

# 基于新型集成学习算法的基岩潜山油藏 储层裂缝开度预测算法

孙致学<sup>1,2</sup>,姜宝胜<sup>1</sup>,肖康<sup>3</sup>,李吉康<sup>1</sup>

(1. 中国石油大学(华东)石油工程学院,山东青岛266580; 2. 非常规油气开发教育部重点实验室 中国石油大学(华东), 山东青岛266580; 3. 中国石油勘探开发研究院非洲研究所,北京100083)

**摘要:**天然裂缝是基岩潜山油藏油气存储及运输的重要场所,而裂缝开度是表征潜山油藏储层品质、油气储量及产能评价的关键参数。为此,提出一种基于集成学习算法的新型裂缝开度预测算法。以B盆地中非乍得某基岩潜山油藏岩心描述、关键井成像测井、裂缝参数解释获取开度数据,以相同深度测井数据作为特征变量构成学习样本。利用K均值聚类算法对学习样本进行降噪,剔除异常数据,以支持向量机回归和XGBoost回归算法为基础模型,再利用随机搜索进行参数优化,通过岭回归算法对基础模型进行集成组合,再进行裂缝开度预测。结果表明所提出的新型集成学习算法比基础模型性能有明显提升。测试集样本预测值与实际值均方根误差为0.047,相关系数达0.931。该算法弥补了单一回归算法不稳定的特点,提高了泛化能力,为裂缝开度预测提供了新思路。

**关键词:**裂缝开度;K均值聚类算法;支持向量回归;XGBoost回归;集成学习

中图分类号:TE122.2

文献标识码:A

## Prediction of fracture aperture in bedrock buried hill oil reservoir based on novel ensemble learning algorithm

SUN Zhixue<sup>1,2</sup>, JIANG Baosheng<sup>1</sup>, XIAO Kang<sup>3</sup>, LI Jikang<sup>1</sup>

(1. School of Petroleum Engineering, China University of Petroleum (East China), Qingdao City, Shandong Province, 266580, China; 2. Key Laboratory of Unconventional Oil & Gas Development, Ministry of Education, China University of Petroleum (East China), Qingdao City, Shandong Province, 266580, China; 3. African Research Institute, PetroChina Research Institute of Petroleum Exploration & Development, Beijing City, 100083, China)

**Abstract:** The natural fractures are important for oil and gas storage and transportation in the bedrock buried hill reservoirs. The fracture aperture is the key parameter for the reservoir quality characterization as well as reserves and productivity evaluation of buried hill reservoirs. In this study, a new fracture aperture prediction algorithm is proposed based on the ensemble learning algorithm. The samples are collected from the bedrock buried hill reservoirs in Basin B of Chad, Central Africa, and their fracture aperture data are extracted from the sample description, key well imaging logging, and fracture parameter interpretation. The same depth logging data are used as the feature variables to constitute the learning sample, and the K-means clustering algorithm is applied to reduce noise of the learning sample and eliminate abnormal data. Based on Support Vector Machine (SVM) regression and XGBoost regression algorithm, and by using random search to optimize model parameters, the fracture apertures are estimated according to the basic models combined by the ridge regression. The results show that the performance of the novel ensemble learning algorithm is better than that of the basic model, the root mean square error between the predicted and actual values of the test set is 0.047, and the correlation coefficient is 0.931. The algorithm improves the instability of the single regression algorithm, improves the generalization ability, and provides a new way for aperture prediction.

**Key words:** fracture aperture; K-means clustering; SVM regression; XGBoost regression; ensemble learning

收稿日期:2019-12-13。

作者简介:孙致学(1979—),男,山东沂南人,博士,副教授,从事复杂油气藏精细描述与开发工作。E-mail:upcszx@upc.edu.cn。

基金项目:国家自然科学基金项目“基于离散-连续介质模型的水-EGS传质传热机理及数值模拟研究”(51774317)和“乍得潜山产能评价及开发技术政策研究”(2019D-3210)。

近年来,全球已在30余个盆地发现了基岩油气资源,基岩油气藏成为重要的勘探开发阵地。相对于常规沉积岩油气储层,基岩储层中天然裂缝类型、产状、特征参数的精准评价更加重要。由于该类油气藏基岩致密、孔渗性极低,天然裂缝系统不仅控制了有效储层发育程度和油气储量规模,同时也是油气开采过程中的重要运移通道。裂缝开度是评价基岩油气藏储层质量的重要参数。相对于裂缝密度,裂缝开度对储层有效渗透率的贡献更为显著,也是影响其产能的主控因素之一<sup>[1-3]</sup>。目前储层天然裂缝开度预测方法主要分为2大类:一类是直接观察法,包括岩心观测、露头识别、电镜观测等;另一类是间接观察法,包括成像测井、数值模拟、经验公式、动态资料分析等。其中露头识别法是获得裂缝开度最直接的途径<sup>[4]</sup>,但地表风化作用使裂缝充填特征发生显著变化,影响测量结果,同时露头区可能遭受后期的改造或掩埋,使得典型裂缝发育的露头不容易获得。岩心中包含着最为直观、详实的裂缝信息,但取心资料往往少且不连续,同时机械应力对岩心的破坏影响天然裂缝开度的测量。随着断层扫描机、阴极射线发光、核磁共振、三维激光扫描技术的发展,裂缝开度表征朝更微观、更立体和更精细的方向发展<sup>[5-6]</sup>。但由于仪器探测能力的限制,无法系统、大规模表征天然裂缝开度。成像测井具有高分辨率且连续测量的特点,能够直观地反映裂缝信息,但由于测量成本高,导致获得的数据非常有限。通过室内数值模拟装置进行裂缝开度模拟分析,测量相对误差较小,但装备适用范围有限,实验参数难以获得,无法真实还原地层条件。进行应力场有限元数值模拟需考虑地质体的岩石物理特征,所需参数较多,难以准确获取。该方法可以在一定程度上预测裂缝分布,但由于模型并不能反映实际地层情况,导致误差较大<sup>[7-8]</sup>。依据渗流力学原理,利用泥浆漏失数据建立裂缝泥浆漏失数学模型,根据钻井资料进行裂缝开

度计算也是当下研究的热点。但该方法受漏失数据的限制,适用范围有限。对大多数油田而言,现有资料中除了少量取心资料外,其余几乎是常规测井资料,因此如何利用常规测井信息建立裂缝的测井响应机理模型,进而计算天然裂缝开度,是不得不面对的实际问题。目前应用测井数据解释天然裂缝仍停留在定性分析水平上<sup>[9-10]</sup>。主要是由于传感器捕获的实时测井数据具有高维、非线性和高噪声性的特点,难以建立与裂缝开度之间的量化关系。机器学习对于解决非线性问题具有先天优势,而集成学习是机器学习领域的研究热点<sup>[11]</sup>,相较于单一机器学习算法,集成学习算法具有更高的精度和更显著的泛化性能。

## 1 样本集构建

研究区位于B盆地中非乍得西南部、中非剪切带中段北侧,大量岩心和井壁取心分析资料揭示B盆地基岩潜山岩性分为变质岩和岩浆岩2大类13个亚类30多种岩石类型,主要由花岗岩、正长岩、闪长岩和二长岩等岩浆岩及混合花岗岩和片麻岩类等正变质岩构成<sup>[12]</sup>。通过对该潜山15口取心井(含新完钻井)岩心裂缝形态、规模及典型特征进行观察及描述,发现研究区油气运移通道包括构造裂缝、网状缝、张剪缝及沿缝溶蚀孔洞,以张剪缝为主(图1)。

### 1.1 学习样本特征

准确且全面的裂缝数据样本是实现模型训练的基础。所用数据集包括测井数据(表1)及相应裂缝开度值,共2140组,由基岩潜山油藏岩心描述、关键井成像测井、岩矿薄片鉴定等数据组成。由裂缝开度分布(图2)可知,样本集中裂缝开度最小值为0.011 mm,最大值为0.544 mm,平均值为0.183 mm,标准差为0.087 mm,裂缝开度主要集中在0.126~0.258 mm。



图1 中非乍得某基岩潜山岩心裂缝照片

Fig.1 Photos of fractured core samples from bedrock buried hill in Chad, Central Africa

表1 测井数据统计结果  
Table1 Statistics of well logging parameters

统计结果	自然电位 (mV)	中子孔隙度 (%)	自然伽马 (API)	深侧向电阻率 ( $\Omega \cdot m$ )	浅侧向电阻率 ( $\Omega \cdot m$ )	体积密度 ( $g/cm^3$ )	声波时差 ( $\mu s/m$ )	密度校正 ( $g/cm^3$ )
平均值	-188.09	8.89	114.05	1 907.47	1 645.11	2.53	60.25	0.06
标准差	5.77	6.45	21.54	1 525.65	1 322.18	0.10	11.72	0.04
最小值	-203.15	1.39	28.57	262.71	114.58	1.89	27.35	0.004
最大值	-179.41	29.36	176.71	9 861.71	8 603.96	2.71	100.89	0.23
中位数	-187.69	6.49	119.63	1 545.52	1 423.02	2.55	56.43	0.04

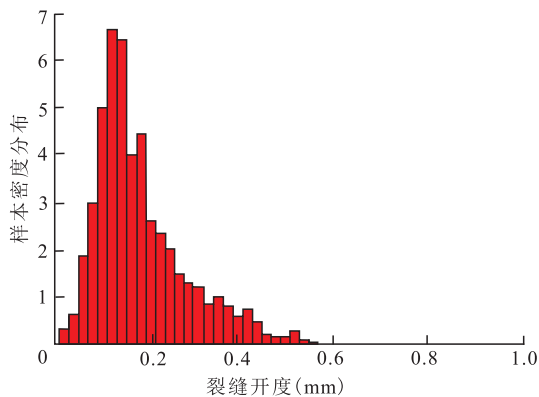


图2 裂缝开度分布

Fig.2 Fracture aperture distribution

## 1.2 Z-score 标准化处理

在进行裂缝预测前,需将学习样本进行Z-score 标准化处理,即将其转换为均值为0、方差为1的分布,其表达式为:

$$x' = \frac{x - \mu}{\delta} \quad (1)$$

如果一个特征的方差比其余特征的方差大许多数量级,那么该特征将会主导整个目标函数,使得模型不能从其余特征学习到数据的特征。相对于min-max 归一化方法,该方法不仅能够去除量纲,还能够均衡考虑所有维度的变量。

## 1.3 K均值聚类算法进行样本去噪

样本数据由于测量仪器及人为因素的干扰,不可避免的引入噪声。为此采取利用K均值聚类算法进行数据过滤的思路,以去除冗余,提高学习样本质量。

K均值聚类算法是基于距离的聚类算法,将距离作为相似性的评价指标,即对象之间的距离越近,相似度越大。而异常点通常距离中心点较远,检测异常点,从而进行样本过滤<sup>[13]</sup>。假设输入的样本向量集合为:

$$X = \{x_1, x_2, \dots, x_j, \dots, x_n\} \quad (2)$$

K均值聚类算法具体步骤包括:①从输入的样

本向量集合中随机选取1个向量作为第1个簇中心点,簇中心集合记为center。②对于满足条件的任意向量,计算与最近簇中心的距离。③计算每个向量被选为簇中心的概率,其表达式为:

$$P(x_j) = \frac{D(x_j)^2}{\sum_{x_j \in \text{center}} D(x_j)^2} \quad (3)$$

④ $P(x_j)$ 最大时对应的向量就是新的簇中心,若新的簇中心改变则重复步骤②—③直到目标函数收敛,聚类结束。

K的取值对聚类算法的效果具有极大影响。若K取值过小,将导致数据粗化,在剔除异常点的同时会误判正常数据,造成有效样本丢失;若K取值过大,将致使聚类结果无法有效收敛,计算时间过长,导致无法有效筛选异常数据。为此采用手肘法来确定K值。手肘法的核心指标是误差平方和,其表达式为:

$$SSE = \sum_{i=1}^K \sum_{p \in C_i} |p - m_i|^2 \quad (4)$$

该方法的核心思想是随着K值的增加,样本数据划分更加精细,各个簇的聚合度逐渐提高,误差平方和逐渐减小。当K值小于真实聚类数时,K值增加会显著增强各个簇的聚合度,误差平方和下降幅度变大。当K值接近真实聚类数时,提高K值各个簇的聚合度变小,误差平方和变化幅度骤减(图3)。

由图3可知,当K取值为5,即当聚类数为5簇时,K均值聚类算法性能最优,过滤异常值能力最强,因此本文聚类数取5。同时计算距离时容易受较大数据的影响而忽略取值较小的数据,需在聚类前进行Z-score 标准化处理。然后通过K均值聚类算法对样本数据进行去噪,找出异常点72组。将异常点剔除,其余2 068组样本数据用于后续算法的训练与测试。

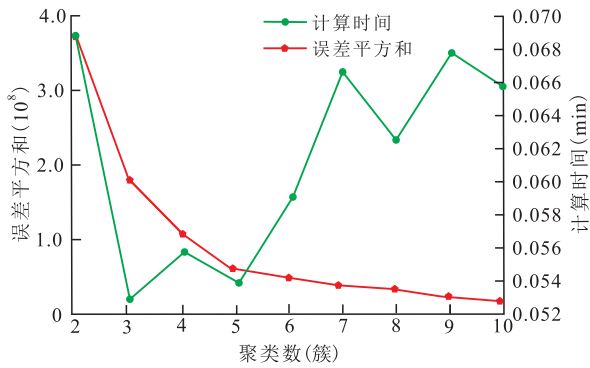


图3 误差平方和、计算时间与聚类数的关系  
Fig.3 Relationship among sum of squared errors, calculation time and cluster number

## 2 预测算法建立

作为机器学习的最新技术,集成学习在智能计算和机器学习领域引起了广泛关注。集成学习不是一种特定的模型而是一种思想,通过结合较简单的基础模型来构建强化模型。本文引入集成学习技术,将2种不同的基础模型结合起来,生成一个更好的模型来预测裂缝开度。

### 2.1 支持向量回归算法

鉴于地质认识及资料丰度的不确定性,以及特征之间具有复杂的非线性关系,应用传统回归模型不能较好地进行裂缝开度预测。而支持向量回归算法可通过核函数将样本数据映射到高维空间,解决非线性问题,同时该算法具有良好的稳定性和泛化能力<sup>[14]</sup>。支持向量回归算法可形式化为:

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m l_e [f(x) - y_i] \quad i = 1, 2, 3, \dots, m \quad (5)$$

引入松弛变量  $\xi_i$  和  $\hat{\xi}_i$ , 可将(5)式写为:

$$\min_{w,b,\xi_i,\hat{\xi}_i} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m (\xi_i + \hat{\xi}_i) \quad i = 1, 2, 3, \dots, m \quad (6)$$

引入拉格朗日算子  $u_i \geq 0, \hat{u}_i \geq 0, \alpha_i \geq 0, \hat{\alpha}_i \geq 0$ , 其拉格朗日函数表达式为:

$$\begin{aligned} L(w, b, \alpha, \hat{\alpha}, \xi, \hat{\xi}, u, \hat{u}) = & \frac{1}{2} \|w\|^2 + \\ & C \sum_{i=1}^m (\xi_i + \hat{\xi}_i) - \sum_{i=1}^m u_i \xi_i - \sum_{i=1}^m \hat{u}_i \hat{\xi}_i + \\ & \sum_{i=1}^m \alpha_i [f(x_i) - y_i - \varepsilon - \xi_i] + \\ & \sum_{i=1}^m \hat{\alpha}_i [f(x_i) - y_i - \varepsilon - \hat{\xi}_i] \quad i = 1, 2, 3, \dots, m \quad (7) \end{aligned}$$

其中:

$$\alpha = (\alpha_1, \alpha_2, \dots, \alpha_m) \quad (8)$$

根据 wolf 对偶的定义,在 KKT 条件下得到拉格朗日对偶形式为:

$$\begin{aligned} \max_{\alpha, \hat{\alpha}} \sum_{i=1}^m y_i (\hat{\alpha}_i + \alpha_i) - \varepsilon (\hat{\alpha}_i + \alpha_i) - \\ \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m (\hat{\alpha}_i - \alpha_i) (\hat{\alpha}_j - \alpha_j) x_i^T x_j \\ i = 1, 2, 3, \dots, m \quad (9) \end{aligned}$$

支持向量回归算法函数表达式为:

$$f(x) = \sum_{i=1}^m (\hat{\alpha}_i - \alpha_i) x_i^T x_j + b \quad i = 1, 2, 3, \dots, m \quad (10)$$

对于非线性问题,可通过非线性变换转化为某个高维空间中的线性问题,即用核函数  $k(x, x_i)$  替换  $x_i^T x_j$  可以实现非线性函数拟合,能较好处理非线性以及高维数的问题,可表示为:

$$f(x) = \sum_{i=1}^m (\hat{\alpha}_i - \alpha_i) k(x, x_i) + b \quad i = 1, 2, 3, \dots, m \quad (11)$$

### 2.2 XGBoost 回归算法

XGBoost 是由一系列回归树组成的强大的预测模型。其核心思想是不断添加回归树,通过生成新树来拟合前一棵树的残差。当训练完成得到  $N$  棵回归树时,将每棵树对应的分数加起来就是该样本的预测值<sup>[15]</sup>,其表达式为:

$$\hat{y}_i = \sum_{k=1}^N f_k(x_i) \quad i = 1, 2, 3, \dots, N \quad (12)$$

XGBoost 目标函数为:

$$L_t = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (13)$$

其中:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w'\|^2 \quad (14)$$

为避免算法拟合过程中的过拟合,算法不能同时训练所有回归树,因此利用固定训练好的回归树,依次添加一棵新树来解决,假设步骤  $t$  的预测值用  $\hat{y}_i^{t-1}$  表示,(12)式可以写为:

$$L_t = \sum_i l(y_i, \hat{y}_i^{t-1} + f_i(x_i)) + \sum_k \Omega(f_k) + Const \quad (15)$$

将其进行二阶泰勒展开为:

$$L_t = \sum_i [g_i f_k(x_i) + \frac{1}{2} h_i f_k^2(x_i)] + \Omega(f_k) \quad i = 1, 2, 3, \dots, n \quad (16)$$

其中:

$$g_i = \partial_{\hat{y}_i^{t-1}} l(y_i, \hat{y}_i^{t-1}) \quad (17)$$

$$h_i = \partial_{\hat{y}_i^{t-1}}^2 l(y_i, \hat{y}_i^{t-1}) \quad (18)$$

则(15)式可以改写为:

$$L_t = \sum_{j=1}^T \left[ \left( \sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left( \sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T$$

$j = 1, 2, 3, \dots, T$  (19)

### 2.3 基于岭回归的集成学习算法

本文所提的裂缝开度预测集成算法以 XGBoost 回归算法和支持向量回归算法为基础模型。每个基础模型均接收输入数据,并给出独立的裂缝开度预测结果,这些预测结果均作为元特征,被馈送到元学习器中(本文的元学习器采用岭回归算法),并给出最终的裂缝开度预测结果(图4)。

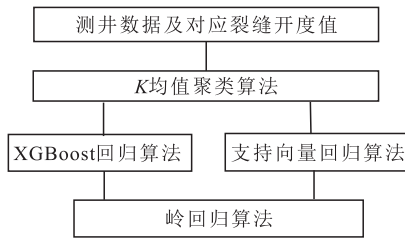


图4 基于岭回归的集成学习算法

Fig.4 Ensemble learning algorithm based on ridge regression

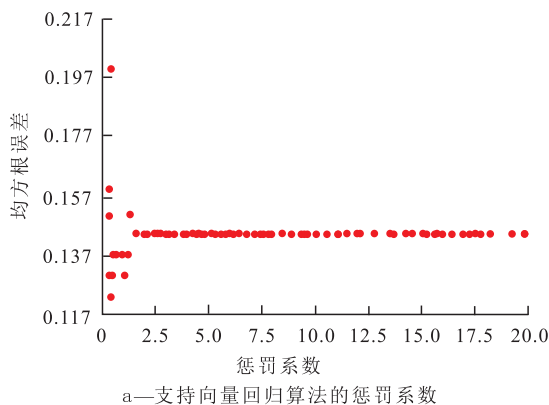
该算法为基础学习器  $\delta_{g=1,2}$  ( $g=1$  为支持向量回归,  $g=2$  为 XGBoost 回归) 对于  $H$  折交叉验证中的每一个待预测的训练样本集合  $D_H$  都有与之对应的训练集预测结果集合  $Z_{Hg}$ 。这样的循环过程完毕后,对于每个基础学习器而言,都有  $H$  对同质基础学习器训练集预测值,将其整合为  $D_{2g}$ 。再将所有基础学习器  $D_{2g}$  整合作为元特征定义为  $D_2$ 。将  $D_2$  馈送到岭回归算法得到加权结果即为最终预测开度值。基于岭回归的集成学习算法的最终表达式为:

$$Y = W\beta + \varepsilon$$
 (20)

其中:

$$\beta = (W^T W + \lambda_L I)^{-1} W^T Y$$
 (21)

在此基础上,经  $K$  均值聚类算法去噪后,应用基



a—支持向量回归算法的惩罚系数

于岭回归的集成学习算法进行裂缝开度预测,笔者将其定义为新型集成学习算法。

## 3 预测算法应用

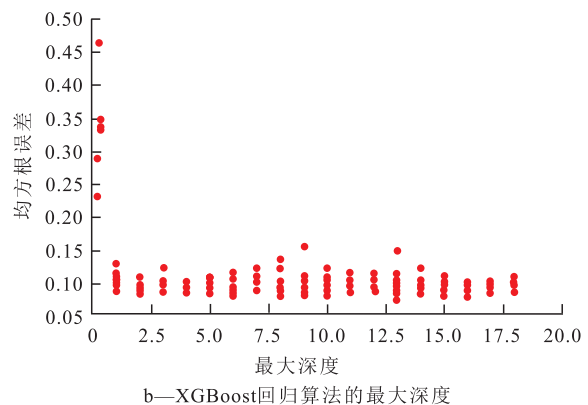
### 3.1 模型参数优化求解

机器学习算法参数的选择直接决定了算法的性能。网格搜索法是当前应用最为广泛的参数优化算法。但该方法依靠穷举所有参数进行优化,计算成本过于庞大,同时对于连续数据需要等间取样,不一定能取得全局最优。故采用随机搜索进行参数优化,该方法主要原理是从指定的分布中采样固定数量的参数设置。与网格搜索法相比,该方法在保障准确度的同时,显著减少计算时间。

根据测试集上模型的均方根误差值来判断基础模型最佳超参数。其中支持向量回归算法的主要超参数为惩罚系数, XGBoost 回归算法的主要超参数为最大深度,其超参数随机优化调参过程如图5所示。在搜索的过程中,超参数快速收敛,并找出最优值。支持向量回归算法的惩罚系数搜索范围为0~20,最优值为0.147,对应均方根误差为0.113; XGBoost 回归算法的最大深度搜索范围为0~18,最优值为13,对应均方根误差为0.076。

### 3.2 模型应用评价分析

确定好模型参数之后,随机选取 80% 经过 Z-score 标准化处理后的样本数据作为训练集共 1 712 组, 20% 的样本数据作为测试集共 428 组来验证模型效果。以均方根误差(RMSE)和真实裂缝开度值与预测裂缝开度值间相关系数( $R^2$ )作为评价标准。将测试集分别代入训练好的支持向量回归算法、XGBoost 回归算法及基于岭回归的集成学习算法中,计算测试集中真实裂缝开度与预测裂缝开度间相关系数(图6)。



b—XGBoost回归算法的最大深度

图5 随机搜索优化调参过程

Fig.5 Parameter adjustment optimization process based on random search

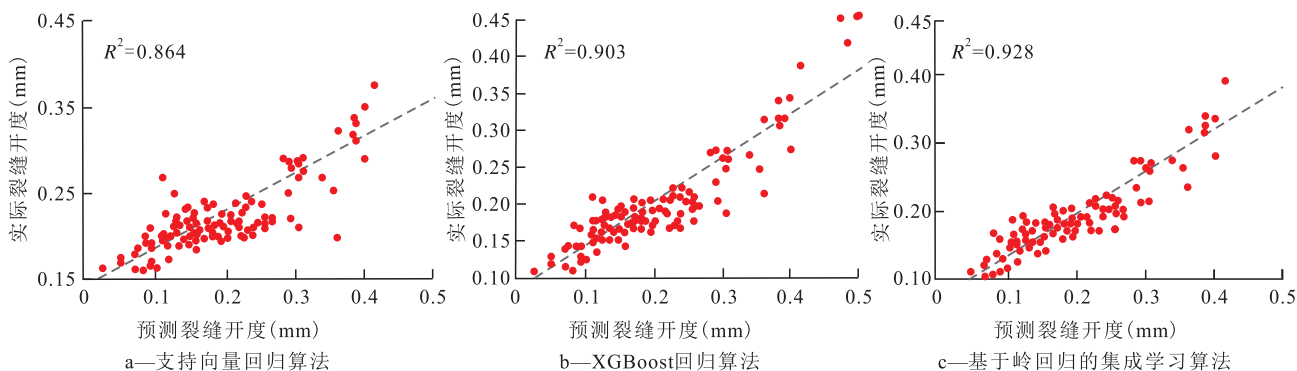


图6 预测裂缝开度与真实裂缝开度交会图

Fig.6 Cross plot of measured and predicted apertures

由图6可以看出,3种算法中,基于岭回归的集成学习算法的 $R^2$ 最高,达0.928。同时为探究 $K$ 均值聚类降噪效果,将样本数据馈送于基于岭回归的集成学习算法中进行训练和测试,并与先前计算结果进行综合对比(表2),发现4组方法中 $K$ 均值-基于岭回归的集成学习算法的RMSE最小, $R^2$ 最大。即该算法的预测裂缝开度值与真实裂缝开度值之间的偏差最小,支持向量回归算法的RMSE最大, $R^2$ 最小。 $K$ 均值聚类算法能够对学习样本进行有效降噪,去除冗余,提高了学习样本的质量。

表2 各算法预测效果对比

Table2 Comprehensive comparison results of prediction effects in various algorithms

算 法	RMSE	$R^2$
支持向量回归	0.113	0.864
XGBoost 回归	0.076	0.903
基于岭回归的集成学习	0.058	0.928
新型集成学习( $K$ 均值-基于岭回归的集成学习)	0.047	0.931

为进一步综合分析该算法的预测效果,将部分样本真实值及各算法预测值进行可视化研究。从图7可以明显看出,支持向量回归算法和XGBoost回归算法预测值整体在真实值上下波动,支持向量回归算法总体变化平稳,但对裂缝开度突变值检测

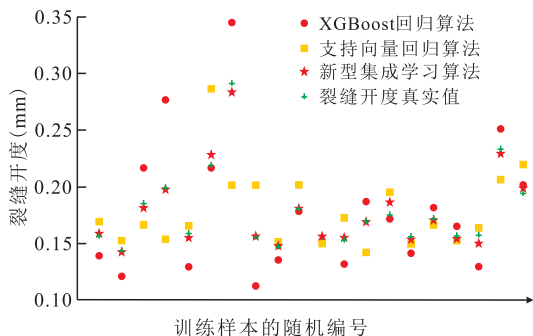


图7 样本观测值

Fig.7 Sample observation values

不明显。XGBoost回归算法对数据敏感,部分数值波动较大。新型集成学习算法的计算结果紧密围绕真实裂缝开度值波动,很好地结合了基础算法的优点,平衡了基础算法的缺点,预测精度明显提升。

## 4 结论

利用测井数据及其对应裂缝开度值,提出基于新型集成学习的基岩潜山油藏储层裂缝开度预测算法。该算法先通过 $K$ 均值将学习样本进行聚类、降噪来提升学习样本质量;以支持向量回归算法和XGBoost回归算法为基础模型,并利用随机搜索进行基础模型参数优化。然后利用岭回归算法对优化好的基础模型进行集成组合。所提出的新型集成学习算法建立的裂缝开度预测模型弥补了单一回归算法不稳定的特点,提升了预测精度,能够充分挖掘测井数据中蕴含的地质信息,为裂缝开度定量预测提供了新的思路。同时该方法可实现自动、快速优化调参,具有广泛的适用性。

### 符号解释

$x'$ ——标准化处理后的样本数据; $x$ ——样本数据; $\mu$ ——样本均值; $\delta$ ——样本方差; $K$ ——聚类数,簇; $X$ ——样本向量集合; $j$ ——样本序号; $n$ ——样本向量总数; $center$ ——簇中心集合; $P(x_j)$ ——簇中心概率; $D(x_j)^2$ ——最近簇中心的距离; $SSE$ ——误差平方和; $i$ ——样本列号; $p$ —— $C_i$ 中的样本点; $C_i$ ——第 $i$ 个簇; $m_i$ —— $C_i$ 中所有样本的均值; $w$ ——法向量; $b$ ——位移项; $C$ ——惩罚系数; $m$ ——样本数量; $l_\epsilon$ ——不敏感损失函数; $\epsilon$ ——软间隔带; $f(x)$ ——支持向量回归算法函数; $y_i$ ——开度实际值; $\xi_i$ 和 $\hat{\xi}_i$ ——松弛变量; $u_i, \hat{u}_i, \alpha_i$ 和 $\hat{\alpha}_i$ ——在第 $i$ 数据下不同的拉格朗日算子; $\alpha$ ——拉格朗日算子 $\alpha_i$ 的合集; $\hat{\alpha}$ ——拉格朗日算子 $\hat{\alpha}_i$ 的合集; $\alpha_j, \hat{\alpha}_j$ ——在第 $j$ 数据下不同的拉格朗日算子; $x_i$ 和 $x_j$ ——输入的第 $i$ 和第 $j$ 个数据; $k(x, x_j)$ ——核函数; $\hat{y}_i$ ——开度预测

值;  $N$ ——回归树的总数, 个;  $f_k$ ——第  $k$  棵回归树算法;  $L_t$ ——步骤  $t$  下的目标函数;  $t$ ——步骤序号;  $\sum_i l(\hat{y}_i, y_i)$ ——预测开度值与真实开度值的差;  $\Omega(f)$ ——惩罚项;  $\gamma$ ——回归树分割的难度系数;  $T$ ——回归树叶子节点个数, 个;  $\lambda$ ——L2 正则系数;  $w'$ ——回归树叶子节点权重;  $\hat{y}_i^{t-1}$ ——步骤  $t$  的预测值;  $Const$ ——常数;  $g_i$ ——损失函数一阶导数;  $h_i$ ——损失函数二阶导数;  $I_j$ ——第  $j$  个叶子的样本集合;  $\delta_{g=1,2}$ ——基础学习器;  $H$ ——交叉验证折数;  $D_H$ —— $H$  折交叉验证中的每一个待预测的训练样本集合;  $Z_{H_g}$ ——与  $D_H$  对应的训练集预测结果集合;  $D_{2g}$ —— $H$  对同质基础学习器训练集预测值;  $D_2$ ——元特征;  $Y$ ——预测的最终开度值;  $W$ ——元特征数据矩阵;  $\beta$ ——岭回归估计值;  $\lambda_L$ ——岭参数;  $I$ ——单位矩阵。

### 参考文献

- [1] 梅丹, 胡勇, 王倩. 裂缝对气藏储层渗透率及气井产能的贡献[J]. 石油实验地质, 2019, 41(5): 769-772.  
MEI Dan, HU Yong, WANG Qian. Experimental study on fracture contribution to gas reservoir permeability and well capacity [J]. Petroleum Geology & Experiment, 2019, 41(5): 769-772.
- [2] 王威, 卢祥国, 吕金龙, 等. 裂缝对致密砂岩储层物性及产气能力影响实验[J]. 大庆石油地质与开发, 2019, 38(4): 160-166.  
WANG Wei, LU Xiangguo, LÜ Jinlong, et al. Experiments of the effects of the fracture on the physical property and gas production capacity for the tight sandstone [J]. Petroleum Geology & Oilfield Development in Daqing, 2019, 38(4): 160-166.
- [3] 吕金龙, 卢祥国, 王威, 等. 裂缝对致密储层渗流能力影响实验研究[J]. 特种油气藏, 2019, 26(2): 141-146.  
LÜ Jinlong, LU Xiangguo, WANG Wei, et al. Experiment research on the effect of fracture on the seepage capacity of tight reservoir [J]. Special Oil & Gas Reservoirs, 2019, 26(2): 141-146.
- [4] MOHAMMAD Javad Afshari Moein, SOMOGYVÁRI M, VALLEY B, et al. Fracture network characterization using stress-based tomography [J]. Journal of Geophysical Research: Solid Earth, 2018, 123(11): 9 324-9 340.
- [5] 杨明清, 王超, 阎治全. 纳秒激光诱导击穿光谱技术在岩屑样品元素分析中的应用[J]. 中国石油勘探, 2018, 23(1): 117-124.  
YANG Mingqing, WANG Chao, YAN Zhiquan. Application of nanosecond laser induced breakdown spectroscopy to elemental analysis of cuttings samples [J]. China Petroleum Exploration, 2018, 23(1): 117-124.
- [6] 陈彦君, 苏雪峰, 王钧剑, 等. 基于 X 射线微米 CT 扫描技术的煤岩孔裂隙多尺度精细表征——以沁水盆地南部马必东区块为例[J]. 油气地质与采收率, 2019, 26(5): 66-72.  
CHEN Yanjun, SU Xuefeng, WANG Junjian, et al. Multi-scale
- [7] LAI J, WANG G, FAN Z, et al. Three-dimensional quantitative fracture analysis of tight gas sandstones using industrial computed tomography [J]. Scientific Reports, 2017, 7(1): 1 825.
- [8] 冯福平, 雷扬, 陈顶峰, 等. 基于有限元数值模拟的致密储层体积压裂效果影响参数分析[J]. 油气藏评价与开发, 2019, 9(1): 29-33, 50.  
FENG Fuping, LEI Yang, CHEN Dingfeng, et al. Parameter analysis of SRV fracturing effect of tight reservoirs based on finite element numerical simulation [J]. Reservoir Evaluation and Development, 2019, 9(1): 29-33, 50.
- [9] 王永东, 雷俊杰, 樊万红, 等. 延长南部低渗透储层天然裂缝定量识别技术[J]. 断块油气藏, 2018, 25(3): 322-327.  
WANG Yongdong, LEI Junjie, FAN Wanhong, et al. Quantitative identification of natural fractures in tight reservoir of southern Yanchang Oilfield [J]. Fault-Block Oil and Gas Field, 2018, 25(3): 322-327.
- [10] 韩刚, 何峰, 张孝珍, 等. 阵列声波测井在储层裂缝识别中的应用——以鄂尔多斯盆地 K 区为例[J]. 油气地质与采收率, 2019, 26(3): 63-69.  
HAN Gang, HE Feng, ZHANG Xiaozhen, et al. Application of array acoustic logging to fracture identification: A case study of Area K in Ordos Basin [J]. Petroleum Geology and Recovery Efficiency, 2019, 26(3): 63-69.
- [11] KRAWCZYK B, MINKU L L, GAMA J, et al. Ensemble learning for data stream analysis: A survey [J]. Information Fusion, 2017, 37(2): 132-156.
- [12] 余朝华, 肖坤叶, 张桂林, 等. 乍得 Bongor 盆地反转构造特征及形成机制分析[J]. 中国石油勘探, 2018, 23(3): 90-98.  
YU Zhaohua, XIAO Kunye, ZHANG Guilin, et al. Analysis on inverted structure characteristics and its forming mechanism in the Bongor Basin, Chad [J]. China Petroleum Exploration, 2018, 23(3): 90-98.
- [13] ARORA P, VARSHNEY S. Analysis of K-means and K-medoids algorithm for big data [J]. Procedia Computer Science, 2016, 78(2): 507-512.
- [14] AL-ANAZI A, GATES I D. A support vector machine algorithm to classify lithofacies and model permeability in heterogeneous reservoirs [J]. Engineering Geology, 2010, 114(3/4): 267-277.
- [15] NGUYEN H, BUI X N, BUI H B, et al. Developing an XGBoost model to predict blast-induced peak particle velocity in an open-pit mine: a case study [J]. Acta Geophysica, 2019, 67(2): 477-490.

编辑 单体珍