

## 结合栈式监督AE与可变加权ELM的回归预测模型

闫静,张雪英,李凤莲,陈桂军,黄丽霞

(太原理工大学信息与计算机学院,太原 030024)

**摘要:**在现代工业生产过程中,许多关键变量与产品质量或生产效率密切相关,关键变量的实时监测是实现利润最大化及节能降耗的有效途径。针对回归预测任务中目标特征提取不全面、预测精度较低等问题,提出一种基于栈式监督自编码器与可变加权极限学习机的回归预测模型。通过堆叠多层自编码器并在每层自编码器中添加回归网络,同时以有监督方式对栈式自编码器(SAE)进行逐层预训练,得到与输出变量相关的特征表示。利用反向传播算法对网络参数进行微调,优化自编码器模型参数。在分析提取特征与输出变量的相关性基础上,对极限学习机(ELM)的输入权值和偏置进行加权得到预测结果。实验结果表明,与基于ELM和SAE-ELM的回归预测模型相比,该模型在多晶硅铸锭的G6产品数据集上的均方根误差降低0.056 7和0.011 2、决定系数提高0.489 3和0.290 3,具有更高的回归预测准确性及更强的鲁棒性与泛化性能。

**关键词:**自编码器;极限学习机;回归预测;深度学习;特征提取

开放科学(资源服务)标志码(OSID):



**中文引用格式:**闫静,张雪英,李凤莲,等.结合栈式监督AE与可变加权ELM的回归预测模型[J].计算机工程,2022,48(8):62-69,76.

**英文引用格式:**YAN J, ZHANG X Y, LI F L, et al. Regression prediction model combining stack supervised AE and variable weighted ELM[J]. Computer Engineering, 2022, 48(8): 62-69, 76.

### Regression Prediction Model Combining Stack Supervised AE and Variable Weighted ELM

YAN Jing, ZHANG Xueying, LI Fenglian, CHEN Guijun, HUANG Lixia

(School of Information and Computer, Taiyuan University of Technology, Taiyuan 030024, China)

**[Abstract]** Many key variables in the modern industrial production are closely related to the product quality or production efficiency. Monitoring such key variables on a regular basis is an effective method of maximizing profits, saving energy, and reducing consumption. This study presents a regression prediction model based on a Stack Supervised Auto-Encoder (SSupAE) and a variable weighted Extreme Learning Machine (vwELM) to address problems such as inaccurate feature extraction and low prediction accuracy in regression prediction tasks. First, the Stacked Auto-Encoder (SAE) is trained in a supervised manner by stacking multi-layer Auto-Encoders (AEs) and adding a regression network to each layer of the AEs, to obtain the output-related feature representation. Then, the parameters of the SSupAE network are fine-tuned by applying back-propagation to optimize the model parameters of the AEs. Finally, the correlation between the extracted features and the output variables are analyzed, and the input weight and bias of the Extreme Learning Machine (ELM) are weighted to obtain the predicted results. The experimental results show that compared with those of the regression prediction model based on the ELM and SAE-ELM, the Root Mean Square Error (RMSE) of the proposed model on the G6 product dataset of polycrystalline silicon ingots is reduced by 0.056 7 and 0.011 2 and the coefficient of determination ( $R^2$ ) is increased by 0.489 3 and 0.290 3, indicating that the proposed model has better regression prediction accuracy, robustness, and generalization performance.

**[Key words]** Auto-Encoder (AE); Extreme Learning Machine (ELM); regression prediction; deep learning; feature extraction  
**DOI:** 10.19678/j.issn.1000-3428.0062416

## 0 概述

回归分析是一种确定两种或两种以上变量间相互

依赖的定量关系的统计分析方法,根据已知过程变量与目标变量间的相关性,建立基于历史过程数据的回归预测模型。由于目标变量总是受一个或多个过程变

基金项目:山西省科技重大专项(20181102008)。

作者简介:闫静(1996—),女,硕士研究生,主研方向为深度学习、数据分析与处理;张雪英(通信作者),教授、博士、博士生导师;李凤莲,教授、博士;陈桂军,讲师、博士;黄丽霞,副教授、博士。

收稿日期:2021-08-19 修回日期:2021-10-07 E-mail: tyzhangxy@163.com

量的影响,且每个过程变量对目标变量的影响不同,因此根据过程变量对目标变量的精准预测可以为过程监控、优化和控制提供重要的实时信息。回归预测模型多数采用不同的非线性结构来提取数据中包含的信息,常用模型包括主成分回归(Principal Component Regression, PCR)<sup>[1]</sup>、偏最小二乘回归(Partial Least Squares Regression, PLSR)<sup>[2]</sup>、人工神经网络(Artificial Neural Network, ANN)<sup>[3]</sup>和支持向量回归(Support Vector Regression, SVR)<sup>[4]</sup>。但对于大量高维、强相关性及高冗余的数据,这些模型的鲁棒性差、预测性能低,而提取输入数据的有效特征表示是建立回归预测模型的关键步骤。多层深度网络能够提取复杂数据的特征,但由于梯度消失和爆炸问题,深度网络并没有比浅层模型表现得更好,直到文献[5]提出通过无监督的逐层预训练和有监督的微调来学习深度网络模型,使得栈式自编码器(Stacked Auto-Encoder, SAE)成为广泛应用于数据分析<sup>[6]</sup>、图像处理<sup>[7]</sup>、语音识别<sup>[8]</sup>、模式识别<sup>[9]</sup>等领域的深度学习<sup>[10]</sup>模型。

深度学习可以通过学习深层非线性网络结构,实现复杂函数逼近,表征输入数据,并利用特征的逐层变换完成最终的预测和识别<sup>[11]</sup>。文献[12]将卷积神经网络(Convolutional Neural Network, CNN)与极限学习机(Extreme Learning Machine, ELM)相结合,提出CNN2ELM模型,用于人脸图像的年龄预测,提高了预测鲁棒性。文献[13]将栈式降噪稀疏自编码器(sDSAE)与ELM相结合,提出sDSAE-ELM算法,利用sDSAE产生ELM的输入权重和隐含层偏置,降低噪声干扰,优化网络结构。文献[14]将SAE与以小波函数为激活函数的ELM结合,提出SAE-WELM模型并将其用于工业铝生产过程中的过度热预测,具有良好的鲁棒性和泛化能力。

针对回归预测问题,对SAE和ELM两部分进行改进再级联是改善回归预测效果的有效方法。文献[13-15]采用SAE进行特征降维或特征提取,取得了较好的效果,但它们未考虑到数据间的相关性,不能反映出目标变量与其他过程变量之间的关系。目前,关于结合改进的SAE和ELM进行回归预测的研究也取得了一定成果,随机确定输入权值和隐含层偏置虽然能够提升网络速度<sup>[16]</sup>,但不能根据输入数据与输出数据间的相关性大小进行合理赋值。本文构建一种基于栈式监督自编码器(Stack Supervised Auto-Encoder, SSupAE)与可变加权极限学习机(variable weighted Extreme Learning Machine, vwELM)的回归预测模型。利用栈式监督自编码器使SAE以有监督的方式进行逐层预训练,提取与目标输出变量相关的高级特征,挖掘数据间的深层关联信息。采用可变权值的方式确定ELM的

输入权值和隐含层偏置,以提升算法的鲁棒性和泛化能力。在多个公共数据集及实际工业生产的多晶硅铸锭数据集上进行实验以验证SSupAE-vwELM模型性能。

## 1 栈式监督自编码器设计

### 1.1 栈式自编码器

自编码器(Auto-Encoder, AE)包括编码和解码两个过程,编码过程将输入 $x$ 通过非线性激活函数映射到隐含层,解码过程将隐含层数据 $h$ 转化为输出值 $z$ ,再重构输入<sup>[17]</sup>。AE网络结构如图1所示。编码过程、解码过程、损失函数的表达式如式(1)~式(3)所示:

$$h = f(x) = s_f(Wx + b) \quad (1)$$

$$z = g(h) = s_g(\tilde{W}h + \tilde{b}) \quad (2)$$

$$J_{AE}(\theta) = \frac{1}{N} \sum_{i=1}^N L(x_i, z_i) \quad (3)$$

其中: $s_f(x)$ 为激活函数; $W$ 和 $b$ 分别为编码器的权值矩阵和偏置向量; $\tilde{W}$ 和 $\tilde{b}$ 分别为解码器的权值矩阵和偏置向量; $N$ 为训练集样本数; $\theta = \{W, \tilde{W}, b, \tilde{b}\}$ 为AE所需优化的参数集; $L(x_i, z_i) = \frac{1}{2} \|x_i - z_i\|^2$ 为重构误差。通过多次迭代使损失函数达到最小,以得到编码器的输出 $h$ ,损失函数越小,隐含层所包含的信息越多,越能够完整地重构出原始数据。由于以上过程没有目标变量的参与,因此AE的训练属于无监督训练。

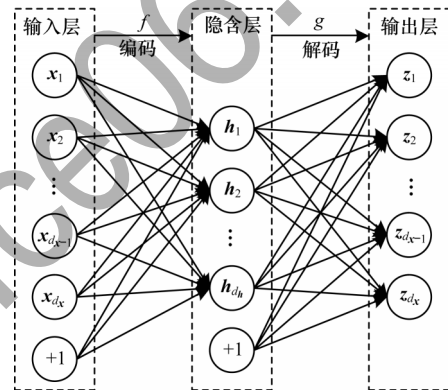


图1 AE网络结构

Fig.1 AE network structure

SAE是通过多层无监督训练的AE逐层堆叠而构造的一种深度网络结构,训练过程分为无监督预训练和有监督微调两个阶段<sup>[18]</sup>,如图2所示。SAE采用无监督的方式逐层预训练来初始化网络参数,在最后一层隐含层后加入BP回归网络进行回归预测,使用目标变量数据 $y$ 对权重和偏置进行整体微调,优化网络结构。

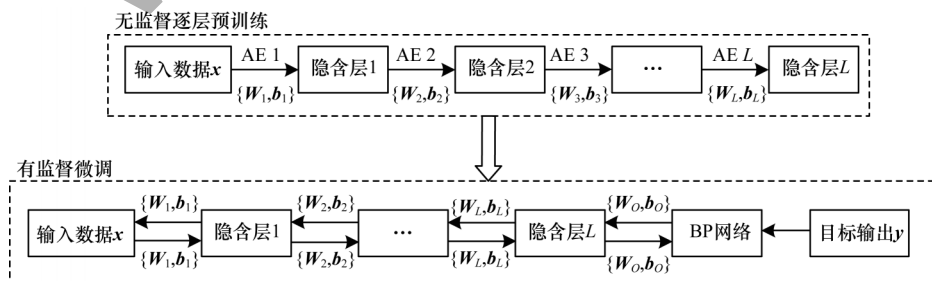


图2 SAE训练过程

Fig.2 SAE training process

1.2 栈式监督自编码器

栈式自编码器的预训练可以逐层学习到输入数据的高级抽象特征,但在实际应用中,SAE的无监督预处理方式未考虑过程变量与目标输出间的相关性,所学习到的特征可能包含与目标输出无关的信息。针对这一问题,提出一种以有监督方式训练的监督自编码器(Supervised Auto-Encoder, SupAE),即在AE编码与解码的基础上添加一层回归网络,AE在解码的同时通过回归网络进行回归预测,使得构成栈式监督自编码器(SSupAE)的每层 SupAE 都以有监督的方式完成预训练,并使该深层网络在学习重构特征的同时将与目标输出变量相关的信息编码到该网络中,挖掘数据的深层特征。

SupAE由编码器、解码器和预测目标输出的回归网络三部分组成,网络结构如图3所示。SupAE的编解码过程与AE相同,其中回归网络预测目标输出值的计算公式如式(4)所示:

$$\hat{y} = s(\mathbf{h}) = s_r(\mathbf{W}_r \mathbf{x} + \mathbf{b}_r) \quad (4)$$

其中:本文使用的回归网络为BP回归网络; $\mathbf{W}_r$ 和 $\mathbf{b}_r$ 分别为BP网络的权值矩阵和偏置向量; $\hat{y}$ 为对目标变量真实值 $y$ 的预测值。

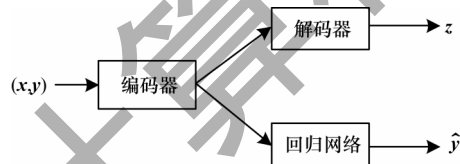


图3 SupAE结构  
Fig.3 SupAE structure

在训练过程中,为了能够对解码重构输入与回归预测输出同时优化,使SupAE获得更好的表示,SupAE的损失函数由重构误差、目标变量的真实值与预测值间的误差两部分组成,通过最小化这两部分的线性组合函数实现对数据的深层挖掘。假设训练集有 $N$ 个样本 $\{\mathbf{x}, \mathbf{y}\} = \{(\mathbf{x}_i, y_i) | \mathbf{x}_i \in \mathbb{R}^{d_x}, y_i \in \mathbb{R}, i = 1, 2, \dots, N\}$ ,其中 $d_x$

表示输入数据的维数,则SupAE的整体代价函数如下:

$$J_{\text{SupAE}(\theta)} = \frac{1}{N} \sum_{\mathbf{x}} ((1 - C)L(\mathbf{x}, \mathbf{z}) + CL(\mathbf{y}, \hat{\mathbf{y}})) \quad (5)$$

$$L(\mathbf{x}, \mathbf{z}) = - \sum_{i=1}^{d_x} [z_i \log_a x_i + (1 - z_i) \log_a (1 - x_i)] \quad (6)$$

$$L(\mathbf{y}, \hat{\mathbf{y}}) = - \sum_{i=1}^N [\hat{y}_i \log_a y_i + (1 - \hat{y}_i) \log_a (1 - y_i)] \quad (7)$$

其中: $L(\mathbf{x}, \mathbf{z})$ 为重构损失函数; $L(\mathbf{y}, \hat{\mathbf{y}})$ 为目标变量的真实值和预测值间的误差损失函数,本文均使用交叉熵损失函数; $C$ 为0~1的常数,用来平衡 $L(\mathbf{x}, \mathbf{z})$ 与 $L(\mathbf{y}, \hat{\mathbf{y}})$ 间的比例。通过平衡重构损失与回归预测值和真实值间的损失来提取输入数据的潜在特征,在一定程度上相当于一种隐式的数据增强,在代价函数中引入 $L(\mathbf{y}, \hat{\mathbf{y}})$ 可以将目标变量值编码到隐含层中,同时将无监督学习的AE转化为有监督学习的AE,使隐含层中包含更多数据的信息,提高模型的泛化能力。

利用反向传播算法结合梯度下降法,更新连接权值 $\mathbf{W}$ 和偏差 $\mathbf{b}$ ,求出使得式(5)达到最小值时的 $\mathbf{W}_{ij}$ 和 $\mathbf{b}_i$ 。更新公式具体如下:

$$\mathbf{W}_{ij} = \mathbf{W}_{ij} - \alpha \frac{\partial J_{\text{SupAE}}(\theta)}{\partial \mathbf{W}_{ij}} \quad (8)$$

$$\mathbf{b}_i = \mathbf{b}_i - \alpha \frac{\partial J_{\text{SupAE}}(\theta)}{\partial \mathbf{b}_i} \quad (9)$$

其中: $\alpha$ 为学习率。通过这种更新权值的方式,获得最优的 $\mathbf{W}$ 和 $\mathbf{b}$ ,使得SupAE隐层学习比较好的隐层表达。

SSupAE是由SupAE通过逐层堆叠构造的一种深度网络结构,如图4所示,其输入是由每个样本对应的过程变量与目标变量值组成,在SupAE进行逐层有监督预训练后,舍弃每层的回归网络和解码器(见图4中点线矩形框部分),以前一个隐含层的输出作为后一个隐含层的输入,通过最小化联合损失函数(见图4中点划线部分),并逐层堆叠以提取包含目标变量信息的高级特征。在最后一个隐含层后添加ELM回归网络,使整个网络再次以有监督的方式进行微调,更新各层的权值和偏置,使该网络达到全局最优。

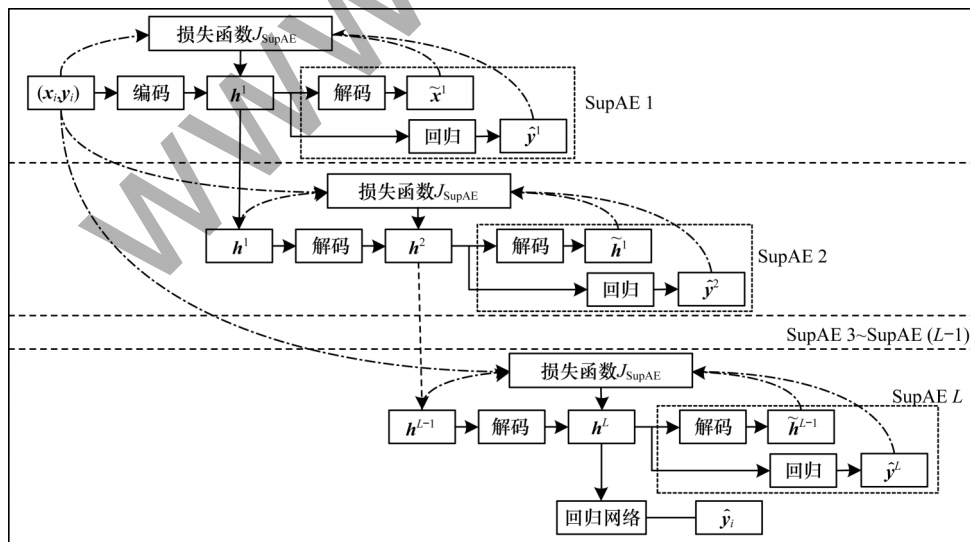


图4 SSupAE结构  
Fig.4 SSupAE structure

## 2 栈式监督自编码器与可变加权极限学习机

### 2.1 极限学习机

ELM是一种单隐层前馈神经网络<sup>[19]</sup>,ELM的网络结构如图5所示。假设训练集有 $N$ 个样本 $\{\mathbf{x}, \mathbf{y}\} = \{(\mathbf{x}_i, \mathbf{y}_i) | \mathbf{x}_i \in \mathbb{R}^{d_x}, \mathbf{y}_i \in \mathbb{R}, i = 1, 2, \dots, N\}$ ,则ELM网络模型可以表示如下:

$$f(\mathbf{x}_i) = \beta \mathbf{g}(\mathbf{W}\mathbf{x}_i + \mathbf{b}) \quad (10)$$

其中: $\mathbf{W}$ 是输入层到隐含层的权值向量; $\mathbf{b}$ 为偏置向量; $\mathbf{g}(\cdot)$ 是激活函数; $\beta$ 是隐含层到输出层的输出权值。ELM的矩阵表达式如式(11)所示。ELM网络的训练过程就是求解式(11)的最小二乘解 $\beta$ ,如式(12)所示。输出权值矩阵 $\beta$ 可由Moore-Penrose广义逆公式求解得到,如式(13)所示。

$$\mathbf{H}\beta = \mathbf{Y} \quad (11)$$

$$\min_{\beta} \|\mathbf{H}\beta - \mathbf{Y}\| \quad (12)$$

$$\beta = \mathbf{H}^{\dagger} \mathbf{Y} \quad (13)$$

其中: $\mathbf{H}$ 表示隐含层的输出矩阵; $\mathbf{Y}$ 表示样本目标输出的真实值矩阵; $\mathbf{H}^{\dagger}$ 是 $\mathbf{H}$ 的广义逆, $\mathbf{H}^{\dagger} = (\mathbf{H}^{\top} \mathbf{H})^{-1} \mathbf{H}^{\top}$ 。

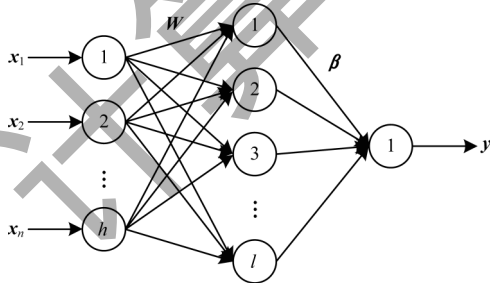


图5 ELM网络结构

Fig.5 ELM network structure

### 2.2 可变加权极限学习机

ELM网络随机确定初始输入权值和偏置,能够提高网络的学习速度,但是在隐含层节点个数一定的情况下,预测精度会受随机性影响<sup>[21]</sup>,因此对权值和偏置进行合理赋值能够提升网络的预测性能。本文将输入变量与目标输出间的相关性融入ELM网络,提出一种根据相关性确定权值与偏置的可变加权极限学习机。

针对回归预测问题,不同的输入变量对目标输出变量的影响不同,对不同变量赋予不同的权值,不仅可以提高ELM训练的精度,而且可以有效提高模型的鲁棒性。对于有 $N$ 个样本的训练数据集 $\{\mathbf{x}, \mathbf{y}\} = \{(\mathbf{x}_i, \mathbf{y}_i) | \mathbf{x}_i \in \mathbb{R}^{d_x}, \mathbf{y}_i \in \mathbb{R}, i = 1, 2, \dots, N\}$ ,样本中第 $j(j \in d_x)$ 个变量与目标变量值的相关系数计算公式如下:

$$\rho_{(j)} = \text{Cov}(\mathbf{x}_{(j)}, \mathbf{y}) / (\sqrt{\text{var}(\mathbf{x}_{(j)})} \cdot \sqrt{\text{var}(\mathbf{y})}) \quad (14)$$

其中: $\mathbf{x}_{(j)}$ 为训练集第 $j$ 个变量的集合,即 $\mathbf{x}_{(j)} = \{\mathbf{x}_{1(j)}, \mathbf{x}_{2(j)}, \dots, \mathbf{x}_{N(j)}\}$ 。协方差和方差的计算公式如下:

$$\text{Cov}(\mathbf{x}_{(j)}, \mathbf{y}) = \sum_{i=1}^N (\mathbf{x}_{i(j)} - \bar{\mathbf{x}}_{(j)}) (\mathbf{y}_i - \bar{\mathbf{y}}) / (N-1) \quad (15)$$

$$\text{var}(\mathbf{x}_{(j)}) = \sum_{i=1}^N (\mathbf{x}_{i(j)} - \bar{\mathbf{x}}_{(j)})^2 / (N-1) \quad (16)$$

$$\text{var}(\mathbf{y}) = \sum_{i=1}^N (\mathbf{y}_i - \bar{\mathbf{y}})^2 / (N-1) \quad (17)$$

其中: $\bar{\mathbf{x}}$ 和 $\bar{\mathbf{y}}$ 分别是输入数据第 $j$ 维变量和目标变量值的平均值。本文采用的可变权值的计算公式如下:

$$\lambda_{(j)} = |\rho_{(j)}| / \sum_{j=1}^{d_x} |\rho_{(j)}| \quad (18)$$

其中: $\lambda_{(j)}$ 表示第 $j$ 维变量的可变权值。

在对ELM的输入权值和偏置进行初始化时,用可变权值 $\lambda_{(j)}$ 分别对相应的输入变量进行加权,则vwELM网络模型表示如下:

$$f(\mathbf{x}_i) = \beta \mathbf{g}(\lambda \mathbf{W}\mathbf{x}_i + \lambda \mathbf{b}) \quad (19)$$

综上,vwELM算法的训练过程如下:

- 1) 计算输入层输入变量与目标输出变量的相关系数,根据相关系数求得每个变量的可变权值。
- 2) 确定隐含层神经元个数,对输入层和隐含层之间的连接权重 $\mathbf{W}$ 和偏置 $\mathbf{b}$ 进行加权初始化。
- 3) 选择一个无限可微的函数作为隐含层神经元的激活函数,计算隐含层的输出矩阵 $\mathbf{H}$ 。
- 4) 根据式(11)计算输出层权值 $\beta$ 。

### 2.3 基于SSupAE-vwELM的回归预测模型

理论上,SSupAE-vwELM算法能比ELM算法实现更精准的预测。一方面,利用SSupAE网络对原始输入数据进行特征提取,所提取的特征包含了目标输出的相关信息。另一方面,vwELM算法通过相关性分析对ELM的权值和偏置加权,既克服了ELM因参数随机赋值产生冗余节点<sup>[22]</sup>,又使其包含了目标输出的相关信息,有利于实现更加精准的预测。

SSupAE-vwELM网络结构如图6所示,将训练好的 $n$ 层SupAE进行堆叠形成SSupAE,以SSupAE的顶层作为vwELM网络的输入进行回归预测。

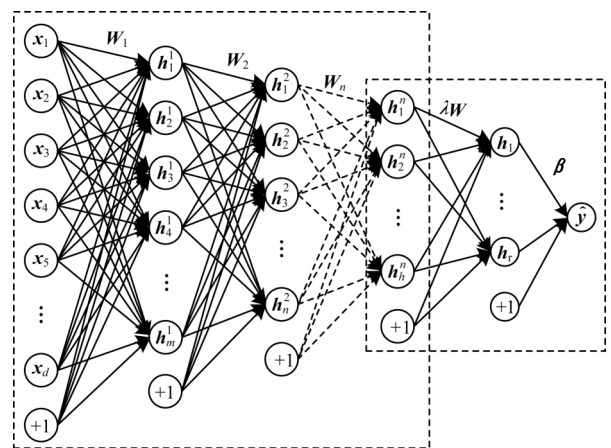


图6 SSupAE-vwELM网络结构

Fig.6 SSupAE-vwELM network structure

1)特征提取。首先针对不同特征维数的数据集  $\{x_i, y_i\}$ , 将原始数据  $x_i$  输入到 SSUpAE 网络中, 对 SSUpAE 网络的每个隐含层节点数设置合适的值, 并对每层的 SupAE 权重和偏置初始化, 分别设置学习率、正则化参数和学习率、丢弃率。在训练中引入目标变量值  $y_i$  使 SSUpAE 以有监督的方式完成训练, 提取输入数据的深层相关特征。

2)回归预测。以 SSUpAE 所提取的特征作为 vwELM 的输入, 根据输入变量与目标变量  $y_i$  值间的相关性计算对应的可变权值, 对 vwELM 的输入权值进行加权, 训练 vwELM 网络, 得到输出权值。

SSUpAE-vwELM 网络训练与测试过程如图 7 所示。

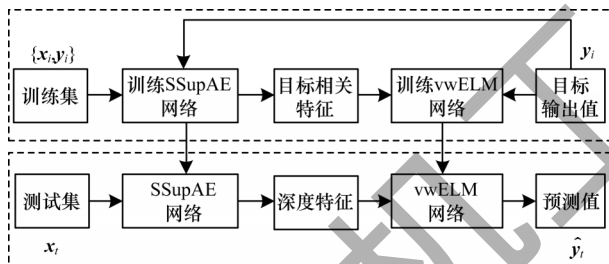


图 7 SSUpAE-vwELM 网络训练与测试过程

Fig.7 Training and testing process of SSUpAE-vwELM

### 3 实验与结果分析

#### 3.1 实验环境与评价指标

应用 MATLAB R2014b 进行实验仿真, 操作系统为 Windows10, 处理器为 Intel Xeon E3-1535M, 内存为 32 GB。采用均方根误差 (Root Mean Square Error, RMSE)、决定系数 ( $R^2$ ) 和程序运行时间 3 个指标对模型回归性能进行评价, RMSE 和  $R^2$  的计算公式如式 (20) 和式 (21) 所示:

$$R_{\text{RMSE}} = \sqrt{\sum_{n=1}^{N_t} (y_n - \hat{y}_n)^2 / (N_t - 1)} \quad (20)$$

$$R^2 = 1 - \frac{\sum_{n=1}^{N_t} (y_n - \hat{y}_n)^2}{\sum_{n=1}^{N_t} (y_n - \bar{y})^2} \quad (21)$$

其中:  $N_t$  为测试结果个数;  $y_n$  和  $\hat{y}_n$  分别为真实值和预测值;  $\bar{y}$  为测试集真实值的平均数。在回归预测中, RMSE 值越小,  $R^2$  值越接近于 1, 预测越精确, 本文通过 RMSE 和  $R^2$  对模型预测结果进行综合对比, 验证模型的预测准确性。

#### 3.2 公共数据集上的实验结果

##### 3.2.1 数据集介绍

为验证本文所提 SSUpAE-vwELM 模型的有效性, 选用 10 个样本大小和属性维度不同的公共数据集, 具体信息如表 1 所示, 其中, Abalone 数据集通过物理测量变量预测鲍鱼年龄, Air Quality 数据集是对意大利某严重污染区域的空气质量进行预测, Boston Housing 数据集通过影响房价的变量预测房价, Concrete 数据集通过混凝土成分预测混凝土的抗压强度, Stocks 数据集是预测 10 家航天公司的股票价格, Bank 数据集是预测客户选择银行的概率,

Computer Activity 数据集是预测电脑 CPU 的运行时间, Kinematics 数据集是预测人体的运动数据, Wine Quality 数据集是预测葡萄牙北部葡萄酒的质量, Yacht Hydrodynamics 数据集是对帆船水力性能的预测。为了解决数据特征属性间数值量纲差异导致的计算问题, 本文将所有数据归一化为  $[0, 1]$ , 并将每个数据集按 8:2 的比例划分训练集和测试集。

表 1 公共数据集信息设置

Table 1 Setting of public dataset information

数据集	样本数	属性维数
Abalone	4 177	8
Air Quality	9 358	14
Boston Housing	506	13
Concrete	1 030	8
Stocks	950	10
Bank	4 499	8
Computer Activity	8 192	21
Kinematics	8 192	8
Wine Quality	4 898	11
Yacht Hydrodynamics	308	6

##### 3.2.2 参数设置

为分析 SSUpAE-vwELM 模型中不同网络参数对整体回归预测性能的影响, 以 Concrete 数据集为例, 对比不同网络层数的 SSUpAE 以及不同隐含层节点数的 vwELM 对整体回归预测准确性的影响。SSUpAE 的输入层节点数与归一化处理后的输入数据特征数保持一致, 设置为 8, 预训练的 batchsize 设置为 80, epoch 设置为 100; 微调的 batchsize 设置为 8, epoch 设置为 1 000。通过 SSUpAE 网络层数对比实验来确定网络结构, 选取 RMSE 及  $R^2$  作为评价指标, 将网络层数从 3 变化到 8, 如图 8 所示, 可以看出 5 层网络结构的 RMSE 最小, 真实值与预测值的拟合度最好, 其中每层的隐含层节点数通过试错法确定, 分别为 40、30、20、10、5。

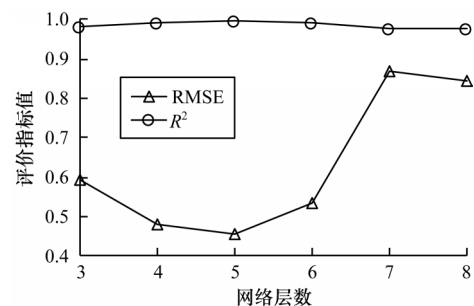


图 8 不同网络层数的 SSUpAE 回归预测性能

Fig.8 Regression prediction performance of SSUpAE at different number of network layers

通过 vwELM 网络隐含层节点数的对比实验确定隐含层节点数, 同样选取 RMSE 及  $R^2$  作为评价指标, 如图 9 所示, 将隐含层节点数从 1 变化到 50, 可以看出隐含层节点数设置为 35 时 RMSE 最小, 真实值与预测值的拟合度最好。

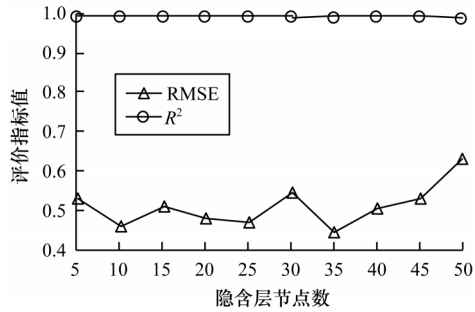


图9 不同隐含层节点数的vwELM回归预测性能  
Fig.9 Regression prediction performance of vwELM at different number of hidden layer nodes

3.2.3 结果分析

为验证SSupAE-vwELM模型的回归预测性能,将ELM、SAE-ELM、SAE-vwELM、SSupAE-ELM及SSupAE-vwELM模型的实验结果在10个公共数据集上进行对比,其中,ELM为未进行特征提取的回归预测模型,SAE-ELM为使用SAE进行特征提取后使用ELM进行回归预测的基础模型,SAE-vwELM为在SAE-ELM模型基础上改进ELM后的模型,SSupAE-ELM为在SAE-ELM基础上改进SAE后的模型。采用五折交叉方式验证模型的预测效果,最终对5次预测结果取平均值。实验结果如表2所示,其中最优结果加粗表示。

表2 公共数据集上的回归预测结果对比  
Table 2 Comparison of regression prediction results on public dataset

数据集	评价指标	ELM	SAE-ELM	SAE-vwELM	SSupAE-ELM	SSupAE-vwELM
Abalone	RMSE	0.967 7	0.604 8	0.401 8	0.291 0	<b>0.277 2</b>
	R <sup>2</sup>	0.898 3	0.924 2	0.935 3	0.941 5	<b>0.942 3</b>
	运行时间/s	<b>7.950 1</b>	8.159 8	8.098 0	9.122 2	12.641 6
Air Quality	RMSE	0.031 7	0.031 1	0.029 1	0.027 8	<b>0.027 1</b>
	R <sup>2</sup>	0.955 4	0.954 5	0.961 8	0.963 0	<b>0.974 2</b>
	运行时间/s	<b>21.153 6</b>	27.932 0	28.393 8	32.796 3	34.987 2
Boston Housing	RMSE	0.710 1	0.530 4	0.423 4	0.374 2	<b>0.280 3</b>
	R <sup>2</sup>	0.882 6	0.930 3	0.932 0	0.934 7	<b>0.944 7</b>
	运行时间/s	<b>2.891 4</b>	3.367 8	3.738 0	3.025 9	4.812 9
Concrete	RMSE	0.822 7	0.801 5	0.800 5	0.778 1	<b>0.722 8</b>
	R <sup>2</sup>	0.979 2	0.980 3	0.983 0	0.983 9	<b>0.986 2</b>
	运行时间/s	<b>5.059 8</b>	10.098 0	10.159 8	12.613 9	14.641 6
Stocks	RMSE	3.024 0	2.433 5	2.323 3	2.302 7	<b>2.027 8</b>
	R <sup>2</sup>	0.972 4	0.981 9	0.983 7	0.984 0	<b>0.994 8</b>
	运行时间/s	<b>3.293 5</b>	10.585 3	10.576 2	10.678 0	12.728 2
Bank	RMSE	0.122 9	0.107 2	0.094 5	0.070 3	<b>0.066 8</b>
	R <sup>2</sup>	0.784 1	0.930 0	0.936 6	0.935 8	<b>0.950 1</b>
	运行时间/s	<b>10.103 9</b>	14.854 7	13.132 8	13.050 5	15.507 4
Computer Activity	RMSE	0.016 2	0.015 6	0.014 8	0.012 8	<b>0.009 0</b>
	R <sup>2</sup>	0.792 8	0.803 8	0.820 7	0.857 4	<b>0.901 2</b>
	运行时间/s	<b>15.842 4</b>	30.932 0	34.794 5	41.997 3	48.603 4
Kinematics	RMSE	1.490 7	1.023 5	1.006 6	1.001 0	<b>0.965 0</b>
	R <sup>2</sup>	0.912 3	0.958 7	0.960 1	0.970 4	<b>0.978 7</b>
	运行时间/s	<b>16.834 3</b>	29.533 2	30.913 2	33.477 4	35.848 8
Wine Quality	RMSE	0.208 7	0.205 8	0.193 7	0.171 5	<b>0.168 8</b>
	R <sup>2</sup>	0.987 1	0.990 7	0.991 9	0.993 6	<b>0.999 1</b>
	运行时间/s	<b>8.313 4</b>	18.327 9	19.618 3	21.214 2	28.336 1
Yacht Hydrodynamics	RMSE	0.739 9	0.513 5	0.340 2	0.335 7	<b>0.308 2</b>
	R <sup>2</sup>	0.983 1	0.988 1	0.988 6	0.992 6	<b>0.996 4</b>
	运行时间/s	1.844 2	2.538 2	2.799 0	<b>1.828 9</b>	1.917 5

从表2可知,在10个公共数据集上SSupAE-vwELM模型相比其他模型的回归预测性能都有所提升。SSupAE-vwELM模型的运行时间长于ELM及SAE-ELM模型的主要原因在于SSupAE的深层

网络不仅能够重构原始数据,而且还将目标输出变量的信息编码到网络中,随着网络层数的增加,其深层非线性网络将原始数据一层一层抽象,所提取的特征更能描述对象本质且提高预测精度,并且通过

结合vwELM回归网络进一步优化了网络结构,提高了网络的鲁棒性和回归预测能力。由此可见,SSupAE-vwELM模型运行时间长说明其相比于ELM及SAE-ELM模型提取的特征更加符合样本本质,鲁棒性更好。以Concrete数据集为例,测试集上部分样本的预测值与真实值的对比结果如图10所示,可以看出除第3个和第9个测试样本外,其余样本的真实值与预测值间的误差很小,可见本文模型的回归预测性能较好。

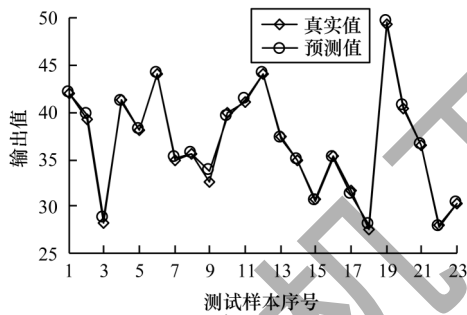


图10 测试集上的预测值与真实值的对比

Fig.10 Comparison between predicted values and actual values on the test set

### 3.3 多晶硅铸锭数据集上的实验结果

为验证SSupAE-vwELM模型的实用性,将其在工业多晶硅铸锭数据集上进行实验。多晶硅作为最主要的光伏产业材料之一,配料数据对多晶硅铸锭的电学性能和生产成本有着重要的影响,但由于每次生产所用配料的批次或重量的差异,会对质量产生影响,因此准确的配料分析和预测模型的建立至关重要<sup>[23]</sup>。少子寿命值即硅锭中少数载流子存活时间,通常被用作评价多晶硅铸锭的质量,根据配料对少子寿命值的准确预测可以有效地指导实际生产。工业上通常用工艺试验来预测产品质量,实现的成本高且难度大。因此,采用深度学习对多晶硅铸锭过程中的少子寿命值进行精准预测对提高产品质量具有重要意义。

本文使用的多晶硅铸锭数据集来源于山西某新能源技术有限公司的实际生产数据,该数据集包括G6和G7两种产品,每种产品包含非免洗原生多晶块料、碎多晶铺底、碎片、中料、提纯锭芯自产、提纯锭芯外购、循环料等7个配料类别,通过属性值评价各种配料的质量,最终所需预测的目标变量为多晶硅的少子寿命值。本文所用到的数据集中G6产品有500个样本,G7产品有391个样本。表3为部分G6数据的示例。

表3 部分G6数据示例

Table 3 Partial G6 data examples

非免洗原生多晶块料	碎多晶铺底	碎片	中料	配料属性值			少子寿命值
				提纯锭芯自产	提纯锭芯外购	循环料	
400	25	50	30	175	225	190	6.45
430	25	25	25	0	60	200	6.51
290	25	0	0	205	30	180	6.62
260	25	20	30	215	190	190	6.74
300	20	50	40	145	130	210	6.70
270	25	20	20	205	30	170	6.95
40	25	25	40	195	215	190	6.56
50	25	50	40	100	30	160	6.70
400	25	0	40	30	0	190	6.79

将多晶硅配料数据按8:2分为训练集与测试集,同样将ELM、SAE-ELM、SAE-vwELM、SSupAE-ELM及SSupAE-vwELM模型的实验结果在G6和G7产品数据集上进行对比,实验结果如表4所示。

表4 多晶硅铸锭数据集上的回归预测结果对比

Table 4 Comparison of regression prediction results on polycrystalline silicon ingot dataset

产品数据集	评价指标	ELM	SAE-ELM	SAE-vwELM	SSupAE-ELM	SSupAE-vwELM
G6	RMSE	0.093 4	0.047 9	0.046 0	0.037 6	<b>0.036 7</b>
	$R^2$	0.121 5	0.320 5	0.355 0	0.604 4	<b>0.610 8</b>
	运行时间/s	<b>1.419 4</b>	3.059 8	3.504 2	2.890 3	4.270 4
G7	RMSE	0.052 5	0.048 0	0.047 0	0.043 3	<b>0.041 7</b>
	$R^2$	0.219 5	0.326 1	0.335 4	0.479 2	<b>0.516 7</b>
	运行时间/s	<b>2.017 6</b>	3.565 3	3.645 1	2.069 2	4.018 7

从表4可以看出,与ELM、SAE-ELM模型相比,SSupAE-vwELM模型虽然运行时间增加,但回归性能在多晶硅铸锭的G6产品数据集中RMSE降低了0.056 7、0.011 2, $R^2$ 提升了0.489 3、0.290 3;在G7产品数据集中RMSE降低了0.010 8、0.006 3, $R^2$ 提升了0.297 2、0.190 6。比较表2和表4中5种模型的预测结果,在多晶硅铸锭数据集上的回归预测结果整体比公共数据集差,主要原因为在实际铸锭生产过程中,记录的不规范和缺失,导致数据中出现异常数据和缺失数据,且每次生产所用的配料的批次或成分的差异,使用同样质量的配料会出现不同少子寿命值的情况,导致最终的预测值与真实值的决定系数较低,但是表2和表4中SSupAE-vwELM模型的预测结果优于其他模型结果的趋势是一致的。

## 4 结束语

为了学习输入数据的显著表征,实现对输出变量的精准预测,本文提出基于SSupAE-vwELM的回归预

测模型。利用SSupAE提取与目标输出变量相关的高级特征,将所提取的特征作为vwELM的输入数据,并根据原始数据的特征表示与输出数据间的相关性大小对ELM的权值和偏置进行加权,解决了回归预测任务中目标特征提取不准确、预测精度低等问题。在多个公共数据集及实际工业生产的多晶硅铸锭数据集上的实验结果表明,与ELM及SAE-ELM模型相比,SSupAE-vwELM模型具有较强的鲁棒性和泛化性能。由于在实验过程中发现SSupAE网络隐含层数、节点数以及vwELM网络隐含层节点数的设置对实验结果影响较大,因此后续将继续研究如何合理准确地设置网络参数,进一步提升SSupAE-vwELM模型的回归性能,使其适用于实际工业生产。

### 参考文献

- [ 1 ] 王鹏新,陈弛,张树誉,等. 基于LAI和VTCI及Copula函数的冬小麦单产估测[J]. 农业机械学报,2021,52(10):255-263.  
WANG P X, CHEN C, ZHANG S Y, et al. Winter wheat yield estimation based on Copula function and remotely sensed LAI and VTCI[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(10):255-263. (in Chinese)
- [ 2 ] 陈思博,潘晓文,刘金福. 基于偏最小二乘法分形计盒维数的冲击定位方法[J]. 振动与冲击,2021,40(2):97-102.  
CHEN S B, PAN X W, LIU J F. Impact localization method based on the partial least squares regression fractal dimension[J]. Journal of Vibration and Shock, 2021, 40(2):97-102. (in Chinese)
- [ 3 ] 王毅红,张建雄,兰官奇,等. 压制生土砖强度的人工神经网络预测模型[J]. 华南理工大学学报(自然科学版),2020,48(7):115-121.  
WANG Y H, ZHANG J X, LAN G Q, et al. Artificial neural network prediction model for compressive strength of compacted earth blocks [J]. Journal of South China University of Technology (Natural Science Edition), 2020, 48(7):115-121. (in Chinese)
- [ 4 ] 闫长斌,汪鹤健,周建军,等. 基于Bootstrap-SVR-ANN算法的TBM施工速度预测[J]. 岩土工程学报,2021,43(6):1078-1087.  
YAN C B, WANG H J, ZHOU J J, et al. Prediction of TBM advance rate based on Bootstrap method and SVR-ANN algorithm[J]. Chinese Journal of Geotechnical Engineering, 2021, 43(6):1078-1087. (in Chinese)
- [ 5 ] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7):1527-1554.
- [ 6 ] HASSAN R U L, LI C G, LIU Y T. Online dynamic security assessment of wind integrated power system using SDAE with SVM ensemble boosting learner[J]. International Journal of Electrical Power & Energy Systems, 2021, 125:106429.
- [ 7 ] 饶利波,庞涛,纪然仕,等. 基于高光谱成像技术结合堆栈自动编码器-极限学习机方法的苹果硬度检测[J]. 激光与光电子学进展,2019,56(11):247-253.  
RAO L B, PANG T, JI R S, et al. Firmness detection for apples based on hyperspectral imaging technology combined with stack autoencoder-extreme learning machine method[J]. Laser & Optoelectronics Progress, 2019, 56(11):247-253. (in Chinese)
- [ 8 ] KHALIL R A, JONES E, BABAR M I, et al. Speech emotion recognition using deep learning techniques: a review[J]. IEEE Access, 2019, 7:117327-117345.
- [ 9 ] HOU L, LUO X Y, WANG Z Y, et al. Representation learning via a semi-supervised stacked distance autoencoder for image classification [J]. Frontiers of Information Technology & Electronic Engineering, 2020, 21(7):1005-1018.
- [ 10 ] HAO X, ZHANG G G, MA S. Deep learning[J]. International Journal of Semantic Computing, 2016, 10(3):417-439.
- [ 11 ] TANYILDIZI H, ŞENGÜR A, AKBULUT Y, et al. Deep learning model for estimating the mechanical properties of concrete containing silica fume exposed to high temperatures [J]. Frontiers of Structural and Civil Engineering, 2020, 14(6):1316-1330.
- [ 12 ] DUAN M X, LI K L, LI K Q. An ensemble CNN2ELM for age estimation [J]. IEEE Transactions on Information Forensics and Security, 2018, 13(3):758-772.
- [ 13 ] 张国令,王晓丹,李睿,等. 基于栈式降噪稀疏自编码器的极限学习机[J]. 计算机工程,2020,46(9):61-67.  
ZHANG G L, WANG X D, LI R, et al. Extreme learning machine based on stacked denoising sparse auto-encoder[J]. Computer Engineering, 2020, 46(9):61-67. (in Chinese)
- [ 14 ] LEI Y X, KARIMI H R, CEN L H, et al. Processes soft modeling based on stacked autoencoders and wavelet extreme learning machine for aluminum plant-wide application[J]. Control Engineering Practice, 2021, 108:104706.
- [ 15 ] YIN J, YAN X F. Stacked sparse autoencoders monitoring model based on fault-related variable selection[J]. Soft Computing, 2021, 25(5):3531-3543.
- [ 16 ] 翟华伟,崔立成,张维石. 一种改进灵敏度分析的在线自适应极限学习机算法[J]. 小型微型计算机系统,2019,40(7):1386-1390.  
ZHAI H W, CUI L C, ZHANG W S. Novel online adaptive algorithm of extreme learning machine based on improved sensitivity analysis [J]. Journal of Chinese Computer Systems, 2019, 40(7):1386-1390. (in Chinese)
- [ 17 ] VINCENT P, LAROCHELLE H, BENGIO Y, et al. Extracting and composing robust features with denoising autoencoders [C]//Proceedings of the 25th international conference on Machine learning. New York, USA: ACM Press, 2008:1-8.
- [ 18 ] 普运伟,郭江,刘涛涛,等. 基于模糊函数等高线与栈式降噪自编码器的雷达辐射源信号识别[J]. 仪器仪表学报,2021,42(1):207-216.  
PU Y W, GUO J, LIU T T, et al. Radar emitter signal recognition based on ambiguity function contour lines and stacked denoising auto-encoders[J]. Chinese Journal of Scientific Instrument, 2021, 42(1):207-216. (in Chinese)
- [ 19 ] HUANG G B, ZHU Q Y, SIEW C K. Extreme learning machine: theory and applications [J]. Neurocomputing, 2006, 70(1/2/3):489-501.

(上接第 69 页)

- [20] YANG L Q, ZHANG J W, WANG X Z, et al. An improved ELM-based and data preprocessing integrated approach for phishing detection considering comprehensive features[J]. Expert Systems With Applications, 2021, 165: 113863.
- [21] HUANG Z Y, YU Y L, GU J, et al. An efficient method for traffic sign recognition based on extreme learning machine[J]. IEEE Transactions on Cybernetics, 2017, 47(4): 920-933.
- [22] 王晓丹, 来杰, 李睿, 等. 多层去噪极限学习机[J]. 吉林大学学报(工学版), 2020, 50(3): 1031-1039.
- WANG X D, LAI J, LI R, et al. Multilayer denoising extreme learning machine[J]. Journal of Jilin University (Engineering and Technology Edition), 2020, 50(3): 1031-1039. (in Chinese)
- [23] 徐静林, 黄丽霞, 张雪英, 等. NRS-SVM 两阶段遗传算法的多晶硅铸锭配料质量分析[J]. 太原理工大学学报, 2021, 52(3): 417-423.
- XU J L, HUANG L X, ZHANG X Y, et al. Quality analysis of polysilicon ingot batching using NRS-SVM two-stage genetic algorithm[J]. Journal of Taiyuan University of Technology, 2021, 52(3): 417-423. (in Chinese)

编辑 陆燕菲