

# 基于机器学习算法的稀土元素 掺杂 TiO<sub>2</sub> 光催化活性分析

吴兴惠<sup>1</sup>, 吴迪<sup>2</sup>, 周玉萍<sup>1</sup>, 史载锋<sup>2</sup>

(1.海南师范大学信息科学与技术学院,海南 海口 571158)

(2.海南师范大学化学与化工学院,海南 海口 571158)

**[摘要]** 机器学习是人工智能及机器学习领域的共同研究热点,其理论和方法已被广泛应用于解决工程应用和科学领域的复杂问题.在化学领域,稀土金属掺杂二氧化钛光催化剂提高光催化活性的研究已有大量研究,但掺杂机理一直存在争论,基于第一性原理的算法复杂且存在误差.为了探索能够不需要化学结构数学模型,以元素电子结构等基础数据为先验知识,通过计算机算法准确预测光催化剂活性,采用线性回归、高斯过程回归、支持向量机回归、K-最近邻法对稀土 TiO<sub>2</sub> 光电性质进行预测研究,并通过实验验证基于逐步回归分析的关键因素分析的有效性与基于关键因素的 k-NN 回归模型的优势.结果表明,使用逐步回归分析得到的关键特征所得的预测性能更好.对比不同的回归方法的预测性能可知, k-NN 的预测性能最好.

**[关键词]** 稀土掺杂二氧化钛,数据分析,机器学习,光催化活性预测

**[中图分类号]** TP181 **[文献标志码]** A **[文章编号]** 1672-1292(2017)03-0087-06

## Photocatalytic Activity Prediction of Rare Earth Doped TiO<sub>2</sub> Based on Machine Learning Algorithm

Wu Xinghui<sup>1</sup>, Wu Di<sup>2</sup>, Zhou Yuping<sup>1</sup>, Shi Zaifeng<sup>2</sup>

(1.School of Computer Science and Technology, Hainan Normal University, Haikou 571158, China)

(2.School of Chemistry and Chemical Engineering, Hainan Normal University, Haikou 571158, China)

**Abstract:** Machine learning is one of the hot points of the research in the area of artificial intelligence and pattern recognition, and its theories and methods have been widely used to resolve complex problems in science and engineering applications. In the chemical area, rare earth doped TiO<sub>2</sub> photocatalysts for improving photocatalytic activity has been massively studied and got many important results. However, the doping mechanism is not fully clear, even though the first-principle algorithm is applied to mechanism simulation, the calculation process is complicated and there are many inevitable errors. To solve the problem, it is expected that the simulation can be done without complicated chemical model, the prediction accuracy can be improved through a simply computer algorithm only based on element's essential data as the prior knowledge. In the present paper, the linear regression, Gaussian process regression and support vector machines regression, k-nearest neighbor algorithm are used to predict the photocatalytic reaction rate constant of rare earth doped TiO<sub>2</sub>. Then, experiments are used to verify the key factor analysis effectiveness based on stepwise multiple regression and advantages of the key factor analysis based on k-NN regression model. Results show that the prediction performance is better with key factors obtained from stepwise regression analysis. Compared with other regression methods, it may be found the prediction performance of k-NN is the best.

**Key words:** rare earth doped TiO<sub>2</sub>, data analysis, machine learning, photocatalytic activity prediction

机器学习是人工智能及机器学习领域的共同研究热点,其理论和方法已被广泛应用于解决工程应用和科学领域的复杂问题<sup>[1-4]</sup>.机器学习在上世纪六十年代末被引入到化学领域,它建立于一个十分直观的

收稿日期:2017-03-18.

基金项目:海南省自然科学基金(20156242).

通讯联系人:史载锋,博士,教授,博士生导师,研究方向:化学. E-mail: zaifengshi@163.com

基本假设,即“物以类聚”,认为性质相近的样本在模式空间中所处的位置相近,它们在空间形成“簇”.机器学习方法具有明显的优点:不需要数学模型,需要的先验知识很少,擅长处理复杂事物和多源数据等.

稀土(RE)元素具有特殊的能级结构,且能吸收紫外光和可见光,因而许多光催化领域的研究者利用稀土掺杂  $\text{TiO}_2$ ,使得后者光催化活性得到显著提高<sup>[5-9]</sup>.但因实验采用的工艺条件各不相同,且影响光催化活性的因素极其复杂,又缺少杂质元素对电子结构影响的详细研究,导致对于掺杂改性的机理说法不一,掺杂改性效果也大相径庭.利用涉及密度泛函理论第一性原理<sup>[10-12]</sup>的计算过于复杂,易导致计算结果产生误差.

本文拟运用机器学习中的机器学习算法分别对稀土金属( $\text{RE} = \text{Y}, \text{Lu}, \text{La}, \text{Ce}, \text{Pr}, \text{Nd}, \text{Pm}, \text{Sm}, \text{Eu}, \text{Gd}, \text{Tb}, \text{Dy}, \text{Ho}, \text{Er}, \text{Tm}, \text{Yb}$ )掺杂  $\text{TiO}_2$  前后的几何结构、能带结构、态密度和电子结构等进行模拟计算,从而揭示稀土掺杂的本质规律.首先,分析影响光催化表观速率常数的关键因素,在机器学习领域可被认为是特征选择.本文将采用相关系数进行相关性分析以及逐步回归决策进行关键特征选择.其次,通过稀土元素基础数据预测其光催化表观速率常数,在机器学习领域被认为是回归问题中的预测.本文将采用线性回归、高斯过程回归、支持向量机回归、 $K$ -最近邻法对稀土  $\text{TiO}_2$  光电性质进行预测研究.最后,通过实验初步验证基于逐步回归分析的关键因素分析的有效性与基于关键因素的  $k$ -NN 回归模型的优势.

通过本文研究,希望能够推动机器学习方法的合理运用,为化学问题的解决提供新的途径.具体来说,一方面,从机器学习方法的角度,有助于应用机器学习算法解决具体问题,分析不同机器学习算法在针对特定问题时的优势与劣势,并探寻更合适的算法以解决问题;另一方面,从化学应用的角度,为进行化学实验提供更丰富的信息,为实验结果提供指导性数据,并能够利用历史数据帮助分析影响结果的关键因素.

## 1 相关机器学习算法原理

### 1.1 特征选择

特征选择<sup>[13-17]</sup>是从一组特征中挑选出一些最有效的特征以降低特征空间维数的过程,是机器学习的关键问题之一.对于机器学习系统,一个好的学习样本是训练分类器的关键,样本中是否含有不相关或冗余信息直接影响着分类器的性能,因此研究有效的特征选择方法至关重要.虽然特征选择方法很多,但针对实际问题的研究还存在很多不足,如何针对特定问题给出有效的方法仍是一个需要进一步解决的问题.

#### 1.1.1 利用统计中的相关系数进行相关性分析,从而选择相关性高的特征

相关系数是用以反映变量之间相关关系密切程度的统计指标.相关系数是按积差方法计算,同样以两变量与各自平均值的离差为基础,通过两个离差相乘来反映两变量之间相关程度,着重研究线性的单相关系数.相关系数的计算公式如下:

$$r_{XY} = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^N (Y_i - \bar{Y})^2}},$$

式中, $n$ 为样本量; $X_i$ 和 $\bar{X}$ 分别为两个变量的观测值和均值. $r_{XY}$ 描述的是两个变量间线性相关强弱的程度. $r_{XY}$ 的取值在-1与+1之间,若 $r_{XY} > 0$ ,表明两个变量是正相关,即一个变量的值越大,另一个变量的值也会越大;若 $r_{XY} < 0$ ,表明两个变量是负相关,即一个变量的值越大,另一个变量的值反而会越小. $r_{XY}$ 的绝对值越大表明相关性越强,要注意的是这里并不存在因果关系.若 $r_{XY} = 0$ ,表明两个变量间不是线性相关,但有可能是其他方式的相关(例如曲线方式).

#### 1.1.2 利用逐步回归决策进行关键特征选择

逐步回归的基本思想是,将变量一个一个引入,引入变量的条件是偏回归平方和经检验是显著的,同时每引入一个新变量后,对已选入的变量要进行逐个检验,将不显著变量剔除,这样保证最后所得的变量子集中的所有变量都是显著的.这样经若干步以后使得“最优”变量子集.

逐步回归决策,首先要建立因变量 $y$ 与自变量 $x$ 之间的总回归方程,再对总的方程及每一个自变量进

行假设检验. 当总的方程不显著时,表明该多元回归方程线性关系不成立;而当某一个自变量对  $y$  影响不显著时,应将其剔除,重新建立不包含该因子的多元回归方程. 筛选出有显著影响的因子作为自变量,并建立“最优”回归方程. 回归方程包含的自变量越多,回归平方和越大,剩余的平方和越小,剩余均方也随之较小,预测值的误差也愈小,模拟的效果愈好. 但方程中的变量过多,预报工作量就会越大,其中有些相关性不显著的预报因子会影响预测的效果. 因此,在多元回归模型中,选择适宜的变量数目尤为重要.

## 1.2 回归模型

### 1.2.1 线性回归

回归分析<sup>[18-21]</sup>中,只包括一个自变量和一个因变量,且二者的关系可用一条直线近似表示,这种回归分析称为一元线性回归分析. 若回归分析中包括两个或两个以上的自变量,且因变量和自变量之间是线性关系,则称为多元线性回归分析. 线性回归的模型是这样的,对于一个样本  $\mathbf{x}_i$ ,其输出值是其特征的线性组合:

$$f(\mathbf{x}_i) = \sum_{m=1}^p w_m \mathbf{x}_{im} + w_0 = \mathbf{w}^T \mathbf{x}_i, \quad (1)$$

式中,  $w_0$  为截距,式(1)中通过增加  $\mathbf{x}_{i0}=1$  将  $w_0$  也吸收到向量表达中,简化了形式,实际上  $\mathbf{x}_i$  有  $p+1$  维度.

线性回归的目标是用预测结果尽可能地拟合目标标签,用最常见的小二乘法作为损失函数:

$$J(\mathbf{w}) = \frac{1}{n} \sum_{i=1}^n (y_i - f(\mathbf{x}_i))^2 = \frac{1}{n} \|\mathbf{y} - \mathbf{X}\mathbf{w}\|^2. \quad (2)$$

### 1.2.2 高斯过程回归

给定训练集  $D$  中的  $n$  个观测点,记为  $D = \{(\mathbf{x}_i, y_i) | i = 1, 2, \dots, n\}$ ,其中  $\mathbf{x}$  表示一个  $D$  维输入向量,  $y$  表示一个输出标量或是依靠因变量的目标输出,所有的  $n$  个列向量的输入就被记为  $D \times n$  维矩阵  $\mathbf{X}$ ,且目标输出也变为向量  $\mathbf{y}$ ,因此记为  $D = (\mathbf{x}, \mathbf{y})$ . 高斯过程回归模型就是对输入向量与目标输出之间的关系  $f$  进行推断,即在给定输入向量时确定目标输出的条件分布<sup>[1]</sup>.

依据高斯过程的定义可以得出,  $f(x_1), f(x_2), \dots, f(x_n)$  服从多元联合高斯分布,其性质完全由均值函数  $m(x_i)$  和协方差函数  $K$  确定,即:

$$\begin{cases} m(x) = E[f(x)], \\ k(x, x') = E[(f(x) - m(x))(f(x') - m(x'))], \end{cases} \quad (3)$$

式中,  $x, x' \in \mathbf{R}^d$  为任意随机变量. 则 GP 也可定义为  $f(x) \sim \text{GP}(m(x), k(x, x'))$ . 通常为了符号上的简洁,会对数据进行预处理,即让其均值函数为 0.

### 1.2.3 支持向量机回归(SVR)

支持向量机回归的基本思想是将输入样本空间非线性变换到另一个特征空间,在这个特征空间中构造回归估计函数,而这种非线性变换是通过定义适当的核函数  $K(x_i, x_j)$  来实现的. 设给定的输入样本  $\mathbf{x}$  为  $p$  维向量,  $n$  个样本及其输出值可表示为:

$$(\mathbf{x}_1, y_1) \cdots (\mathbf{x}_n, y_n) \in \mathbf{R}^p \times \mathbf{R}, \quad (4)$$

则 SVR 的学习问题就是一个二次规划问题. 通常采用 Vapnik 的  $\varepsilon$  不敏感损失函数来表示,即指定容许误差  $\varepsilon$ ,若样本  $\mathbf{x}$  误差为  $\xi$ ,则当  $|\xi| \leq \varepsilon$  时不计损失,否则损失计为  $|\xi| - \varepsilon$ . 回归函数表示为:

$$f(\mathbf{x}) = \sum_{i=1}^n (\mathbf{a}_i - \mathbf{a}_i^*) K(\mathbf{x}, \mathbf{x}_i) + b, \quad (5)$$

式中,  $\mathbf{a}$  和  $\mathbf{a}^*$  为求解的  $p$  维向量,  $\varnothing(\mathbf{x})$  为从样本空间到高维特征空间的映射函数,  $K(\mathbf{x}_i, \mathbf{x}_j) = \varnothing(\mathbf{x}_i) \cdot \varnothing(\mathbf{x}_j)$  为核函数,表示为两个  $\varnothing(\mathbf{x})$  的点积. 可以看出,核函数的选择对于 SVR 的预测性能有着重要影响. 核函数的采用使得映射函数  $\varnothing(\mathbf{x})$  不必明确求出,使求解非线性回归成为可能. 式(5)中对应于权值  $(\mathbf{a} - \mathbf{a}^*)$  不为 0 的样本  $\mathbf{x}_i$  称为支持向量. 显然,支持向量的数目决定了计算的复杂度,且与预测精度存在较强关联.

### 1.2.4 K-最近邻法(KNN)

KNN 是一种基于类比的算法,其基本思想是在多维空间中找到与待测样本最近邻的  $K$  个点 ( $1 \leq K \leq n-1$ , 其中  $K=1$  对应最近邻法,采用留一法时  $K=n-1$  对应全局预测),并根据这  $K$  个点的类别来判断未知

样本的类别. 类似的, KNN 也可用于回归估计, 即以这  $K$  个点作为训练样本来计算待测点的值. 其邻近性用欧氏距离来定义, 设两个样本分别为  $\mathbf{x}_i = \{x_{i1}, x_{i2}, \dots, x_{im}\}$  与  $\mathbf{x}_j = \{x_{j1}, x_{j2}, \dots, x_{jm}\}$ , 则欧氏距离为:

$$\text{Dist}(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_{a=1}^m (x_{ia} - x_{ja})^2}. \quad (6)$$

SVR 以留一法实施预测时选取训练样本是基于全局的, 即待测样本之外所有的样本参与训练. 由于样本集的异质性, 基于全局预测的精度并非总是最优: 例如某样本集明显地分为两个亚类, 则预测第一亚类的某个样本时, 若训练集中包含有第二亚类样本, 将对预测结果产生干扰; 此时, 合适  $K$  值的 KNN 预测精度可能高于全局预测. KNN 的缺点在于尽管可以利用系统聚类、非线性映射等方法来估算  $K$  值大小, 但要先验地给出每一个待测样本的最优  $K$  值仍然是非常困难的<sup>[12~14]</sup>.

## 2 关键因素分析与表观速率常数预测

表观速率常数是描述稀土元素掺杂  $\text{TiO}_2$  光催化反应的动力学常数. 对同一光催化反应, 表观速率常数的大小不仅与稀土元素掺杂量、反应温度、反应物浓度等相关, 还与稀土元素的多种化学特征有关. 通过化学实验测量并计算出不同稀土元素在同一类化学反应中的表观速率常数, 然后通过数据分析考查各化学特征数据与速率常数之间的关系.

## 3 实验

本文以真实的化学数据进行实验, 以验证机器学习算法在稀土元素数据分析中的有效性与先进性. 首先, 通过运用机器学习算法, 对已有的稀土元素的相关化学特征与测量出的表观速率常数进行分析, 提供对表观速率常数具有重要影响的关键因素. 本文主要采用两种方法, 一是利用相关系数进行相关性分析; 二是利用逐步回归方法进行特征选择. 然后, 运用线性回归、 $k$ -NN 回归、支持向量回归与高斯过程回归 4 种机器学习回归算法, 对稀土元素的表观速率常数进行预测.

稀土元素的基本化学特征数据来源于化学元素周期表等化学专业数据册. 为获得表观速率常数, 本文采用球磨法分别制备 Sc、Y、La、Nd、Tb 等 17 种稀土金属元素掺杂  $\text{TiO}_2$  光催化剂, 以 300W 中压汞灯为光源, 对亚甲基蓝溶液进行光催化降解. 根据亚甲基蓝浓度随光照时间的变化, 计算对应每一种光催化剂的一级反应速率常数, 即表观速率常数, 作为本文数据分析中的输出特征依据.

### 3.1 数据处理

本文所获得的原始的稀土元素(17 种稀土元素数据)包括其单质、氧化物与氢氧化物等相关化学特征, 共有 36 个实验测得的数字特征, 以及通过实验测得的表观速率常数. 原始数据包含许多冗余信息与噪声信息, 本文去掉未完全观测的特征与无信息的特征, 保留 19 个可利用的完全观测特征, 最终得到的可以用于分析的稀土元素数据的描述如表 1 所示.

为验证所运用的方法在关键因素分析与表观速率常数预测方面的可行性, 随机地将数据分为训练集与测试集, 其中训练集包含 14 种稀土元素, 测试集包含 3 种稀土元素. 为去除实验的偶然性, 随机地对数据进行 10 种训练集与测试集的分割, 并针对每一种分割进行实验, 统计最终的平均实验结果.

### 3.2 关键因素分析

采用计算相关系数与逐步回归两种方法来分析影响稀土元素的表观速率常数的关键因素, 即影响数据输出的关键特征. 首先, 根据相关系数的计算公式, 分别得出元素的每一个特征与元素的表观速率常数之间相关性, 然后根据相关系数的绝对值的大小进行排序, 筛选出相关性最高的几个特征.

通过数据分析得到的关键因素, 一方面可通过专业知识进行分析或通过化学实验进行进一步验证, 另一方面可作为已知信息进行利用以实现相关预测任务, 并通过预测的结果验证所选关键因素的合理程度.

表 1 稀土元素数据描述

Table 1 Rare earth element data descriptions

名称	描述
稀土元素	已标注 13 种(包括 Sc, Y, La, Nd, Sm, Eu, Gd, Tb, Ho, Er, Tm, Yb, Lu)
元素特征	19 个完全观测特征(包括“密度”, “熔点”, “沸点”等)
输出特征	表观速率常数(取值范围 0.039~0.086)



表 2 展示了利用两种方法得到的影响元素表观速率常数的关键因素. 从表 2 可以看出,两种不同的方法得到的关键因素有所不同,但也有共同之处(如升华热). 通过该类数据分析方法得到的关键因素可以为化学实验的相关分析提供指导性信息.

3.3 表观速率常数预测

从数据与应用的角度分析,关键因素分析能够更加深入地分析数据,提供隐藏在数据中的信息;从表观速率常数预测模型的角度分析,由于数据点的数目远少于特征的数目,所以选出关键因素更有利于预测任务的进行. 本文分别使用全部特征与关键特征实现表观速率常数的预测任务,并通过实验分析使用不同的关键特征对预测结果的影响与不同的机器学习预测算法的性能.

为证明关键因素分析对于预测模型的必要性,本文首先使用所有特征进行模型训练,得到表观速率常数的预测结果,如表 3 所示. 然后分别使用如上所述的两种关键因素进行模型训练,得到表观速率常数的预测结果,如表 4 所示. 表 3 与表 4 展示的是多次实验预测结果的平均相对误差与标准差. 这里需要指出的是,表 3 中所展示的关于支持向量回归与高斯过程回归的预测结果并无评价意义,因为这两种方法在使用全部特征时已经过拟合,且不同测试点的预测结果是同一个常数.

表 2 影响稀土元素表观速率常数的关键因素

Table 2 The key factors on the apparent rate constant of rare earth elements

利用相关系数得到的关键因素	利用逐步回归得到的关键因素
熔化热	升华热
升华热	沸点
RE <sup>3+</sup> 磁矩	氢氧化物沸点
剪切模量	密度

表 3 使用全部特征进行表观速率常数预测结果

Table 3 The prediction results of the apparent rate constant using the whole features

方法名称	预测结果(平均相对误差±标准差)
支持向量回归	0.149 8±0.077 6(过拟合)
高斯过程回归	0.108 8±0.062 7(过拟合)
线性回归	—(参数过多)
K-NN 回归	0.011 2±0.062 3

表 4 使用关键特征进行表观速率常数预测结果

Table 4 The prediction results of the apparent rate constant using the key features

方法名称	利用相关系数关键特征预测结果 (平均相对误差±标准差)	利用逐步回归关键特征预测结果 (平均相对误差±标准差)
支持向量回归	0.149 8±0.077 6	0.149 8±0.077 6
高斯过程回归	0.185 6±0.055 0	0.143 5±0.056 6
线性回归	0.561 7±0.376 0	0.141 8±0.057 9
K-NN 回归	0.112 4±0.045 3	0.105 1±0.043 0

观察表 4 中的预测结果,对比两种关键因素分析方法可以发现,使用逐步回归分析得到的关键特征得到的预测性能更好. 对比不同的回归方法的预测性能可以发现,k-NN 的预测性能最好,这可能是因为对于支持向量回归与高斯过程回归这种较为复杂的模型,数据点过少时会影响模型的性能.

4 总结

本文分别使用全部特征与关键特征实现表观速率常数的预测任务,并通过实验分析使用不同的关键特征对预测结果的影响与不同的机器学习预测算法的性能. 对比两种关键因素分析方法可以发现,使用逐步回归分析得到的关键特征所得预测性能更好. 对比不同的回归方法的预测性能可以发现,k-NN 的预测性能最好,这可能是因为对于支持向量回归与高斯过程回归这种较为复杂的模型,数据点过少时会影响模型的性能. 因此,针对该类数据点较少的计算,需要对现有机机器学习算法进行改进,以期提高预测计算的准确度.

[参考文献](References)

[1] SNELSON E,RASMUSSEN C E,GHAHRAMANI Z. Warped Gaussian process[C]//Proc of the NIPS 16. Vancouver,British Columbia,Canada,2004.

[2] NGUYEN T D,PETERS J. Incremental online sparsification for model learning in realtime robot control[J]. Neurocomputing, 2011,74(11):1 859-1 867.

- [3] PETELIN D, KOCIJAN J. Control system with evolving Gaussian process model[C]//Proc of IEEE Symposium Series on Computational Intelligence, 2011.
- [4] MUSIZZA B, PETELIN D, KOCIJAN J. Accelerated learning of Gaussian process models[C]//Proc of the 7th EUROSIM Congress on Modelling and Simulation. Praha, CZ, VCVUT, 2010.
- [5] SHI H X, ZHANG T Y, AN T C, et al. Enhancement of photocatalytic activity of nano-scale  $\text{TiO}_2$  particles co-doped by rare earth elements and heteropolyacids[J]. Journal of colloid and interface science, 2012, 380(1): 121–127.
- [6] DI S C, GUO Y P, LÜ H W, et al. Microstructure and properties of rare earth  $\text{CeO}_2$ -doped  $\text{TiO}_2$  nanostructured composite coatings through micro-arc oxidation[J]. Ceramics international, 2015, 41(5A): 6 178–6 186.
- [7] VIGNESH C B, DANIEL R R, JOHN N K. Assessment of mechanisms for enhanced performance of Yb/Er/titania photocatalysts for organic degradation; role of rare earth elements in the titania phase[J]. Applied catalysis B( environmental ), 2017, 202: 156–164.
- [8] HÜSNÜ A Y, MUHSIN Ç. The effect of rare earth element doping on the microstructural evolution of sol-gel titania powders[J]. Journal of alloys and compounds, 2017, 695: 1 336–1 353.
- [9] YUAN W J, ZHU Q Y, DENG C J, et al. Effects of rare earth oxides additions on microstructure and properties of alumina-magnesia refractory castables[J]. Ceramics international, 2017, 43(9): 6 746–6 750.
- [10] AOIFE L, ANNA I, MICHAEL N. A first principles investigation of  $\text{Bi}_2\text{O}_3$ -modified  $\text{TiO}_2$  for visible light Activated photocatalysis: the role of  $\text{TiO}_2$  crystal form and the  $\text{Bi}^{3+}$  stereochemical lone pair[J]. Materials science in semiconductor processing, 2014, 25: 59–67.
- [11] LI S J, QIU H, WANG C D, et al. Highly efficient  $\text{NaTaO}_3$  for visible light photocatalysis predicted from first principles[J]. Solar energy materials and solar cells, 2016, 149: 97–102.
- [12] ZENG X Y, XIAO X Y, ZHANG W P, et al. Interfacial charge transfer and mechanisms of enhanced photocatalysis of an anatase  $\text{TiO}_2(001)$ - $\text{MoS}_2$ -graphene nanocomposite: A first-principles investigation[J]. Computational materials science, 2017, 126: 43–51.
- [13] FU A M, WANG X Z, HE Y L, et al. A study on residence error of training an extreme learning machine and its application to evolutionary algorithms[J]. Neurocomputing, 2014, 146: 75–82.
- [14] ZHENG W B, FU X P, YING Y B. Spectroscopy-based food classification with extreme learning machine[J]. Chemometrics and intelligent laboratory systems, 2014, 139: 42–47.
- [15] WOUBISHET Z T, ESKO S. Machine learning for durability and service-life assessment of reinforced concrete structures: recent advances and future directions[J]. Automation in construction, 2017, 77: 1–14.
- [16] LIU Y, BI J W, FAN Z P. Multi-class sentiment classification; the experimental comparisons of feature selection and machine learning algorithms[J]. Expert systems with applications, 2017, 80: 323–339.
- [17] JOSÉ M M M, PABLO E M, CARLO B, et al. Prediction of the hemoglobin level in hemodialysis patients using machine learning techniques[J]. Computer methods and programs in biomedicine, 2014, 117(2): 208–217.
- [18] DI K C, LI W, YUE Z Y, et al. A machine learning approach to crater detection from topographic data[J]. Advances in space research, 2014, 54(11): 2 419–2 429.
- [19] MA C, ZHANG H H, WANG X F. Machine learning for Big Data analytics in plants[J]. Trends in plant science, 2014, 19(12): 798–808.
- [20] TAYFUN K, CAMBAZOGLU B B, CEVDET A, et al. A machine learning approach for result caching in web search engines[J]. Information processing and management, 2017, 53(4): 834–850.
- [21] CYRIL V, GILLES N, SOTERIS K, et al. Machine learning methods for solar radiation forecasting; a review[J]. Renewable energy, 2017, 105: 569–582.

[责任编辑: 严海琳]