

基于地层成分和钻进参数的钻速预测模型

熊虎林¹, 李 谦²

(1.新疆地质矿产勘查开发局第九地质大队,新疆 乌鲁木齐 830000; 2.成都理工大学环境与土木工程学院,四川 成都 610059)

摘要:钻速预测是优化钻进施工过程中的重点难点。本文以实际采集的数据资料(包含 4 大类共 28 种不同的数据类型)出发,建立了以地层成分和钻进参数相结合为基本的钻速预测模型。为了提高模型的预测精度,将原始数据首先进行了标准化处理,消除了其量纲和数量级对模型预测可能产生的影响。由于原始数据种类过多,将原始数据先后进行了聚类分析和因子分析,提取其中有效信息最终将其缩减为 11 个输入参数和 1 个输出参数的预测模型。利用人工神经网络技术对这个模型进行了非线性拟合,结果显示本预测模型能够将误差控制在 10% 以内,具有一定的指导生产实践的能力。

关键词:钻速预测模型;地层成分;钻进参数;多元统计分析;神经网络

中图分类号:P634.9 **文献标识码:**A **文章编号:**1672-7428(2018)10-0195-07

ROP Prediction Model Based on Formation Composition and Drilling Parameters/XIONG Hu-lin¹, LI Qian²(1.No.9 Geological Team, Xinjiang Geology and Mineral Resources Exploration and Development Bureau, Urumqi Xinjiang 830000, China; 2.College of Environment and Civil Engineering, Chengdu University of Technology, Chengdu Sichuan 610059, China)

Abstract: Prediction is an important and difficult target during the drilling optimization process. Drawn upon the field data(comprising of a total of 28 types in 4 categories), a ROP prediction model is developed with combination of formation composition and drilling parameters. In order to increase the prediction accuracy and avoid the potential impact caused by its dimension and order of magnitude, the raw data is standardized. In addition, due to the excessive volume of the raw data, it is processed with polymer analysis and factor analysis to pick effective data, and then reduced to a prediction model of 11 input parameters and 1 output parameter. The technology of artificial neural network is used to perform the non-linearly fitting work, and the results indicate that the prediction error is less than 10%, which means the proposed prediction model can somehow provide guidance for field operations.

Key words: ROP prediction model; formation composition; drilling parameters; multivariate statistical analysis; neural network

0 引言

近年来,由于资源的持续开采,埋深浅、易开采的矿产油气资源储量急剧降低,故埋深大、难开采的矿产油气资源是当前钻探钻井领域所面对的主要施工对象。为了提高施工效率,优化钻进技术受到了极大的关注,其中对钻速的预测是整个钻进施工过程中重难点。由于实际钻进施工条件复杂、影响因素极多且难以考虑完全,钻速的预测判定一直是优化钻进过程中的主要难题^[1-3]。近年来,随着计算机及网络技术的发展,已有相关研究人员将大数据算法应用到钻速预测领域。如景宁等应用层次分析法与前馈神经网络建立了 AHP-BP 组合模型,利

用岩石抗压强度、钻头尺寸、钻头类型、泵压、转速和钻井液密度预测了实际钻速,将相对误差控制在了 10% 以内^[4]。武成刚等利用主成分分析法将岩石抗压强度、抗拉强度、密度、声波波速和平均节理间距 5 个常见的岩石物理力学参数建立 3 个新变量,建立了新变量与钻速的预测模型^[5]。李昌盛利用 Bourgoyne & Youngs 方程为理论基础,从元坝地区抽取 300 组录井数据利用多元回归的方法建立了钻速预测模型,平均误差 < 13%^[6]。

尽管目前已有不少的研究者在钻速预测领域进行研究,但经过调研,目前在钻速预测分析领域尚未出现基于地层成分和钻进参数共同组合的钻速预测

收稿日期:2018-07-13

作者简介:熊虎林,男,汉族,1986 年生,工程师,勘查技术与工程专业,从事钻井施工工作,新疆乌鲁木齐市西山路 67 号,xiionghulin@126.com。

模型相关研究。而从实际钻进施工出发,地层成分直接在很大程度上决定了地层的岩性及其工程力学强度,同时目前大多数临井地质资料记录多以地层成分为主,因此基于地层成分的钻速预测模型相对于目前常用的以岩石力学强度更为基础和实用。除此以外,钻进参数的选择也是实际钻速控制的一个主要因素,因此本文以新疆乌鲁木齐市米东区页岩气钻进米参1井实际数据为资料,建立了基于地层成分和钻进参数组合的钻速预测模型。

1 原始数据分析

1.1 数据收集与标准化

基于米参1井的数据资料,现场收集到的数据资料详尽繁杂,从中提取的应用于建立本次钻速预测模型的数据类型如表1所示,共4大类,28个不同的数据类型,通过对钻井井史资料的筛选,总共选取了98组不同深度的数据进行了分析。

表1 初始选用数据类型

数据类型	选用数据
矿物含量/%	石英、钾长石、斜长石、方解石、铁白云石、白云石、伊蒙间层、伊利石、高岭石、绿泥石
常量元素/%	二氧化硅、三氧化二铝、三氧化二铁、氧化钙、氧化镁、氧化钾、氧化钠、二氧化钛、五氧化二磷、氧化锰、氧化亚铁
钻进参数	钻压(kN)、转速(r/min)、排量(L/s)、泵压(MPa)、机械钻速(m/h)
地层深度/m	顶深、底深

收集的原始数据由于各指标之间计量单位和数量级的不同,导致各指标差异很大,无法直接进行综合分析。因此,为提高分析的精度和速度,解决各指标数据间不可综合性问题,需要对各指标进行数据标准化处理。在数据分析领域中,数据标准化是为了去除量纲,将不同数量级的数据变成同一数量级,消除数量级的影响。如表1所示可以明显看出原始数据的量纲差异。同理,在原始数据中,数量级的差距非常明显,如地层深度数据最大为2100.18 m,常量元素二氧化硅最大值为66.37%,机械钻速最小0.1 m/h。这种数量级的差距若不进行数据标准化处理,数据分析建模过程中小数量级的数据影响则会被大数量级的数据影响掩盖,不能完全体现出数据对结果的影响^[7]。本文中使用的常规的标准差法对原始数据进行预处理,即以每一数据与其平均值之差除以该指标的标准差(式1),经过处理后的所有数据如表2所示,不仅消除了数据本身量纲,也将数据大小全部投影到[-1,1]之间,最大程度的保留了数据自身的差异性。

$$x_i' = \frac{x_i - \bar{x}}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} \quad (1)$$

式中: x_i' ——处理后的数据指标; x_i ——处理前的指标; \bar{x} ——处理前指标的平均值; n ——样品数量。

1.2 数据相关性分析

表2 原始数据标准化前后对比(部分)

地层深度/m	机械钻速/(m·h ⁻¹)		钻压/kN		转速/(r·min ⁻¹)		排量/mL		泵压/MPa		
	前	后	前	后	前	后	前	后	前	后	
1500.00	0.009	0.68	-0.041	60	0.052	90	0.107	28	0.023	6.5	-0.080
1511.18	0.011	0.31	-0.090	40	-0.052	60	-0.061	20	-0.117	3.5	-0.177
1520.00	0.013	0.37	-0.082	50	-0.000	60	-0.061	24	-0.047	10.0	0.033
1539.69	0.017	0.18	-0.107	50	-0.000	50	-0.118	20	-0.117	11.5	0.082
1550.58	0.019	0.67	-0.043	40	-0.052	60	-0.061	20	-0.117	11.0	0.065
1561.00	0.021	0.77	-0.030	60	0.052	90	0.107	28	0.023	11.0	0.065
1568.68	0.023	0.27	-0.095	30	-0.104	50	-0.118	20	-0.117	10.0	0.033
1580.00	0.025	0.63	-0.048	60	0.052	90	0.107	28	0.023	10.0	0.033

注:计算机进行数据处理时小数位数不止两位,在此为简化只列出两位小数。

在钻速预测模型分析建模时,即使是近几年新兴的数据挖掘或神经网络等算法,都是直接利用数据建立模型,但这些方式忽略了一个重要的问题,各数据之间的线性相关性。本文建立的模型分析的数据之间存在线性相关的现象,在数理统计领域被称为多变量的多重共线性现象。这种数据指标之间的

多重共线性会对后期建立钻速预测模型带来影响,严重影响到相关参数的估计,错误的扩大或隐藏部分参数对整个模型的影响^[8]。因此数据之间的多重共线性会扩大模型误差,破坏模型的稳健性。这种现象对钻速预测的危害十分严重,但其存在却又十分普遍。基于数理统计的基本定义,本文讨论的相

相关性系数的计算方法如式(2)所示。

$$r(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2)$$

式中: x_i, y_i ——变量 X, Y 的数据; \bar{x}, \bar{y} ——变量 X, Y 的平均值。

相关性计算结果区间为 $[-1, 1]$, 若计算结果越接近 ± 1 , 则表示计算的两个变量之间具有越强的线性相关性。如表 1 所示原始数据在采集过程中分为了明显的三大类, 各类型数据之间没有相关性, 因此本节的相关性分析主要集中在三大类数据内部各指标之间。分属不同类型的数据指标由于代表物理意义完全不同则不进行相关性计算, 视为完全不相关。结果显示, 尽管理论上三大类数据各单项指标之间不存在任何实际意义上的相关性, 部分数据之间相关性仍旧非常高, 超过了传统上认为的相关性系数阈值 ± 0.6 。因此可认为尽管目前尚未有研究显示各地层成分和钻进参数之间的相关性, 但这个相关性仍旧实际存在。表 3 所示为三大类数据中相关性系数接近或超过阈值的数据参数, 未超过阈值的数据参数省略未列出。

表 3 原始数据高相关性数据组合

数据类型	相关参数组合	相关性系数
矿物含量	石英-钾长石	-0.560
	石英-斜长石	-0.634
	石英-白云石	-0.595
	钾长石-伊利石	-0.678
	斜长石-方解石	-0.545
	斜长石-伊利石	-0.624
	白云石-伊利石	-0.511
常量元素	二氧化硅-氧化钙	-0.688
	二氧化硅-氧化镁	-0.655
	三氧化二铝-氧化钾	0.673
	三氧化二铝-氧化钛	0.936
	三氧化二铁-氧化亚铁	0.648
	氧化钙-氧化钛	-0.517
	氧化镁-氧化钠	0.630
	氧化钾-氧化钛	0.560
钻进参数	钻压-转速	0.690

从表 3 中可以看出, 三大类数据中各数据指标之间或多或少都存在相关性, 其中矿物含量和常量元素之间存在相关性的数据指标组合数量较多, 钻进参数数据类中只有钻压和转速存在较高的相关性。这种现象可以解释为在自然条件下指定区域中

地层生成的条件存在一定规律性, 因此其矿物成分和常量元素之间必然存在某种联系, 这种联系就是相关性存在的理论基础。而钻进参数多来自于现场施工人员的经验, 人工可调的形式使参数之间的相关性显著降低。

1.3 数据建模分析流程

由表 1 可知, 在已知地层条件的前提下可提取用以预测钻速的数据繁杂, 若直接利用所有数据进行建模分析则可能耗费大量的计算资源, 也并不能保证建模的准确性。由表 3 可知, 选用的数据之间存在较高的相关性, 这种相关性的存在也会为钻速预测模型的建立带来较为明显的误差。因此本文拟建立的钻速预测模型分析流程如图 1 所示, 在对原始数据进行标准化处理依次进行聚类分析、因子分析和数据建模 3 个主要流程, 建模完成后再进行验证和应用。其中聚类分析和因子分析均是了解决原始数据类型太多且存在相关性的问题而设置的分析流程, 数据建模则是利用了非线性拟合较为准确的人工神经网络技术进行建模。

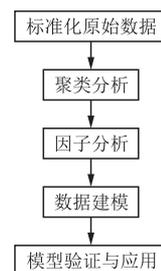


图 1 钻速预测模型建模分析流程

2 聚类分析

2.1 聚类分析简介

建模过程中首先进行的聚类分析是为了解决原始数据种类过多的问题。作为数据分析的重要工具, 聚类分析的主要目的是寻找原始数据集中的“自然分组”, 分好组以后可认为同组的数据指标具有相同的作用和功能, 从而在数据建模分析中同组只需选择其中最具代表性的数据指标即可^[9]。通过聚类分析, 可在表 1 所示的三大类数据中(钻进参数数据本身较少, 无需聚类分析)各自分类, 从中选取具有代表性的数据指标进行下一步的因子分析, 从而达到了减少原始数据指标类型的目的。

本文拟选择最直接的直接聚类法进行聚类分析, 其基本思想为先把聚类对象单独视为一类, 依次

将距离最小的两个对象合并为新类,后计算新类与其他类的距离。重复进行上述步骤直到所有类均合并为一类。本文进行的聚类分析属于变量间的聚类分析(R型聚类统计量),描述变量 X 、 Y 间的距离 d_{XY} 以式(3)所示的夹角余弦来度量。

$$d_{XY} = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{(\sum_{i=1}^n x_i^2)(\sum_{i=1}^n y_i^2)}} \quad (3)$$

2.2 数据聚类结果分析

利用数据分析软件 SPSS 自带的工具箱完成矿物含量和常量元素原始数据的聚类分析,结果如图 2、图 3 所示。从聚类结果中可以看出各数据之间的相似程度不同,而相似程度较高的数据可只取其中之一即可。以筛除 40%左右的数据变量为原则,可将变量间的距离阈值取为 10,则经过聚类分析的筛选结果如表 4 所示,矿物含量从 10 个数据指标筛选

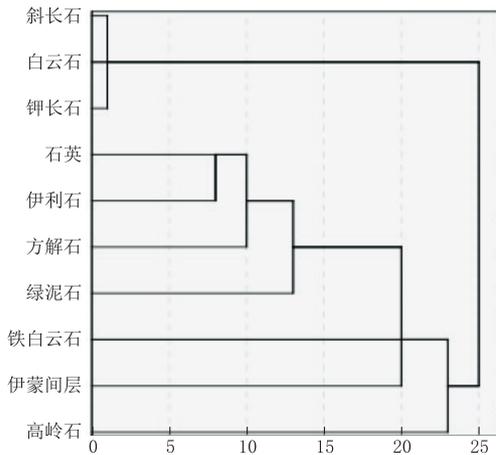


图 2 矿物含量聚类分析图

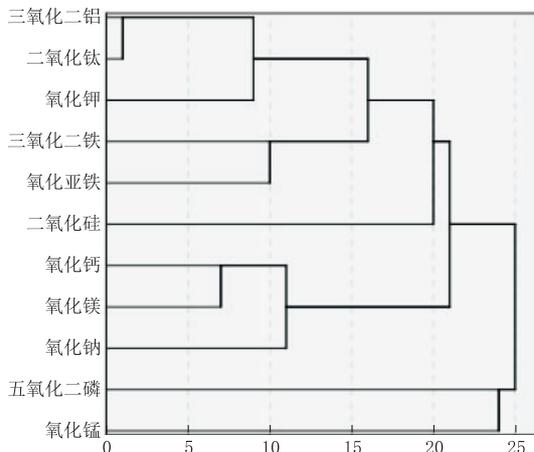


图 3 常量元素聚类分析图

表 4 经过聚类分析筛选出的代表性的数据类型

数据类型	筛出数据
矿物含量	石英、钾长石、铁白云石、伊蒙间层、高岭石、绿泥石
常量元素	二氧化硅、三氧化二铁、氧化钙、氧化钠、二氧化钛、五氧化二磷、氧化锰

为 6 个,常量元素从 11 个数据指标筛选为 7 个。

3 因子分析

3.1 因子分析简介

完成原始数据类型的筛选后,因子分析的主要目的是为了消除如表 3 所示的数据间的相关性。因子分析模型属于多元统计学科中处理大量数据的一种方法,目的在于从研究原始变量或参数之间相关系数矩阵的内部结构出发,把一些具有错综复杂关系的变量归结为少数几个公共因子。其基本思想是根据相关性的大小将变量分组,使得同组内变量之间相关性较高,不同组的变量相关性较低,每组变量代表一个基本结构,用一个不可观测的综合变量表示,这个基本结构称为公共因子。如式(4)所示,因子分析的模型可以表示为在 n 个原始指标包含少数 m ($m < n$) 个因子与特殊因子的线性组合^[10]。

$$\begin{cases} f_1 = a_{11}X_1 + a_{12}X_2 + \dots + a_{1n}X_n \\ f_2 = a_{21}X_1 + a_{22}X_2 + \dots + a_{2n}X_n \\ \vdots \\ f_m = a_{m1}X_1 + a_{m2}X_2 + \dots + a_{mn}X_n \end{cases} \quad (4)$$

式中: X_1, X_2, \dots, X_n ——原始数据类型; f_1, f_2, \dots, f_m ——公共因子; $(a_{ij})_{n \times m}$ ——因子得分矩阵,用以表示第 i 个变量在第 j 个指标中所占比例。

3.2 因子分析指标检验

根据因子分析的流程,首先需要判断用这些数据作因子分析是否合适,因此需要进行 KMO 检验和 Bartlett 球状检验。KMO (Kaiser - Meyer - Olkin) 检验统计量是用于比较变量间简单相关系数和偏相关系数的指标。当相关系数平方和远远大于偏相关系数平方和时, KMO 值接近 1, 意味着变量间的相关性强,适合作因子分析;反之 KMO 接近 0, 表明原有变量间相关性弱,不合作因子分析。常用的 KMO 度量标准是 0.9 以上表示非常适合; 0.8 表示适合; 0.7 表示一般; 0.7 以下均不适合。Bartlett 球形度检验以各变量的相关系数矩阵为基础,是用于检验各变量数据是否相互独立。若 Bartlett 球形度检验的显著性 $Sig > 0.05$, 则表示各变

量各自独立相关性低,无需进行因子分析。

基于 SPSS 软件对三大类输入数据(矿物含量和常量元素采用经过聚类分析筛出的指标,为消除数据相关性在因子分析时将钻进参数数据类也加入)进行 KMO 检验和 Bartlett 球状检验结果如表 5 所示。从表 5 中可以看出,矿物含量、常量元素和钻进参数三大类指标均适合进行因子分析,KMO 指标均超过 0.7,Bartlett 指标也未超过 0.05。

表 5 因子分析指标检测值

检验方法	矿物含量	常量元素	钻进参数
KMO 检验	0.757	0.797	0.701
Bartlett 检验	2.52×10^{-8}	1.44×10^{-24}	5.31×10^{-24}

3.3 提出综合因子

利用 SPSS 对三大类数据进行因子分析,首先需要确定综合因子数量。每一个综合因子的选择都是原始数据的体现,但要达到综合并且降低指标数量的需求,综合因子在一定程度上会牺牲部分原始数据的信息,在因子分析中以特征值对样本方差的累积贡献率进行度量。累积贡献率为 100% 时认为综合因子包含了原始数据的全部信息,但此时综合因子数量与原始数据数量相同,因此并不能达到降低数据种类的目的。因此为了最大程度的协调综合因子数量与包含原始数据信息的关系,在本次因子

$$a_k = \begin{bmatrix} 0.569 & -0.460 & 0.121 & -0.085 & 0.131 & 0.241 \\ 0.266 & 0.085 & -0.048 & 0.032 & 0.747 & -0.538 \\ 0.198 & -0.002 & 0.860 & -0.059 & -0.152 & -0.376 \\ -0.107 & -0.066 & -0.102 & 0.989 & -0.062 & -0.227 \end{bmatrix} \quad (5)$$

分析中,如式(6)所示,表示常量元素的因子得分矩阵 a_c 。自左至右每一列代表的原始数据分别为二氧化硅、三氧化二铁、氧化钙、氧化钠、二氧化钛、五氧化二磷、氧化锰,自上而下代表了 4 个综合因子。可看

$$a_c = \begin{bmatrix} -0.522 & 0.175 & 0.429 & 0.167 & -0.109 & 0.155 & -0.108 \\ -0.170 & 0.740 & 0.031 & 0.017 & 0.463 & -0.006 & 0.031 \\ -0.122 & 0.025 & 0.069 & -0.486 & -0.067 & 0.795 & -0.121 \\ 0.264 & 0.137 & 0.067 & 0.118 & -0.211 & -0.073 & 0.896 \end{bmatrix} \quad (6)$$

出综合因子 1 主要由二氧化硅和氧化钙的含量确定,综合因子 2 主要由三氧化二铁和二氧化钛的含量确定,综合因子 3 主要由氧化钠和五氧化二磷的含量确定,综合因子 4 主要由氧化锰的含量确定。

如式(7)所示,表示钻进参数的因子得分矩阵自左至右每一列代表的原始数据分别为钻压、转速、排量和泵压,自上而下代表了 2 个综合因子。可看出综合因子 1 主要由钻压和转速值确定,综合因子 2 则主要由排量和泵压确定。

分析中将综合因子的累积贡献率阈值设定为 80%,即选定的综合因子将包含 80% 原始数据的信息,综合因子数量确定如表 6 所示。

表 6 综合因子计算分析结果

数据类型	矿物含量	常量元素	钻进参数
综合因子数量	4	4	2
累积贡献率/%	82.083	83.016	80.733

确定综合因子数量后,即可从式(4)所示的因子得分矩阵中确定每一个综合因子的转换方程,从而不仅确定了综合因子的计算方法,也在一定程度上确定该综合因子的物理意义。三大类数据的综合因子计算结果如式(5)~式(7)所示。式(5)~式(7)所示因子得分矩阵的每一行代表了一个综合因子,每一列则代表了一个原始数据,每一项数据的大小则代表了该原始数据在综合因子中所占的比重。

如式(5)所示,表示矿物含量的因子得分矩阵 a_k 自左至右每一列代表的原始数据分别为石英、钾长石、铁白云石、伊蒙间层、高岭石、绿泥石,自上而下代表了 4 个综合因子。可看出综合因子 1 主要由石英和钾长石含量确定,综合因子 2 主要由高岭石和绿泥石的含量确定,综合因子 3 主要由铁白云石含量确定,综合因子 4 主要由伊蒙间层确定。

出综合因子 1 主要由二氧化硅和氧化钙的含量确定,综合因子 2 主要由三氧化二铁和二氧化钛的含量确定,综合因子 3 主要由氧化钠和五氧化二磷的含量确定,综合因子 4 主要由氧化锰的含量确定。

$$a_p = \begin{bmatrix} 0.491 & 0.454 & 0.239 & 0.145 \\ 0.203 & -0.070 & -0.471 & 0.709 \end{bmatrix} \quad (7)$$

4 神经网络建模

4.1 神经网络简介

通过因子分析,三大类原始数据种类数进一步降低,矿物含量数据类型降低为4类,常量元素数据类型也降低为4类,钻进参数数据类型降低为2类。这10种数据指标相关性两两为零,因此本文讨论的钻速预测模型即可简化为有10个输入参数、1个输出参数(机械钻速)的数学模型。这种多输入变量的非线性拟合模型用人工神经网络技术能够取得比较好的拟合效果。

人工神经网络(Artificial neural network,简称ANN)技术自20世纪40年代起源以来,经历计算机硬件、软件和计算理论的蓬勃发展,至今已形成一种能进行并行分散处理、具有非线性映射、自适应学习和较强容错性的智能分析工具。其中基于误差反向传播算法(Back-Propagation)的BP神经网络应用最为广泛。如图4所示为BP算法的典型结构,主要是由一个输入层、一个或多个隐含层、一个输出层组成。各层由若干个神经节点构成,每一个节点能够根据输入值、作用函数和阈值计算得到输出值。网络整个学习过程由分为信息正向传播和误差反向传播两个过程。信息的正向传播过程中输入信息依次经过输入层、隐含层、输出层,输出层处理后作为最终结果输出;而在误差的反向传播过程中,网络自动对比输出结果与期望输出值的误差,若误差过大则逐层将误差信息反馈给输出层和隐含层,系统自行调整隐含层和输出层阈值并重新计算,如此循环直至输出误差满足要求为止^[11]。

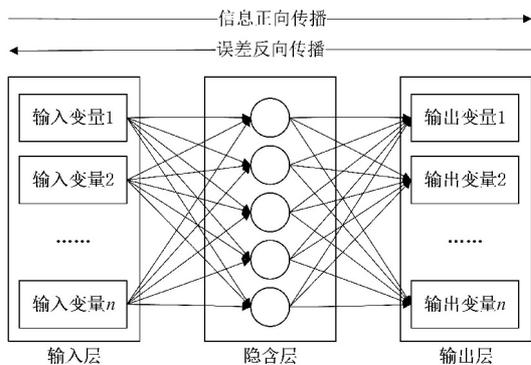


图4 BP神经网络算法结构

4.2 神经网络建模及效果分析

在保证训练数据量的前提下,训练后的神经网络对多输入单输出非线性模型具有极高的拟合和预测精度。根据前文的参数设置,经过标准化、聚类分析和因子分析处理的98组原始数据转化为一个98行×11列的输入矩阵 M_{in} ,其中每一行数据代表一

个深度的矿物含量、常量元素、钻进参数数据,而经过标准化的机械钻速则生成一个98行×1列的目标矩阵 M_{out} 。利用Matlab自带的BP神经网络工具箱即可完成神经网络的训练和验证。整个过程中系统自动将98行数据随机分为3类,75%的数据(73个)作为训练数据;15%的数据(15个)作为验证数据,用以在训练过程中实时验证结果,修订输出值;10%的数据(10个)作为测试数据,用以测试训练完成后的训练模拟精度。训练结果如图5所示,可以看出无论是用以训练、验证和测试集,神经网络的预测精度都很高。以相关系数 R 进行预测精度的度量,训练集的相关系数 $R=0.97744$,验证集的相关系数 $R=0.93144$,测试集的相关系数 $R=0.81583$,综合所有数据的相关系数 $R=0.90138$,可知用神经网络进行钻速预测的精度误差达到了10%以内,能够满足现场实际使用的需要。

5 结论

本文以新疆乌鲁木齐市米东区页岩气钻进米参1井实际数据为资料,建立了基于地层成分和钻进参数组合的钻速预测模型。原始数据资料涵盖了大量的地层成分信息(总共包含地层深度,矿物含量包含10类数据,常量元素11类数据,钻进参数4类数据,机械钻速),若直接采用所有的数据进行预测,一方面可能加大运算难度,另一方面也可能降低预测精度,故本文采用了标准化、聚类分析、因子分析3个阶段的数据分析对原始数据进行了处理。标准化将所有的原始数据投影到 $[-1,1]$ 区间,在保留所有数据差别信息的同时消除了原始数据的量纲和数量级影响,降低了对建立模型的误差。聚类分析则通过分析原始数据间的相似程度对其进行归类,用以消除数据类似的相关数据,降低原始数据类别。通过聚类分析,矿物含量数据类型降低为6类,常量元素类型降低为7类。因子分析则为了消除原始数据之间的相关性,进一步提高模型预测的精度。通过因子分析,矿物含量和常量元素数据类型各提取了4个综合因子,钻进参数数据类型提取了2个综合因子。基于数据处理,利用BP神经网络建立包含地层信息和钻进参数信息的钻速预测模型,该模型包含11个输入参数(地层深度、4个矿物含量综合因子、4个常量元素综合因子、2个钻进参数综合因子)和1个输出参数(机械钻速)。通过对98组数据

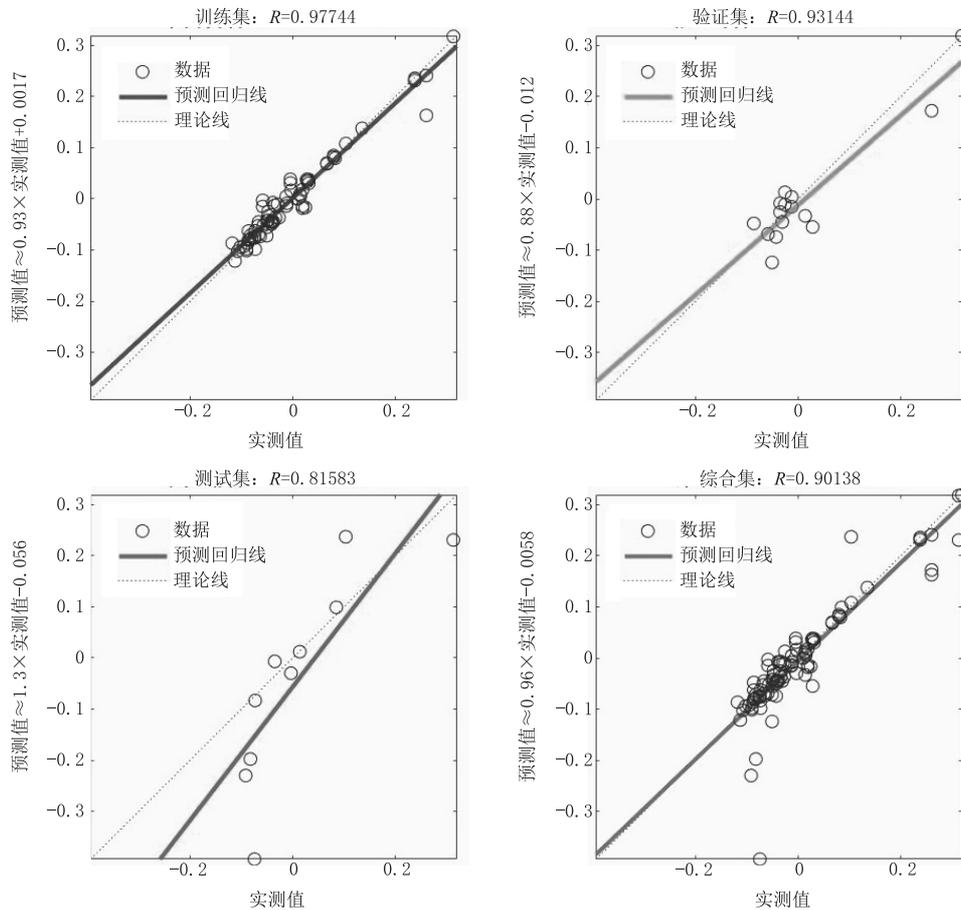


图 5 神经网络建模效果分析

的训练、验证和测试,利用神经网络建立的钻速预测模型误差能够降低到 10% 以内,能够满足常规使用的需要。这种基于地层信息和钻进参数预测模型的建立能够为相邻地区的地质勘探钻进施工提供必要的参考信息,有一定的指导施工价值。

参考文献:

- [1] 石祥超,万尚贤,孟英峰,等.气体钻井钻速方程的建立与验证[J].天然气工业,2010,(11):56-58.
- [2] 张立刚,吕华恩,李士斌,等.钻井参数实时优选方法的研究与应用[J].石油钻探技术,2009,(4):35-38.
- [3] 林元华,宗玉宇,梁政,等.石油钻井机械钻速预测研究进展[J].石油钻探技术,2004,(1):10-13.
- [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,(7):17-20.

- [5] 武成刚,赵明,郭志文.基于主成分分析法的钻速预测研究[J].矿业研究与开发,2015,(10):84-86.
- [6] 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,(7):1740-1744.
- [7] 韩胜娟.SPSS 聚类分析中数据无量纲化方法比较[J].科技广场,2008,(3):229-231.
- [8] 谢小韦.浅析多元线性回归中多重共线性问题的三种解决方法[J].科技信息,2009,(28):117-118.
- [9] 王骏,王士同,邓赵红.聚类分析研究中的若干问题[J].控制与决策,2012,(3):321-328.
- [10] 梁斌,卓梅霞.基于 SPSS 统计软件的因子分析法及实证分析[J].河西学院学报,2011,(5):45-49.
- [11] 李谦,段隆臣,张大伟,等.基于模拟月壤挖取采样扭矩试验及建模[J].地球科学——中国地质大学学报,2013,38(6):1363-1370.