

引用格式: 聂云丽, 高国忠. 基于随机森林的页岩气“甜点”分类方法[J]. 油气藏评价与开发, 2023, 13(3): 358-367.

NIE Yunli, GAO Guozhong. Classification of shale gas “sweet spot” based on Random Forest machine learning[J]. Petroleum Reservoir Evaluation and Development, 2023, 13(3): 358-367.

DOI: 10.13809/j.cnki.cn32-1825/te.2023.03.011

基于随机森林的页岩气“甜点”分类方法

聂云丽^{1,2}, 高国忠^{1,2}

(1. 长江大学非常规油气省部共建协同创新中心, 湖北 武汉 430100; 2. 长江大学地球物理与石油资源学院, 湖北 武汉 430100)

摘要: 为了解决页岩气“甜点”分类识别涉及指标多、需要根据个人经验判别、耗时耗力的问题, 提出了一种基于随机森林模型的页岩气“甜点”分类方法。首先, 选取长宁区的10口井数据, 利用肯德尔相关分析筛选出用于识别的11种特征。然后再分别采用单棵决策树和随机森林方法进行预测, 得到页岩气“甜点”识别结果。最后, 对预测结果分类并进行算法参数优化。实际应用结果表明, 单棵决策树预测精度虽可以达到97.7%, 但呈现过拟合趋势, 且剪枝之后拟合精度大大降低到只有70.7%; 采用的随机森林方法避免了单棵决策树的缺陷, 并且预测的精度达到98%, 而且, 计算代价小, 能有效降低时间损耗、节省人力成本。证明随机森林机器学习方法结合多源信息是实现页岩气“甜点”识别预测的一种有效手段。

关键词: 页岩气; “甜点”; 机器学习; 决策树; 随机森林

中图分类号: TE132

文献标识码: A

Classification of shale gas “sweet spot” based on Random Forest machine learning

NIE Yunli^{1,2}, GAO Guozhong^{1,2}

(1. Cooperative Innovation Center of Unconventional Oil and Gas, Yangtze University (Ministry of Education & Hubei Province), Wuhan, Hubei 430100, China; 2. College of Geophysics and Petroleum Resources, Yangtze University, Wuhan, Hubei 430100, China)

Abstract: The classification and identification of shale gas “sweet spot” involves a variety of different factors, which requires personal experience, and is usually time and resources consuming. In order to solve this problem, an efficient and effective classification and identification method for shale gas “sweet spot” based on the Random Forest method is proposed. Firstly, data from ten wells in Changning area are selected and eleven features are selected for “sweet spot” classification by the Kendall correlation. Then, the single decision tree and the Random Forest method are used for the “sweet spot” classification and identification. Finally, the results are verified and the Random Forest parameters are optimized. The experimental results show that although the prediction accuracy of a single decision tree can reach 97.7%, it shows a trend of overfitting, and the fitting accuracy is greatly reduced by only 70.7% after pruning. The Random Forest method avoids the disadvantage of the single decision tree method, and the prediction accuracy reaches 98%. Moreover, the computational cost is low, which can effectively reduce the time loss and save the labor cost. As a result, the proposed Random Forest machine learning method with multi-source information is an effective shale gas “sweet spot” classification and identification method.

Keywords: shale gas; “sweet spot”; machine learning; decision tree; Random Forest

收稿日期: 2022-03-02。

第一作者简介: 聂云丽(1997—), 女, 在读硕士研究生, 主要从事地球物理大数据方面的研究。地址: 湖北省武汉市蔡甸区大学城路111号长江大学地球物理与石油资源学院, 邮政编码: 430100。E-mail: 823113137@qq.com

通信作者简介: 高国忠(1974—), 男, 博士, 教授, 主要从事测井技术、定向钻井、电磁勘探、地球物理正反演、大数据和人工智能等方面的研究。地址: 湖北省武汉市蔡甸区大学城路111号长江大学地球物理与石油资源学院, 邮政编码: 430100。E-mail: ggao@yangtzeu.edu.cn

基金项目: 长江大学非常规油气省部共建协同创新中心开放基金“水力压裂裂缝的扩展机理及数值模拟研究”(UOG2022-05)。

全球非常规油气资源量占油气总储量的80%^[1],页岩气更高效的生产在我国能源战略安全中发挥着重要作用。页岩气“甜点”通常指页岩气勘探开发的最佳富集层位。ZOU等^[2]认为低成本实现单井高收益采收是“甜点”评价的主要目的。中国的页岩气勘探开发获得重大进展,成为北美以外首次完成大规模商业开发的国家^[3-4]。因此,深入研究页岩气“甜点”预测方法大规模经济开发具有重要意义^[5-6]。总之,页岩气“甜点”的预测是页岩气低成本勘探和高收益的关键。

近年来,国内外的研究人员也对页岩气的“甜点”预测问题进行了很多研究。MA^[7]分析了页岩气富集规律及其商业化生产的关键因素;HASHMY等^[8]提出具有良好的流动特征和储层属性是“甜点”经济评价的关键;李小明等^[9]综合测井解释效果,从岩石学、有机地球化学、矿物组成、孔隙类型及结构、物性及含气性等方面对荆门探区五峰组—龙马溪组页岩储层特征和“甜点”层段进行详细评价;夏宏泉等^[10]利用测井资料从烃源岩特性、储层物性和完井质量方面优选了10个具有代表性的地质和工程“甜点”参数;WU等^[11]通过岩心薄片鉴定标定测井相及沉积微相进行流动单元划分评价“甜点”参数,并进行最佳“甜点”预测;WANG等^[12]对四川焦石坝区块进行研究得出影响页岩“甜点”因素主要有脆性指数、天然裂纹、水平应力差系数、断裂韧性、成岩作用等;陈桂华等^[13]用页岩埋深、页岩储层净厚度、页岩中有机质丰度、页岩中有机质成熟度和页岩的脆性度这5个指标来描述页岩气“甜点”,指出页岩储层快速评价中重点要确定“五度”;马永生等^[14]认为页岩气“甜点”评价包含两部分:页岩气参数评价如TOC(总有机碳含量)、有机物类型、成熟度、页岩厚度、含气含量等及可压裂性评价;ALSHAKHS等^[15]、CHORN等^[16]利用TOC、孔隙度、脆性等参数分别评价了北美地区Barnett页岩以及澳大利亚坎宁盆地Goldwyer组页岩。前人对页岩气“甜点”评价方法总结为:①利用单一因素进行分析确定页岩气优选区,考虑情况单一;②利用对各种指标进行打分,直接人为给定页岩气“甜点”区,主观性相对太大;③各专家按照自己的计算得到对应的权重,但是权重归一化难以得到满足,准确性可能不可靠。总结国内外的相关研究成果,可以将影响页岩气“甜点”的因素分为3类,即有机质质量、储层品质和完井质量^[17-19]。因此,如何从

多源信息中“优中选优,甜中选甜”是页岩气“甜点”预测的关键。

由于页岩气“甜点”受许多因素影响,简单的线性分类不足以实现高精度“甜点”预测,解决这种非线性分类问题正是随机森林的优势。前人在算法方面的研究主要有:周广照等^[20]优选出研究区总含气量的影响因素,运用BP神经网络模型进行页岩储层含气性评估;王彬等^[21]通过人工神经网络输入美国页岩气储层的7个重要参数,输出盈亏平衡点(天然气销售价格),优先选择页岩气优势开发区域;汪敏等^[22]运用MAMO算法对页岩气储层进行预测,与多标签学习算法MLKNN、BPMLL、GLOCAL和QUIRE相比,MAMO算法在页岩气储层综合品质预测精度均值分别提升了45个百分点、68个百分点、68个百分点和51个百分点。样本和随机特征的多样性可以促使随机森林更好地发挥优势,与常规评价方法相比,随机森林预测页岩储层特性能够节约成本并快速获取高精度的页岩属性及分布特征。因而基于随机森林的页岩储层属性预测,会成为页岩气“甜点”评价的重要方法^[23]。

为解决上述问题,提出了一种基于随机森林模型的页岩气的“甜点”分类方法,利用多特征变量可以提高随机森林的可用特征,降低特征冗余干扰,达到高精度分类。以长宁区的10口井数据为例,首先将井的各类数据关联,运用肯德尔相关分析筛选出用于识别“甜点”的11种特征;然后采用单棵决策树和随机森林算法进行“甜点”分类,并调节参数;最后应用随机森林算法获取高精度的分类结果,并与决策树分类结果进行对比,分析评价随机森林对页岩气“甜点”的分类优势。通过实际数据验证该方法的有效性,该方法可以运用到实际页岩气“甜点”预测分类中,对页岩气的开采提供有效的技术支撑。

1 研究数据

研究区地处四川省泸州市境内,长宁背斜区域构造上位于川东南隔挡式断褶带与娄山断褶带的交汇处^[24],浅部地层(二叠系—白垩系)部分遭受剥蚀。受川东南构造演化的影响,长宁背斜构造总体表现为东西展布,呈左雁列排列,多种构造带发育。五峰组—龙马溪组沉积时期,长宁处于还原环境的深水陆棚有利相带,沉积了一套横向展布稳定的黑色富

有机质页岩优质储层。研究区储层受多期复杂构造运动作用,导致微裂缝及断层发育^[25-26]、储层空间类型多样,且受页岩气成因及来源的差异性影响,使得储层非均质性极强。相较于中浅层而言,研究区深层页岩气储层具有赋存时代老、埋藏深、储层致密、压裂难度大等特点^[27-28],并且影响页岩气“甜点”的要素很多,传统方法难以将多要素精细表征优质页岩气“甜点”类型。因此,采用随机森林算法从多因素中快速对页岩气“甜点”类型进行识别。

选取的特征参数按照三质量分类^[17-19],包括有机质质量:TOC、总含气量(压力系数为1.0)、总含气量(压力系数为2.0);储层品质:孔隙度,厚度(m);完井质量^[29-30]:泊松比、杨氏模量(GPa)、垂向应力(MPa)、破裂压力(MPa)、脆性矿物含量(含碳酸盐岩+碎屑岩)、脆性矿物含量(纵横波比)。有研究表明^[30-31]:不同压力系数对产能影响较大,因此,选择2种地层压力系数的总含气量计算值。页岩脆性对压裂施工影响很大^[32-33],因此,从矿物学和岩石物理学2个不同角度综合表征岩石脆性指数。且加入同一指标的不同计算值可以更好地区分页岩气“甜点”类型。

页岩气“甜点”通常需要满足页岩厚度大、总有机碳含量高、含气量较高、脆性矿物含量高,低泊松比、高弹性模量等条件,依据页岩气的三质量参数,将页岩气“甜点”优劣依次分为I型、II型和III型3大类。决策树模型和随机森林模型选取长宁区的10口

井数据,由11个特征构成的466条样本数据制作成实验数据集。将样本数据按照7:3(训练集:测试集)进行划分,其中训练集样本279个,测试集样本187个,训练集与测试集样本的构成要素一致。表1给出了数据集样本特性。

2 计算原理

2.1 肯德尔相关性分析

肯德尔相关系数(τ)主要是检验2个随机变量之间的统计依赖性^[34],其中 τ 的取值 $[-1, 1]$ 。假设2个随机数据特征孔隙度和厚度分别用 X 和 Y 表示, (X_i, Y_i) 与 (X_j, Y_j) 是随机向量孔隙度和厚度的样本,随机向量 (X, Y) 的Kendall相关系数 τ ^[35]定义为 (X_i, Y_i) 与 (X_j, Y_j) 的和谐概率减去它们的不和谐概率,见式(1):

$$\tau(X, Y) = P[(X_i - X_j)(Y_i - Y_j) > 0] - P[(X_i - X_j)(Y_i - Y_j) < 0] \quad (1)$$

运用肯德尔相关性分析法对11个影响页岩气“甜点”类型的特征进行计算分析,计算的11个特征之间的 τ 值如图1,可以看到本研究筛选出来的各个特征之间的相关性都很小,仅总含气量与孔隙度之间 τ 值达到0.56,垂向应力与破裂压力之间 τ 值达到0.68,其余特征参数多重共线性特征很弱。图1中也表明对页岩气“甜点”具有较大影响的特征有:TOC、总含气量、孔隙度。

TOC、总含气量、孔隙度的数据对页岩气“甜点”

表1 样本特性
Table 1 Characteristics of sample

类型	有机质质量			储层品质			完井质量				
	TOC/ %	总含气量 (压力系数为1.0)/ (m ³ /t)	总含气量 (压力系数为2.0)/ (m ³ /t)	孔隙度/ %	厚度/ m	泊松比	杨氏 模量/ GPa	垂向 应力/ MPa	破裂 压力/ MPa	脆性矿物 含量(含碳酸盐 岩+碎屑岩)/%	脆性矿物 含量(纵横 波比)/%
最小值	2.30	0.90	1.30	2.70	3.0	0.190	31.1	50.10	43.90	43.00	42.80
I 平均值	7.25	4.60	6.80	5.00	735.0	0.240	25 444.4	81.30	91.20	62.80	65.35
最大值	12.20	8.30	12.30	7.30	1 467.0	0.290	50 857.7	112.50	138.50	82.60	87.90
最小值	1.70	2.30	3.50	3.20	11.0	0.150	36.2	54.10	40.60	46.60	42.10
II 平均值	3.05	3.95	5.75	4.60	474.0	0.225	25 519.2	83.35	94.50	63.15	61.45
最大值	4.40	5.60	8.00	6.00	937.0	0.300	51 002.3	112.60	148.40	79.70	80.80
最小值	1.00	1.30	2.00	0.30	7.0	0.240	38.6	51.80	46.70	48.40	61.50
III 平均值	2.85	1.50	4.30	1.45	177.9	0.290	31 901.2	82.25	110.65	55.55	75.00
最大值	4.70	1.70	6.60	2.60	348.8	0.340	63 763.9	112.70	174.60	62.70	88.50

敏感性较大,因此,分类之前,需要对敏感的参数进行分析。将TOC、总含气量(压力系数为1.0)、总含气量(压力系数为2.0)、孔隙度4个特征与“甜点”类型的关系进行分析,其中TOC、总含气量、孔隙度及“甜点”类型的关系如图2。可以看到直接利用TOC、含气量和孔隙度区分“甜点”类型时,并不能做到很精确,甚至出现误判。因此,将这4个特征单独运用决策树和随机森林算法进行进一步实验分析。

2.2 决策树构建相关计算

决策树的构建,首先需要进行特征选择。主要运用信息增益和基尼指数(gini)来进行特征选择。ID3(Iterative Dichotomiser 3)算法是构建决策树的模型之一,通过计算每个特征的信息增益,使用信息增益高的特征作为划分标准。CART(Classification and Regression Tree)分类回归树^[36]能处理分类问题。构

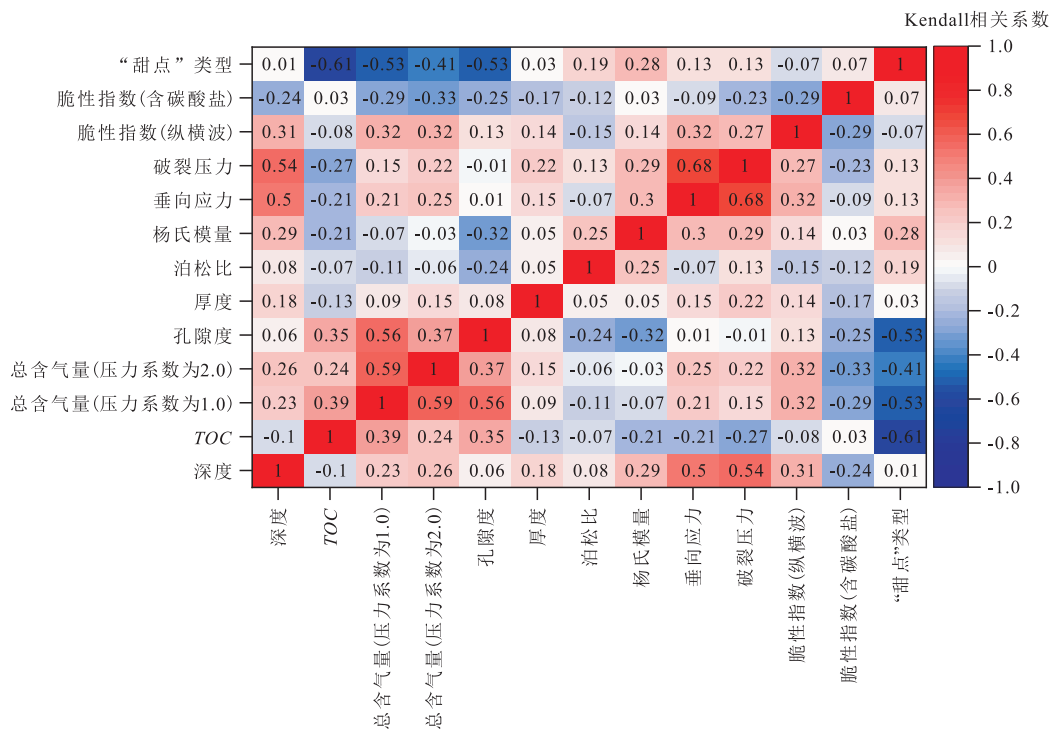


图1 Kendall 相关系数计算各个特征之间的关系
Fig. 1 Kendall correlation coefficient for all parameters

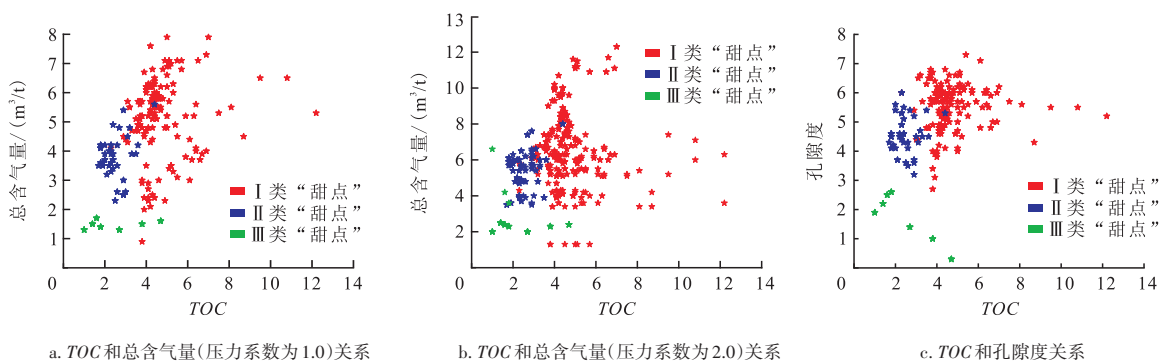


图2 TOC、总含气量(压力系数为1.0)、总含气量(压力系数为2.0)、孔隙度及“甜点”类型的关系

Fig. 2 Relationship between TOC, total gas content (pressure coefficient: 1.0), total gas content (pressure coefficient: 2.0), porosity and “sweet spot” type

建的CART分类树,使用基尼指数来选择最好的数据分割特征。

以孔隙度特征为例,说明信息增益和基尼指数的计算。信息增益^[37]表示得知孔隙度属性 X 的信息而使得页岩气“甜点”类型信息的不确定性程度减少。特征孔隙度 X 对训练数据集 D 的信息增益 $g(D, X)$,见式(2),孔隙度的信息增益越大,说明使用属性孔隙度划分后的页岩气“甜点”分类样本越纯,越有利于分类。

$$g(D, X) = B(D) - B(D|D) \quad (2)$$

基尼指数 $\text{gini}(D)$ ^[38]表示集合 D 的不确定性,基尼指数 $\text{gini}(D, X=a)$ 表示集合 D 经特征孔隙度 $X=a$ 分割后的不确定性。基尼指数越低,从集中选择的样本被错误分类的概率就越低,即收集的纯度越高;反之,收集的纯程度就越低^[39]。假设样本共有 k 类,样本点属于第 k 类的概率为 p_k ,则概率分布的基尼指数为式(3):

$$G(p) = \sum_{k=1}^k p_k(1 - p_k) = 1 - \sum_{k=1}^k p_k^2 \quad (3)$$

如果根据特征孔隙度 X 是否取某一可能值 a ,将样本集 D 分为 D_1 和 D_2 ,如式(4):

$$\begin{aligned} D_1 &= \{(x, y) \in D | X(x) = a\}, \\ D_2 &= D - D_1 \end{aligned} \quad (4)$$

则在特征孔隙度 X 的条件下,集合 D 的基尼指数为:

$$\text{gini}(D, X = a) = \frac{|D_1|}{|D|} \text{gini}(D_1) + \frac{|D_2|}{|D|} \text{gini}(D_2) \quad (5)$$

在决策树的构建时,按照孔隙度特征的计算方式,将其余特征分别进行计算,不断迭代构建决策树。生成符合规律的和高效的页岩气“甜点”的分类规则。

2.3 随机森林预测流程

随机森林作为一种高度灵敏的机器学习算法^[40],是有监督的学习方法,属于装袋算法(Bootstrap aggregating,简称Bagging)。随机森林算法通过Bagging集成学习方法将多个决策树集成,根据所有决策树投票获得最终结果。随机森林具有高精度,泛化性好和鲁棒性强等特点,且能有效抑制过拟合,在分类问题中得到了广泛应用。

实验中将原始训练样本集采用装袋算法有放回

的抽样,组成含有 n 个样本的新的训练集,重复 T 次得到 T 个训练集,对于每个训练集,从其特征变量中随机选取 m 个特征,剔除掉训练集中没有被选择的特征对应的参数,利用新的训练集构建决策树,不剪枝,在每个训练集上采用随机森林分类算法独立地训练出 T 个预测结果,最终采用投票的方式,投票多的决策树最终进行决策。

3 实验验证和效果分析

3.1 决策树构建验证

图3给出了运用ID3算法构造得到的决策树。该树是选择信息增益作为划分的依据,从输出的结果可以看到决策树tree 1选择了 TOC 、总含气量(压力系数为1.0)、孔隙度、总含气量(压力系数为2.0)、破裂压力、泊松比、厚度、杨氏模量特征,并且 TOC 作为根节点,深度为7,得到12个叶节点和11个非叶节点,有12个分支,其中划分为I型“甜点”的有4个分支,划分为II型“甜点”的有6个分支,划分为III型“甜点”的有2个分支。

图4给出了运用CART算法构造得到的页岩气“甜点”分类决策树。该树是选择基尼指数作为划分的依据,构造的决策树tree 2是把前面选择的所有指标全部都运用,甚至有些指标我们需要运用多次。首先选择了孔隙度作为根节点,深度为10,得到20个叶节点和19个非叶节点,产生20个分支,其中划分为I型“甜点”产生11个分支,划分为II型“甜点”有7个分支,划分为III型“甜点”有2个分支。

假设分类树划分得太细,噪声数据就会过拟合,应该通过修剪来解决。决策树的预修剪通过提前停止树的构建来进行。一旦停止,该节点将成为包含具有最频繁元组的类的叶子。在该文中定义了高度并在决策树达到该高度时停止生长。图5给出了剪枝后得到的决策树tree 3。剪枝之后,决策树仅仅选择了4个特征对类别进行划分,分别是孔隙度、垂向应力、总含气量(压力系数为1.0)和 TOC ,并且决策树选择孔隙度作为根节点,深度为4,得到5个叶节点和4个非叶节点,产生5个分支,其中划分为I型“甜点”产生3个分支,划分为II型“甜点”有1个分支,划分为III型“甜点”有1个分支。



注:entropy为信息熵;samples为该节点的样本数;value表示各种类所占的数量;class为该“甜点”划分的类别。

图3 运用ID3算法实现得到的决策树 tree 1
Fig. 3 Decision tree 1 from ID3 algorithm

3.2 随机森林算法模型验证

将页岩气“甜点”的样本数据集运用随机森林模型进行研究,并将参数进行不断的调整,当最佳参数选择运用信息增益,每棵决策树最大深度6,每棵决策树使用的变量占比中最大指标选择0.5,最大样本切片4,决策树数量达到15棵,能得到随机森林最好的决策,得到如图6a所示的随机森林模型精度图。从图6中可以得到随机森林模型在决策树的数量达到15时,训练集准确性达到100%,并且随机森林是

优异的机器学习方法,能有效抑制过拟合,十分适合页岩气“甜点”的分类预测。对照实验选择了其中一口井的数据进行随机森林模型分析如图6b所示的随机森林模型精度图,可以得到随机森林模型在决策树的数量达到80时,训练集准确性能达到100%,且在决策树数量为80以前,随机森林的精度总是表现得起伏不定。运用单口井少量的数据,需要更多的决策树才能使得随机森林做出正确的、稳定的分类,也证明了决策树的个数增加,则参与训练的样本重复利用率越高,模型更加健壮。



注:gini为基尼指数。

图4 运用CART算法实现得到的决策树 tree 2
Fig. 4 Decision tree 2 from CART algorithm

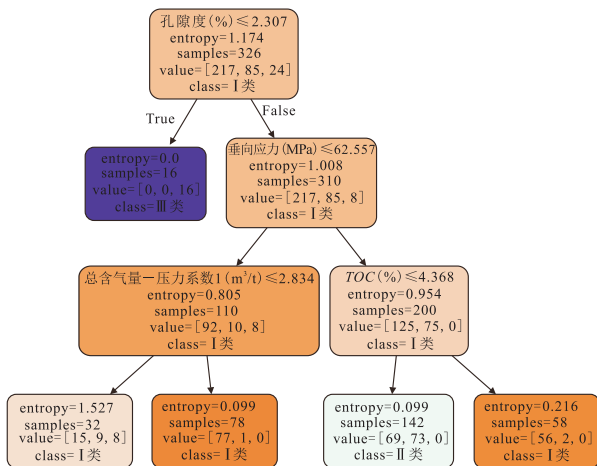


图5 剪枝后的决策树 tree 3
Fig. 5 Decision tree 3 after pruning

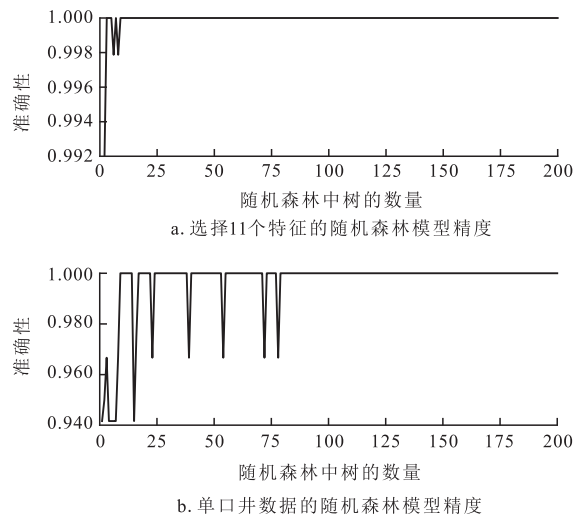


图6 随机森林模型精度
Fig. 6 Precision chart for random forest model

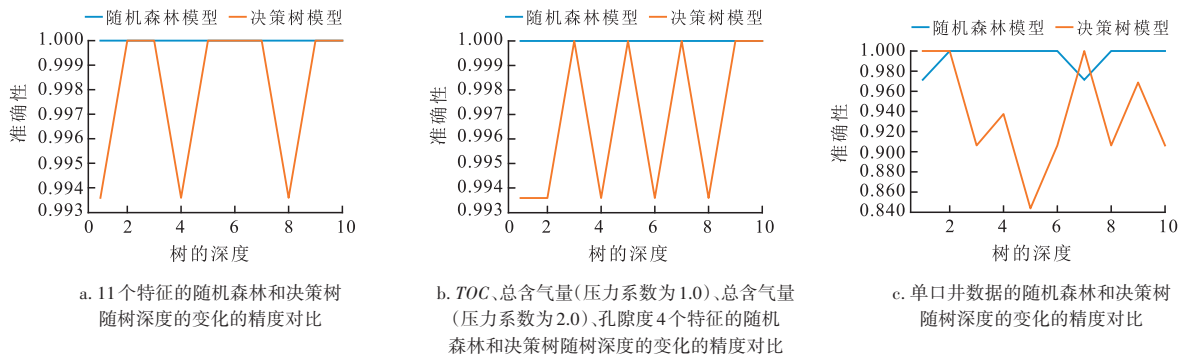


图7 随机森林和决策树随树深度变化的精度对比

Fig. 7 Comparison of accuracy of random forest and decision tree with tree depth

3.3 模型结果对比

从决策树对页岩气“甜点”分类测试集的结果得到ID3算法、CART算法和预剪枝对“甜点”的预测准确率分别达到了97.1%、97.8%和70.7%，其中ID3算法和CART算法的拟合程度分别达到99%和100%。由此认为针对这个该区的数据仅仅运用单个决策树算法是不足以满足的，所以最终选择应用随机森林进行“甜点”分类，表2给出了决策树和随机森林的预测结果比照，图7给出的随机森林和决策树随着树深度变化的精度对比。图7a是11个特征的随机森林和决策树随树深度的变化的精度对比图，图7b是TOC—孔隙度—总含气量(压力系数为1.0)—总含气量(压力系数为2.0)4个特征的随机森林和决策树随树深度的变化的精度对比，图7c是单口井数据时随机森林和决策树随树深度的变化的精度对比。从图7可以看出随机森林能够非常高效地对页岩气“甜点”进行预测，其预测效果非常好，也不会产生过拟合。同时图7b和图7c也表明随机森林对于小数据或者低维数据，可能不能产生很好的分类效果，更加适合处理高维度和数据量大的页岩气“甜点”分类问题。从随机森林测试集的分类结果可以看出，对I型、II型、III型的f1-score分别达到0.98、0.96、1.00。因此，可以得出所提出的随机森林模型非常适合页岩气“甜点”的预测评价。该模型的精度足够给出可靠的预测结果。

4 结论

1) 随机森林算法可降低人为主观性的影响，提

表2 决策树与随机森林预测效果对比

Table 2 Comparison of prediction effectiveness between decision tree and random forest

模型	甜点类型	precision (精确率)	recall (召回率)	f1-score (f1评分)	support (支持度)
决策树	I型	1.00	0.96	0.98	122
	II型	0.91	1.00	0.95	49
	III型	1.00	1.00	1.00	16
随机森林	I型	0.97	1.00	0.98	118
	II型	1.00	0.92	0.96	53
	III型	1.00	1.00	1.00	16

升页岩气“甜点”识别预测效果。其分类的结果更加准确客观，可以实际应用于页岩气“甜点”分类的预测评价中。

2) 基于肯德尔相关分析得到的11个主要特征，用于决策树的页岩气“甜点”分类时，取得了较高的分类精度，但是分割得越细拟合程度越高，以至于过拟合。采用剪枝的方法将拟合程度降低，但是精度只达到70.7%，不能达到预期的效果。而对于随机森林算法，参与分类的特征种类越多，其分类的精度和准确度越高。从选取的实际数据分类结果来看，随机森林分类模型的精度达到98%，说明该算法在页岩气“甜点”分类预测评价中有非常大的应用潜力。

参考文献

- [1] ZOU C N, YANG Z, ZHU R K, et al. Geologic significance and optimization technique of sweet spots in unconventional shale systems[J]. Journal of Asian Earth Sciences, 2019, 178: 3-19.
- [2] ZOU C N, YANG Z, ZHANG G S, et al. Conventional and unconventional petroleum “orderly accumulation”: Concept and

- practical significance[J]. *Petroleum Exploration and Development*, 2014, 41(1): 14-30.
- [3] 邹才能,董大忠,王玉满,等.中国页岩气特征、挑战及前景(二)[J]. *石油勘探与开发*, 2016, 43(2): 166-178.
ZOU Caineng, DONG Dazhong, WANG Yuman, et al. Shale gas in China: Characteristics, challenges and prospects (II) [J]. *Petroleum Exploration and Development*, 2016, 43 (2): 166-178.
- [4] 赵文智,贾爱林,位云生,等.中国页岩气勘探开发进展及发展展望[J]. *中国石油勘探*, 2020, 25(1): 31-44.
ZHAO Wenzhi, JIA Ailin, WEI Yunsheng, et al. Progress in shale gas exploration in China and prospects for future development [J]. *China Petroleum Exploration*, 2020, 25(1): 31-44.
- [5] 廖东良.页岩气层“双甜点”评价方法及工程应用展望[J]. *石油钻探技术*, 2020, 48(4): 94-99.
LIAO Dongliang. Evaluation methods and engineering application of the feasibility of “Double Sweet Spots” in shale gas reservoirs[J]. *Petroleum Drilling Techniques*, 2020, 48(4): 94-99.
- [6] 何希鹏.四川盆地东部页岩气甜点评价体系与富集高产影响因素[J]. *天然气工业*, 2021, 41(1): 59-71.
HE Xipeng. Sweet spot evaluation system and enrichment and high yield influential factors of shale gas in Nanchuan area of eastern Sichuan Basin[J]. *Natural Gas Industry*, 2021, 41(1): 59-71.
- [7] MA X H. Enrichment laws and scale effective development of shale gas in the southern Sichuan Basin[J]. *Natural Gas Industry B*, 2019, 6(3): 240-249.
- [8] HASHMY K, ABUEITA S, BARNETT C, et al. Log-based identification of sweet spots for effective fracs in shale reservoirs [C]// Paper SPE-149278-MS presented at the Canadian Unconventional Resources Conference, Calgary, Alberta, Canada, November 2011.
- [9] 李小明,柳吉荣,吝文,等.荆门探区五峰组—龙马溪组页岩储层特征及甜点层段评价[J]. *煤田地质与勘探*, 2021, 49(6): 1-11.
LI Xiaoming, LIU Jirong, LIN Wen, et al. Characteristics of the shale gas reservoirs and evaluation of sweet spots in Wufeng Formation and Longmaxi Formation in Jingmen exploration area [J]. *Coal Geology & Exploration*, 2021, 49(6): 1-11.
- [10] 夏宏泉,赖俊,李高仁,等.基于测井资料的页岩油储层甜点预测[J]. *西南石油大学学报(自然科学版)*, 2021, 43(4): 199-207.
XIA Hongquan, LAI Jun, LI Gaoren, et al. Sweet spot prediction of shale oil reservoir based on logging Data[J]. *Journal of Southwest Petroleum University(Science & Technology Edition)*, 2021, 43(4): 199-207.
- [11] WU H Z, XIONG L, GE Z W, et al. Fine characterization and target window optimization of high-quality shale gas reservoirs in the Weiyuan area, Sichuan Basin[J]. *Natural Gas Industry B*, 2019, 6(5): 463-471.
- [12] WANG J B, FENG M G, YAN W, et al. Influence factors and evaluation methods for shale reservoir fracability in Jiaoshiba Area[J]. *Fault-Block Oil & Gas Field*, 2016, 23(2): 216-220.
- [13] 陈桂华,肖钢,徐强,等.页岩油气地质评价方法和流程[J]. *天然气工业*, 2012, 32(12): 1-5.
CHEN Guihua, XIAO Gang, XU Qiang, et al. A method and workflow for shale oil and gas geological evaluation[J]. *Natural Gas Industry*, 2012, 32(12): 1-5.
- [14] 马永生,蔡勋育,赵培荣.中国页岩气勘探开发理论认识与实践[J]. *石油勘探与开发*, 2018, 45(4): 561-574.
MA Yongsheng, CAI Xunyu, ZHAO Peirong. China's shale gas exploration and development: Understanding and practice[J]. *Petroleum Exploration and Development*, 2018, 45 (4): 561-574.
- [15] ALSHAKHS M, REZAEI R. Sweet-spot mapping through formation evaluation and property modelling using data from the Goldwyer Formation of the Barwire Terrace, Canning Basin[J]. *Petroleum*, 2019, 5(1): 13-29.
- [16] CHORN L, YARUS J, DEL ROSARIO-DAVIS S, et al. Identification of shale sweet spots using key property estimates from log analysis and geostatistics[C]// Paper URTEC-1580188-MS presented at the SPE/AAPG/SEG Unconventional Resources Technology Conference, Denver, Colorado, USA, August 2013.
- [17] ALDRICH J B, SEIDLE J P. Sweet spot identification and optimization in unconventional reservoirs[J]. *Mountain Geologist*, 2018, 52(3): 5-12.
- [18] PAN R F, GONG Q, YAN J, et al. Elements and gas enrichment laws of sweet spots in shale gas reservoir: A case study of the Longmaxi Fm in Changning Block, Sichuan Basin[J]. *Natural Gas Industry B*, 2016, 3(3): 195-201.
- [19] 张筠,葛祥,王志文.洛带气田遂宁组致密储层的快速产能评价[J]. *测井技术*, 2007, 191(4): 342-346.
ZHANG Yun, GE Xiang, WANG Zhiwen. Quick evaluation on productivity of compacted clastic reservoir in Suining Formation of Luodai Gas Field[J]. *Well Logging Technology*, 2007, 191(4): 342-346.
- [20] 周广照,李显明,黄斌,等.优化BP神经网络在川西上三叠统陆相页岩含气性预测中的应用[J]. *矿物岩石*, 2017, 37(3): 90-96.
ZHOU Guangzhao, LI Xianming, HUANG Bin, et al. Application of optimized BP Networks to gas content PR prediction of continental shale in Upper Triassic of Western Sichuan Basin[J]. *Mineralogy and Petrology*, 2017, 37(3): 90-96.
- [21] 王彬,汤勇,劭睿,等.利用神经网络优选页岩气有利开发区域[J]. *重庆科技学院学报(自然科学版)*, 2015, 17(6): 33-35.
WANG Bin, TANG Yong, MENG Rui, et al. Optimize the exploitation advantageous area of shale gas reservoirs by artificial neural network[J]. *Journal of Chongqing University of Science and Technology (Natural Sciences Edition)*, 2015, 17 (6): 33-35.
- [22] 汪敏,冯婷婷,闵帆,等.页岩气储层预测的多标签主动学习算法[J]. *计算机应用*, 2022, 42(2): 646-654.

- WANG Min, FENG Tingting, MIN Fan, et al. Multi-label active learning algorithm for shale gas reservoir prediction[J]. Journal of Computer Applications, 2022, 42(2): 646-654.
- [23] 钱辰, 杨少春, 许子君. 基于机器学习的页岩气甜点评价及其应用综述[C]//油气田勘探与开发国际会议论文集. 西安: 西安石油大学, 2019: 573-583.
- QIAN Chen, YANG Shaochun, XU Zijun. Review of evaluation of shale gas sweet spots and its application based on machine learning[C]// Proceedings of the International Conference on Oil and Gas Exploration and Development. Xi'an: Xi'an Shiyu University, 2019: 573-583.
- [24] 陈胜, 赵文智, 欧阳永林, 等. 利用地球物理综合预测方法识别页岩气储层甜点——以四川盆地长宁区块下志留统龙马溪组为例[J]. 天然气工业, 2017, 37(5): 20-30.
- CHEN Sheng, ZHAO Wenzhi, OUYANG Yonglin, et al. Comprehensive prediction of shale gas sweet spots based on geophysical properties: A case study of the Lower Silurian Longmaxi Fm in Changning block, Sichuan Basin[J]. Natural Gas Industry, 2017, 37(5): 20-30.
- [25] 刘伟, 梁兴, 姚秋昌, 等. 四川盆地昭通区块龙马溪组页岩气“甜点”预测方法及应用[J]. 石油地球物理勘探, 2018, 53(S2): 211-217.
- LIU Wei, LIANG Xing, YAO Qiuchang, et al. Shale gas sweet spot identification in Longmaxi, Sichuan Basin[J]. Oil Geophysical Prospecting, 2018, 53(S2): 211-217.
- [26] 万远飞, 秦启荣, 范宇, 等. 长宁背斜龙马溪组页岩裂缝发育特征及期次解析[J]. 特种油气藏, 2021, 28(1): 59-66.
- WAN Yuanfei, QIN Qirong, FAN Yu, et al. Development characteristics of shale fractures in longmaxi formation of changning anticline and the stage analysis[J]. Special Oil & Gas Reservoirs, 2021, 28(1): 59-66.
- [27] 葛勋, 郭彤楼, 马永生, 等. 四川盆地东南缘林滩场地区上奥陶统五峰组-龙马溪组页岩气储层甜点预测[J]. 石油与天然气地质, 2022, 43(3): 633-647.
- GE Xun, GUO Tonglou, MA Yongsheng, et al. Prediction of shale reservoir sweet spots of the Upper Ordovician Wufeng-Longmaxi Formations in Lintanchang area, southeastern margin of Sichuan Basin[J]. Oil & Gas Geology, 2022, 43(3): 633-647.
- [28] 朱维兵, 庞青松, 张朝界. 页岩气旋转式井壁取心器转向机构设计与优化[J]. 石油机械, 2021, 49(4): 51-57.
- ZHU Weibing, PANG Qingsong, ZHANG Chaojie. Design and optimization of steering device for shale gas rotary sidewall coring[J]. China Petroleum Machinery, 2021, 49(4): 51-57.
- [29] 彭寿昌, 查小军, 雷祥辉, 等. 吉木萨尔凹陷芦草沟组上“甜点”段页岩油储层演化特征及差异性评价[J]. 特种油气藏, 2021, 28(4): 30-38.
- PENG Shouchang, ZHA Xiaojun, LEI Xianghui, et al. Evolution characteristics and difference evaluation of shale oil reservoirs in the upper sweet spot interval of Lucaogou Formation in Jimusaer Sag[J]. Special Oil & Gas Reservoirs, 2021, 28(4): 30-38.
- [30] 曾义金, 陈作, 卞晓冰. 川东南深层页岩气分段压裂技术的突破与认识[J]. 天然气工业, 2016, 36(1): 61-67.
- ZENG Yijin, CHEN Zuo, BIAN Xiaobing. Breakthrough in staged fracturing technology for deep shale gas reservoirs in SE Sichuan Basin and its implications[J]. Natural Gas Industry, 2016, 36(1): 61-67.
- [31] LIANG X, WANG G C, XU Z Y, et al. Comprehensive evaluation technology for shale gas sweet spots in the complex marine mountains, South China: A case study from Zhaotong national shale gas demonstration zone[J]. Natural Gas Industry B, 2016, 3(1): 27-36.
- [32] 沈骋, 吴建发, 付永强, 等. 页岩气井长水平段压裂一体化动态评估——以长宁国家级页岩气示范区为例[J]. 天然气工业, 2022, 42(2): 123-132.
- SHEN Cheng, WU Jianfa, FU Yongqiang, et al. Integrated dynamic evaluation of long lateral fracturing in shale gas wells: A case study on the Changning National Shale Gas Demonstration Area[J]. Natural Gas Industry, 2022, 42(2): 123-132.
- [33] 杨光, 田伟志, 吕江, 等. 威远构造W202区块龙马溪组龙11亚段页岩气储集层岩石学特征[J]. 特种油气藏, 2021, 28(2): 34-40.
- YANG Guang, TIAN Weizhi, LYU Jiang, et al. Petrological characteristics of shale gas reservoirs in Long11 sub-member of Longmaxi Formation in W202 Block of Weiyuan structure[J]. Special Oil & Gas Reservoirs, 2021, 28(2): 34-40.
- [34] 陈永秀. 相关系数含义的理解[J]. 中国考试, 2011, 231(7): 15-19.
- CHEN Yongxiu. Methods for calculating the correlation coefficient[J]. Journal of China Examinations, 2011, 231(7): 15-19.
- [35] KAMNITUI N, GENEST C, JAWORSKI P, et al. On the size of the class of bivariate extreme-value copulas with a fixed value of Spearman's Rho or Kendall's Tau[J]. Journal of Mathematical Analysis and Applications, 2019, 472(1): 920-936.
- [36] 刘军. 基于叶枝比率的决策树构建算法[J]. 信息安全, 2013, 146(2): 9-12.
- LIU Jun. Algorithm of constructing decision tree based on the leaf and branch ratio[J]. Netinfo Security, 2013, 146(2): 9-12.
- [37] QUINLAN J R. Induction of decision trees[J]. Machine Learning, 1986, 1(1): 81-106.
- [38] PORTER B W, BAREISS R, HOLTE R C. Concept learning and heuristic classification in weak-theory domains[J]. Artificial Intelligence, 1990, 45(1-2): 229-263.
- [39] 陈敏雅, 石蕾. 基于SVM多分类决策树的研究综述[J]. 电脑知识与技术, 2008, 8(8): 1427-1429.
- CHEN Minya, SHI Lei. Study the survey in multi-class classifier based on SVM decision tree[J]. Computer Knowledge and Technology, 2008, 8(8): 1427-1429.
- [40] 曹正凤, 谢邦昌, 纪宏. 一种随机森林的混合算法[J]. 统计与决策, 2014, (4): 7-9.
- CAO Zhengfeng, XIE Bangchang, JI Hong. A hybrid algorithm of random forest[J]. Statistics & Decision, 2014, (4): 7-9.