

大气腐蚀数据降维最优维度研究

潘纪情, 付冬梅, 杨焘, 刘磊明

(北京科技大学 自动化学院, 北京 100083)

摘要: **目的** 确定金属大气腐蚀数据降维的最优维度。**方法** 分别采用PCA、MDS、Isomap和LLE四种方法对大气腐蚀数据进行降维处理, 并采用集成学习算法建立预测模型。针对不同的降维方法和近邻点个数计算, 使用MAPE (Mean Absolute Percentage Error, 相对百分误差绝对值的平均值) 对预测结果进行评价, 将最佳预测速率所对应的维度作为最优维度。**结果** 不同的降维方法和近邻参数作用下, 最优维度从2维到7维不等。流形学习方法对大气腐蚀数据进行降维的MAPE均小于线性降维方法。**结论** 适合每种降维方法的最优维度可能是不同的, 通过对MAPE进行对比, 得到大气腐蚀数据在各种降维方法的最优维度。

关键词: 大气腐蚀数据; 降维方法; 最优维度; 流形学习; 集成学习

DOI: 10.7643/issn.1672-9242.2020.03.019

中图分类号: TB114

文献标识码: A

文章编号: 1672-9242(2020)03-0114-06

Optimal Dimension of Dimensionality Reduction of Atmospheric Corrosion Data

PAN Ji-qing, FU Dong-mei, YANG Tao, LIU Lei-ming

(School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China)

ABSTRACT: The work aims to determine the optimal dimension for the dimensionality reduction of metals' atmospheric corrosion data. The four methods such as PCA, MDS, Isomap and LLE were used for the dimensionality reduction of atmospheric corrosion data, and an ensemble learning algorithm was used to establish the prediction model. For different dimensionality reduction methods and the calculation of the number of neighbors, the mean absolute percentage error (MAPE) was used to evaluate the prediction results, and the dimension corresponding to the best prediction rate was used as the optimal dimension. Under the action of different dimensionality reduction methods and the neighbor parameters, the optimal dimension ranged from 2 to 7 dimensions. Manifold learning method was used for the dimensionality reduction of atmospheric corrosion data, and the resulting MAPE was less than that of the linear dimensionality reduction method. The optimal dimension for each dimensionality reduction method may be different. Finally, the optimal dimension of the atmospheric corrosion data processed by the four dimensionality reduction methods is obtained through the comparison of the MAPE values.

KEY WORDS: atmospheric corrosion data; dimensionality reduction method; optimal dimension; manifold learning; ensemble learning

收稿日期: 2018-08-30; 修订日期: 2018-10-30

Received: 2018-08-30; Revised: 2018-10-30

基金项目: 国家重点研发计划(2017YFB0702104); 博士后科学基金(2017M620615); 中央高校基本科研业务费(FRF-TP-16-082A1)

Fund: The National Key R&D Program of China (2017YFB0702104); China Postdoctoral Science Foundation(2017M620615); Fundamental Research Funds for the Central Universities (FRF-TP-16-082A1)

作者简介: 潘纪情(1991—), 男, 青海人, 硕士研究生, 主要研究方向为大气腐蚀数据挖掘。

Biography: PAN Ji-qing (1991—), Male, from Qinghai, Master, Research focus: atmospheric corrosion data mining.

通讯作者: 付冬梅(1963—), 女, 辽宁人, 博士, 教授, 主要研究方向为数据挖掘与共享、智能算法与控制、图形分析与处理。

Corresponding author: FU Dong-mei (1963—), Female, from Liaoning, Ph. D., Professor, Research focus: intelligent data and sharing, mining algorithm and control, graphical analysis and processing.

在大气环境中,金属材料会与大气环境介质发生化学作用,从而出现腐蚀问题^[1]。腐蚀问题给人类带来巨大的损失,大气腐蚀造成的金属材料损失占全部腐蚀的 1/2 以上,而全球腐蚀的经济损失每年超过万亿美元^[2]。我国由于金属腐蚀造成的经济损失也很严重,每年有 10⁷ t 钢材因腐蚀而变成废弃物^[3]。因此研究腐蚀现象具有极其重要的意义,通过人工智能和数据挖掘的方法对腐蚀数据进行处理分析,探究大气腐蚀环境因素和腐蚀速率之间的关系,为腐蚀防护提供有价值的信息。大气腐蚀是由多种环境因素共同作用的反应^[4],自然环境材料腐蚀采集周期漫长^[5],通常需要以年为单位获取数据,这就造成了大气腐蚀数据高维度、小样本的特点。针对数据的这种特点,对大气腐蚀数据进行降维处理,以降低数据维度及其样本数量之间的比例差异,为腐蚀速率预测建模提供优质数据。

降维是指通过某种数学变换将原始高维数据属性空间转变到低维的子空间^[6],使用低维子空间嵌入的特征来表示高维数据。常用的降维方法包括线性降维和非线性降维。线性降维方法要求数据集中的各个变量之间独立无关,即存在于线性结构之中。典型的线性降维方法有:主成分分析^[7](Principal Component Analysis, PCA)和多维缩放^[8](Multiple Dimensional Scaling, MDS)等。非线性降维方法则不对数据作线性结构限制,应用更广泛,常用的是基于特征分解的流形学习。流形学习假定样本是嵌入在高维空间中的低维流形数据,流形是局部与欧式空间同胚的空间。尽管数据样本在高维空间十分复杂,但是样本局部仍然具备欧式空间的性质,可以在局部建立流形映射关系,并逐步将这种映射关系推广到全局,建立低维的流形数据,通过这些低维流形数据就可以进一步分析样本^[9]。常用的流形学习方法有局部线性嵌入^[10](Locally Linear Embedding, LLE)和等度量映射^[11](Isometric Mapping, Isomap)等。通过降维处理技术,将高维数据在其低维空间中表征出来,以提高数

据在建模等数据挖掘任务上的准确性。

集成学习通过构建并结合多个学习器来完成学习任务^[12],集成学习的一般结构是:先产生一组“个体学习器”,再使用某种策略将它们结合起来。集成学习通过对多个学习器进行结合,可以获得比单一学习器显著优越的泛化能力。根据个体学习器的生成方式不同,可以将集成学习分成两大类:一种是个体学习器之间的依赖关系强,必须串行生成序列方法;另一种是个体学习器不存在强依赖关系,可以并行化方法。前者的代表为 Boosting,后者的代表是 Bagging 和 Random Forest。在 Boosting 中,最为著名的代表是 AdaBoost^[13]。AdaBoost 通过对样本权重和基学习器权重进行调节,将训练的重点集中在比较难训练的数据样本上,最终实现强学习器。

在数据降维处理到预测建模的过程中,重点问题是降维后低维维度的确定。有关低维维度的选择,在以往的工程实践中都是靠专家经验来选取的。低维维度的选择对后期的预测建模有影响,恰当维度的低维数据能够反映出数据结构特点,过低的维度会过度丢失原数据信息,过高的维度数据又会使冗余信息过多,而不能良好地体现出数据结构中的主要特征。文中采用了 PCA、MDS、LLE、Isomap 四种降维方法,分别对大气腐蚀数据进行降维,对低维数据使用 AdaBoost 方法进行腐蚀速率预测,对不同维度预测结果的影响进行了讨论。

1 大气腐蚀数据收集和预处理

文中的大气腐蚀数据来源于国家材料腐蚀平台^[14],选取了北京、广州、江津、青岛、琼海、万宁、武汉等 7 个城市 1984 年 10 月至 1985 年 9 月和 1998 年 11 月到 1999 年 10 月的大气腐蚀数据,特征属性包括平均温度、最高温度、最低温度、平均湿度、平均气压、最高气压、最低气压等,部分数据见表 1。

表 1 北京地区部分大气腐蚀数据展示
Tab.1 Display of part of atmospheric corrosion data in Beijing

试验地点	试验时间	平均温度/ ℃	极端最高 温度/℃	极端最低 温度/℃	平均相 对湿度/ %	平均 气压/ hPa	最高 气压/ hPa	最低 气压/hPa	瞬时法 SO ₂ / (mg·cm ⁻³)	瞬时法 HCl/ (mg·cm ⁻³)	连续法 H ₂ S/(mg· 100 ⁻¹ m ⁻² ·d ⁻¹)	连续法硫酸盐 化速率/(mg· 100 ⁻¹ m ⁻² ·d ⁻¹)
北京	1998/11/1	3	21	-10	64	1014.1	1024.3	1003.5	0.0449	0.018	0.0428	0.3062
北京	1998/12/1	-2	9	-10	45	1018.4	1027.4	1001.2	0.0456	0	0.0502	0.2562
北京	1999/1/1	-15	8	-10	69	1016.4	1029.4	1001	0.0448	0.009	0.0429	0.4131
北京	1999/2/1	18	9.3	-11	60	1018.8	1029	1001.4	0.125	0.0042	0.0557	0.8036
北京	1999/3/1	19	26	-10	88	1010	1023.2	996	0.166	0.0084	0.0558	0.7007
北京	1999/4/1	25	33	-1	83	1006.6	1016.7	996.4	0.1703	0.0043	0.0636	0.68

度量单位的不同会影响数据分析,例如离子浓度的描述使用 mol/L 和 mg/(m²·d) 的单位进行描述,将

会导致不同的结果。对于同一种变量,如果采取较小的单位来表示,就会使该属性具有较大的数量级,在

数据中占有较大的权重。规范化数据给每个属性赋予相同的权重^[15],排除量纲对于数据的影响。文中采用最小-最大规范化对数据进行处理,即:

$$v'_i = \frac{v_i - X_{\min}}{X_{\max} - X_{\min}} \quad (1)$$

式中: X_{\min} 为属性 X 的最小值; X_{\max} 为属性 X 的最大值; v_i 是属性 X 的第 i 个值; v'_i 是第 i 个值归一化后的值。将所有的属性值全部都归一化到 $[0,1]$ 中,使数据无量纲化,能够有效避免数值问题。将数据都统一化到一个参考系中,便于后期的数据处理工作。

文中选取的是 Q235 金属材料的大气腐蚀数据,数据跨度从 1984 年 10 月至 1985 年 9 月和 1998 年 11 月到 1999 年 10 月,腐蚀速率来自 7 个城市每年的记录,共有 14 条腐蚀速率数据。由于每 12 个月对应一个腐蚀速率,对于特征属性,取这 12 个月数据归一化后的平均值作为特征数值。最终得到 14 条 21 维的大气环境腐蚀数据。

2 降维方法

2.1 PCA

PCA^[7]是最为常用的一种线性降维方法,PCA 的主要思路是使原始数据和基于投影重构的低维数据之间的距离最小,使式(2)取得最小值:

$$\sum_{i=1}^m \left\| \sum_j^{d'} z_{ij} w_j - x_i \right\|_2^2 \quad (2)$$

式中: z_i 代表样本点 x_i 在低维空间中的投影, $z_i = z_{i1}, z_{i2}, \dots, z_{id'}$; w_j 表示投影的低维坐标系; m 表示样本量个数; d' 表示低维维度。

PCA 算法的步骤有以下三步:对样本进行中心化处理: $x_i \leftarrow x_i - \frac{1}{m} \sum_{i=1}^m x_i$; 计算所有样本的协方差矩阵 XX^T ; 对协方差矩阵做特征值分解,取最大的 d' 个特征值所对应的特征向量。第三步中求解得到的特征向量即低维数据。

2.2 MDS

MDS 算法^[8]也是一种常用的线性降维方法,MDS 要求在原始空间中样本之间的距离在低维空间中依然得以保留。设有 m 个样本,在原始空间中构建的距离矩阵为 $D \in R^{m \times m}$, D 中的每个元素为 $d_{ij} = |x_i - x_j|$,降维之后为 $|z_i - z_j| = d_{ij}$ 。通过距离来构建降维后样本的内积矩阵 B ,通过计算 B 矩阵的特征值和特征向量来得到低维数据。

1) 首先计算以下几个距离:

$$d_{ij}^2 = \|z_i\|^2 + \|z_j\|^2 - 2z_i^T z_j \quad (3)$$

$$d_i^2 = \frac{1}{m} \sum_{j=1}^m d_{ij}^2 \quad (4)$$

$$d_{.j}^2 = \frac{1}{m} \sum_{i=1}^m d_{ij}^2 \quad (5)$$

$$d_{..}^2 = \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^m d_{ij}^2 \quad (6)$$

2) 计算降维后的样本内积矩阵 B ,其中 B 矩阵中的每个元素由式(4)-(6)求得:

$$b_{ij} = -\frac{1}{2}(d_{ij}^2 - d_i^2 - d_j^2 + d_{..}^2) \quad (7)$$

3) 得到矩阵 B 之后,对矩阵 B 进行特征值分解,将 B 矩阵的前 d' 个特征值构成对角矩阵 $d(\lambda_1, \lambda_2, \dots, \lambda_{d'})$, λ 值按从大到小排列,其相应的特征向量矩阵为 $V, V A^{\frac{1}{2}} \in R^{m \times d'}$,每行是一个样本的低维坐标。

2.3 Isomap

MDS 算法^[11]中的距离是基于欧式距离进行计算的,但在具有流形结构的高维空间中,计算欧式直线距离是具有误导性的。在流形空间上,直线距离是不可达的,因此需要引入测地线距离来计算两点间的距离。测地线是指曲面上侧地曲率处处为 0 的曲线,在局部意义上意味着两点之间的最短距离^[16]。对每个样本点寻找近邻点,建立一个近邻连接图。在近邻图中,近邻点之间是存在有限距离连接的,而非近邻点之间的连接为无穷大。那么这个时候就使用图论中经典的 Dijkstra 算法^[17]来计算近邻图中任意两个点之间的距离,获得所有样本点之间的距离 $d(x_i, x_j)$ 。将其作为 MDS 的输入,则 MDS 算法的输出便是低维空间数据。

1) 采用 k 近邻法寻找样本点 x_i 的近邻点, x_i 与 k 个近邻点之间的距离为两点间的欧式距离,和非近邻点之间的距离为无穷大。

2) 使用 Dijkstra 算法计算任意两个样本点之间的距离 $d(x_i, x_j)$ 。

3) 将 $d(x_i, x_j)$ 作为 MDS 算法的输入,降维结果是原数据在低维空间中投影值。

2.4 LLE

LLE^[10]要求在降维前后样本间的线性关系保持一致。LLE 算法主要分三个步骤:

1) 通过 k 近邻法选择近邻点,通过近邻点将样本点线性重构出来。

$$x_i = w_{i1}x_1 + w_{i2}x_2 + \dots + w_{ij}x_j \quad (8)$$

2) 通过近邻点构建误差函数,计算样本点进行线性重构的系数 w_i :

$$\begin{cases} \min_{w_1, w_2, \dots, w_m} \sum_{i=1}^m \left\| x_i - \sum_{j \in Q_i} w_{ij} x_j \right\| \\ \text{s.t.} \sum_{j \in Q_i} w_{ij} x_j = 1 \end{cases} \quad (9)$$

3) 在低维空间的映射过程中,保持重构系数 w_i 不变,使高维空间中每一点 x_i 对应低维空间的 z_i ,则:

$$\min_{y_1, y_2, \dots, y_m} \sum_{i=1}^m \left\| y_i - \sum_{j \in Q_i} w_{ij} y_j \right\| \quad (10)$$

构建系数矩阵 $W_{ij} = w_{ij}$ ，得到

$$M = (I - W)^T (I - W) \tag{11}$$

对 M 矩阵进行特征值分解，最小的 d' 个特征值对应的特征向量组成的矩阵就是式 (11) 的解，也就是低维数据。

3 预测模型：AdaBoost 算法

AdaBoost 是 Boosting 系列算法中的代表，其产生一系列的基学习器，各个基学习器的训练集由之前基学习器的表现决定，把已经被判断错误的样本以较大的概率出现在新的基学习器中，新的基学习器能够更好地处理判断错误的样本。AdaBoost 可以用于解决二分类、多分类和回归等问题，而在大气腐蚀数据中主要进行回归预测，AdaBoost 算法中用来回归预测的 R2 算法^[18]包括以下几个步骤。

- 1) 对每个样本初始化相同的权值 $\frac{1}{m}$ ，样本的概率分布为 $D_t(i) = \frac{1}{m}$ ，设定平均损失函数为 $\bar{L}_t = 0$ 。
- 2) 当平均损失函数 $\bar{L}_t < 0.5$ 时，建立回归模型 $f_t(x) \rightarrow y$ 。
- 3) 针对每个样本计算损失函数 $L_t(i) = \frac{|f_t(x_i) - y_i|}{D}$ ，其中 $D = \sup |f_t(x_i) - y_i|, i = 1, 2, \dots, m$ 。
- 4) 计算平均损失函数 $\bar{L}_t = \sum_{i=1}^m L_t(i) D_t(i)$ 。
- 5) 计算 $\beta_t = \frac{\bar{L}_t}{1 - \bar{L}_t}$ ，该数值越低，表示在预测中的置信度越高。
- 6) 更新样本权值 $D_{t+1}(i) = \frac{D_t \beta_t^{1-L_t(i)}}{Z_t}$ ，其中 Z_t 是一个规范化因子。
- 7) 得到最后输出 $f_{fin(x)} \inf [y \in Y: \sum_{t: f_t(x) \leq y}$

$$\lg \frac{1}{\beta_t} \geq \frac{1}{2} \sum_t \lg \frac{1}{\beta_t}]。$$

4 实验仿真

对大气腐蚀数据进行整理之后，选取金属 Q235 对应年份的腐蚀速率为预测目标，采取 MAPE 作为预测模型的衡量标准，MAPE 的计算公式为：

$$R_{MAPE} = \sum_{t=1}^m \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times \frac{100}{n} \tag{12}$$

式中： y_t 为实际值； \hat{y}_t 为预测值； m 为样本数量。

应用四种降维方法对大气腐蚀数据进行降维处理，将得到的低维数据作为预测模型的输入进行回归预测。回归预测的模型选择 AdaBoost.R2，在该模型中需要设定的参数包括基回归器对象、基回归器数量、学习率、损失函数。在训练模型过程中，使用 SVR 作为基回归器，数量设置为 50，学习率为 1，损失函数选择线性损失函数。考虑到样本数量和训练集训练集的分配问题，采用留一交叉验证法，将 13 个样本点的 Q235 金属腐蚀速率作为训练集，留一个样本点作为测试集，依次选择不同的样本点作为测试集进行交叉验证。

近邻点的个数 k 是流形学习算法的重要参数，不同的 k 值对降维效果会产生不同的影响。文中对同一降维方法中相同近邻点个数、不同维度的 MAPE 值（见表 2）进行对比，将最小 MAPE 值对应的维度作为该近邻点个数下的最优维度，针对线性降维仅仅从不同维度的 MAPE 值来衡量即可。通过表 2 可以总结出在不同情况下的最优维度，见表 3。

从表 3 中可以看出，使用不同降维方法和选择不同近邻点个数的時候，会产生不同的最优维度。由此可见，不同的降维方法以及不同的近邻参数会产生不同的最优维度，各种情况下 MAPE 的变化趋势如图 1 所示。

表 2 不同降维方法及其不同低维数据建模的 MAPE 值

Tab.2 Different dimensionality reduction methods and MAPE values of low dimensional data modeling

d'	Isomap					MDS	PCA	LLE				
	3	4	5	6	7	*	*	3	4	5	6	7
2	16.091	16.59	14.63	13.359	19.178	12.775	14.525	14.791	15.847	12.511	14.597	13.964
3	12.58	13.464	15.912	14.995	13.255	16.559	17.304	14.103	16.395	17.53	15.763	16.732
4	15.259	15.145	17.526	15.844	15.594	15.178	17.278	15.919	14.102	14.41	13.338	14.837
5	13.315	15.305	16.513	14.234	15.426	18.45	14.828	13.644	12.649	15.605	13.568	17.569
6	13.998	16.338	15.041	16.289	14.891	13.758	16.058	12.375	16.499	13.776	16.156	11.798
7	13.214	14.536	12.415	12.33	16.832	19.78	15.052	14.262	15.578	14.389	17.946	16.531
8	14.359	15.409	14.742	14.575	17.315	14.327	15.662	12.898	14.502	15.028	14.618	14.646
sum	98.81	106.78	106.77	101.626	112.491	110.827	110.707	97.992	105.572	103.249	105.986	106.077

表3 不同情况下的最优维度
Tab.3 Optimal dimensions under different conditions

	Isomap					MDS	PCA	LLE				
	3	4	5	6	7			3	4	5	6	7
最优维度	3	3	7	7	3	2	2	6	5	2	4	6

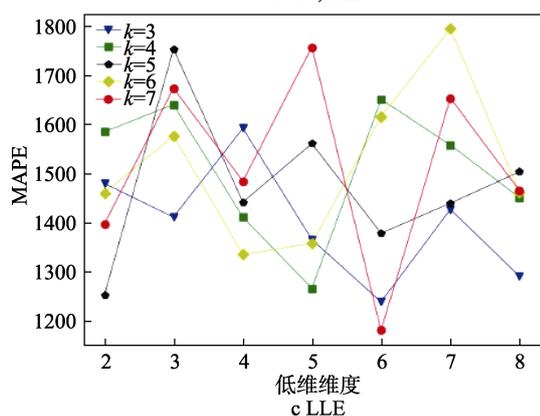
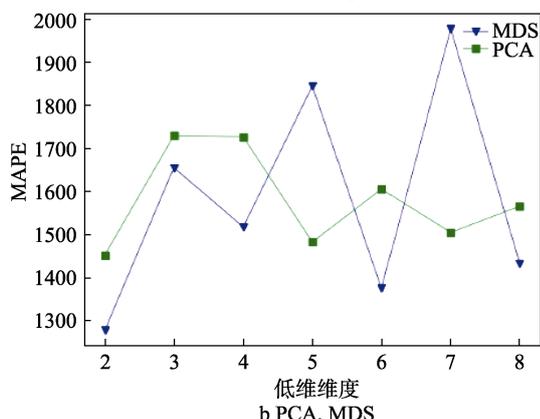
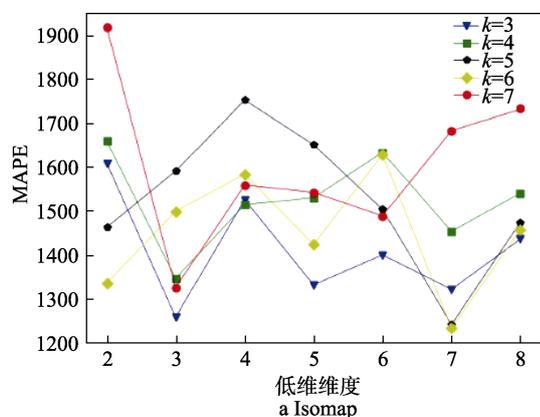


图1 不同降维方法和不同近邻个数情况下 MAPE 的变化趋势

Fig.1 Different dimensionality reduction methods and the change trend of MAPE with different neighbor points

从图1中可以看出,不同的降维方法和近邻参数作用下,最优维度从2维到7维不等。由此可见,并不是低维维度越高或者越低,预测的效果就越好,大部分的最优维度是在维度值变化中产生的,这反映了维度对于预测精度的影响。

为了方便对比不同降维方法之间的差异,将不同

维度下的 MAPE 进行求和,将两种流形学习方法的 MAPE 取平均值来评价其效果,见表4。可以看出,两种流形学习方法对大气腐蚀数据进行降维的 MAPE 均小于线性降维方法。由此表明,基于大气腐蚀现象影响因素较多,且各个因素之间相互影响,采取流形学习的降维方法要优于线性降维方法。

表4 不同降维方法的 MAPE 和对比
Tab.4 MAPE and contrast of different dimensionality reduction methods

	Isomap	MDS	PCA	LLE
总 MAPE	105.295	110.827	110.707	103.775

5 结论

- 1) 通过不同的降维方法对大气腐蚀数据进行降维处理,并使用 AdaBoost 集成学习模型对金属 Q235 的腐蚀速率进行预测,通过对预测结果的分析得到每种降维方法的最优维度。
- 2) 降维维度值影响建模精度,过低或过高的维度值都不利于建模精度的提高。
- 3) 针对大气腐蚀数据降维处理,流形学习的降维效果明显好于线性降维方法。

在采用降维方式对其他国家材料腐蚀平台中的数据进行处理时,关于降维维度值的选择可以参考本中的计算结论,避免由于维度选择不佳导致的模型精度问题。

参考文献:

[1] 曾佳俊,周学杰,吴军,等.金属材料大气腐蚀试验相关性与寿命预测研究现状[J].腐蚀科学与防护技术,2015,27(1):90-94.
ZENG Jia-jun, ZHOU Xue-jie, WU Jun, et al. Correlation and Life Prediction of Atmospheric Corrosion Tests of Metal Materials[J]. Corrosion Science and Protection Technology, 2015, 27(1): 90-94.

[2] LI X, ZHANG D, LIU Z, et al. Materials Science: Share Corrosion Data[J]. Nature News, 2015, 527(7579): 441.

[3] 傅欣欣,董俊华,韩恩厚,等.低碳钢 Q235 在模拟酸雨大气腐蚀条件下的电化学阻抗谱监测[J].金属学报,2014,50(1):57-63.
FU Xin-xin, DONG Jun-hua, HAN En-hou, et al. Electrochemical Impedance Spectroscopy Monitoring on Mild Steel Q235 in Simulated Industrial Atmospheric Corrosion Environment[J]. Acta Metallurgica Sinica, 2014,

- 50(1): 57-63.
- [4] 刘凯吉. 大气腐蚀环境的分类及腐蚀性评定[J]. 全面腐蚀控制, 2015, 29(10): 26-27.
LIU Kai-ji. Classification and Evaluation of Environment of Atmospheric Corrosion[J]. Total Corrosion Control, 2015, 29(10): 26-27.
- [5] 鲁庆. 基于数据挖掘的材料自然环境腐蚀预测研究[D]. 北京: 北京科技大学, 2015.
LU Qing. Research on Prediction of Material Natural Environment Corrosion Based on Data Mining[D]. Beijing: University of Science and Technology Beijing, 2015.
- [6] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
ZHOU Zhi-hua. Machine Learning[M]. Beijing: Tsinghua University Press, 2016.
- [7] WOLD S, ESBENSEN K, GELADI P. Principal Component Analysis[J]. Chemometrics and Intelligent Laboratory Systems, 1987, 2(1-3): 37-52.
- [8] COX T F, COX M A A. Multidimensional scaling[M]. Boca Raton: CRC Press, 2000.
- [9] 李波. 高维数据的流形学习分析方法[M]. 武汉: 武汉大学出版社, 2016.
LI Bo. Manifold Learning Analysis Method for High-dimensional Data[M]. Wuhan: Wuhan University Press, 2016.
- [10] ROWEIS S T, SAUL L K. Nonlinear Dimensionality Reduction by Locally Linear Embedding[J]. Science, 2000, 290(5500): 2323-2326.
- [11] TENENBAUM J B, SILVA DE V, LANGFORD J C. A Global Geometric Framework for Nonlinear Dimensionality Reduction[J]. Science, 2000, 290(5500): 2319-2323
- [12] ZHOU Z H. Ensemble Methods: Foundations and Algorithms[M]. Boca Raton: Chapman & Hall/CRC, 2012.
- [13] FREUND Y, SCHAPIRE R E. A Decision-theoretic Generalization of On-line Learning and an Application to Boosting[J]. Journal of Computer And System Sciences, 1997, 55(1): 119-139.
- [14] HAN J W, KAMBER M, PEI J. Data Mining: Concepts and Techniques 3rd Edition[M]. Beijing: China Machine Press, 2012: 113-114
- [15] 曹元鹏, 周大可, 杨欣, 等. 基于测地线采样的三维表情人脸识别[J]. 吉林大学学报(信息科学版), 2015, 33(4): 429-434.
CAO Yuan-peng, ZHOU Da-ke, YANG Xin, et al. Expression Invariant 3D Face Recognition Based on Geodesic Curve Features[J]. Journal of Jilin University (Information Science Edition), 2015, 33(4): 429-434.
- [16] QING G, ZHENG Z, YUE X. Path-planning of Automated Guided Vehicle Based on Improved Dijkstra Algorithm[C]// Control And Decision Conference (CCDC). IEEE, 2017.
- [17] SOLOMATINE D P, SHRESTHA D L. AdaBoost RT: A Boosting Algorithm for Regression Problems[C]// Proceedings 2004 IEEE International Joint Conference. IEEE, 2004.