

基于 GA-BPNN 混合智能模型的钻速预测^{*}邱腾煌^{1,2} 钱玉宝^{1,2} 季 威³(1. 长江大学机械工程学院 荆州 434023; 2. 油气钻采工程湖北省重点实验室(长江大学) 荆州 434023;
3. 中海油田服务股份有限公司 三河 065201)

摘 要: 石油勘探和开发领域中,准确预测机械钻速对于提高钻井效率和降低工程风险至关重要。准确的机械钻速预测为制定钻井方案、评估钻井风险提供重要依据,但对于复杂的非线性的钻井系统,传统的钻速方程和机器学习方法无法全面考虑影响机械钻速的因素。本文基于一种遗传算法优化的反向传播神经网络的机械钻速预测模型,以中国南海某油田历史钻井数据为基础,通过 SG 平滑处理,归一化处理和 Pearson、Spearman 和 Kendall 相关系数综合分析进行特征参数选择的数据预处理,与 BP、RBF、MEA-BP 神经网络模型以及 ELM、RF、SVM、KNN 等传统机器学习方法进行比较验证。实验结果表明,GA-BP 的 R^2 为 0.967,预测值与实测值具有良好的一致性,比标准 BP 神经网络预测 R^2 精确提高了 17.64%,也较其他模型具有更准确的预测结果。这种混合智能预测模型能够准确预警和预防钻井事故,为指导油田钻井施工参数提供有效数据,从而提高钻井施工的经济效益。

关键词: SG 平滑处理;机械钻速;相关系数;反向传播神经网络;遗传算法;机器学习

中图分类号: TN919.5;TE355 **文献标识码:** A **国家标准学科分类代码:** 520.2060

Drilling rate prediction based on GA-BPNN hybrid intelligent model

Qiu Tenghuang^{1,2} Qian Yubao^{1,2} Ji Wei³

(1. College of Mechanical Engineering, Yangtze University, Jingzhou 434023, China;

2. Hubei Key Laboratory of Oil and Gas Drilling and Production Engineering (Yangtze University), Jingzhou 434023, China;

3. China Oilfield Services Limited, Sanhe 065201, China)

Abstract: In the field of oil exploration and development, accurate prediction of mechanical drilling rate is crucial for improving drilling efficiency and reducing engineering risks. Accurate mechanical drilling rate prediction provides an important basis for formulating drilling plans and assessing drilling risks. However, traditional drilling rate equations and machine learning methods cannot fully consider the factors affecting the mechanical drilling rate in complex nonlinear drilling systems. This paper presents a mechanical drilling rate prediction model based on a genetic algorithm optimized backpropagation neural network (GA-BPNN), using historical drilling data from an oil field in the South China Sea. The data preprocessing includes SG smoothing, normalization, and comprehensive feature parameter selection through Pearson, Spearman, and Kendall correlation coefficients. The model is compared and verified with BP, RBF, MEA-BP neural network models, and traditional machine learning methods such as ELM, RF, SVM, and KNN. The experimental results show that the GA-BP has an R^2 of 0.967, and the predicted values are in good agreement with the measured values, with an accuracy increase of 17.64% compared to the standard BP neural network prediction R^2 , and more accurate predictions than other models. This hybrid intelligent prediction model can accurately predict and prevent drilling accidents, provide effective data for guiding oil field drilling construction parameters, thereby improving the economic benefits of drilling construction.

Keywords: SG smoothing; mechanical drilling rate; correlation coefficients; backpropagation neural network; genetic algorithm; machine learning

0 引言

随着全球能源需求的持续增长和传统油气资源的逐渐枯

竭,油气勘探领域逐步向低渗透、深层超深层、深水以及页岩油气等非常规资源拓展^[1-2],而这些资源的开采难度大、成本高,因此,提高钻井效率和降低开采成本成为亟待解决的问题。

油气开采领域中,机械钻速(rate of penetration, ROP)的预测作为衡量钻井效率的关键参数^[3],用于描述钻头在地层中钻进的速度,以米每小时(m/h)为单位,其预测精度直接影响到钻井方案的制定和风险评估。目前有许多直接和间接的传统方法来预测和优化 ROP,这些方法主要基于影响 ROP 的许多变量的不同相互作用,但该过程主要依赖于经验公式和地质参数的简单计算^[4]。然而现有的许多 ROP 数学模型大多涉及大量物理实验,在面对复杂多变的地质环境时,往往难以提供足够的预测精度和适应性,存在现场实际局限性^[5]。

随着大数据与人工智能的兴起,数据驱动的预测分析在其他领域取得了显著成功,其在石油和天然气开采中的应用也引起了广泛关注^[6-7]。目前国内外一些研究者已经开始尝试将机器学习技术应用于油气开采,例如,赵颖等^[8]以南海某井的钻井参数建立基于极限学习机(extreme learning machine, ELM)的机械钻速预测模型,通过优化模型参数可有效避免钻井事故的发生。宋先知等^[9]分别基于决策树(decision tree, DT)、随机森林(random forest, RF)、支持向量机(support vector machine, SVM)和全连接神经网络(fully connected neural network, FCNN)建立了机械钻速智能预测模型并通过预测结果和模型复杂度对比分析,得出 SVM 有更稳定的预测精度。神经网络在机器学习中也广泛应用于机械钻速的预测, Ahmed 等^[3]讨论了人工神经网络(artificial neural network, ANN)、ELM、SVM 等模型在钻速预测中的可行性,进一步推广了人工智能的应用范围。刘胜娃等^[10]基于海量数据,建立 ANN 技术的定向井机械钻速预测模型,验证了充足而精准的测量数据能使神经网络预测准确度显著提高。这些研究大多局限于初步探索阶段,尚未形成系统化的理论体系和应用模式。在实际应用中仍存在一定的局限性和不足,一些研究在预测 ROP 时使用了大量钻井变量,其中一些变量与机械钻速没有直接的物理关系,有些是推论,从而使模型填充了冗余或不相关的数据。尤其是在处理复杂地质条件时,单一的机器学习方法往往容易陷入局部最优,且预测结果的稳定性较差,因此优化传统的人工智能预测模型是当前的研究重点。

本文以遗传算法(genetic algorithm, GA)优化 BP 神经网络(backpropagation neural network, BPNN)的算法结合 GA 全局搜索能力和反向传播算法自适应学习能力的优点^[11],能够针对各种特征和多种影响因素创建独特的混合智能预测模式形成适应钻采过程的 ROP 预测方法^[12],可以更好地适应不同的地质环境和工况变化,进一步提高机械钻速预测精度。

1 数据预处理

本研究所使用的数据源自中国南海某油田的现场钻井施工数据,总计 1 040 个数据点,覆盖井深 1 253~3 339 m 范围,数据采集频率为每 2 m 一次。对于数据中的缺失值,

采用插值法进行补全,确保数据的完整性和连续性。数据涉及的参数包括机械钻速、井深、转速、钻压、钻井液排量、漏斗黏度、钻头压降、井径、泥浆质量以及温度等。在预测模型构建过程中,数据预处理和特征选择至关重要,因此,在数据分析和模型构建之前,必须充分清洗和处理数据,以确保模型的准确性和可靠性^[13]。

1.1 Savitzky-Golay 卷积平滑法

由于钻井数据在采集过程中容易受到噪声干扰,影响模型训练效果。为此,本文采用萨维茨基-戈莱平滑滤波器(savitzky-golay, SG)对原始数据进行降噪处理^[14-15]。该方法通过在局部窗口内对数据进行多项式拟合,有效减少了随机波动,同时保留了数据的原始趋势,特别适合处理非周期性和非线性噪声样本,图 1 展示了 SG 卷积平滑法的工作原理。

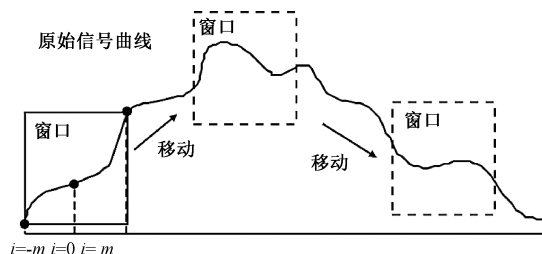


图 1 SG 卷积平滑法的窗口移动原理

Fig. 1 Principle of window movement in SG convolution smoothing method

SG 平滑法效果的程度与选取窗口宽度的不同有关系,并且满足多种不同情况的去噪需求。在位置 k 处的平滑值 \bar{X}_k 是通过对原始信号 X 在窗口范围内的数据进行平均得到的,如式(1)所示。

$$\bar{X}_k = \bar{X}_k = \frac{1}{2n+1} \sum_{i=-n}^n X_{k+i} \quad (1)$$

一般设定平滑的窗口宽度是 $w = 2n+1$, 为正奇数,滑动窗口中包含的数据点为 $x = (-n, -n+1, 0, 1, n-1, n)$, 最后使用 $k-1$ 次多项式对平滑窗口内的所有数据点进行高阶多项式拟合,如式(1)所示。

$$f_w = \sum_{i=0}^q a_i x_i = a_0 + a_1 x + a_2 x^2 + \cdots + a_{k-1} x^{k-1} \quad (2)$$

式中: q 为多项式次数, $q = k-1$; a_i 为第 i 个数据的系数; x_i 为窗口中第 i 个数据点。

窗口移动一次可以写出 w 个方程,最小二乘法拟合残差 e ,如式(2)所示。

$$\begin{pmatrix} f_{-n} \\ f_{-n+1} \\ \vdots \\ f_n \end{pmatrix} = \begin{pmatrix} 1 & -n & \cdots & (-n)^{q-1} \\ 1 & -n+1 & \cdots & (-n+1)^{q-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & n & \cdots & (n)^{q-1} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{k-1} \end{pmatrix} + \begin{pmatrix} e_{-n} \\ e_{-n+1} \\ \vdots \\ e_n \end{pmatrix} \quad (3)$$

即 $f = XA + E$, A 的最小二乘解 $A' = (X^T X)^{-1} X^T f$, f 为降噪前的信号,故降噪后 $f' = XA' = X(X^T X)^{-1} X^T f$ 。

窗口宽度(w)决定了平滑点数,过小的窗口宽度无法有效平滑噪声,而过大的窗口宽度则可能导致信号特征的丢失。多项式阶次(q)决定了拟合的复杂度,较高的阶次可以更好地

拟合细节,但过高的阶次会导致过拟合,从而保留噪声。

为系统地评估 w, q 对降噪效果影响,选择3种不同的窗口宽度(11、19、29)和同时为避免阶次过高,出现过拟合的状态,选择两种不同的多项式阶次(2、3)进行实验,比较机械钻速与井深之间的曲线变化。不同参数组合的对比曲线如图2所示。

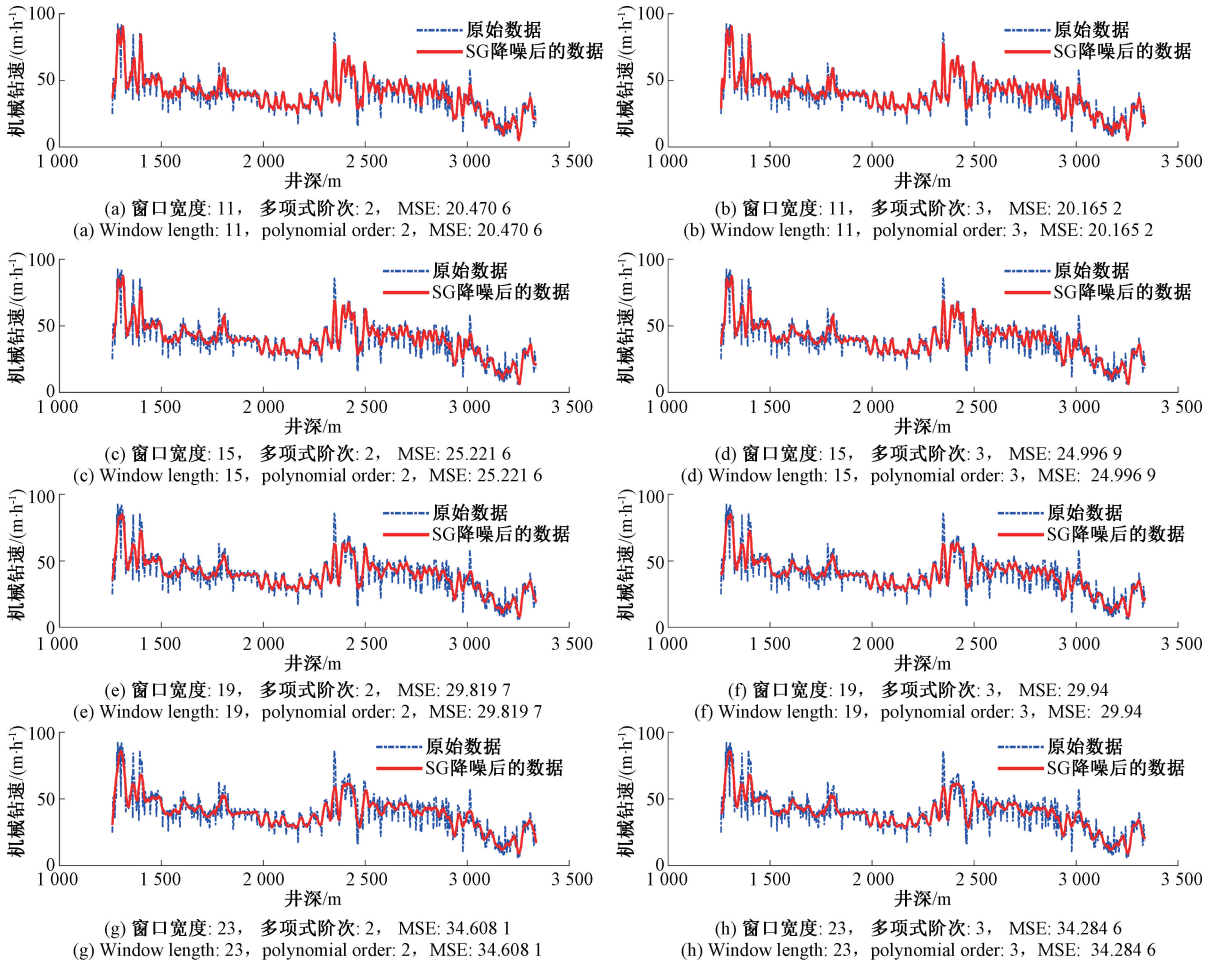


图2 SG不同参数组合数据降噪处理结果

Fig. 2 SG data denoising processing results of different parameter combinations

SG平滑处理在不同窗口宽度和多项式阶次下对钻井测量数据的影响。当窗口宽度为11且多项式阶次为3时,均方误差(mean squared error, MSE)为20.165 2,滤波效果较好,但异常点未有效去除。窗口宽度为15且多项式阶次为3时, MSE为24.996 9,平滑效果和信号保留达到了较好的平衡。进一步增大窗口宽度至19和23时, MSE显著上升,信号特征丢失较多,且过度平滑导致关键数据节点被抹去,降低了数据的准确性和可靠性。

综上所述,图3所示的窗口宽度为15、多项式阶次为3的SG平滑降噪后的数据变化图,能够有效减少钻井测量数据中的随机波动和不必要的噪声,数据曲线更加平滑、清晰,同时不会显著影响原有的数据特性,从而使得数据特征更容易识别和分析,表明这种组合在信号平滑和细节

保留之间达到了较好的平衡。

1.2 数据归一化

各钻井参数之间的数值大小在数量级上的存在较大差异,为提高模型数据处理的泛化能力,需要对原始数据进行归一化处理,通过数据归一化,可以将不同特征的值域缩放到相同的范围,避免特征之间的尺度差异对相关性分析以及后续的结果的影响。采用归一化公式为:

$$x_{norm} = \frac{(y_{max} - y_{min})(x - x_{min})}{x_{max} - x_{min}} + y_{min} \quad (4)$$

式中: x_{norm} 为归一后的数字; x 为归一前的数值; x_{max} 、 x_{min} 分别为数据归一前的最大值、最小值; y_{max} 、 y_{min} 分别为数据归一后的最大值、最小值。

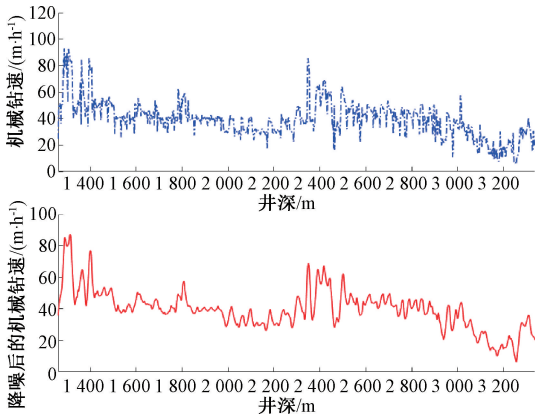


图 3 SG 平滑降噪前后的数据变化对比图

Fig. 3 Comparison chart of data before and after SG smoothing denoising

1.3 相关性分析

为了进一步提高 ROP 预测模型的准确性,通过对 ROP 与钻井参数之间的相关性进行分析,选择出最具影响力的特征,降低模型的复杂性,从而优化模型的性能。根据数据的特点不同,所采用的度量变量之间的关联程度的统计量也不相同。Pearson 线性相关、Spearman 等级相关和 Kendall 等级相关是最常用的相关系数。

1) Pearson 线性相关系数

Pearson 相关系数用来度量具有线性关系的两个变量之间相关关系的密切程度及其相关方向。常以符号 r 表示样本相关系数, $\rho_{X,Y}$ 表示总体相关系数。

设随机变量 X, Y 取自两组数据有限的离散样本,样本长度为 n ,则计算两组数据间的协方差的方法为:

$$Cov(X, Y) = \frac{\sum_{i=1}^n (x_i - E(X))(y_i - E(Y))}{n} \quad (5)$$

式中: $E(X)$ 和 $E(Y)$ 分别代表数据 X 和 Y 期望值。

Pearson 的总体相关系数公式为:

$$\rho_{X,Y} = \frac{Cov(X, Y)}{\sqrt{D(X)} \sqrt{D(Y)}} = \frac{E[(X - EX)(Y - EY)]}{\sqrt{D(X)} \sqrt{D(Y)}} \quad (6)$$

式中: $Cov(X, Y)$ 是随机变量 X, Y 的协方差, $D(X)$ 和 $D(Y)$ 分别代表 X 和 Y 的方差。

Pearson 的样本相关系数公式为:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (7)$$

式中: X_i, Y_i 分别为两个样本第 i 个数据的值, \bar{X} 与 \bar{Y} 分别为两个样本的平均值, n 为样本总数。

2) Spearman 等级相关系数

Spearman 系数相当于 Pearson 相关系数的非参数形

式,它根据数据的秩而不是数据的实际值计算,适用于有序数据和满足正态分布假设的等间隔数据。

随机变量 X, Y 之间的 Spearman 相关系数记为 r_s , 假设原始的数据 x_i, y_i 已经按从大到小的顺序排列,记 $R(x_i)$ 是 x_i 在 x 中的大小排名名次, $R(y_i)$ 是 y_i 在 y 中的大小排名名次, $\overline{R(x)}$ 是 x 名次均值, $\overline{R(y)}$ 是 y 名次均值, n 为数据对个数,则 Spearman 秩相关系数为:

$$r_s = \frac{\sum_{i=1}^n (R(x_i) - \overline{R(x)})(R(y_i) - \overline{R(y)})}{\sqrt{\sum_{i=1}^n (R(x_i) - \overline{R(x)})^2} \sqrt{\sum_{i=1}^n (R(y_i) - \overline{R(y)})^2}} = \frac{6 \sum_{i=1}^n (R(x_i) - R(y_i))^2}{n(n^2 - 1)} \quad (8)$$

3) Kendall's tau 等级相关系数

Kendall's tau 等级相关系数基于 (i, j) 同序对的计数(其中 $i < j$),同序是指 $x_{a,i} - x_{a,j}$ 和 $x_{b,i} - x_{b,j}$ 具有相同的符号。对于矩阵中第 a 列 x_a 和第 b 列 x_b , Kendall's tau 等级相关系数定义为:

$$\tau = \frac{2K}{m(m-1)} \quad \text{其中 } K = \sum_{i=1}^{m-1} \sum_{j=i+1}^m \xi \quad (9)$$

$$\xi = \begin{cases} 1, & (x_{a,i} - x_{a,j})(x_{b,i} - x_{b,j}) > 0 \\ 0, & (x_{a,i} - x_{a,j})(x_{b,i} - x_{b,j}) = 0 \\ -1, & (x_{a,i} - x_{a,j})(x_{b,i} - x_{b,j}) < 0 \end{cases}$$

在上述定义中, m 是数据集中观测值的数量,或者说是矩阵中每列的元素个数。

对机械钻速(ROP)与钻采过程中的关键钻井参数进行了细致的相关性分析。通过计算 Pearson、Spearman 和 Kendall 相关系数,分析结果的可视化展示分别如图 4(a)、(b)、(c)。

表 1 详细的量化了各参数与 ROP 之间的线性和非线性关系,展示了 ROP 与多个钻井参数之间的相关性分析结果,包括 Pearson、Spearman 和 Kendall 相关系数。相关性数值的绝对值越接近于 1,说明两个参数之间相关性越大;越接近于 0,说明两个参数相关性越弱。

通过分析 Pearson 相关系数,可以初步确定与 ROP 有较强相关性的参数。例如,井深(-0.58)、钻头压降(-0.61)和温度(-0.58)与 ROP 之间表现出较强的负相关关系,这表明这些参数在预测模型中可能是重要的特征。

为了确保特征选择的稳定性,需结合 Spearman 和 Kendall 相关系数的结果。例如,转速的 Pearson 相关系数为 0.25,但 Spearman 和 Kendall 相关系数分别为 0.30 和 0.21,这显示出一定的一致性,说明转速可能对 ROP 有一定的影响。

对于相关系数绝对值较小的参数,可以考虑排除。例如,漏斗黏度的相关系数在所有度量下都接近于 0,表明其与 ROP 的相关性较弱,可以在特征选择时排除。

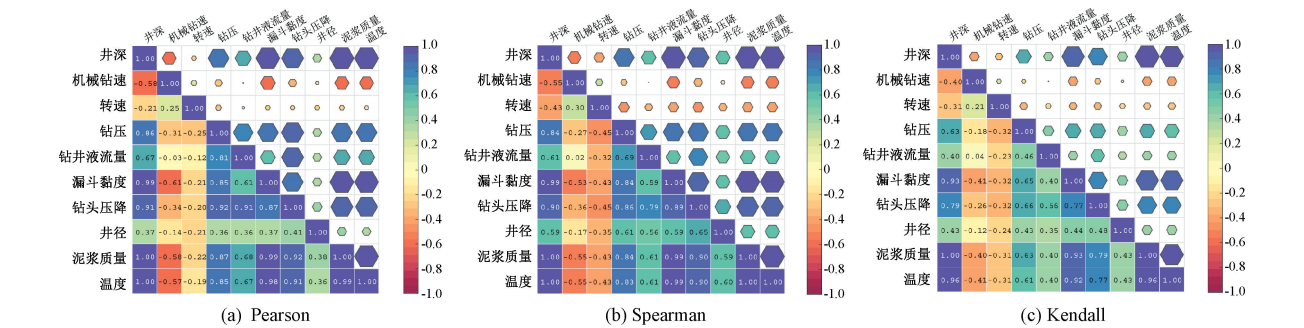


图 4 相关性分析
Fig. 4 Comparison chart of data before and after SG smoothing denoising

表 1 机械钻速(ROP)与钻井参数的相关性分析结果

相关系数 (滤波后)	井深/ m	转速/ (r•min ⁻¹)	钻压/ kN	漏斗黏度/ (s•qt ⁻¹)	钻头压降/ psi	井径/ in	泥浆质量 (循环系统/t)	温度/ ℃	转速/ (r•min ⁻¹)
Pearson 相关系数	-0.58	0.25	-0.31	-0.03	-0.61	-0.34	-0.14	-0.58	-0.57
Spearman 相关系数	-0.55	0.30	-0.27	0.02	-0.53	-0.36	-0.17	-0.55	-0.55
Kendall 相关系数	-0.40	0.21	-0.18	0.04	-0.41	-0.26	-0.12	-0.40	-0.41

除了统计上的相关性,还需考虑参数在实际钻井过程中的意义和可操作性。尽管泥浆质量的相关系数较低,但在实际操作中仍需监控和调整,因此在特征选择时需综合考虑。

综上所述,通过对各钻井参数与 ROP 之间相关性的

分析,可以有效选择出对 ROP 预测有显著影响的参数。井深、钻速、钻压、钻头压降、井径、泥浆质量、以及温度这 7 个参数作为机械钻速预测模型的输入变量,具体结果如表 2 所示。

表 2 筛选出的钻井参数的整体统计表

统计类型	井深/ m	转速/ (r•min ⁻¹)	钻压/ kN	钻头压降/ psi	井径/ in	泥浆质量 (循环系统/t)	温度/ ℃
平均值	2 299	126.93	15.3	1870.11	12.09	332.16	64.02
标准差	600.73	8.76	7.75	661.08	0.19	65.8	11.19
最小值	1 260	54	0	755.29	11.73	222.68	43
均值	2 299	130	15.4	1 567.24	12.08	329.94	63
最大值	3 338	131	36.66	2 908.68	15.14	455.45	81

2 方法介绍

2.1 BP 神经网络

BP 神经网络是一种多层前馈神经网络^[16],具有 3 层或 3 层以上的神经结构。每一层由若干神经元组成,通过使用误差逆向传播算法进行训练,使用梯度下降法和反向传播的学习规则不断迭代、调整网络的权值和阈值,训练出误差平方和最小的前馈神经网络,具有很好的学习能力和数据处理能力^[17]。

图 5 展示 4 层 BP 神经网络的训练过程,包括正向传播和逆向反馈两个阶段。在正向传播过程中,训练样本对应的钻井参数(包括井深、钻速、钻压、钻头压降、井径、泥

浆质量、温度)的影响权值分别对应着 $x_1、x_2、x_3、\cdots、x_7$,通过各隐含层依次计算后传递到输出层,再与期望值 y 进行比对,如果计算偏差超过设定的阈值,则进入逆向反馈阶段。在反馈过程中,以特定形式将输出结果反向传递到输入层,从而获得各神经元的误差信号,并使用这些信号修正各级神经元的权值。通过多次训练迭代,可以得到机械钻速的预测结果。

然而,BP 神经网络的初始权值和阈值通常是通过随机函数生成的,训练过程完全依赖误差函数对随机生成的初始值进行调整。如果权值和阈值的选择不当,将会对结果产生影响,导致预测精度不稳定。此外,梯度下降法容易陷入局部最优状态,导致无法获取最佳权值和阈值。因

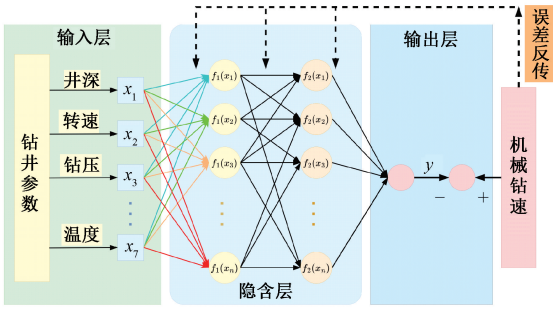


图 5 BP 神经网络模型结构

Fig. 5 BP neural network model structure

此,在 BP 神经网络训练过程中,权值和阈值的选择需要采

用合适的优化策略来避免权值阈值陷入局部最优^[18]。

2.2 遗传算法优化的 BP 神经网络

遗传算法优化的 BP 神经网络(genetic algorithm-back propagation neural network,GA-BPNN)是一种结合遗传算法和反向传播算法的神经网络训练方法。其中 GA^[19]是一种通过迭代运算将问题参数编码为染色体,并模拟达尔文生物进化论的自然选择和遗传学机制的进化模型来实现交换种群中染色体信息以生成符合优化目标的算法。GA-BPNN 机械钻速预测模型主要分为原始数据的预处理、GA 优化寻找最优的初始权值和阈值、BP 神经网络的机械钻速模型训练和预测,3 个主要部分组成,流程图如图 6 所示。具体步骤如下:

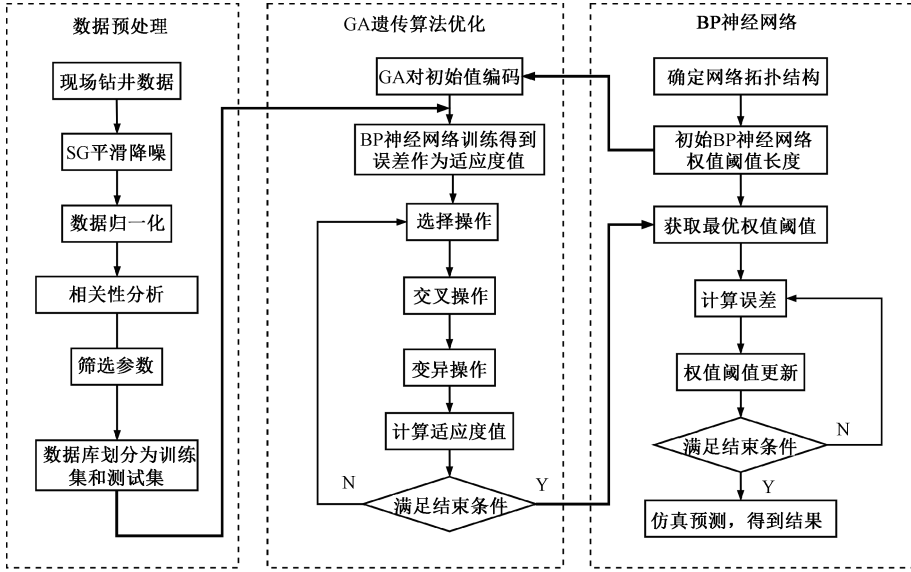


图 6 GA-BPNN 预测模型流程图

Fig. 6 BP neural network model structure

步骤 1)通过对降噪处理后的数据进行数据归一化,降低钻井参数数值大小的干扰;

步骤 2)分析钻井参数之间的性,筛选特征变量,简化模型;

步骤 3)根据输入输出参数个数确定 BP 神经网络结构:在 BP 神经网络模型中,输入层为每组数据的 7 个影响因子,输出层为机械钻速,构成 7 个节点的输入层和 1 个节点的输出层,初始化网络参数;

步骤 4)将 BP 神经网络的权重、阈值等参数使用二进制编码成染色体,如图 7(a)所示;

步骤 5)通过随机产生若干条染色体来构成初始种群,如图 7(b)所示;

步骤 6)采用轮盘赌法从当前种群中选择一部分染色体,作为父代进行交叉和变异。根据染色体适应度的大小进行选择的方法被称为轮盘赌法,即适应度越大的染色体被选中的概率也越大,如图 7(c)所示;

步骤 7)父代染色体之间进行单点交叉,生成新的子代

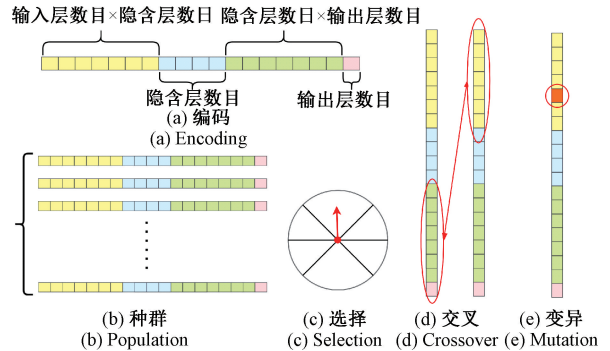


图 7 BP 神经网络参数的遗传编码处理示意图

Fig. 7 Schematic diagram of genetic encoding processing for BP neural network parameters

染色体,如图 7(d)所示;

步骤 8)对经过交叉操作产生的子代染色体,以一定概率进行基因突变,即将某些基因位置上的二进制码进行变异,如图 7(e)所示。

步骤 9)将父代和子代染色体组合起来形成新的种群,重新计算适应度值;

步骤 10)当达到指定迭代次数或者满足误差要求等条件时,停止遗传算法的迭代,输出优化后的 BP 神经网络参数。否则,返回步骤 5);

步骤 11)最终输出适应度最高的染色体所对应的 BP 神经网络参数作为最优解,即完成 GA 优化 BP 神经网络算法的过程。

通过将最优权值和阈值代入神经网络模型中,BP 可通过前向传播计算每个神经元的输出,然后利用误差反向传播算法调整每个神经元的权值和阈值,直到误差达到最小值。此过程使得网络能够更好地拟合训练数据并泛化^[17]到测试数据上。

2.3 模型对比分析

为验证 GA-BPNN 模型的优越性,本文对机械钻速预

测中常用机器学习方法以及神经网络分类预测模型进行对比分析。这些模型包括传统的标准 BP 神经网络模型、基于径向基函数的 RBF(radial basis function,RBF)神经网络模型、以及传统的机器学习方法^[20-21]:ELM、RF、SVM、K 邻近算法(k-nearest neighbors,KNN),并使用一种基于心智模型优化算法的思维进化算法(mind evolutionary algorithm,MEA)优化 BP 神经网络^[22]进行预测,在 MATLAB 2020a 软件上进行仿真,硬件配置的操作系统为 Windows 10,由 i7-12700F 处理器、4.8 GHz 处理频率和 32 GB 内存组成。为避免过拟合现象,采用留一法验证模型性能,随机从数据集中选取 80%(832 个点)作为训练集,剩余 20%(209 个点)作为测试集,模型预测结果如图 8 所示,通过对比分析这些模型的预测结果和预测误差等指标,可以更全面地评估 GA-BPNN 机械钻速模型在实际应用中的表现。

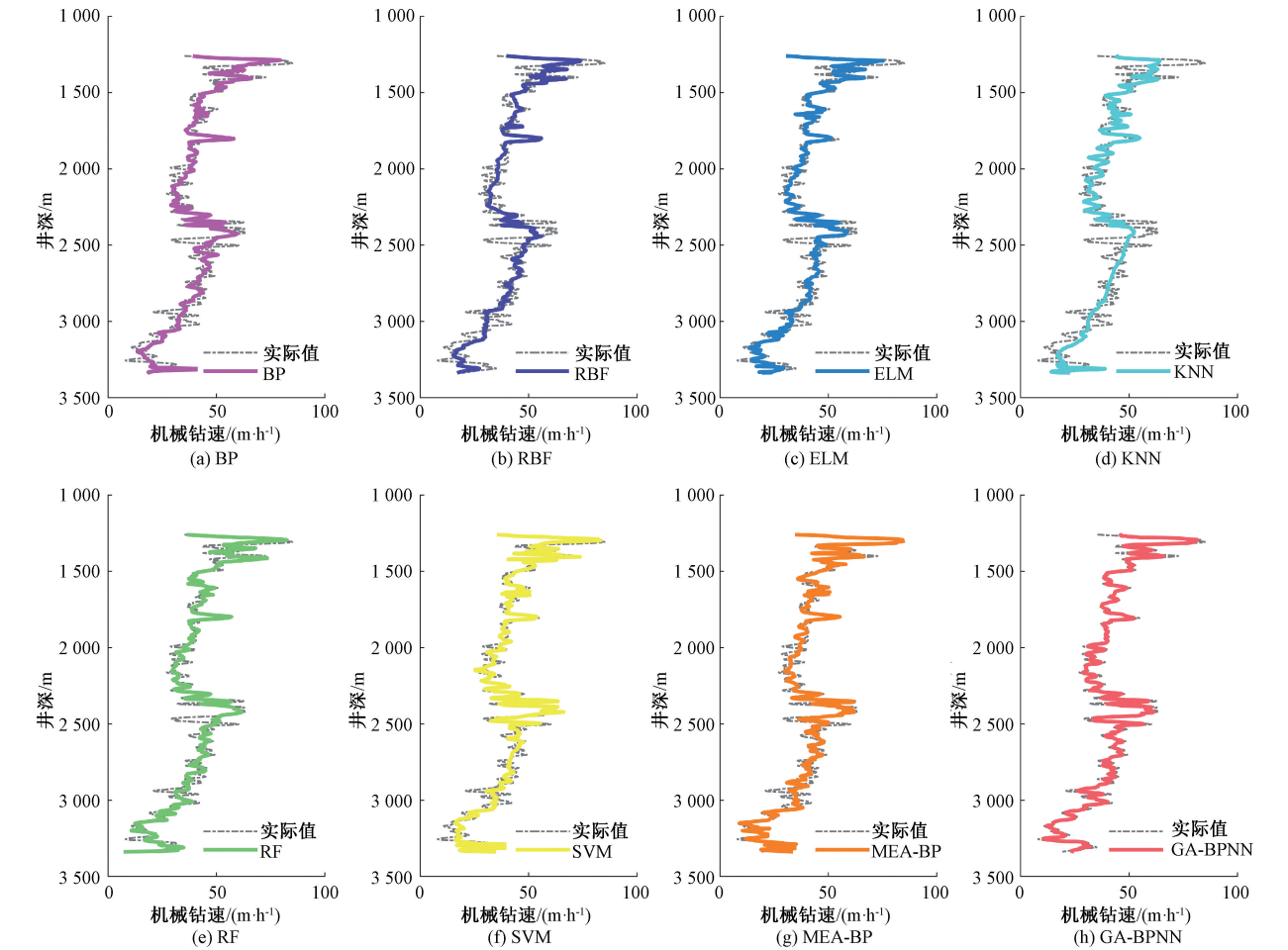


图 8 不同算法预测机械钻速预测值与实际值散点拟合图

Fig. 8 Scatter plot of predicted and actual values of mechanical drilling speed for different algorithms

图 9 展示了使用不同算法(BP、RBF、ELM、KNN、RF、SVM、MEA-BP、GA-BPNN)对机械钻速进行预测的预测值与实际值进行对比的散点图。每个子图中,灰色点表示

训练集,彩色点表示测试集。图中的蓝色虚线表示理想情况下预测值与实际值完全一致的情况,即拟合度为 100%。如果散点图中的点高度重合并几乎呈现出一条直线,则说

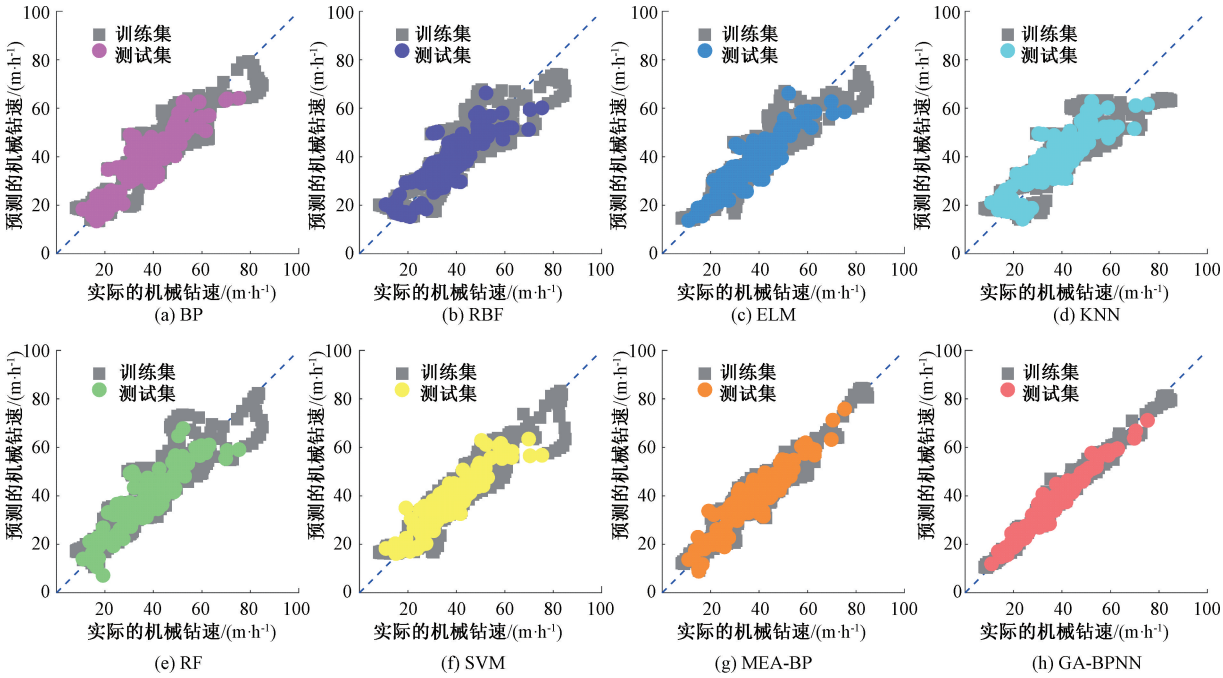


图 9 不同算法预测机械钻速与实际钻速的拟合对比图

Fig. 9 Scatter plot of predicted and actual values of mechanical drilling speed for different algorithms

明该算法的拟合度较高,预测效果较好。

预测结果显示,GA-BPNN 模型的预测值与实际值的散点图显示高度重合,几乎呈现出一条直线;其次是 MEA-BP 模型的预测结果,表明其对机械钻速的预测与实际数值之间存在显著的一致性。其他几种模型的散点拟合图分布较为散乱,拟合效果明显不如 GA 或 MEA 优化后的 BP 神经网络。

为更系统更具体的评估 GA 优化的神经网络模型的有效性和准确性,使用拟合优度 R^2 和均方根误差(RMSE)这两个指标来评估几种机械钻速预测模型的性能,评价结果详如表 3 所示。

表 3 多种预测方法的预测结果评价表
Table 3 Evaluation table of prediction results from various prediction methods

预测方法	训练集 R^2	测试集 R^2	训练集 RMSE	训练集 RMSE
BP	0.814	0.822	5.373	4.666
RBF	0.768	0.770	6.003	5.299
ELM	0.855	0.867	4.753	4.033
KNN	0.719	0.769	6.611	5.313
RF	0.817	0.818	5.325	4.712
SVM	0.882	0.859	4.279	4.157
MEA-BP	0.909	0.873	3.752	3.943
GA-BP	0.982	0.967	1.697	2.010

R^2 是一个衡量回归模型拟合程度的指标,其取值范围在 $[0,1]$,越接近 1 说明模型的拟合效果越好,即模型预测值越接近实际测量值。

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - x_i)^2}{\sum_{i=1}^n (x_i - x_i)^2} \tag{10}$$

式中: y_i 为预测值, x_i 为实际值。

RMSE 是衡量实际数据与预测数据之间误差的标准差,它也是回归模型常用的性能指标。RMSE 越小,代表模型的预测精度越高。

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \tag{11}$$

式中: y_i 为预测值, x_i 为实际值。

GA-BPNN 模型预测的训练集的 R^2 为 0.982, RMSE 为 1.696 7;测试集的 R^2 为 0.967, RMSE 为 2.010,均比较 MEA-BP 预测结果更为精准,误差更小,即遗传算法比思维进化的优化效果更佳,在多种预测模型中性能最优。与标准的 BP 神经网络的预测拟合优度 R^2 提高 17.64%,预测精度有显著的提升,由此可见,GA-BPNN 既能充分利用遗传算法的全局搜索特性,又能对 BP 的权值和阈值进行一定程度的优化处理,改进 BP 神经网络对初始权重依赖的问题,从而大大提高 BP 神经网络的预测准确性。

3 结 论

本研究基于一种基于遗传算法优化的反向传播神经

网络(GA-BPNN)模型预测中国南海某油田的机械钻速。通过对钻井数据进行 SG 平滑降噪处理,并利用 Pearson、Spearman 和 Kendall 相关系数综合分析,筛选出了影响钻速的关键参数,并建立了一个高精度的预测模型。该模型的决定系数高达 96.67%,预测误差低至 2.010 m/h,显著优于传统 BP 神经网络及其他六种常用预测方法。GA-BPNN 模型的高预测精度为钻井作业提供了更为准确和可靠的决策支持,有助于优化钻井方案,降低油气开采成本,并提高钻探效率。

参考文献

- [1] ZHANG N N, WANG Q, WANG J J, et al. Characteristics of oil and gas discoveries in recent 20 years and future exploration in the world[J]. China Petroleum Exploration, 2018, 23(1): 44-53.
- [2] 胡文瑞, 鲍敬伟. 石油行业发展趋势及中国对策研究[J]. 中国石油大学学报(自然科学版), 2018, 42(4): 1-10.
HU W R, BAO J W. Development trends of oil industry and China's countermeasures[J]. Journal of China University of Petroleum (Edition of Natural Science), 2018, 42(4): 1-10.
- [3] AHMED O S, ADENIRAN A A, SAMSURI A. Computational intelligence based prediction of drilling rate of penetration: A comparative study[J]. Journal of Petroleum Science and Engineering, 2019, 172: 1-12.
- [4] 景宁, 樊洪海, 纪荣艺, 等. 基于数据挖掘技术的深井钻速预测方法研究[J]. 石油机械, 2012, 40(7): 17-20.
JING N, FAN H H, JI R Y, et al. Data mining technology-based research on the prediction method of deepwell rop[J]. China Petroleum Machinery, 2012, 40(7): 17-20.
- [5] HEGDE C, DAIGLE H, MILLWATER H, et al. Analysis of rate of penetration(ROP) prediction in drilling using physics-based and data-driven models[J]. Journal of Petroleum Science & Engineering, 2017, 159: 295-306.
- [6] 许奕东, 李飞. 人工智能背景下测量仪器技术发展探讨[J]. 电子测量技术, 2023, 46(23): 1-6.
XU Y D, LI F. Discussion on the development of measuring instrument technology under the background of artificial intelligence [J]. Electronic Measurement Technology, 2023, 46(23): 1-6.
- [7] BELLO O, TEODORIU C, YAQOUB T, et al. Application of artificial intelligence techniques in drilling system design and operations: A state of the art review and future research pathways[C]. Lagos, Nigeria: SPE, 2016.
- [8] 赵颖, 孙挺, 杨进, 等. 基于极限学习机的海上钻井机械钻速监测及实时优化[J]. 中国海上油气, 2019, 31(6): 138-142.
ZHAO Y, SUN T, YANG J, et al. Extreme learning machine-based offshore drilling ROP monitoring and real-time optimization [J]. China Offshore Oil and Gas, 2019, 31(6): 138-142.
- [9] 宋先知, 裴志君, 王潘涛, 等. 基于支持向量机回归的机械钻速智能预测[J]. 新疆石油天然气, 2022, 18(1): 14-20.
SONG X ZH, PEI ZH J, WANG P T, et al. Intelligent prediction for rate of penetration based on support vector machine regression[J]. Xinjiang Oil & Gas, 2022, 18(1): 14-20.
- [10] 刘胜娃, 孙俊明, 高翔, 等. 基于人工神经网络的钻井机械钻速预测模型的分析与建立[J]. 计算机科学, 2019, 46(S1): 605-608.
LIU SH W, SUN J M, GAO X, et al. Analysis and establishment of drilling speed prediction model for drilling machinery based on artificial neural networks[J]. Computer Science, 2019, 46(S1): 605-608.
- [11] 赵奕昕, 李薇, 朱家明. 基于 GA-BP 神经网络的极端降水定量分析与预测[J]. 内蒙古师范大学学报(自然科学汉文版), 2022, 51(6): 576-581.
ZHAO Y X, LI W, ZHU J M. Prediction and quantitative analysis of extreme precipitation based on GA-BP neural network[J]. Journal of Inner Mongolia Normal University (Natural Science Edition), 2022, 51(6): 576-581.
- [12] HUANG SH H, TIAN L, ZHANG J SH, et al. Support vector regression based on the particle swarm optimization algorithm for tight oil recovery prediction [J]. ACS Omega, 2021, 6(47): 32142-32150.
- [13] BRENJKAR E, BINIAZ DELIJANI E, KARROUBI K. Prediction of penetration rate in drilling operations: a comparative study of three neural network forecast methods[J]. Journal of Petroleum Exploration and Production Technology, 2021, 11(2): 805-818.
- [14] 胡顺石, 黄春晓, 杨斌, 等. 自适应加权 Savitzky-Golay 滤波重构 MODIS 植被指数时间序列[J]. 测绘科学, 2020, 45(4): 105-116.
HU SH SH, HUANG CH X, YANG B, et al. Reconstruction of MODIS vegetation index time series by adaptive weighted Savitzky-Golay filter[J]. Science of Surveying and Mapping, 2020, 45(4): 105-116.
- [15] 倪朋勃, 吴昊晟, 毛敏, 等. 基于 SG-KNN 的随钻地层流体组分光谱感知技术研究[J]. 钻采工艺, 2022,

- 45(5): 45-50.
- NI P B, WU H SH, MAO M, et al. Research on spectral perception technology of formation fluid composition while drilling based on SG-KNN [J]. Drilling & Production Technology, 2022, 45(5): 45-50.
- [16] 于珍珍, 邹华芬, 于德水, 等. 融合田间水热因子的甘蔗产量 GA-BP 预测模型[J]. 农业机械学报, 2022, 53(10): 277-283.
- YU ZH ZH, ZOU H F, YU D SH, et al. Sugarcane yield GA-BP prediction model incorporating field water and heat factors [J]. Transactions of the Chinese Society of Agricultural Machinery, 2022, 53(10): 277-283.
- [17] 李梦妍, 于文震. 基于改进小波神经网络和灰色模型的装备性能参数预测 [J]. 电子测量技术, 2016, 39(3): 18-22.
- LI M Y, YU W ZH. Prediction of equipment performance parameters based on improved wavelet neural network and grey model [J]. Electronic Measurement Technology, 2016, 39(3): 18-22.
- [18] 李琪, 屈峰涛, 何璟彬, 等. 基于 BAS-BP 的钻井机械钻速预测模型[J]. 西安石油大学学报(自然科学版), 2021, 36(6): 89-95.
- LI Q, QU F T, HE J B, et al. Prediction model of mechanical ROP during drilling based on BAS-BP[J]. Journal of Xi'an Shiyou University (Natural Science Edition), 2021, 36(6): 89-95.
- [19] 周中, 邓卓湘, 陈云, 等. 基于 GA-BP 神经网络的泡沫轻质土强度预测[J]. 华南理工大学学报(自然科学版), 2022, 50(11): 125-132.
- ZHOU ZH, DENG ZH X, CHEN Y, et al. Strength prediction of foam light soil based on GA-BP neural network [J]. Journal of South China University of Technology (Natural Science Edition), 2022, 50(11): 125-132.
- [20] 李昌盛. 基于多元回归分析的钻速预测方法研究[J]. 科学技术与工程, 2013, 13(7): 1740-1744.
- LI CH SH. Study of method for predict rate of penetration based of multiple regression analysis [J]. Science Technology and Engineering, 2013, 13(7): 1740-1744.
- [21] 王娜娜, 栗文义, 李建菽. 基于变分模态分解-布谷鸟搜索-支持向量回归的变压器油中溶解气体浓度预测方法[J]. 电子测量技术, 2024, 47(4): 10-17.
- WANG N N, LI W Y, LI J Q. Prediction of dissolved gas content in transformer oil based on variational mode decomposition-cuckoo search-support vector regression model [J]. Electronic Measurement Technology, 2024, 47(4): 10-17.
- [22] 张立刚, 苗振华, 黄小刚, 等. 基于 MEA-BP 神经网络的钻井机械钻速预测[J]. 自动化与仪表, 2022, 37(11): 87-92.
- ZHANG L G, MIAO ZH H, HUANG X G, et al. Prediction of drilling ROP based on MEA-BP neural network [J]. Automation & Instrumentation, 2022, 37(11): 87-92.

作者简介

邱腾煌, 硕士研究生, 主要研究方向为石油机械、数据挖掘、人工智能。

E-mail: 943595923@qq.com

钱玉宝(通信作者), 博士, 副教授, 硕士生导师, 主要研究方向为海洋石油装备研制、大数据与人工智能应用。

E-mail: 109184785@qq.com

季威, 学士, 主要研究方向为固井装备以及工具设计研究、自动化控制。

E-mail: jiwei4@cosl.com.cn