% \quad (9)$$

式中： N ——腐蚀样本个数；

y_i ——第 i 个样本的腐蚀速率实际值；

y_i^* ——第 i 个样本的腐蚀速率预测值。

4 实例验证

海底管道腐蚀主要与海洋环境中的物理、化学和生物因素有关，其中 SRB (硫酸盐还原菌) 和真菌等微生物对管道的腐蚀作用主要体现在细菌的新陈代谢上，微生物产生硫化物，从而改变海水的氧含量和 pH 值，故不再对微生物进行单独考虑。基于此，选择盐度、温度、溶解氧、氧化还原电位、pH 值、流速、硫化物和海洋生物附着度等 8 个因素作为影响海底管道外腐蚀速率的主要因素。对某海底管道沿线设置 50 组挂片进行实海挂片试验，挂片尺寸 100 mm×100 mm×5 mm，在 200 天后得到管道沿线不同海洋环境下的腐蚀数据，部分数据见表 1。

对表 1 中的数据进行归一化处理后代入 KPCA 模型，对数据采用相关性矩阵和规划分析计算特征值和方差百分比，计算结果见表 2 和图 1。

表 1 某海底管道沿线腐蚀数据统计结果

序号	盐度/%	温度/℃	溶解氧/(mg·L ⁻¹)	氧化还原电位/mV	pH 值	流速/(m·s ⁻¹)	硫化物/(μg·L ⁻¹)	海洋生物附着度/%	腐蚀速率/(μA·cm ⁻¹)
1	3.152	27.64	6.52	386	5.2	0.341	347	85	16.85
2	2.852	27.85	8.49	364	7.1	0.265	355	67	16.97
3	3.408	17.64	7.55	138	8.5	0.124	342	74	17.04
4	3.102	25.54	6.17	350	8.1	0.458	366	50	12.47
...
50	2.045	24.74	7.15	215	7.5	0.184	331	63	11.85

表 2 KPCA 分析结果

主成分	特征值	方差/%	累积/%
1	2.418 8	34.55	34.55
2	1.737 27	24.82	59.37
3	1.063 99	15.20	74.57
4	0.766 68	10.95	85.52
5	0.605 48	8.65	94.17
6	0.341 21	4.87	99.05
7	0.066 57	0.55	99.60
8	0.032 45	0.40	100

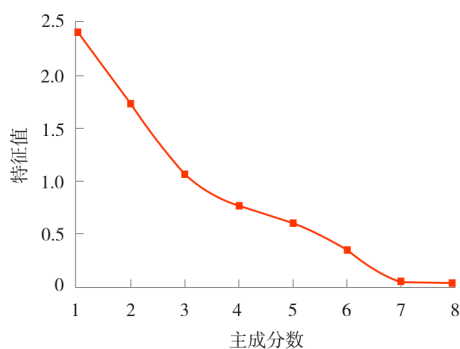


图 1 主成分分析碎石图

从图 1 可知, 前 5 个主成分之间的特征值差异较大, 且累积贡献率为 94.17%, 超过 90%, 因此可以选取前 5 个主成分代替原先的 8 个影响因素, 根据主成分的系数矩阵对数据进行重组, 样本数据由之前的 8 维变为 5 维, 重组后的样本数据见表 3。

根据 Kolomogorov 定理, 确定隐含层的节点数为 15, 选取 Sigmoid 函数作为激活函数, 对 GA 种群进行初始化操作, 设置种群个数 50, 交叉

表 3 重组后的样本数据

序号	主成分 1	主成分 2	主成分 3	主成分 4	主成分 5
1	1.910 1	1.671 31	-0.097 37	-1.081 83	-1.080 86
2	1.516 64	0.391 93	0.044 11	-0.191 76	-0.853 34
3	-3.227 82	2.433 36	-0.500 8	-0.153 92	0.361 39
4	-2.594	0.978 25	-0.508 84	-1.079 54	-0.695 32
...
50	-1.182 18	-1.969 19	-0.240 01	0.006 58	-0.874 04

因子区间 $[0.5, 1]$, 变异因子区间 $[0.002, 0.005]$, 最大迭代次数 200, 对 GA 算法的进化过程进行迭代计算, 迭代过程如图 2 所示。随着迭代次数的增加, 适应度函数 ARV 不断减小, 改进前后的 GA 算法分别在 152 次和 48 次达到收敛条件, 稳态误差分别为 0.186 1 和 0.062 4, 说明改进后的 GA 算法收敛速度更快, 寻优效果更好。

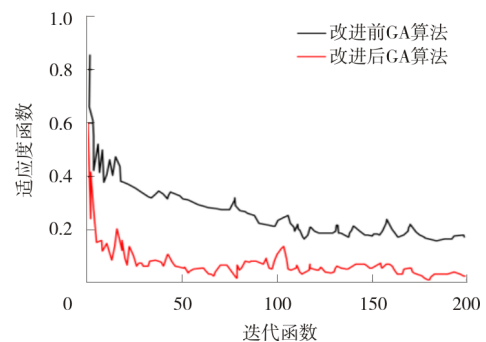


图 2 GA 算法迭代过程示意图

为验证 KPCA-GA-ELM 算法的预测精度, 将预测结果与 ELM 模型、KPCA-ELM 模型和 BP 模型进行对比, 验证集的预测结果见表 4, 不同预测模型的相对误差如图 3 所示。

表 4 不同预测模型的预测结果

序号	实际值/ ($\mu\text{A}\cdot\text{cm}^{-1}$)	ELM		KPCA-ELM		BP		KPCA-GA-ELM	
		预测值/ ($\mu\text{A}\cdot\text{cm}^{-1}$)	相对误差/%	预测值/ ($\mu\text{A}\cdot\text{cm}^{-1}$)	相对误差/%	预测值/ ($\mu\text{A}\cdot\text{cm}^{-1}$)	相对误差/%	预测值/ ($\mu\text{A}\cdot\text{cm}^{-1}$)	相对误差/%
1	16.85	17.03	1.07	16.97	0.71	18.42	9.32	16.84	0.06
4	12.47	13.54	8.58	12.87	3.21	14.64	17.40	12.34	1.04
17	17.52	17.99	2.68	17.1	2.40	19.54	11.53	17.23	1.66
25	8.69	8.14	6.33	8.45	2.76	7.63	12.20	8.79	1.15
36	12.34	11.47	7.05	12.03	2.51	11.01	10.78	12.64	2.43
48	10.67	10.16	4.78	10.34	3.09	9.67	9.37	10.73	0.56

由表 4 和图 3 可知, BP 模型的预测精度最差, 最大相对误差 17.40%; ELM 模型由于权值

范数较小, 减少了对网络中所有参数的迭代过程, 模型相对可获得更好的泛化性能, 故预测精

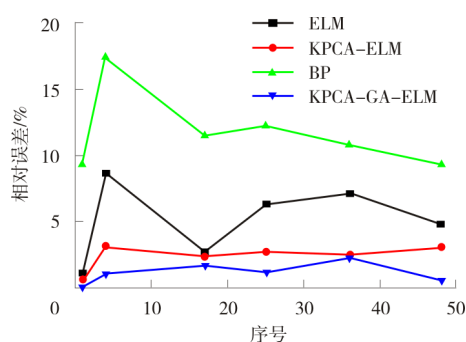


图 3 不同预测模型的相对误差对比

度有所提高,最大相对误差 8.58%;KPCA-ELM 模型对影响因素进行了数据降维,避免了无效数据的使用和训练,节省了算力,预测精度进一步提高,最大相对误差 3.21%;而 KPCA-GA-ELM 模型通过对 ELM 模型参数的不断寻优,预测精度最高,其验证集样本内的相对误差均小于其余 3 种模型,最大相对误差为 2.43%。

进一步分析 4 种模型 MSE、MAPE 和 R^2 ,见表 5。其中,KPCA-GA-ELM 模型 MSE 和 MAPE 分别为 0.148 33 和 1.15,均比其余 3 种模型小,且相关系数 $R^2=0.998\ 8$,在模型中最大,说明 KPCA-GA-ELM 模型的预测性能和鲁棒性较好,可以用来预测海底管道外腐蚀情况,可为有效评估管道完整性和制定维护策略提供实际指导。

表 5 不同预测模型评价指标对比

模型	评价指标		
	MSE	MAPE/%	R^2
ELM	0.608 33	5.08	0.986 48
KPCA-ELM	0.303 33	2.44	0.995 84
BP	1.525	11.76	0.972 19
KPCA-GA-ELM	0.148 33	1.15	0.998 8

5 结 论

(1) 采用 KPCA 对影响海底管道外腐蚀的海洋环境因素进行了筛选,结合改进的 GA 模型对 ELM 模型的输入权值和隐含层偏差进行优化处理,经实例验证,KPCA-GA-ELM 模型最大相

对误差 2.43%,MSE 和 MAPE 分别为 0.148 33 和 1.15,预测精度最高;

(2) 在迭代的过程中,ELM 模型的隐含层节点数和激活函数采用经验确定,但这两个参数对模型的计算结果影响较大,今后应将其作为研究方向进一步优化模型算法。

参考文献:

- [1] 骆正山,袁宏伟.基于误差补偿的 GM-RBF 海底管道腐蚀预测模型[J].中国安全科学学报,2018,28(3):96-101.
- [2] 俞树荣,李建华,李淑欣,等.埋地管道腐蚀剩余寿命预测概率模型[J].中国安全科学学报,2008,18(6):11-15.
- [3] 唐晓时,春涛,曹光,等.海岸土壤环境对油气管道局部腐蚀的影响[J].中国腐蚀与防护学报,2016,36(3):191-196.
- [4] 张新生,吕品品.考虑随机效应的腐蚀管道贝叶斯退化分析[J].中国安全科学学报,2019,29(8):73-80.
- [5] 毕傲睿,骆正山,王小完,等.基于土壤腐蚀主成分的金属管道退化维纳过程研究[J].材料保护,2018,51(1):37-42.
- [6] 王齐,胡林林.基于 BP 神经网络的油气长输管道土壤腐蚀性预测[J].当代化工,2016,45(9):2198-2200.
- [7] 者娜,杨剑锋,刘文彬,等.基于 KPCA 和 SVM 的工艺管道腐蚀速率预测[J].腐蚀与防护,2019,40(1):56-60.
- [8] 孙艳丽,杨娜,张正涛,等.基于核主元分析和支持向量机的结构损伤识别研究[J].应用基础与工程科学学报,2018,26(4):888-900.
- [9] HUANG G B,ZHU Q Y,SIEW C K. Extreme learning machine:theory and applications[J]. Neurocomputing,2006 (70):489-501.
- [10] 范伟,林瑜阳,李钟慎.遗传算法优化的 BP 神经网络压电陶瓷蠕变预测[J].电机与控制学报,2018,22(7):91-96.
- [11] SRINIVAS M, PATNAIK L M. Adaptive probabilities of crossover and mutation in genetic algorithms [J]. IEEE Transactions on Systems,Man,and Cybernetics, 1994,24(4):656-667.

作者简介:高 帅(1992—),男,陕西咸阳人,大学本科,工程师,2014年毕业于中国石油大学(华东)油气储运工程专业,现主要从事海洋油气开采及处理,海底管道运输、防腐及安全管理工作。

收稿日期:2021-07-05

编辑:黄蔚莉