

基于粒子群优化相关向量机的岩层可钻性预测

韩丽丽

(中国地质科学院勘探技术研究所, 河北 廊坊 065000)

摘要: 本文提出了一种基于粒子群优化相关向量机(PSO-RVM)的岩层可钻性预测方法。该方法选取岩层埋深 H 、声波时差 AC 、电阻率 ρ_d 、岩层密度 ρ 和泥质含量 V_{sh} 等 5 个参数作为评价岩层可钻性的基本参数。以某油田 Du4 钻井为例, 采用 PSO-RVM 方法、多元回归方法和 RVM 方法对岩层可钻性进行评价。计算结果表明, PSO-RVM 模型的预测结果与实测数据非常接近, 其预测精度明显高于多元回归方法和 RVM 方法, 说明本文提出的方法具有一定的优越性和较高的精度, 可以较好地应用于钻井工程中岩层可钻性预测。

关键词: 粒子群算法; 支持向量机算法; 岩层可钻性

中图分类号: P634.1 文献标识码: A 文章编号: 1672-7428(2016)03-0023-04

Drillability of Rock Formations Assessment by Relevance Vector Machine Based on Particle Swarm Optimization/HAN Li-li (The Institute of Exploration Techniques, CAGS, Langfang Hebei 065000, China)

Abstract: This paper presents a relevance vector machine based on particle swarm optimization method (PSO-RVM) for assessing the drillability of rock formations. Five parameters, including the depth of rock formations (H), acoustic travel time (AC), electrical resistivity (ρ_d), density of rock formations (ρ) and the shaliness of rock formations (V_{sh}), are selected as the basic parameters in the PSO-RVM model. The Du4 well drilling in an oil field is chosen as an example and the PSO-RVM method, multiple regression method and relevance vector machine (RVM) method are used to assess the rock formation drillability of the well. The results suggest that the predict results of PSO-RVM method accord well with the measured data and the prediction accuracy is significantly higher than that of multiple regression method and the RVM method. It is shown that PSO-RVM method can be applied in the prediction of rock formation drillability with its advantages and high accuracy.

Key words: particle swarm optimization; support vector machine algorithm; rock drillability

0 引言

岩层可钻性是衡量岩石抵抗钻具破坏能力的一项指标^[1]。该指标被广泛应用于地质勘探和油气钻井方面, 其对于确定钻井工艺、合理选择钻头、预测钻速等方面具有重要指导意义。目前, 岩层可钻性的获取方法主要有室内和现场岩石钻进试验法和测井资料预测法^[2]。由于开展室内试验和现场试验需要投入大量的人力物力资源, 而对于深部钻探, 经济损耗更为严重, 因此, 室内和现场岩石钻进试验法的应用受到限制。基于声波测井数据, 国内很多学者采用非线性方法对岩层可钻性进行预测, 取得了很好的效果, 这些方法主要有多元回归、BP 神经网络、模糊综合评判和支持向量机(SVM)等方法^[3-6]。研究表明, 岩层可钻性与岩层埋深和测井参数(如声波时差、电阻率、岩层密度和泥质含量)等存在复杂的非线性关系^[7-8], 这对于传统预测模

型构成了严峻的挑战。

相关向量机(RVM)是一种基于贝叶斯(Bayesian)框架的回归预测模型, 与传统非线性方法相比, 该方法具有训练样本少、预测精度高的优点^[9-11]。本文采用粒子群(PSO)算法对相关向量机算法的参数进行优化, 提出一种基于粒子群优化相关向量机(PSO-RVM)的岩层可钻性预测方法。

1 相关向量机算法和粒子群算法

1.1 相关向量机算法

相关向量机(RVM)是 Tipping^[10]提出的一种基于稀疏 Bayesian 概率模型的机器学习语言, 利用该方法可以进行回归预测。目前, 该方法是机器学习领域的研究热点之一。

假定一训练样本输入集 $\{x_i\}_{i=1}^N$ 和相应的输出集 $\{t_i\}_{i=1}^N$, 则 RVM 的回归模型为^[10-11]:

$$\begin{aligned} t_i &= y(x; w) + \varepsilon_i \\ &= w^T \varphi(x) + \varepsilon_i \\ &= \sum_{i=1}^N w_i K(x, x_i) + w_0 + \varepsilon_i \end{aligned} \quad (1)$$

式中: $w = (w_0, w_1, w_2, \dots, w_N)^T$, $\varphi(x) = [1, K(x, x_1), K(x, x_2), \dots, K(x, x_N)]^T$; ε_i ——满足 Gaussian 分布 $N(0, \sigma^2)$ 的附加噪声; w_i ——模型的权值; $K(x, x_i)$ ——核函数。

根据以往研究,本文选用 RBF 高斯核函数^[10]:

$$K(x, x_i) = \exp \frac{-||x - x_i||^2}{\sigma^2} \quad (2)$$

式中: σ ——又称为 RBF 高斯核函数的宽度。

假设样本之间相互独立,则可以通过公式(3)来计算样本的似然估计概率^[10]:

$$p(t|w, \sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left(-\frac{1}{2\sigma^2} ||t - \Phi w||^2\right) \quad (3)$$

式中: $t = (t_1, t_2, \dots, t_N)^T$, $\Phi = [\varphi(x_1), \varphi(x_2), \dots, \varphi(x_N)]^T$ 。

在 Bayesian 框架下,采用极大似然估计法可以计算出模型的权值 w 。但为了避免模型出现过分学习情况, RVM 为 w 定义了 Gaussian 先验概率分布来进行约束^[12-14]:

$$p(w, \alpha) = \prod_{i=0}^N N(w_i | 0, \alpha_i^{-1})^{-N/2} \quad (4)$$

式中: α —— $N+1$ 超维参数,且 $\alpha = (\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_N)^T$ 。

根据 Bayesian 准则,后验概率可通过式(5)计算^[10]:

$$p(w, \alpha, \sigma^2 | t) = \frac{p(t | w, \alpha, \sigma^2) p(w, \alpha, \sigma^2)}{p(t)} \quad (5)$$

给定一个新的样本 x^* , 则预测相应目标 t^* 为^[11]:

$$\begin{aligned} p(t^* | t) &= \int (t^* | w, \alpha, \sigma^2) p(w, \alpha, \sigma^2 | t) dw d\alpha d\sigma^2 \\ &= \int (t^* | w, \sigma^2) p(w, \alpha, \sigma^2 | t) dw d\alpha d\sigma^2 \end{aligned} \quad (6)$$

根据 Bayesian 原理,可利用先验概率分布和似然分布来计算所有未知参数的后验概率分布^[15]:

$$p(w | t, \alpha, \sigma^2) = (2\pi)^{-\frac{N+1}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(w - \mu)^T \Sigma^{-1}(w - \mu)\right] \quad (7)$$

式中: Σ 、 μ ——分别为后验协方差和均值,且满足

$$\Sigma = (\sigma^{-2} \Phi^T \Phi + A)^{-1} \quad (8)$$

$$\mu = \sigma^{-2} \Sigma \Phi^T t \quad (9)$$

式中: $A = \text{diag}(\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_N)$ 。

最后,通过最大似然方法便可计算得到超参数 α 和方差 σ^2 。给出一个新的测试样本 x^* , 其预测值为:

$$t^* = \mu^T \varphi(x^*) \quad (10)$$

RVM 的回归是一个迭代过程。在计算过程中,经过多次迭代,一大部分 α_i 会趋近无限大,其对应的权值则趋于 0; 仅有少数 α_i 会趋近于某一定值,其相应的 x_i 便称之为相关向量。

1.2 粒子群算法(PSO)

粒子群算法(PSO)是由 Kennedy 和 Eberhart^[16] 在 1995 年提出的一种基于群体智能的全局优化算法。该算法被广泛地应用于参数优化、模式识别等方面,具有参数少和收敛速度快等优点。所谓粒子是指待研究问题的一个解,粒子通过其位置 $X_i = (x_{i1}, x_{i2}, \dots, x_{id})$ 和速度 $V_i = (v_{i1}, v_{i2}, \dots, v_{id})$ 来定义。通过随机改变粒子的速度,可以调整粒子的位置,即改变粒子当前的解。在调整粒子位置的过程中,粒子间通过相互学习,可以得到每个粒子的个体最优解 $P_i = (p_{i1}, p_{i2}, \dots, p_{id})$ 和群体最优解 $P_g = (p_{g1}, p_{g2}, \dots, p_{gd})$ 。粒子速度和位置定义为^[16]:

$$v_{id} = \beta v_{id} + \delta_1 r_1 (p_{id} - x_{id}) + \delta_2 r_2 (p_{gd} - x_{id}) \quad (11)$$

$$x_{id} = x_{id} + v_{id} \quad (12)$$

式中: d —— $1 \sim D$ 之间取值, D 为向量维数; β ——惯性权重; δ_1 、 δ_2 ——学习因子,本文取 $\delta_1 = \delta_2 = 2$; r_1 、 r_2 ——随机变量,在 $[0, 1]$ 之间取值。

本文采用的 β 随着搜索次数的增加在 $[0.4, 0.9]$ 之间线性递减^[12]:

$$\beta = \beta_{\max} - \text{iter} \times \frac{\beta_{\max} - \beta_{\min}}{\text{iter}_{\max}} \quad (13)$$

式中: $\beta_{\max} = 0.9$ 、 $\beta_{\min} = 0.4$; iter ——目前迭代次数, iter_{\max} 表示总的迭代次数。

2 基于 PSO 优化的相关向量机算法

在 RVM 算法中, RBF 高斯核的宽度 σ 参数对模型回归预测的效果起到关键作用。本文提出的 PSO-RVM 核心思想是利用粒子群算法对 RVM 算法中的 σ 参数进行优化,从而使整个模型达到最

优。具体步骤如下。

(1)粒子的初始化,文中粒子指的便是要求解的参数 σ 。将 σ 作为粒子的位置向量,并随机定义初始位置和速度。

(2)计算粒子的适应度。

(3)比较粒子的当前位置适应度值与其经历过的最好位置的适应度值,若前者较好,则将其作为当前的最好位置。

(4)比较每个粒子的个体最佳适应度值与粒子群体的最优适应度值,若前者较好,则更新粒子群体的最优适应度值。

(5)根据方程式(11)和(12)更新粒子的速度和位置。

(6)如果满足结束条件,则程序运行结束,否则,返回步骤(2)继续运行。

3 地层可钻性级值预测

研究表明,岩层可钻性与岩层物化特性密切相关,而岩层的物化性质可以通过声波时差 AC 、电阻率 ρ_d 、岩层密度 ρ 和泥质含量 V_{sh} 来反映^[3,7]。其中,声波时差反映了岩层拉伸、剪切等变形性质;电阻率反映了岩层的含水性、密实程度;岩层密度和泥质含量是影响岩石强度的直接参数;岩层的应力状态与岩层埋深密切相关^[7]。因此,结合以往研究成果,本文选择岩层埋深 H 、声波时差 AC 、电阻率 ρ_d 、岩层密度 ρ 和泥质含量 V_{sh} 等 5 个参数作为评价岩层可钻性的基本参数。利用粒子群优化相关向量机方法,构建 PSO - RVM 岩层可钻性预测模型,模型结构如图 1 所示。

型训练样本(如表 1 所示),选取剩余的 11 个数据作为预测样本(见表 2)。

表 1 训练样本数据

样本编号	岩层埋深/ m	声波时差 $AC/(\mu s \cdot m^{-1})$	地层密度 $\rho/(g \cdot cm^{-3})$	电阻率 $\rho_d/(\Omega \cdot m)$	泥质含量 $V_{sh}/\%$	实测可钻性极 值 K_d
X ₁	4194.0	150.3	2.74	2496.0	1.7	4.452
X ₂	4198.3	174.5	2.63	1246.3	0.0	7.769
X ₃	4200.0	144.0	2.81	1955.4	1.3	7.966
X ₄	4212.1	147.6	2.80	3981.2	1.3	7.003
X ₅	4217.9	159.8	2.74	3031.4	0.2	5.676
X ₆	4231.3	177.8	2.67	3975.4	0.0	3.070
X ₇	4237.3	190.0	2.59	1513.8	0.0	4.693
X ₈	4249.3	165.0	2.72	927.7	23.7	5.128
X ₉	4258.6	170.6	2.68	2063.6	0.5	6.421
X ₁₀	4269.8	159.8	2.69	40982.5	0.3	6.057
X ₁₁	4275.5	169.6	2.65	1955.3	0.4	5.942

表 2 预测样本数据

样本编号	岩层埋深/ m	声波时差 $AC/(\mu s \cdot m^{-1})$	地层密度 $\rho/(g \cdot cm^{-3})$	电阻率 $\rho_d/(\Omega \cdot m)$	泥质含量 $V_{sh}/\%$	实测可钻性极 值 K_d
Y ₁	4195.9	165.7	2.65	778.4	1.2	6.848
Y ₂	4198.9	166.7	2.76	1095.5	0.7	7.309
Y ₃	4205.5	155.8	2.77	2395.4	0.2	6.548
Y ₄	4213.1	155.5	2.78	3724.8	0.9	8.355
Y ₅	4221.0	144.7	2.61	14561.6	2.4	6.358
Y ₆	4235.0	172.2	2.61	1482.7	0.4	4.669
Y ₇	4239.3	191.9	2.57	1716.2	0.0	3.845
Y ₈	4257.1	177.8	2.67	1792.0	0.2	5.805
Y ₉	4269.3	156.2	2.67	40984.6	0.8	6.007
Y ₁₀	4272.8	156.8	2.72	4449.7	1.2	5.716
Y ₁₁	4278.4	157.2	2.78	4597.9	4.5	7.079

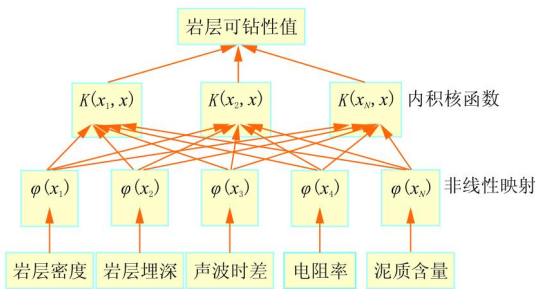


图 1 岩层可钻性预测 PSO - RVM 模型

4 实例应用

本文选取了文献[3]提供的某油田 Du4 钻井实测数据作为算例。从该井共收集到 22 个样本数据,本文随机选取其中的 11 个数据作为 PSO - RVM 模

本文采用 Matlab 软件编制 PSO - RVM 计算程序,设置种群大小(粒子个数)为 50,粒子维数为 2,最大迭代次数为 80。为了说明 PSO - RVM 模型的优越性,选用多元回归方法和 RVM 方法作为对比研究。计算结果如表 3 和图 2 所示。 K_d 为实测可钻性值,而 K_{d1} 、 K_{d2} 和 K_{d3} 分别代表了多元回归模型、RVM 模型和 PSO - RVM 模型的计算结果。从图 2 可知,PSO - RVM 模型预测结果与实测数据非常接近,两条曲线基本重合,而多元回归模型和 RVM 模型的预测结果与实测数据存在一定差距。根据表 3 可知,PSO - RVM 模型预测结果的平均绝对误差仅为 0.0295,平均相对误差为 0.51%;而多元回归模型和 RVM 模型预测结果的平均绝对误差分别为 0.6427、0.2875,平均相对误差分别为 10.43%、4.60%。所以,相比于多元回归模型和

RVM模型,本文提出的 PSO-RVM模型具有精度高的优点,可推广应用于钻井工程中岩层可钻性预测。

表3 预测结果对比

样本 编号	可钻性级值				绝对误差			相对误差/%		
	K_d	K_{d1}	K_{d2}	K_{d3}	K_{d1}	K_{d2}	K_{d3}	K_{d1}	K_{d2}	K_{d3}
Y_1	6.848	6.196	6.289	6.893	0.6520	0.5590	0.0450	9.52	8.16	0.66
Y_2	7.309	6.565	7.452	7.336	0.7440	0.1430	0.0270	10.18	1.96	0.37
Y_3	6.548	6.929	6.979	6.497	0.3810	0.4310	0.0510	5.82	6.58	0.78
Y_4	8.355	6.815	7.985	8.376	1.5400	0.3700	0.0210	18.43	4.43	0.25
Y_5	6.358	6.084	6.374	6.381	0.2740	0.0160	0.0230	4.31	0.25	0.36
Y_6	4.669	5.551	4.533	4.672	0.8820	0.1360	0.0030	18.89	2.91	0.06
Y_7	3.845	4.394	3.650	3.911	0.5490	0.1950	0.0660	14.28	5.07	1.72
Y_8	5.805	5.502	5.497	5.812	0.3030	0.3080	0.0070	5.22	5.31	0.12
Y_9	6.007	5.574	5.684	5.971	0.4330	0.3230	0.0360	7.21	5.38	0.60
Y_{10}	5.716	6.413	6.003	5.695	0.6970	0.2870	0.0210	12.19	5.02	0.37
Y_{11}	7.079	6.464	6.685	7.104	0.6150	0.3940	0.0250	8.69	5.57	0.35

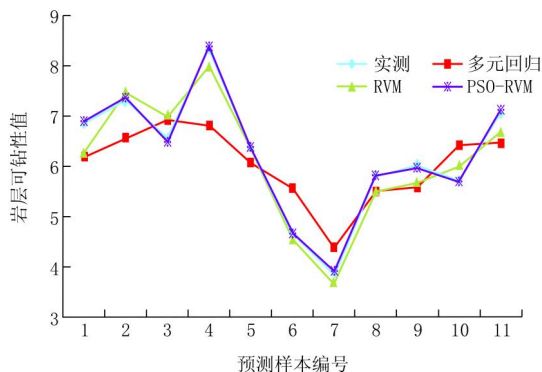


图2 PSO-RVM模型、多元回归模型和RVM模型预测结果对比

5 结论与建议

(1) 本文提出一种基于粒子群优化相关向量机(PSO-RVM)的岩层可钻性预测方法,该方法选取岩层埋深 H 、声波时差 AC 、电阻率 ρ_d 、岩层密度 ρ 和泥质含量 V_{sh} 等5个参数作为评价岩层可钻性的基本参数。

(2) 将 PSO-RVM 方法应用于某油田 Du4 钻井岩层可钻性评价,计算结果表明,该方法的预测结果与实测数据非常接近,具有很高的预测精度,可以较好地应用于钻井工程中岩层可钻性预测。

(3) 为了说明模型的优越性,选用多元回归方

法、RVM方法与 PSO-RVM方法进行对比研究,结果表明,PSO-RVM方法的预测精度明显高于多元回归方法和 RVM方法。

参考文献:

- [1] 沙林秀,邵小华,张奇志,等.基于 Bayesian 多分支岩石可钻性值估计[J].中国石油大学学报(自然科学版),2014,38(3):73-79.
- [2] 郑德帅,高德利,冯江鹏.实钻条件下井底岩石可钻性预测模型研究[J].岩土力学,2012,33(3):859-863.
- [3] 刘之的,夏宏泉,陈平.基于灰色 GM(0,N)法的测井预测岩石的可钻性研究[J].天然气工业研究,2004,24(11):76-78.
- [4] 薛亚东,高德利.基于人工神经网络的实钻地层可钻性预测[J].石油钻采工艺,2001,23(1):26-29.
- [5] 李俊萍,段隆臣,李谦.模糊综合评判法在岩石可钻性分级中的应用[J].地质科技情报,2012,31(1):127-130.
- [6] 马海,王延江,魏茂安,等.地层可钻性级值预测新方法[J].石油学报,2008,29(5):761-765.
- [7] 邢军,姜谔男,邱景平,等.基于 DE-SVM 的岩层可钻性预测研究[J].东北大学学报(自然科学版),2010,31(9):1345-1348.
- [8] 马海,王延江,胡睿,等.基于相关向量机的地层可钻性级值预测[J].中国石油大学学报(自然科学版),2010,34(2):67-70.
- [9] 吴军,臧艳彬,陈星星.基于模式识别理论的卡钻类型判别分析[J].探矿工程(岩土钻掘工程),2015,42(10):31-34.
- [10] Samui J, Lansivaara T, Kim D. Utilization relevance vector machine for slope reliability analysis[J]. Applied Soft Computing, 2011,11(5):4036-4040.
- [11] Samui P. Slope stability analysis: a support vector machine approach[J]. Environmental Geology, 2008,56:255-267.
- [12] Tipping M E. Sparse Bayesian learning and relevance vector machine[J]. Journal of Machine Learning Research, 2001, (1): 211-244.
- [13] Wang X D, Ye M Y, Duanmu C J. Classification of data from electronic nose using relevance vector machines[J]. Sensors and Actuators B: Chemical, 2009,140(1):143-148.
- [14] Xu C, Dai F C, Xu X W, Lee Y H. GIS-based support vector machine modeling of earthquake-triggered landslide susceptibility in the Jianjiang River watershed, China[J]. Geomorphology, 2012,145-146:70-80.
- [15] 李慧杰,刘亚南,卫志农,等.基于相关向量机的短期风速预测模型[J].电力自动化设备,2013,33(10):28-32.
- [16] Kennedy J, Eberhart RC. Particle swarm optimization[C]. In: Proceedings of the IEEE international conference on neural networks, Perth, Australia, 1995, pp 1942-1948.