

# 基于人工智能的钻速预测模型数据有效性下限分析

李 谦<sup>1</sup>, 曹彦伟<sup>1</sup>, 朱海燕<sup>2</sup>

(1. 成都理工大学环境与土木工程学院, 四川 成都 610059; 2. 成都理工大学能源学院, 四川 成都 610059)

**摘要:** 钻速预测对于优化钻探工艺、降低作业成本、实现科学钻探具有重要意义,它是钻探钻井作业的一项重要内容。基于人工智能的钻速预测精度令人瞩目,但该技术需求的海量数据对传统钻探钻井作业的要求较高。为明确使用人工智能建立钻速预测模型的最少数据量,本文基于中国南海10口井的21917条数据进行了分析。通过相关性分析,所有的输入参数可被划分为高、中、低相关性3大类。通过逐步引入参数建立预测模型和对比预测精度,发现当引入的参数数量足够时,3种相关性参数均可建立起高精度( $\geq 85\%$ )的预测模型。引入参数的相关性越高,建立高精度预测模型需求的参数数量越少。通过逐步扩大取样间隔的方式,对比发现所有的预测模型均呈现随取样间隔的增大、预测模型的准确性降低的规律。而预测模型建模的取样间隔下限可通过寻找精度降低时的拐点获得。经过验证,在数据维度与取样精度均为下限时,基于3种相关性参数建立的BP神经网络预测模型仍然能够获得较高的预测精度。

**关键词:** 钻速预测;BP神经网络;数据需求;有效性下限;人工智能

**中图分类号:** P634.9    **文献标识码:** A    **文章编号:** 2096-9686(2021)03-0021-10

## Discussion on the lower limit of data validity for ROP prediction based on artificial intelligence

LI Qian<sup>1</sup>, CAO Yanwei<sup>1</sup>, ZHU Haiyan<sup>2</sup>

(1. College of Environment and Civil Engineering, Chengdu University of Technology, Chengdu Sichuan 610059, China;  
2. College of Energy, Chengdu University of Technology, Chengdu Sichuan 610059, China)

**Abstract:** Prediction of ROP is of great significance for optimizing drilling technology, reducing operating costs, and realizing scientific drilling, and it is also an important part of drilling operations. The accuracy of ROP prediction with artificial intelligence technology is remarkable, but the massive data required by this technology puts forward higher requirements for the traditional drilling operations. To determine the minimum amount of data for ROP modeling based on artificial intelligence, analysis was carried out based on 21917 data samples collected from 10 wells at South China Sea. Through correlation analysis, all input parameters were divided into three categories: high, medium and low correlation. By gradually introducing parameters to establish a prediction model to compare the accuracy, it was found that when the number of parameters was sufficient, the parameters in all three categories can be used to establish a high-precision ( $\geq 85\%$ ) prediction model; however, the higher the correlation of the parameters, the less the number of the parameters required to set up a high accuracy prediction model. When the sampling interval is gradually expanded, comparison found that the accuracy of all the prediction models decreased with the sampling interval increased. The lower limit of the data sampling interval for setting up the prediction model can be obtained through finding out the downward inflection point of prediction accuracy. It is verified that the BP neural network prediction model based on any of the three correlation parameters can still obtain high prediction accuracy when both data

收稿日期:2020-10-30    DOI:10.12143/j.ztgc.2021.03.003

基金项目:中海石油(中国)有限公司项目“南海西部油田上产2000万方钻完井关键技术研究”(编号:CNOOC-KJ135ZDXM38ZJ05ZJ)子课题“乐东10区超高温高压井综合提速技术研究”

作者简介:李谦,男,汉族,1987年生,讲师,地质资源与地质工程专业,博士,从事钻井智能化分析与应用研究工作,四川省成都市成华区二仙桥东三路1号,liqian2014@cdut.edu.cn。

引用格式:李谦,曹彦伟,朱海燕.基于人工智能的钻速预测模型数据有效性下限分析[J].钻探工程,2021,48(3):21-30.

LI Qian, CAO Yanwei, ZHU Haiyan. Discussion on the lower limit of data validity for ROP prediction based on artificial intelligence [J]. Drilling Engineering, 2021, 48(3): 21-30.

dimension and sampling accuracy are at the lower limits.

**Key words:** ROP prediction; BP neural network; data requirements; lower limit of validity; artificial intelligence

## 0 引言

钻探钻井作业是固体矿产与油气资源勘探过程中至关重要的环节。钻速作为评估钻探钻井作业最有效的指标之一,钻速预测对于优化钻探工艺、降低作业成本、实现科学钻探具有重要意义,它是钻探钻井作业的一项重要内容。国内外众多学者对钻速的评估与预测进行了大量的研发。但由于地层与工艺的复杂性,影响钻速的因素繁多,同时相关领域理论体系的不完备性,这些因素之间至今未建立较为令人信服的通用分析模型。各国学者基于不同的现场参数与钻速的关系,先后建立了多种不同的钻速预测模型。

截止目前,常规的钻速预测方案有理论推导型与数据挖掘型两种形式。理论推导型预测方程主要基于钻头结构、受力平衡、机械比能等理论进行推演,如孟英峰等<sup>[1]</sup>、Chen等<sup>[2]</sup>均提出基于机械比能理论可实现钻速的预测与优化;邹德永等<sup>[3]</sup>提出基于钻头力学平衡的迭代可实现定向钻井中PDC钻头的钻速预测;刘军波等<sup>[4]</sup>认为将钻头转速的影响带入传统的三维钻速方程将提高钻速预测的精度。

这类研究中也结合数值仿真模拟进行的分析,如Saksala等<sup>[5]</sup>提出将基于粘塑性和损伤力学的本构模型引入有限元模拟,结合接触力学模型建立了较为准确的钻速-钻头-岩石相互作用预测模型。

理论推导型模型固然可以获得显著的解析解,但建模过程中为简化边界条件而引入的大量假设条件,无疑为模型的应用带来较大的限制。因此,利用现场监测数据,从中进行挖掘和分析的钻速预测技术近年来备受瞩目。同时得益于近年来机器学习技术的发展,越来越多的钻速预测方式引入人工智能的相关概念(如表1所示),均获得了较高的预测精度。随着人工智能技术的深度介入,钻速预测的效率开始引起关注。Bataee等<sup>[6]</sup>将神经网络的预测精度与常规模型的预测精度进行了对比,发现同参数条件下神经网络的预测精度可达98.5%,均高于其余方案;Hegde等<sup>[7-8]</sup>评估了试验驱动型模型与使用人工智能的数据驱动型模型的建模精度,发现数据驱动的模型相较于其他模型可降低预测误差达12%。

表1 近年基于数据挖掘与人工智能的钻速预测方案汇总(部分)

Table 1 ROP prediction algorithm based on data mining and artificial intelligence in recent years (partial)

作者	模型参数	使用算法
Arabjamaloei <sup>[9]</sup>	井深、钻压、转速、钻头水力参数、钻头磨损级别、钻井液循环当量 密度、水头高度	遗传算法
Yi <sup>[10]</sup>	实时伽马测井数据、钻压、转速、泵量	启发式进化算法
Ansari <sup>[11]</sup>	钻压、转速、钻头类型、钻头尺寸、钻头磨损、井斜角、钻井液密度、 粘度、泵压、泵量	竞争支持向量机回归
Eskandarian <sup>[12]</sup>	钻压、钻井液密度	特征排名、随机森林和多层感知器
Ahmed <sup>[13]</sup>	转速、钻压、扭矩、泵压、泵量、钻井液密度、漏斗粘度、固相含量	神经网络
Gan <sup>[14]</sup>	泵压、泵量、扭矩、转速、钻压、井深	混合支持向量机回归
Bhatnagar <sup>[15]</sup>	钻压、转速、钻井液类型、岩石单轴抗压强度	神经网络
Moraveji <sup>[16]</sup>	井深、钻压、转速、喷嘴水力、岩石屈服波速比	响应面与蝙蝠算法结合
Al-Abduljabbar <sup>[17]</sup>	钻压、转速、泵量、扭矩、钻井液密度、粘度、岩石单轴抗压强度	健壮模型(Robust model)
Elkhatatny <sup>[18]</sup>	泵量、泵压、转速、钻压、扭矩、岩石单轴抗压强度、岩石密度、波速	自适应微分估算神经网络

在人工智能与大数据产业蓬勃发展的当前,对数据的渴望前所未有。而在钻探钻井行业中,现场作业的多影响因素及条件复杂性无疑对精确模型

的需求更高,同时带来对更多数据的采集与分析需求。但值得注意的是,随着对数据采集类型与数据量的激增,对施工作业中的数据采集作业会引起施

工成本与难度的增加,过多的数据采集操作也会潜在影响常规钻探钻井施工。当数据采集量超过实际需求时,原本用以降低成本提高效率的数据采集分析反而将拖累钻探钻井施工,达到事倍功半的效果。因此,钻探钻井数据建模中的数据采集规模与精度的下限判定是避免这种情况产生的必要条件。故本文基于中国南海某片区10口井的相关数据,对通过使用神经网络建立钻速预测方程需求的数据规模下限问题进行了分析与探讨。

## 1 数据来源与相关性分析

### 1.1 数据来源与规模

本文使用的数据源于中国南海某片区,经过数据采集、整合、清洗、缺失值补充与错误值删除后,获得源自10口井,涵盖包括井眼位置、施工工艺、钻井液性能、地质条件、钻头钻具共5大类44种不同类型的21917条数据生成初始数据集。如表2所示为初始数据集所包含的所有数据类型。

表2 初始数据集包含的数据变量及缩写  
Table 2 Data variables and abbreviations in the initial data set

类 型	编 号	变 量 名	单 位	缩 写	
井眼位置	1	井号		No.	
	2	深度	m	D	
施工工艺	3	井径	mm	d	
	4	钻压	t	WOB	
	5	大钩载荷	t	WOH	
	6	转速	r/min	RPM	
	7	扭矩	klbf·ft	T	
	8	泵压	psi	SPP	
	9	泵量	L/min	Q	
	10	钻进时间	h	BT	
	11	泵送时间	h	PT	
	12	入口钻井液密度	g/cm <sup>3</sup>	MI	
	13	出口钻井液密度	g/cm <sup>3</sup>	MO	
	14	入口钻井液温度	°C	TI	
	15	出口钻井液温度	°C	TO	
	钻井液性能	16	屈服值	Pa	YP
		17	密度	g/cm <sup>3</sup>	MW
18		漏斗粘度	s	MV	
19		塑性粘度	mPa·s	PV	
20		3转读数		D3	

续表

类 型	编 号	变 量 名	单 位	缩 写
	21	6转读数		D6
	22	10 s静切力	Pa	SS
	23	10 min静切力	Pa	SM
	24	滤失量	mL/30 min	FL
	25	泥饼厚度	mm	MT
	26	pH		PH
	27	N		N
	28	K	Pa·s <sup>-n</sup>	K
	29	氯离子含量	mg/L	CLC
	30	钙离子含量	mg/L	CAC
	31	膨润土含量	kg/m <sup>3</sup>	SOC
	32	固相含量	%	SO
	33	含沙量	%	SA
地质条件	34	地震速度	m/s	EV
	35	孔隙压力	g/cm <sup>3</sup>	PP
	36	破裂压力	g/cm <sup>3</sup>	FP
	37	上覆压力	g/cm <sup>3</sup>	OP
	38	岩性		TYP
钻头钻具	39	喷嘴数量	个	NN
	40	喷嘴等效直径	mm	NS
	41	液力提速马达		ST
	42	钻头内排磨损分级		WI
	43	钻头外排磨损分级		WO
目标参数	44	钻速	m/h	ROP

### 1.2 基本数据集的相关性分析

在数据建模分析前,对输入输出数据进行相关性分析是判定建模数据是否有用的必备步骤。本次研究使用Person相关系数计算法,令 $\rho_{ab}$ 为变量 $a, b$ 的Person相关系数,主要计算方法如式(1)所示,其计算结果区间为 $[-1, 1]$ ,主要的判定标准如所表3所示。

$$\rho_{ab} = \frac{\text{cov}(a, b)}{\sigma_a \sigma_b} = \frac{\sum_{i=1}^n (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_{i=1}^n (a_i - \bar{a})^2} \sqrt{\sum_{i=1}^n (b_i - \bar{b})^2}} \quad (1)$$

式中: $\text{cov}(a, b)$ ——变量 $a, b$ 的协方差矩阵; $\sigma_a, \sigma_b$ ——变量 $a, b$ 各自的标准差; $a_i, b_i$ ——变量 $a, b$ 数据集中第 $i$ 个变量值; $\bar{a}, \bar{b}$ ——变量 $a, b$ 平均值; $n$ ——变量 $a, b$ 的数据集大小。

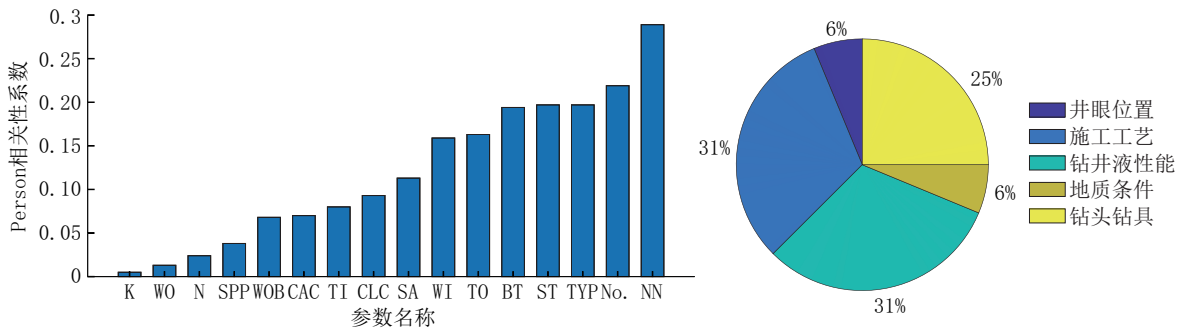
分别计算如表2所示前43个输入参数与输出参数钻速之间的相关性统计如图1中柱状图所示,

表3 相关性判定标准  
Table 3 Correlation criteria

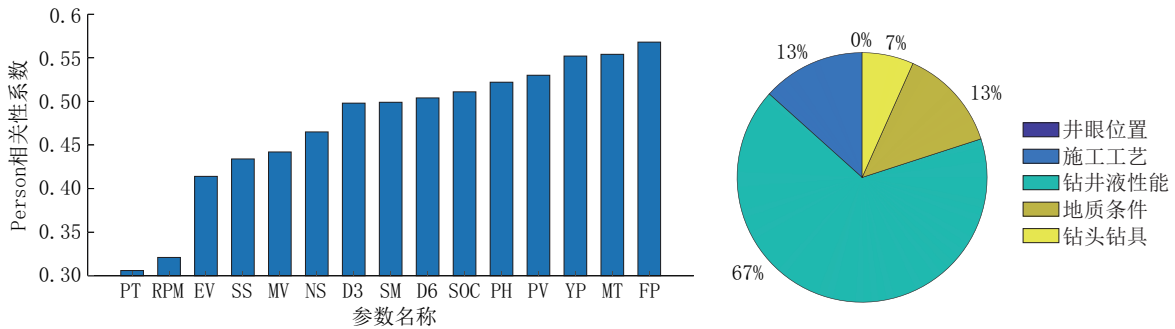
值域	相关性
$\rho_{ab} > 0$	$a, b$ 变量正相关
$\rho_{ab} < 0$	$a, b$ 变量负相关
$ \rho_{ab}  \geq 0.6$	$a, b$ 变量高度相关
$0.3 \leq  \rho_{ab}  < 0.6$	$a, b$ 变量中度相关
$ \rho_{ab}  < 0.3$	$a, b$ 变量低相关或不相关

可知在当前初始数据集中,低相关参数包含16种,其中8种相关性系数 $<0.1$ ;中相关系数包含15种,其中9种相关性系数位于 $[0.5, 0.6]$ 区间;高相关系

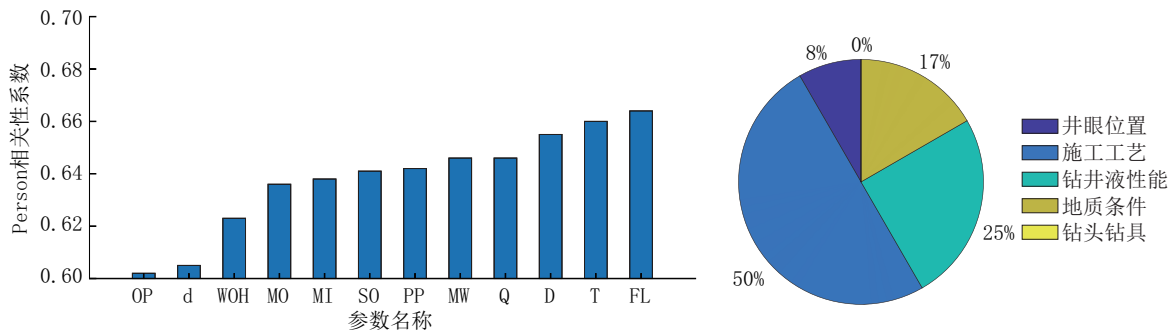
数包含12种,相关性系数全部位于 $[0.6, 0.7]$ 区间。对3种相关性级别的参数进行类别统计如图1中饼图所示,可见高相关性参数中施工工艺与钻井液性能占据75%的比例,中相关性参数中钻井液性能占据67%的比例。参考图1柱状图的高相关区参数详细清单,可发现高相关性参数中的施工工艺类参数具体包括扭矩(T)、泵量(Q)、出入口钻井液密度(MI、MO)、大钩载荷(WOH)和井径(d)。严格来说,其中泵量、出入口钻井液密度也可视为钻井液性能参数。综上所述,该初始数据集中钻井液性能是影响钻速的主要参数类型。



(a)低相关性参数(相关性系数 $<0.3$ )计算结果与分布



(b)中相关性参数(相关性系数 $0.3\sim 0.6$ )计算结果与分布



(c)高相关性参数(相关性系数 $>0.6$ )计算结果与分布

图1 初始数据集相关性计算结果

Fig.1 Correlation calculation results of the initial data set



## 2 数据有效性下限分析

### 2.1 数据分组、训练与精度评价

本文拟采用BP神经网络结合10折交叉验证的数据分组方式完成不同维度与取样精度的模拟分析。因本次研究数据总量较大(21917条),综合考虑建模精度,选用10折交叉验证法进行数据分组。将该初始数据集划分为相斥的10个互斥子集如式(2)所示。为避免数据划分过程中引入额外的偏差而对最终结果产生影响,数据分组以“分层采样”的形式进行划分,保证分组完成后训练集与测试集的数据分布与原数据集相同。故每个子集 $D_i$ 均需要从初始数据集(令其为ini\_data\_all)中分层采样得到。在本次训练建模中,将以井号和地层深度2个参数为主(2个参数的结合可视为整个区块的地层分布与位置)进行分层采样。完成分层采样后,如图2所示,每次用其中的9个子集的并集作为训练集,余下的作为测试集。最终形成10组训练/测试集,从而可进行10次训练和测试,最终返回10个测试结果均值作为建模精度的评估。

$$\begin{cases} ini\_data\_all = D_1 \cup D_2 \cup \dots \cup D_{10} \\ D_i \cap D_j = \emptyset, i \neq j, i \in [1, 10], j \in [1, 10] \end{cases} \quad (2)$$

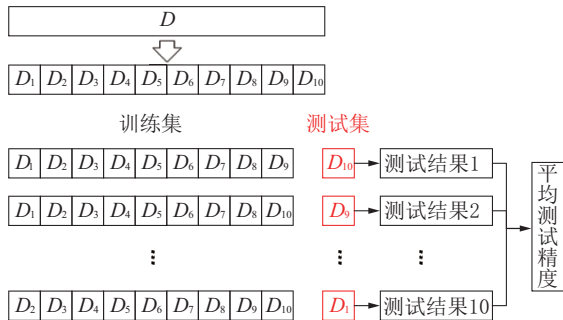


图2 10折交叉验证基本原理  
Fig.2 Basic principle of 10-fold cross-validation

钻速预测属于典型的多输入单输出的非线性拟合,因此建模方式选用为在该方面拟合精度非常高的BP神经网络,建立网络结构如图3所示。该网络包含3层结构,输入层、隐藏层和输出层。每层均包含若干神经元,其中只有相邻层的所有神经元两两连接,同层与不相邻层的神经元均完全不相连。输入与输出层的神经元数量分别与输入、输出变量相同,分别负责网络的数据输入与计算结果的输出。隐藏层的神经元数量则会影响到分析速率与精

度,如果隐藏层结点数过少,网络不能具有必要的学习能力和信息处理能力。反之,若过多,会大大增加网络结构的复杂性。参考相关研究基础,本次研究中选用隐藏层神经元数量 $p = 10$ 。

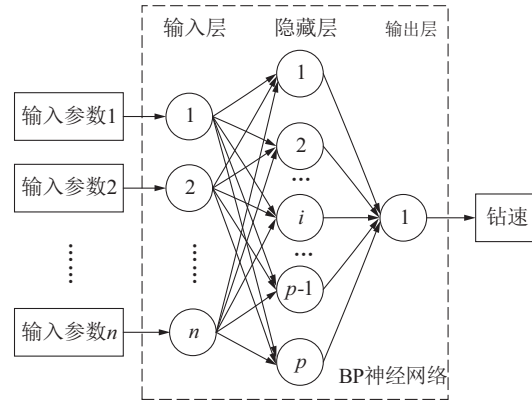


图3 本次研究使用的BP神经网络基本结构  
Fig.3 Basic structure of the BP neural network used in this study

本次研究中BP神经网络建模精度选择以预测输出 $y_{i-output}$ 与真实值 $y_{i-true}$ 之间的绝对误差RMSE与拟合精度 $R^2$ 为精度度量,其各自计算方法如式(3)所示,当计算结果 $|RMSE|=0, R^2=1$ 时,BP神经网络的预测输出与真实值完全匹配,达到最高精度。

$$\begin{cases} RMSE = \sum_{i=1}^m (y_{i-output} - y_{i-true}) / \sqrt{m} \\ R^2 = 1 - RMSE^2 / \{ \sum_{i=1}^m [(y_{i-output} - y_{i-true})^2] / m \} \end{cases} \quad (3)$$

式中: $y_{i-output}$ ——第 $i$ 个元素BP神经网络的预测输出; $y_{i-true}$ ——第 $i$ 个元素真实值; $m$ ——元素数量。

### 2.2 建模数据维度下限分析

建模数据维度是指建立钻速预测模型的参数类型数量,而建模数据维度的下限则可被定义为达到足够建模预测精度需求的最少参数数量。对满足精度的建模参数数量下限的研究,可最大程度地降低现场对参数检测需求或后期分析数据需求,避免由于对建模数据量的盲目增长需求而带来的生产与分析成本的急剧上升。结合图1所示的相关性分析结果,本次研究分别针对高、中、低相关区的建模参数组合进行。具体分析方案设计为依次从高、中、低相关性参数组中相关性最低的参数开始,逐渐增加引入BP神经网络输入参数的数量,分别计算并统计引入不同输入参数数量后的网络建模预

测精度,从而获取钻速预测精度随引入参数数量不同的改变趋势。

基于高、中、低3种不同相关性的参数进行建模数据维度变化的钻速预测建模精度分析,分析结果如图4所示,其变化趋势完全类似。随着引入参数数量的提升,BP神经网络预测钻速与真实钻速的绝对误差  $RMSE$  值逐渐下降,同时预测精度  $R^2$  也逐渐上升。可观察到的是高、中、低3种相关性参数在引入3个及以上的参数数量后,无论其预测误差还是预测精度均得到跃升,当相关性越低时,其跃升的幅度越高。引入参数数量3可视为预测建模精度的跃升阈值。

除跃升阈值以外,还可观察到不同相关性的参数的建模精度随引入数量的提升范围存在上限,低

相关性参数在引入15个参数后建模精度达到上限约为92%,中相关性参数在引入13个参数后建模精度达到上限约92%,高相关性参数在引入11个参数后达到上限约94%。精度上限的存在说明初始数据集中参数包含的信息量存在上限,这也与相关性计算结果中最高相关性未超过0.7相符。

横向对比3种相关性参数组(如图4所示),在引入相同数量的参数时,参数的相关性越高,最终钻速预测的误差越低,预测精度也越高。但值得注意的是,这种区别并不如相关性指数的区分度那么明显。在仅引入单参数时,低相关性参数钻速预测最大误差达到13.31 m/h,同时其精度仅有19.28%。对比之下,单参数引入时中相关性参数预测的最大误差为11.78 m/h,预测精度49.92%,高相关性参数预测的最大误差为10 m/h,预测精度为67.58%。可见在引入较少参数时,即使相关性很高的参数预测误差也很大,预测精度也较低。同理,将同相关性所有参数引入后,低相关性参数能够取得的最低预测误差为5.42 m/h,预测精度为91.70%;中相关性参数能够取得的最低预测误差为5.41 m/h,预测精度为91.72%;高相关性参数能够取得的最低误差为4.59 m/h,预测精度为94.12%。三者之间差距并不大,这说明当引入参数数量达到一定程度后,即使是相关性较低的参数,通过增加引入参数的数量,BP神经网络也能取得较高的预测精度与较低的预测误差。

参数间相关性的不同对钻速预测精度的影响则表现为到达指定精度需要的参数数量下限不同。如图5所示,以工业常用的最低精度标准85%为限,基于本数据集,低相关性参数需要引入9个,中相关性参数需要引入6个,而高相关性参数仅需要引入4个。若将精度指标下限提升至90%,则低相关性、中相关性与高相关性参数则需分别引入12个、10个与9个。由此可见,随精度指标下限的提高,不同相关性参数的引入数量(维度下限)也在上升,且不同相关性参数引入数量差距在逐渐减少。这是由于原始数据建模精度存在上限所致,这表征原始数据集中包含的真正影响钻速的信息存在上限。该观点在如图4所示的建模数据维度下限试验中也得到验证,当引入参数数量超过10以后,高、中、低3种相关性参数的预测精度均达到上限,不再随数据维度的上升而改变。

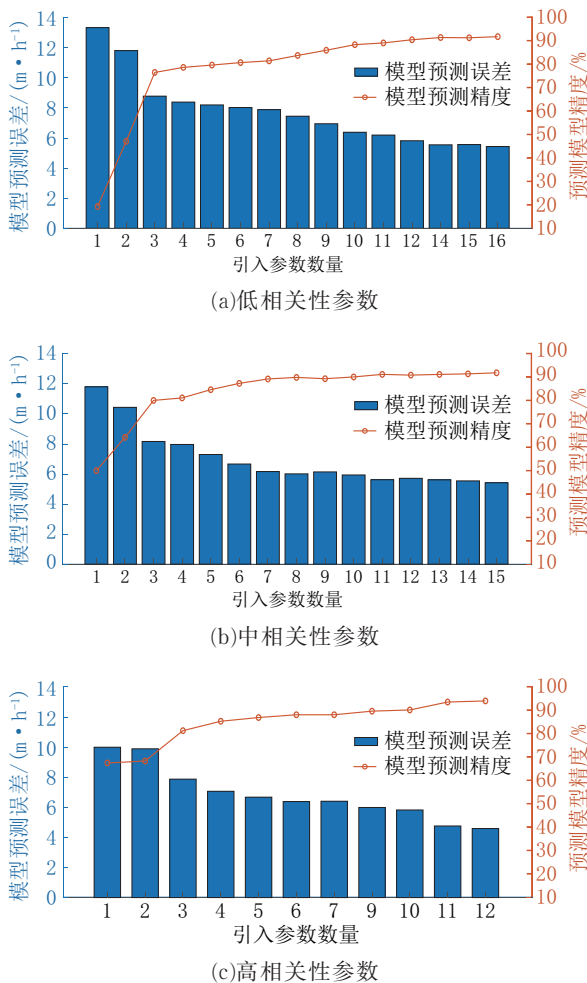


图4 钻速预测模型建模数据维度下限试验结果

Fig.4 Test results of the lower limit of the data dimension for the ROP prediction model

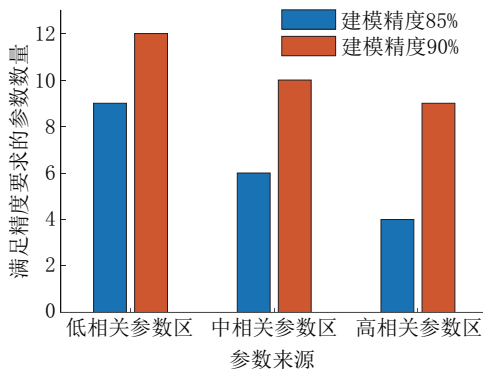


图5 不同相关性参数达到需求预测精度的维度下限

Fig.5 Lower limit of the dimension of different correlation parameters for the required accuracy

### 2.3 建模数据精度下限分析

由维度下限分析可知,当引入数据维度足够多时,即使是低相关参数也可取得较高的预测精度。在此基础上,本研究进一步分析在数据维度保持不变时,维持建模精度所需的数据取样间隔精度下限。对取样精度下限的研究,可减少数据分析时对数据量的需求,提高分析效率,避免对分析数据量的疑惑而引起对分析结论的怀疑。同时,也可降低实际钻探钻井采样时对录井、监测数据的要求。尤其是使用成本较高的随钻测量/测井(MWD/LWD)技术时,对采样精度的降低可提高设备的使用寿命,进一步降低使用成本。本次分析以图5所示85%和90%两组精度下的参数组合为蓝本,逐渐增加取样间隔,从基本数据集默认的1 m一个样本开始,逐步扩大到100 m一个样本。通过对不同精度样本进行建模误差和精度分析,观察是否能够找到建模精度下限。

高、中、低3种相关性精度下限分析结果如图6所示。整体来看,无论是相关性的区别(高、中、低)还是初始精度的区别(85%、90%),在取样间隔逐渐增大时呈现相同的变化趋势,即随着取样间隔的增大,钻速预测模型的误差同步增大,建模精度也逐渐下降。同时值得注意的是,如表4所示,在不同相关性与初始精度的数据结果统计中发现,当取样精度超过10 m时,所有不同相关性和初始精度条件下的建模预测误差增长幅度与建模下降率明显上升。说明取样精度10 m左右可视为建模取样精度的下限值。当取样精度超过10 m时,会引起精度的急剧下降,对建模及后续使用带来极大的分析

误差。

结合表4与图6相关数据,可发现在改变取样间隔时,建模参数的相关性相较初始精度更加敏感。低相关参数在取样精度间隔逐渐增大时会更明显的反映出建模误差增大与建模精度减小的情况,其变化幅度明显大于中相关与高相关参数。当数据取样间隔由1 m增长到100 m时,85%初始精度的低相关性参数建模误差由6.94 m/h增长到15.83 m/h,增长率为128%,同时建模精度由85.9%下降到52.2%,下降幅度33.7%;而90%初始精度的低相关性参数建模误差由5.80 m/h增长到12.47 m/h,增长率为115%,同时建模精度由90.4%下降至58.7%,下降幅度31.5%。相对而言,中相关性与高相关性参数的变化幅度稍小。85%初始精度的中相关性参数建模误差增长率为88%,精度下降22.2%;90%初始精度的中相关性参数建模误差增长率95%,精度下降29.9%。85%初始精度的高相关性参数建模误差增长率为52.8%,精度下降14.1%;90%初始精度的高相关性参数建模误差增长率86.8%,精度下降19.9%。由此说明,当建模数据相关性增大时,相应的由于建模取样数据间隔增大而导致的建模误差会减小,这说明基于高相关性参数建立的模型更加稳定与健壮,受数据取样精度的影响更小。

### 2.4 效果验证

结合数据建模维度与精度下限分析结果,分别以10 m的取样精度,以85%的初始精度建立基于低相关性参数(引入9参数)、中相关性参数(引入6参数)、高相关性参数(引入4参数)BP神经网络预测模型,将预测钻速与实测钻速对比如图7所示。由图7可知,无论初始数据的相关性如何,在数据维度与精度均达到下限值时,BP神经网络都能取得较高的预测精度。其中各模型主要的缺陷与隐患在于数据模型的稳定性,低相关性模型的稳定性将低于中、高相关性预测模型,建模参数发生变化时可能会引起较大的预测误差波动。

## 3 结论

基于中国南海某片区10口井相关参数的初始数据集(包含井眼位置、施工工艺、钻井液性能、地质条件、钻头钻具5大类44种不同参数共21917条数据),本文根据Person相关系数的定义,计算了43

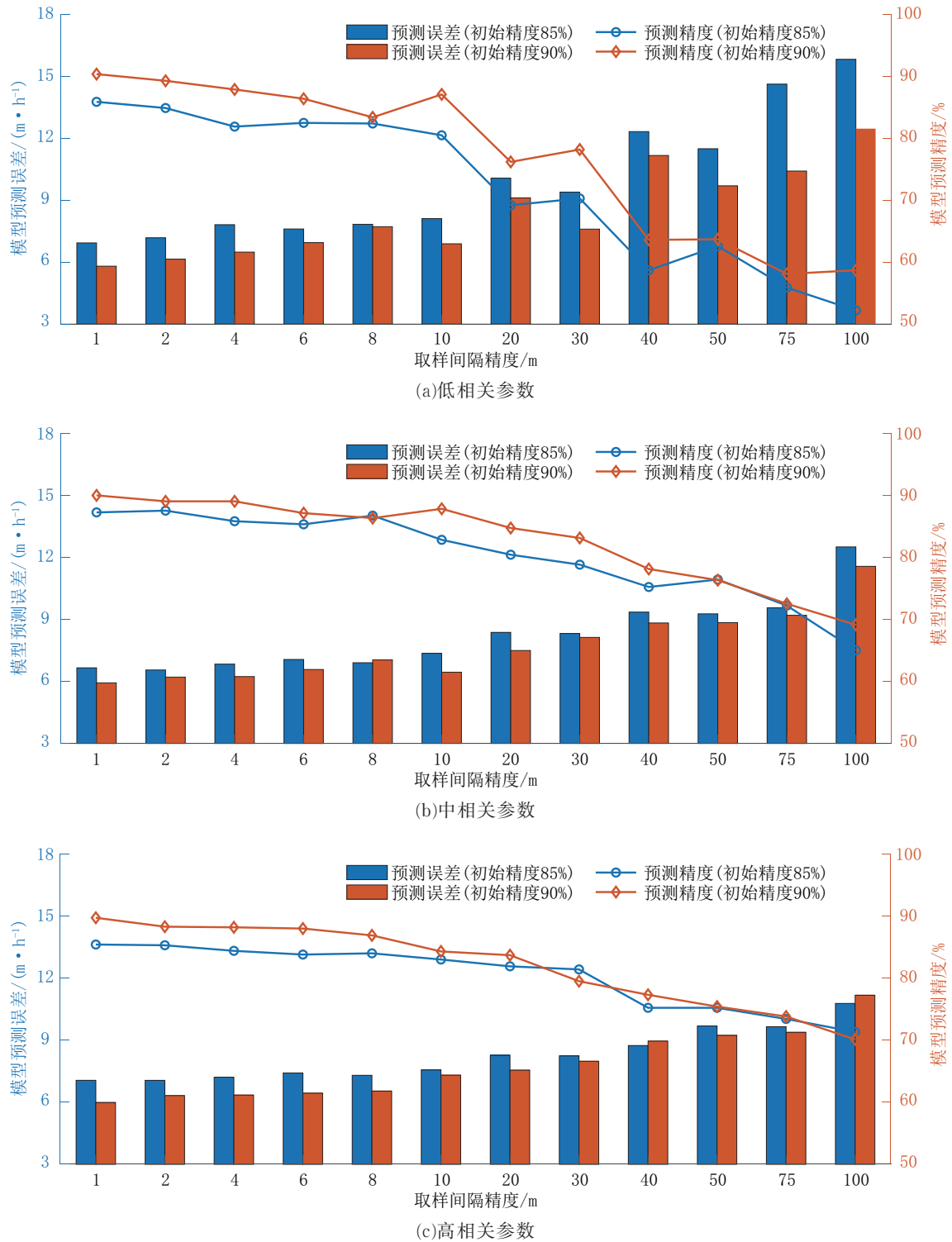


图6 钻速预测取样精度下限分析结果

Fig.6 Analysis results of the lower limit of sampling accuracy for ROP prediction

种不同输入参数与实际钻速之间的相关性,并将所有输入参数划分为低相关性参数(含16种参数)、中相关性参数(含15种参数)和高相关性参数(含12

种参数)。通过使用BP神经网络,结合10折交叉验证法,分别就低、中、高相关性参数建立了不同的钻速预测方程,并计算了各自的预测误差与预测精



表 4 精度下限建模试验分析结果

Table 4 Analysis results of modeling test for the lower accuracy limit

相关性划分	初始精度/%	取样精度 1 m		取样精度 10 m		取样精度 40 m		取样精度 100 m	
		RMSE/ (m·h <sup>-1</sup> )	R <sup>2</sup> /%	RMSE/ (m·h <sup>-1</sup> )	R <sup>2</sup> /%	RMSE/ (m·h <sup>-1</sup> )	R <sup>2</sup> /%	RMSE/ (m·h <sup>-1</sup> )	R <sup>2</sup> /%
低相关参数	85.9	6.94	85.9	8.11	80.5	12.32	58.7	15.83	52.2
	90.4	5.80	90.4	6.89	87.1	11.17	63.6	12.47	58.7
中相关参数	87.2	6.65	87.2	7.36	82.8	9.35	75.2	12.50	65.0
	90.0	5.93	90.0	6.44	87.8	8.82	78.1	11.56	69.1
高相关参数	85.4	7.07	85.4	7.58	83.0	8.76	75.2	10.80	71.3
	90.0	5.99	90.0	7.32	84.3	8.97	77.3	11.19	70.1

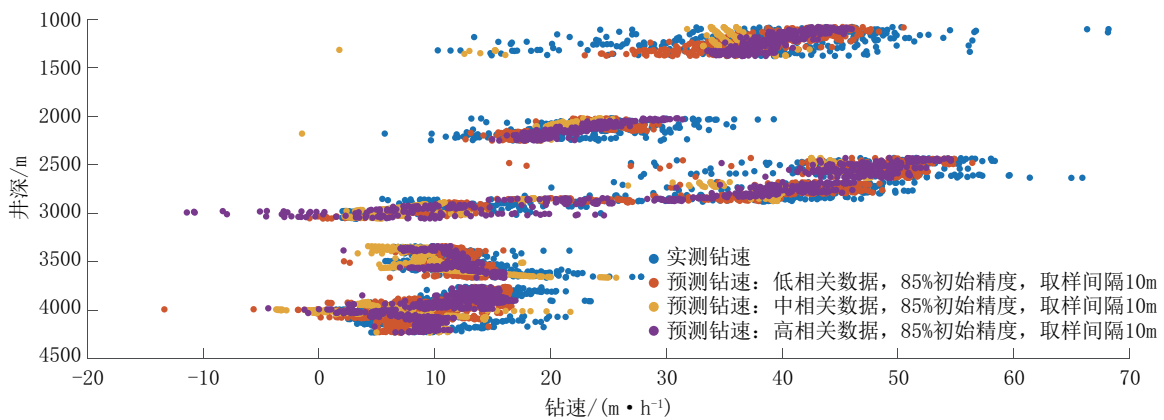


图 7 基于最低维度与取样精度下限的建模预测效果

Fig.7 Prediction results of the model based on the lowest data dimension and sampling accuracy

度,获得如下结论:

(1)就建模数据维度上看,BP神经网络建模预测的精确度将随引入参数数量的上升而上升。参数的相关性决定了引入参数数量的下限。以工业常用的85%精度为下限,低相关性参数需要引入9个,中、高相关性参数则分别需要引入6个与4个,当精度下限提升到90%,相应的需要引入的低、中、高相关性参数分别提升为12个、10个与9个。

(2)在引入足够的参数后,无论引入参数的相关性高低,都可取得大于90%的预测精度。但不同参数能够达到的预测精度存在上限,本数据集中低、中、高相关性参数建模精度的上限分别为92%、92%与94%左右。该上限值与引入参数包含的有效信息相关,预测精度达到上限后继续引入参数并不会提高预测精度。

(3)在确定初始预测精度较高(≥85%)的前提下,建模参数取样的精度间隔增大,会引起建模参数数量的降低和其中包含信息的丢失,从而会引起

预测精度的下降。通过分别对高、中、低相关性参数组进行建模验证可知,当数据取样间隔超过10m后会引入建模精度的急剧下降,故本文数据集的钻速预测模型的建模取样精度下限为10m。

(4)针对不同相关性与初始精度的预测建模对比后发现,参数相关性的降低会增大由于取样间隔增大而导致的建模误差,即使用低相关性参数建立模型需要的取样精度下限应略高于高相关性参数建立模型时的取样要求,以防建模过程中产生的波动。

(5)验证结果表明,在引入足够的参数数量与取样间隔精度后,低相关性、中相关性、高相关性参数均可建立足够准确的BP神经网络钻速预测方程。

参考文献(References):

[1] 孟英峰,杨谋,李皋,等.基于机械比能理论的钻井效率随钻评价及优化新方法[J].中国石油大学学报(自然科学版),2012,

- 36(2):110-114.
- MENG Yingfeng, YANG Mou, LI Gao, et al. New method of evaluation and optimization of drilling efficiency while drilling based on mechanical specific energy theory[J]. *Journal of China University of Petroleum (Edition of Natural Science)*, 2012, 36(2):110-114.
- [2] Chen X, Fan H, Guo B, et al. Real-time prediction and optimization of drilling performance based on a new mechanical specific energy model[J]. *Arabian Journal for Science and Engineering*, 2014, 39(11):8221-8231.
- [3] 邹德永,王家骏,卢明,等.定向钻井PDC钻头三维钻速预测方法[J].*中国石油大学学报(自然科学版)*,2015,39(5):82-88.  
ZOU Deyong, WANG Jiajun, LU Ming, et al. A novel prediction method of PDC bit's ROP in directional drilling[J]. *Journal of China University of Petroleum (Edition of Natural Science)*, 2015, 39(5):82-88.
- [4] 刘军波,韦红术,赵景芳,等.考虑钻头转速影响的新三维钻速方程[J].*石油钻探技术*,2015,43(1):52-57.  
LIU Junbo, WEI Hongshu, ZHAO Jingfang, et al. A new 3D ROP equation considering the rotary speed of bit[J]. *Petroleum Drilling Techniques*, 2015, 43(1):52-57.
- [5] Saksala T, Fourmeau M, Kane P, et al. 3D finite elements modelling of percussive rock drilling: Estimation of rate of penetration based on multiple impact simulations with a commercial drill bit[J]. *Computers and Geotechnics*, 2018, 99:55-63.
- [6] Bataee M, Irawan S, Kamyab M. Artificial neural network model for prediction of drilling rate of penetration and optimization of parameters[J]. *Journal of the Japan Petroleum Institute*, 2014, 57(2):65-70.
- [7] 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 and Engineering*, 2017, 159:295-306.
- [8] Hegde C, Gray K E. Use of machine learning and data analytics to increase drilling efficiency for nearby wells[J]. *Journal of Natural Gas Science and Engineering*, 2017, 40:327-335.
- [9] Arabjamaloei R, Shadzadeh S R. Modeling and optimizing rate of penetration using intelligent systems in an Iranian southern oil field (Ahwaz Oil Field) [J]. *Petroleum Science and Technology*, 2011, 29:1637-1648.
- [10] Yi P, Kumar A, Samuel R. Realtime rate of penetration optimization using the shuffled frog leaping algorithm [J]. *Journal of Energy Resources Technology, Transactions of the ASME*, 2015, 137:32902.
- [11] Ansari H R, Sarbaz Hosseini M J, Amirpour M. Drilling rate of penetration prediction through committee support vector regression based on imperialist competitive algorithm[J]. *Carbonates and Evaporites*, 2017, 32(2):205-213.
- [12] Eskandarian S, Bahrami P, Kazemi P. A comprehensive data mining approach to estimate the rate of penetration: Application of neural network, rule based models and feature ranking [J]. *Journal of Petroleum Science and Engineering*, 2017, 156: 605-615.
- [13] Ahmed, Ali, Elkatatny, et al. New artificial neural networks model for predicting rate of penetration in deep shale formation [J]. *Sustainability*, 2019, 11(22):6527.
- [14] Gan C, Cao W, Wu M, et al. Prediction of drilling rate of penetration (ROP) using hybrid support vector regression: A case study on the Shennongjia area, Central China[J]. *Journal of Petroleum Science and Engineering*, 2019, 181:106200.
- [15] Bhatnagar A, Khandelwal M. An intelligent approach to evaluate drilling performance [J]. *Neural Computing and Applications*, 2012, 21(4):763-770.
- [16] Moraveji M Keshavarz, Naderi M. Drilling rate of penetration prediction and optimization using response surface methodology and bat algorithm[J]. *Journal of Natural Gas Science and Engineering*, 2016, 31:829-841.
- [17] Al-Abduljabbar A, Elkatatny S, Mahmoud M, et al. A robust rate of penetration model for carbonate formation[J]. *Journal of Energy Resources Technology*, 2018, 141(4).
- [18] Elkatatny S. Development of a new rate of penetration model using self-adaptive differential evolution-artificial neural network [J]. *Arabian Journal of Geosciences*, 2019, 12(2).

(编辑 周红军)