

Application of Particle Swarm Optimization Support Vector Machine in Reservoir Prediction

Chenhao Xiong¹, Fu Zhao¹, Binchuan Qin¹, Zhengxiang Lv^{1,2}

¹College of Energy, Chengdu University of Technology, Chengdu Sichuan

²State Key Laboratory of Oil-Gas Reservoir Geology & Exploitation, Chengdu University of Technology, Chengdu Sichuan

Email: 294183271@qq.com, 1944949956@qq.com, 971286917@qq.com, 417330439@qq.com

Received: Dec. 25th, 2019; accepted: Jan. 8th, 2020; published: Jan. 15th, 2020

Abstract

In view of the problem that it is difficult to predict reservoirs in tight sandstone reservoirs of Enping formation in area of Zhu I depression, SVC is used to predict porosity of reservoirs by logging interpretation. The correctness of the model is improved by combining SVC with particle swarm optimization. The reservoir types of Enping formation are classified by typical well test and production test analysis. 80% of the samples of each type of reservoir are selected as modeling data, and random cross validation is carried out in each type of reservoir to get the cross-validation score of the model. Then the particle optimization method is used to improve the cross-validation score of the model to get the best prediction model. Then the prediction model is tested with the samples that are not involved in the modeling. From this model, the well log interpretation map can be made, that is, the location of the well section where the effective reservoir is located can be directly seen and the thickness of effective reservoir can also be calculated conveniently.

Keywords

Support Vector Classification (SVC), Particle Swarm Optimization (PSO), Reservoir Classification, Log Interpretation

粒子群寻优支持向量机在储层类型预测中的应用

熊晨皓¹, 赵福¹, 覃斌传¹, 吕正祥^{1,2}

¹成都理工大学能源学院, 四川 成都

²成都理工大学“油气藏地质及开发工程”国家重点实验室, 四川 成都

Email: 294183271@qq.com, 1944949956@qq.com, 971286917@qq.com, 417330439@qq.com

收稿日期: 2019年12月25日; 录用日期: 2020年1月8日; 发布日期: 2020年1月15日

摘要

针对珠一坳陷恩平组致密砂岩储层难以预测的问题, 本文在使用基于支持向量机分类模型的储层类型预测方法的基础上, 结合粒子群寻优法对模型进行校正与完善, 提高了模型的正确率。通过典型井测试与试采分析法对恩平组储层的样本进行了储层类型分类, 在每种类型的储层样本中选取80%的样本作为建模数据, 并在每种储层类型内进行乱序交叉验证, 得到模型的交叉验证分数, 然后用粒子寻优的方法提高模型交叉验证分数, 得到最佳预测模型, 再用未参与建模的样本对预测模型进行检验。由此模型作出测井解释图, 即可以直观的看出有效储层的所在井段的位置, 也能方便地计算出有效储层的厚度。

关键词

储层预测, 支持向量机, 粒子群寻优, 测井解释

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

储层预测是储层评价研究中的重要问题, 但由于岩心的取芯和测试费用较高, 岩心测量数据一般较少且不连续。测井解释研究工作之一就是结合已有岩心数据与测井数据来求取全井段的地质属性。一般来说, 生产单位目前用的测井模型为一元线性回归模型[1], 例如对储层孔隙度的预测, 即在四性分析的基础上选择与孔隙度相关性最好的一种测井参数来建立模型, 这种模型能够简单快捷地对储层孔隙度进行预测, 但预测精度不高且无法应用于非均质性较强的地层。本文所用的储层分类预测方法, 不用对储层的孔隙度和渗透率具体值进行预测, 而是在对样本进行储层分类的基础上, 利用预测模型, 直接对未知储层进行分类预测, 求得有效储层, 从而跳过对储层孔隙度以及渗透率具体值预测的基础上再根据预测值进行分类的繁琐步骤。

为了应对如致密砂岩储层这种非均质性较强的地层, 目前测井解释模型已经运用了多种非线性方法, 如偏最小二乘法[2]、逐步回归法[3]、核岭回归法[4]、人工神经网络法[5]等。相比于人工神经网络对样本的需求量较大[6], 支持向量机(SVM)模型以其对样品数量需求小, 可解决高纬度非线性问题以及可避免神经网络模型易陷入局部极小值的特点逐渐进入人们的视野[7]。支持向量机是 Corinn Cortes 与 Vapnik [8] 在 90 年代首次提出的一种基于统计学习理论和结构风险最小化原理的机器学习方法, 它的主要特点是能够深度挖掘数据之间的关系并在模型的复杂性和模型的学习能力之间寻求最佳折中。近几年来, 许多研究者在支持向量机对于储层的孔隙度预测方面取得了较大的研究成果。乐友喜等[9]利用支持向量机方法对胜利油田某区块进行砂体孔隙度预测并与人工神经网络方法进行对比, 证明了人工神经网络与支持向量机相比在样本较少时会出现欠学习现象而使预测结果不理想。滕新保[10]等将岩性信息加入支持向量机

模型, 并证明了径向基函数(RBF)对支持向量回归模型最好。

一般而言, 人们在对储层孔隙度运用支持向量机方法进行预测时, 通常使用网格搜索法来寻找适合模型的最优化参数 c 和 δ^2 , 这种穷举搜索方法往往需要人们在准确度和消耗时间之间作出妥协, 做不到既准确又迅速。张向君等[11]通过交互检验方法优选支持向量机的惩罚因子及核函数参数, 提高支持向量机储层预测精度。李建军等[12]通过拐点法的数值实验选择出高斯函数。然后针对石油地质勘探的实际问题, 将支持向量机运用测井曲线预测储层参数。本文则提出利用粒子群寻优的方法, 不必地毯式搜索也可以快速、准确地找到适合模型的最优参数。

位于珠江口盆地的珠一坳陷, 其古近系从深到浅依次为文昌组、恩平组、珠海组, 其中的文昌组和恩平组储层非常发育, 以致密砂岩储层为主。在文昌组沉积时期, 盆地主要为断陷湖泊沉积环境, 从而形成了中深湖 - 浅湖 - 辫状三角洲沉积体系; 而在恩平组时期, 盆地处于断拗转换阶段, 主要形成滨浅湖 - 河流、三角洲 - 沼泽沉积体系[13]。储层岩石结构主要以中砂岩和粗砂岩为主, 孔隙度分布在 7%~20%, 渗透率分布在 0.02~1000 mD, 层间的非均质性较强。由于每口井之间测井年代不同, 所用测井仪器不同, 测井曲线也有一定的缺失, 所以本文在测井曲线标准化的基础上, 采用针对地区典型井分层位的方法建立测井预测模型。此种方法建立的预测模型具有较强的针对性, 而且改进后的支持向量回归算法也能解决模型细分后样本数量减少的问题。

2. 原理分析

2.1. 支持向量分类原理

支持向量机分类原理与支持向量回归原理相似, 即是将低维空间中的数据映射到高维空间中, 并在高维空间中寻找一个回归超平面, 使所有数据到该超平面的距离最近。对于已知数据集

$$\{x_1, y_1, x_2, y_2, \dots, x_i, y_i\}, (i=1, 2, \dots, n) \quad (1)$$

在任意给定数 $\varepsilon > 0$ 情况下, 存在超平面

$$f(x) = (w \cdot x) + b \quad (2)$$

使得所有数据到该超平面的距离最近。其中, ε 为惩罚程度, w 为权重向量, $(w \cdot x)$ 为权重向量 $w \in R^n$ 与向量 $x \in R^n$ 的内积, b 为偏置系数, 基于结构风险最小化原则, 引入松弛变量 ξ_i, ξ_i^* , 分别代表上下边界的松弛因子, 即可得到支持向量机公式[14], 它要解决的原始最优化问题为

$$\min \frac{1}{2}(w \cdot w) + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (3)$$

$$\text{s.t.} \begin{cases} y_i - (w \cdot x_i) - b \leq \varepsilon + \xi_i \\ (w \cdot x_i) + b - y_i \leq \varepsilon + \xi_i^* \\ \varepsilon \geq 0, i=1, 2, \dots, n \end{cases} \quad (4)$$

式(3)中常数 C 为惩罚系数。通过引入拉格朗日乘子 $a_i (a_i^* \geq 0)$, 并根据带有不等式约束的极值问题的 KKT 条件, 将原始最优化问题转化为其对偶问题[15]

$$\begin{cases} \max \varphi(a) = \sum_{i=1}^n (a_i - a_i^*) y_i - \frac{1}{2} \sum_{i,j=1}^n (a_i - a_i^*) (a_j - a_j^*) (x_i \cdot x_j) \\ \text{s.t.} \sum_{i=1}^n (a_i - a_i^*) = 0 \\ a_i, a_i^* \geq 0 \end{cases} \quad (5)$$

通过引入核函数 $K(x_i, x_j)$ 来代替对偶问题函数式(5)中的内积 $(x_i \cdot x_j)$ ，可将低维空间中非线性不可分样本映射到高维空间中，使其线性可分。求解对偶问题即可得到

$$f(x) = \sum_{i=1}^n (a_i - a_i^*) K(x_i, x) + b \quad (6)$$

其中的实验涉及的致密砂岩是具有较强非均质性的储层，故核函数使用非线性效果较好的径向基函数[5]，公式为

$$K(x_i, x) = \exp\left\{-\frac{(x_i - x)^2}{2\sigma^2}\right\} \quad (7)$$

式(7)中的 σ 为核宽度系数，用于控制高维度空间中样本分布的复杂程度。将核函数代入，即可得到本次实验所用支持向量机模型[16]公式

$$f(x) = \sum_{i=1}^n (a_i - a_i^*) \exp\left[-\frac{(x_i - x)^2}{2\sigma^2}\right] + b \quad (8)$$

2.2. 粒子群寻优原理

支持向量回归模型中的参数 c 和 δ^2 对模型影响较大且难以寻得最优值，本文选用粒子群寻优原理，寻找模型参数 c 和 δ^2 的全局最优解，并以交叉验证的模型的评估分数来判断模型的优劣。

粒子群寻优算法(PSO)是一种群体智能优化算法，它通过在空间中模拟鸟群觅食行为寻找最优值。假设在一个 N 维的空间中，共有 n 个粒子 $X = (X_1, X_2, \dots, X_n)$ ，每个粒子代表问题的一个潜在解。通过不断地迭代，粒子根据个体之间的相互作用及信息共享寻找最优值。设第 i 个粒子的速度为 $v_i = (v_{i1}, v_{i2}, \dots, v_{iN})$ ，其个体极值为 $P_i = (P_{i1}, P_{i2}, \dots, P_{iD})$ ，种群的历史最优值为 $P_g = (P_{g1}, P_{g2}, \dots, P_{gD})$ 。每次迭代中，粒子速度和位置更新公式[17]如式(9)、式(10)所示：

$$v_{id}^{k+1} = \omega v_{id}^k + c_1 r_1 (P_{id}^k - X_{id}^k) + c_2 r_2 (P_{gd}^k - X_{id}^k) \quad (9)$$

$$X_{id}^{k+1} = X_{id}^k + v_{id}^{k+1} \quad (i=1, 2, \dots, m; d=1, 2, \dots, D) \quad (10)$$

式(11)中的 v_i^k 表示第 i 个粒子的第 k 次迭代速度， X_i^k 表示第 i 个粒子的第 k 次迭代状态量， k 表示迭代次数， r_1 和 r_2 为 $[0,1]$ 之间的随机数，用于保持群体的多样性，适当调整学习因子 c_1 和 c_2 ，扩大搜索空间避免陷入局部极值。

3. 改进支持向量回归预测模型

3.1. 数据标准化

由于多条测井曲线的刻度和量纲均有所不同，所以在进行储层类型预测建模时，必须要对所有参与建模的测井曲线进行数据标准化处理，来消除由于量纲和刻度不同造成的大量刚数据在模型中占比过大等影响。为了消除 E2 井的井径扩大段出现的异常最值等离群数据的影响，本次标准化采用的是 z-score 标准化，其公式如式(11)所示：

$$y_{ij} = \frac{x_{ij} - x_i}{s_i} \quad (11)$$

式(11)中的 x_{ij} 为实际测井曲线上对应的测井深度的测井值， y_{ij} 为 x_{ij} 的标准化后的值， x_i 为 E2 井恩平组所有采样点测井曲线的算术平均值， s_i 为对应的测井曲线标准差。本文通过 Python3.6 软件编程实现对 GR、DT、DEN、CNL、RT 等测井曲线的批量标准化。

3.2. 样本分类与选择

3.2.1. 储层下限分类

为了保证建模样本的多样性,避免出现模型偏向低孔隙度或高孔隙度,本文采用有利储层下限,对样本进行分类,再从各分类中进行选取的方法来保证样本的多样性。本文主要采用了典型井测试与试采分析法,选择具有取心资料的测试或者累产低产、中产、高产层进行物性分析,分别确定出这些典型层位的物性特征,在此基础上求取了各类产层的孔隙度和渗透率均值,这些典型层位的物性均值即为物性下限。根据上述方法中确定的储集物性下限,可将 E 地区恩平组储层按照储集物性的好坏依次分为 I、II、III、IV 四类储层,各类储层的孔隙度和渗透率如表 1 所示。

Table 1. Reservoir classification and evaluation of Enping formation in Zhu I depression

表 1. 珠一坳陷恩平组储层分类评价表

储层分级	孔隙度(%)	渗透率(mD)	储层评价
I	>17	>22	好
II	14.5~17	5~22	较好
III	12~14.5	2~5	较差
IV	<12	<2	差或非储层

其中 I 类为在自然开采条件下能够获得较好产能的储层; II 类为自然开采条件下基本能够获得工业产能的储层; III 类储层在自然开采条件下不能达到工业产能,但稍加改造后就能够产出油气; IV 类储层为含油性很差的储层或是非储层。

3.2.2. 样本选取

将珠一坳陷 27 口重点井恩平组的 204 个实测岩心样本依照表 1 分为四类。其中 III 类储层样本数最少,只有 7 个,随机选取其中的 80%作为模型样本,即 5 个 I 类储层样本。依次再在另三类储层每种类型随机挑选 80%的样本,剩下样本作为训练样本,模型样本与训练样本的分配如表 2 所示。

Table 2. Sample distribution of Enping formation in well E2

表 2. E2 井恩平组样本分配表

储层分级	I	II	III	IV
模型样本数	24	46	5	88
测试样本数	6	11	2	22

3.3. 模型建立

本文采用 27 口重点井恩平组的 204 个实测岩心样本中挑选出的 163 个实测岩心样本,剩下 41 个实测样本作为测试样本进行交叉验证。所有样本数据根据测井响应识别的原则,选取代表地层特征的 5 个参数,其中 GR 为伽马, RD 为深侧向电阻率, DT 为声波, DEN 为体积密度, CNL 为中子密度。

为了应对目标储层非均质性较强的特点,本文选用支持向量机的非线性核——径向基函数,作模型的核函数,其中支持向量机的惩罚系数 c 和径向基函数参数 δ^2 由粒子群寻优法来求最优解。粒子群寻优设置粒子为 30 个,以惩罚系数 c 作为横坐标,径向基函数参数 δ^2 作为纵坐标,初始位置为 c 和 δ^2 都在 0.1~10 之间的一个随机数,由于 c 和 δ^2 的有效范围在 $10^{-8} \sim 10^8$ 之间,所以为了减少粒子的盲目探索,初期粒子的每次移动乘以 10 作为一个步长,当确定大致范围后再缩小每次移动的步长。本次建模使用 Python3.6.3 编写模型处理程序,模型建立流程如图 1 所示。

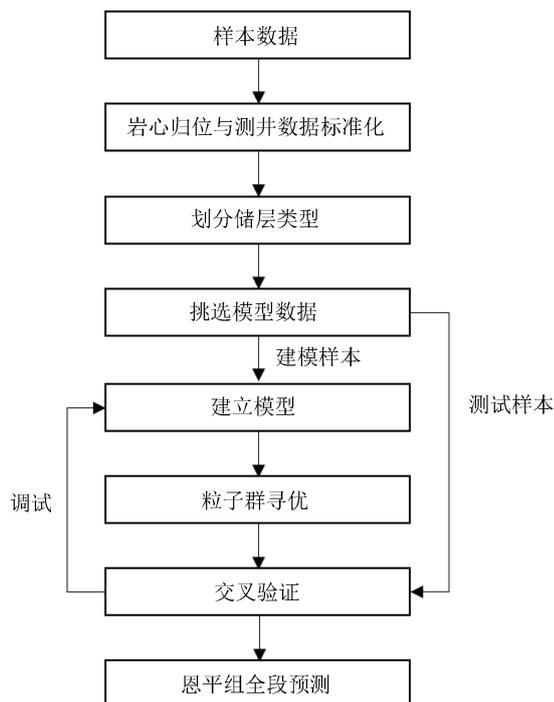


Figure 1. Establishment process of Enping prediction model
图 1. 恩平组预测模型建立流程

在岩心归位与测井曲线标准化的基础上，对归位后的岩心依据物性数据进行储层分类，然后对每种类型的储层选取其中 80% 的数据进行建模，运用粒子群寻优算法求得支持向量机的模型参数参数 c 和 δ^2 ，并用根据交叉验证分数实时调试粒子群，最终寻得模型最优参数 c 和 δ^2 并建立模型。将剩下的 20% 岩心数据用于检验模型的预测正确率，若模型的总体预测正确率大于 80% 即可说明模型达到精度要求，最后将达到精度要求的模型用于恩平组全段的储层预测。

3.4. 预测结果及分析

对模型参数 c 和 δ^2 的粒子群寻优过程如图 2 所示，图中的横坐标为粒子群寻优迭代次数，本文设定的最大迭代次数为 200 次，图中的纵坐标为目前为止寻得的最佳参数在交叉验证时的模型评估分数 (cross_value_score)，其值越接近 1，则模型准确度越高。整个寻优过程在迭代到 151 次时达到最优值 0.8695，用时 4209 秒，寻得的 $c = 5.107053011237296$ ， $\delta^2 = 3.2348451984267936$ 。

本文同时也使用了常用的网格搜索法进行寻优并与粒子群寻优进行对比。由于网格搜索法原理为使用穷举法来将所用的参数都运行一遍，如果采用和粒子群寻优相同精度，运算量就过于庞大，所以本文使用的网格搜索法在编程时设定的步进间隔大小为 1。但即便如此，运算时长依然超过了 4 小时以上。得到的模型评估分数最优值为 0.8633，寻得的 $c = 5.110$ ， $\delta^2 = 3.235$ 。与网格搜索法相比，虽然两种模型差距较小，粒子群寻优需要人为输入的参数更少，寻优过程全部由程序完成，且粒子群寻优算法的收敛速度较快。

经过对模型的优化，再将得到的模型用于对未参与建模、也未参与检验的 E2 井恩平组 85 个实测岩心样本进行预测，得到的预测储层分类结果与实测储层分类结果对比图如图 3 所示，从图中可以看到两幅图的孔隙度和渗透率均为岩心实测值，其中绿色样本点为 IV 类储层，蓝色样本点为 III 类储层，红色样本点为 II 类储层，黄色样本点为 I 类储层。85 个实测样本中各类储层预测的正确率如表 3 所示。

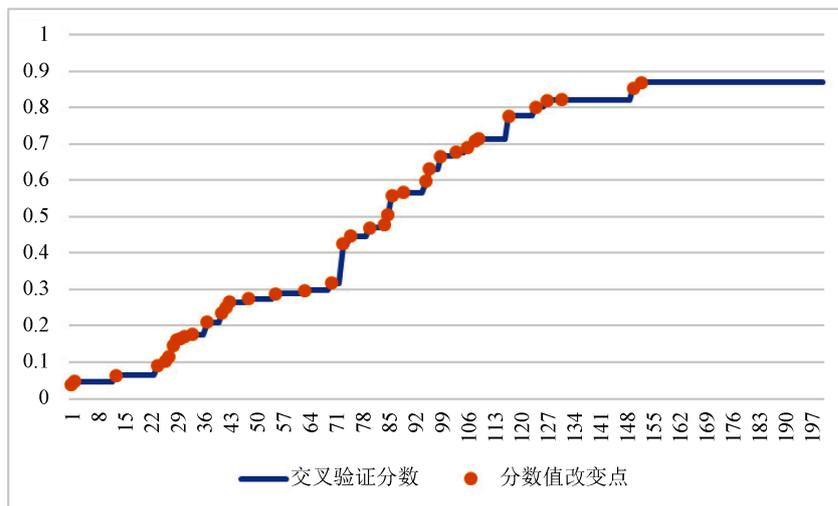


Figure 2. Process chart of particle swarm optimization

图 2. 粒子群寻优过程图

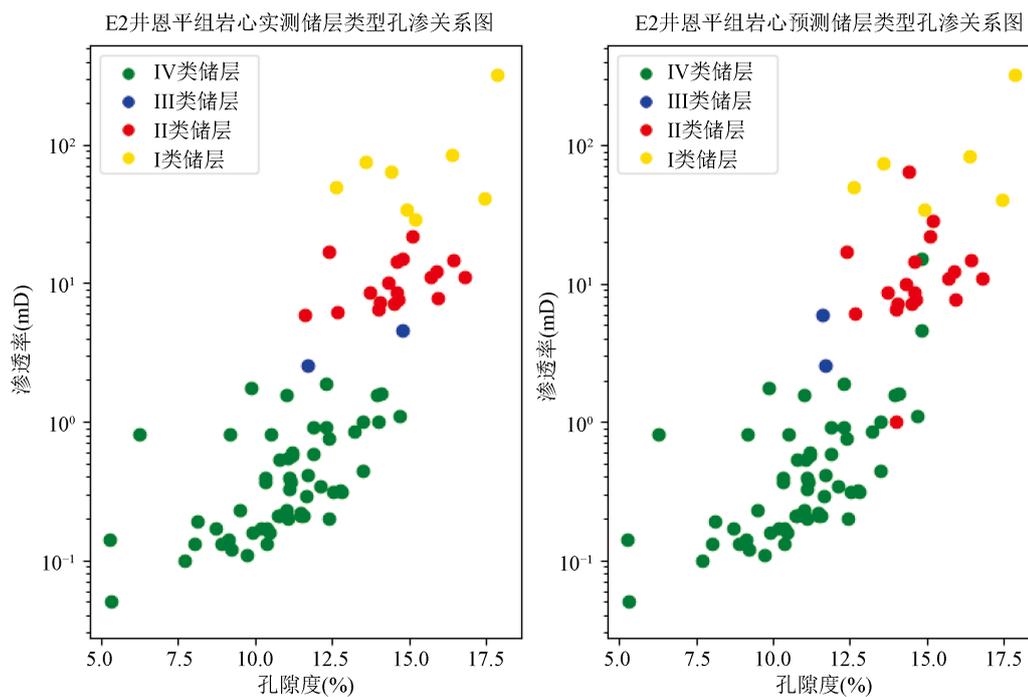


Figure 3. Comparison of the relationship between the measured core and the predicted reservoir type porosity and permeability in Enping formation of well E2

图 3. E2 井恩平组岩心实测与预测储层类型孔渗关系对比图

Table 3. Statistical table for prediction accuracy of reservoir type of measured samples in well E2

表 3. E2 井实测样本储层类型预测正确率统计表

储层分级	I	II	III	IV
实测样本数	8	18	2	57
预测正确样本数	6	16	1	56
正确率	75%	88.9%	50%	98.3%

由表 3 可以看出, 尽管模型的精度较高, 但模型对于 III 类储层高孔隙度依然不太敏感, 预测精度较低, 模型出现了精确度有偏差的情况。对于此种情况, 可能是由于建模的时候 III 类储层的建模样本数较少的原因造成。

将建立好的预测模型对 E2 井恩平组全段砂岩进行储层预测, 并在 Resform 软件中成图, 可以得到储层类型的测井解释图如图 4 所示。

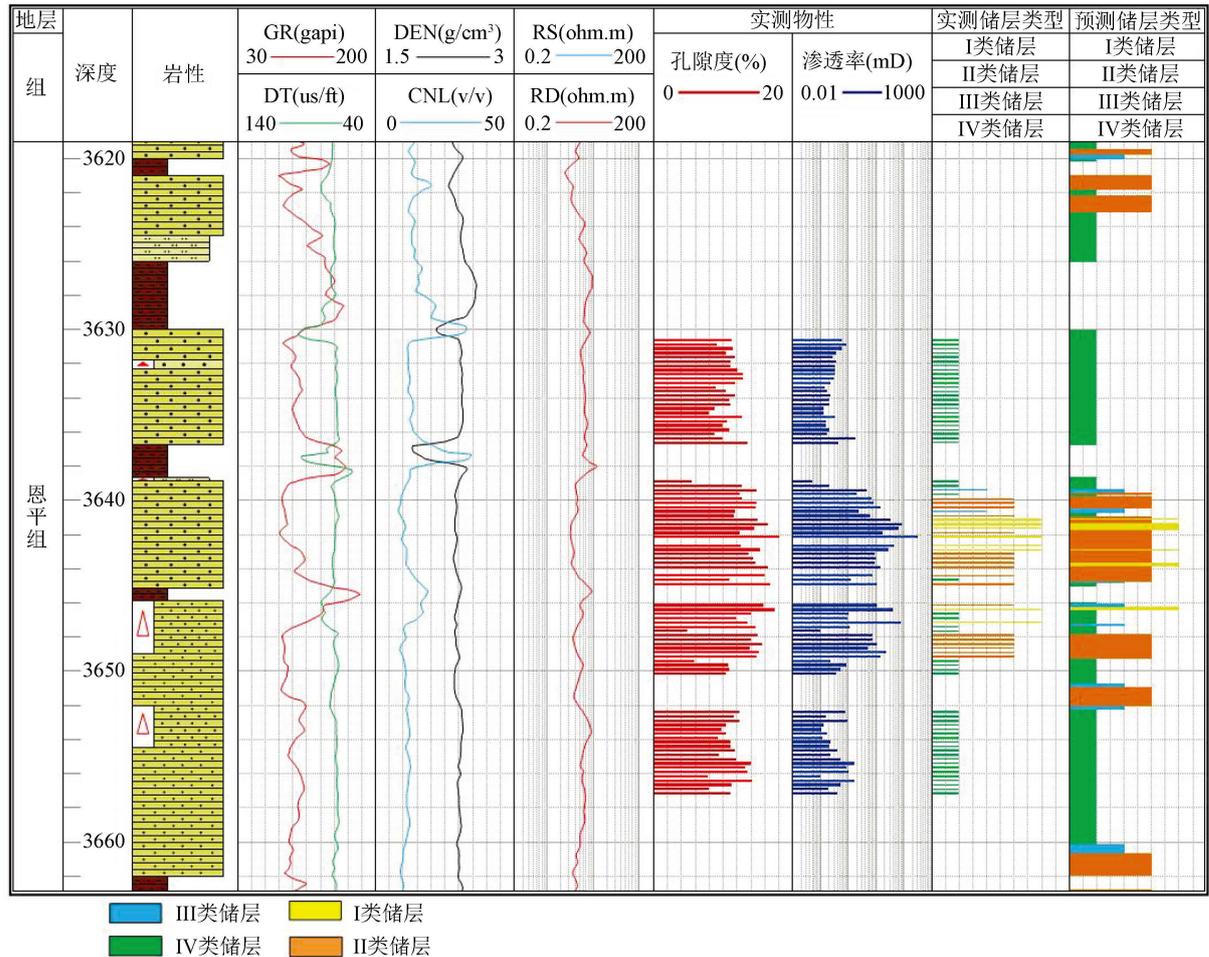


Figure 4. Reservoir type prediction map of Enping formation in well E2

图 4. E2 井恩平组储层类型预测图

图 4 中的 I 类储层由长度为 4 的黄色线段组成, II 类储层由长度为 3 的橙色线段组成, III 类储层由长度为 2 的蓝色线段组成, IV 类储层由长度为 1 的绿色线段组成。此种预测方法以及显示方式, 可以较为精确的看到优质储层所在井段以及精准计算出优质储层的厚度。

4. 结论

1) 粒子群寻优的收敛速度快, 与传统网格搜索法相比, 具有速度快、准确率高的特点[18]。将其运用到对支持向量机回归模型的参数 c 和 δ^2 的寻优中, 可以得到相对误差较低, 精确度较高的预测模型。

2) 根据储层下限对实测样本进行分类, 可以在建立预测模型时确保建模样本的多样性。在检验模型时, 对测试样本正确率进行分类统计, 可以看出模型对于每类储层的预测精确度, 对模型精确度不高的

储层可以适量增加该类储层在建模样本中所占的比例。

3) 将模型运用于检验井恩平组全段砂岩的储层类型预测, 可以较为准确的得到优质储层所在位置及优质储层的厚度。

参考文献

- [1] 徐壮, 石万忠, 翟刚毅, 等. 涪陵地区页岩总孔隙度测井预测[J]. 石油学报, 2017, 38(5): 533-543.
- [2] 胡作维, 李云. 基于偏最小二乘法评价低渗透砂岩储层质量[J]. 特种油气藏, 2013, 20(5): 36-39.
- [3] 范雯. 逐步回归分析方法在储层参数预测中的应用[J]. 西安科技大学学报, 2014, 34(3): 350-355.
- [4] 陈文浩, 王志章, 董少群, 等. 核岭回归方法解释致密砂岩储层孔隙度[J]. 测井技术, 2015, 39(6): 710-714.
- [5] 吕晓光, 杜庆龙. 应用神经网络模型进行油层孔隙度, 渗透率预测[J]. 大庆石油地质与开发, 1996(3): 27-31.
- [6] Iqbal, M. (2005) Numerical Solutions of Linear Ill-Posed Problems. *Integral Transforms & Special Functions*, **16**, 29-37. <https://doi.org/10.1080/1065246042000271992>
- [7] Phillips, D.L. (1962) A Technique for the Numerical Solution of Certain Integral Equations of the First Kind. *Journal of the ACM*, **9**, 84-97. <https://doi.org/10.1145/321105.321114>
- [8] Cortes, C. and Vapnik, V. (1995) Support-Vector Networks. *Machine Learning*, **20**, 273-297. <https://doi.org/10.1007/BF00994018>
- [9] 乐友喜, 袁全社. 支持向量机方法在储层预测中的应用[J]. 石油物探, 2005, 44(4): 388-392.
- [10] 滕新保, 张宏兵, 曹呈浩, 等. 一种新的砂泥岩孔隙度估计模型及其应用[J]. 河海大学学报: 自然科学版, 2015, 43(4): 346-350.
- [11] 张向君, 张晔. 基于支持向量机的交互检验储层预测[J]. 石油物探, 2018, 57(4): 597-600.
- [12] 李建军, 伦墨华. 基于支持向量机的石油勘探预测[J]. 科技通报, 2018(4): 79-83.
- [13] 施和生, 雷永昌, 吴梦霜, 等. 珠一坳陷深层砂岩储层孔隙演化研究[J]. 地学前缘, 2008, 15(1): 169-175.
- [14] 孟倩, 马小平, 周延. 改进的粒子群支持向量机预测瓦斯涌出量[J]. 矿业安全与环保, 2015(2): 1-5.
- [15] 张燕君, 王会敏, 付兴虎, 等. 基于粒子群支持向量机的钢板损伤位置识别[J]. 中国激光, 2017(10): 197-203.
- [16] 帅勇, 宋太亮, 王建平. 考虑全过程优化的支持向量机预测方法[J]. 系统工程与电子技术, 2017, 39(4): 931-940.
- [17] Xiong, W.L. and Xu, B.G. (2006) Study on Optimization of SVR Parameters Selection Based on PSO. *Journal of System Simulation*, **9**, 2442-2445.
- [18] 付超, 林年添, 张栋, 等. 多波地震深度学习的油气储层分布预测案例[J]. 地球物理学报, 2018, 61(1): 293-303.