

文章编号:1009-9603(2022)01-0168-07

DOI:10.13673/j.cnki.cn37-1359/te.2022.01.021

# 基于CBFS-CV算法的煤层气井 压裂效果主控因素识别

闵超<sup>1,2</sup>,张馨慧<sup>1,2</sup>,杨兆中<sup>3</sup>,李小刚<sup>3</sup>,代博仁<sup>1,2</sup>

(1.西南石油大学 理学院,四川 成都 610500; 2.西南石油大学 人工智能研究院,四川 成都 610500;  
3.西南石油大学 油气藏地质及开发工程国家重点实验室,四川 成都 610500)

**摘要:**准确识别煤层气井压裂效果的主控因素,进而有效指导重复压裂方案优化,是煤层气井提升重复压裂产能的关键。依托研究区块的地质及工程大数据,利用基于Copula互信息的特征选择和交叉验证算法(CBFS-CV)识别影响压裂效果的主控因素,并结合梯度提升回归模型进行产能预测检验,形成了一种改进的煤层气井压裂效果主控因素识别算法。该算法可有效减少冗余性特征且增大相关性,并确定最佳特征数目。结果表明:煤体结构、储层参数(含气量、含气饱和度和临界储比)和施工排量参数(最大施工排量)是影响研究区块压裂效果的3个主控因素,通过梯度提升回归模型验证CBFS-CV算法所识别出的主控因素的预测符合率达88%,证明了该算法的有效性。利用结果对该区块典型井进行主控因素分析,采用氮气泡沫解堵方案解决煤体结构差、煤粉堵塞等问题,现场施工后日产气量由288 m<sup>3</sup>/d增至805 m<sup>3</sup>/d,压裂效果明显改善。

**关键词:**煤层气;压裂效果;主控因素;Copula互信息;回归模型;大数据

中图分类号:TE319

文献标识码:A

## Identification of main controlling factors of fracturing performance in coalbed methane wells based on CBFS-CV algorithm

MIN Chao<sup>1,2</sup>, ZHANG Xinhui<sup>1,2</sup>, YANG Zhaozhong<sup>3</sup>, LI Xiaogang<sup>3</sup>, DAI Boren<sup>1,2</sup>

(1. School of Science, Southwest Petroleum University, Chengdu City, Sichuan Province, 610500, China; 2. Institute for Artificial Intelligence, Southwest Petroleum University, Chengdu City, Sichuan Province, 610500, China; 3. State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation, Southwest Petroleum University, Chengdu City, Sichuan Province, 610500, China)

**Abstract:** Accurately identifying the main controlling factors of the fracturing performance in coalbed methane (CBM) wells and then effectively guiding the optimization of repeated fracturing schemes are the keys to improving the repeated fracturing productivity in CBM wells. Relying on the geological and engineering data in the research block, the feature selection based on Copula mutual information and cross-validation (CBFS-CV) algorithm was adopted to identify the main controlling factors that affect the fracturing performance. In combination with the gradient boosting regression model for productivity prediction and inspection, an improved identification algorithm was formed for CBM wells. This algorithm can effectively reduce the redundant features and increase correlation, and thus determine the optimal number of features. The results show that the coal structure, reservoir parameters (gas content, gas saturation, critical reservoir ratio), and operation displacement parameters (maximum operation displacement) are the three main controlling factors that affect the fracturing performance in the research block. The gradient boosting regression model verifies that the prediction coincidence rate of the main controlling factors identified by the CBFS-CV algorithm reaches 88%, which proves the effectiveness of the algorithm. Moreover, the main controlling factors of the typical well in this block were analyzed based on the above results, and

收稿日期:2021-07-08。

作者简介:闵超(1982—),男,四川新都人,教授,博士,从事最优化方法与不确定理论在油气田开发中的应用研究。E-mail:minchao@swpu.edu.cn。

基金项目:国家科技重大专项“多层复杂煤体结构区煤储层直井压裂技术研究”(2016ZX05044-004-002),四川省科技计划项目“四川页岩气产业发展质量综合监测和评价技术研究与应用示范”(2020YFG0145)。

the plugging removal solution with nitrogen foam was applied to the problems of poor coal structure and coal powder plugging. After field operation, the daily gas production increased from 288 m<sup>3</sup>/d to 805 m<sup>3</sup>/d, and the fracturing performance was significantly improved.

**Key words:** coalbed methane; fracturing performance; main controlling factor; Copula-based mutual information; regression model; big data

中国煤层气储量巨大,埋深为2 000 m以内的煤层气资源占有量位列世界前三,是非常规能源的重要组成部分。渗透率低、地应力梯度分布不均以及普遍欠压实是中国煤层气储层的三大特点<sup>[1]</sup>,故需对储层进行改造以提升产气效果,进而实现工业化的开采。水力压裂可使煤层气井达到工业化开采的条件<sup>[2]</sup>,但由于压裂前期的选井技术不足,容易导致初次压裂规模不够和有效支撑裂缝延展性差等问题。此外,在生产过程中还存在压裂液残渣过多或破胶不彻底引起的煤粉堵塞问题<sup>[3]</sup>,使得煤层气井的初次压裂效果较差。重复压裂是为了提升单井产气量,使压裂增产的有效期变长,因此,提高重复压裂井的压裂效果是解决煤层气开发的关键问题之一。

目前针对煤层气井压裂效果主控因素识别的研究方法主要包括层次分析法<sup>[4]</sup>、数值模拟分析法<sup>[5-7]</sup>、单因素分析法<sup>[8]</sup>和灰色关联分析法<sup>[9-11]</sup>等,但以上方法都有一定的局限性。层次分析法的单因素评判需要人为打分,多因素情况下权重不易确定;数值模拟分析法依赖于数据本身,无法对其他客观条件进行模拟;单因素分析法忽略了特征指标之间的相互作用;灰色关联分析法对样本数据没有要求,但分析过程中涉及到最优值确定的主观性判断。而基于Copula互信息的特征选择算法(CBFS)是使用Copula互信息来度量因素与因素以及因素与类别之间的关系,精确地刻画了变量之间的关系,确定了各因素之间的相互依赖关系,有效地解决了互信息累加导致某些因素被高估的问题,增加了算法的稳定性。

基于大数据和特征选择<sup>[12]</sup>,笔者提出一种改进的煤层气井压裂效果主控因素的识别算法(CBFS-CV),该算法利用Copula互信息最大限度地减少了冗余性,增加了因素与指标的相关性,弥补了最佳特征数目不确定的缺点<sup>[13]</sup>。典型井案例分析表明,这对辅助煤层气井的重复压裂优化设计、提升压裂效果提供了支撑作用。

## 1 CBFS-CV算法

特征选择是指在众多特征中选择出有代表性

的特征,其可有效缓解维数灾难,降低学习任务难度,是机器学习过程中一个重要的步骤。特征选择模型主要包括3种,即Embedded,Filter和Wrapper。SNEHALIKA等提出的CBFS算法属于Filter模型评价标准的信息标准类<sup>[14]</sup>,该算法使用Copula互信息作为评价函数,通过最大相关性和最小冗余性标准输出特征子集,但难以确定特征数目 $k$ 的取值。针对这一问题,在基于Copula互信息的特征选择算法基础上加入了交叉验证算法,结合梯度提升回归模型进行预测检验,形成了一种改进特征选择算法(CBFS-CV)。现有的特征选择算法大多依赖于数据集,但数据集的噪声会影响所选择的特征子集,以基于Copula互信息作为评价函数,最小化被选特征和未被选特征之间的信息量,最大化类别和未被选特征之间的Copula互信息,采用Copula函数研究随机过程的相关性<sup>[15]</sup>,把联合分布函数和各自的边缘分布函数连接在一起,有效减少冗余性,增大其相关性。

Copula函数又称连接函数,最早由SKLAR提出即Sklar定理<sup>[16]</sup>:假设 $X_1, X_2, \dots, X_n$ 是 $n$ 个随机变量,其各自边缘分布函数为 $F_1(x_1), F_2(x_2), \dots, F_n(x_n)$ ,联合分布函数为 $F(x_1, x_2, \dots, x_n)$ ,则有一个Copula函数 $C(\cdot)$ ,使得:

$$F(x_1, x_2, \dots, x_n) = C(F_1(x_1), F_2(x_2), \dots, F_n(x_n)) \quad (1)$$

令 $c$ 表示Copula密度函数,对该联合分布函数求 $n$ 阶偏导数,则表达式为:

$$f(x_1, x_2, \dots, x_n) =$$

$$c(F_1(x_1), F_2(x_2), \dots, F_n(x_n)) \prod_i f_i(x_i) \quad (2)$$

互信息和信息熵均是信息论的重要概念。离散型随机变量 $X$ ,设其可能取值为 $\{x_1, x_2, \dots, x_n\}$ ,其信息熵表达式为:

$$H(X) = -\sum_{i=1}^n p(x_i) \log_2 p(x_i) \quad (3)$$

随机变量 $X, Y$ 的互信息表达式为:

$$I(X, Y) = \iint f(x, y) \log \frac{f(x, y)}{f(x)f(y)} dx dy \quad (4)$$

将(2)式代入(4)式后,可得随机变量 $X, Y$ 基于

Copula的互信息表达式为:

$$I(X, Y) = \iint c(u, v) \log c(u, v) du dv = -H(C(u, v)) \quad (5)$$

其中  $u, v$  分别代表边缘分布函数  $F(x), F(y)$ 。由(5)式可以看出,随机变量的联合分布类似于其所对应的 Copula 分布的负熵。

CBFS算法选择特征的标准为选择与类变量相关性最大,最小化被选特征与未被选特征之间的互信息,故被选择特征子集表达式为:

$$S = \arg \max_S I(S, Y_{\text{label}}) = \arg \max_{f_i \in F-S} I(f_i, Y_{\text{label}} | S) \quad (6)$$

当特征  $f_1, f_2, \dots, f_s$  被选择时,第  $s+1$  个特征被选择需满足条件:

$$f_{s+1} = \arg \max_{f_i \in (F-S)} [I_c(f_i; Y_{\text{label}}) - I_c(f_i; f_1; f_2; \dots; f_s)] \quad (7)$$

将(5)式代入(7)式,可得:

$$f_{s+1} = \arg \max_{f_i \in (F-S)} [-H(C(P(f_i), P(Y_{\text{label}}))) + H(C(P(f_i), P(f_1), \dots, P(f_s)))] \quad (8)$$

CBFS-CV算法的具体步骤为:①初始化特征子集  $S$ 。②初始化最相关特征,计算总特征集合  $F$  与类别变量  $Y_{\text{label}}$  的基于 Copula 的互信息值,并将最大值赋值于  $S$ 。③查找最小冗余特征子集,计算未被选特征  $f_i$  与被选特征  $(f_1, f_2, \dots, f_s)$  之间的基于 Copula 的互信息值,并将信息值存储于集合  $D$  中。④查找最大相关特征子集,计算未被选特征  $f_i$  与类别  $Y_{\text{label}}$  之间的基于 Copula 的互信息,并将信息值存储于集合  $R$  中。⑤对集合  $R$  与集合  $D$  作差,并将差值存储于集合  $E$  中,从  $E$  中选取互信息值最大的特征,合并到所选特征子集  $S$  中。⑥循环执行步骤③—⑤  $k$  次,得到特征子集  $S$ 。⑦进行交叉验证,在  $S$  中依次选择不同数量特征,组成特征子集,以准确度为评判标准对特征子集进行排序,确定准确度最高或最符合条件所对应的特征子集及  $k_{\text{best}}$  值。

## 2 煤层气井压裂效果主控因素识别

选取的 X 区块位于沁水盆地东南部,该地区总体上为一单斜构造<sup>[17]</sup>,区块面积达 763.21 km<sup>2</sup>,投产井数 1 013 口,主要目的煤层为二叠系山西组的 3 号煤层,动用地质储量为 228.29×10<sup>8</sup> m<sup>3</sup>,煤体主要以碎裂煤和碎粒煤为主,其次为原生煤和砂质泥岩。多

数煤层压裂井采用套管注入和清水压裂液,支撑剂使用 50/100 目细砂、20/40 目中砂以及 12/20 目粗砂。

煤层气井的压裂效果受多种因素共同影响,主要表现在工程因素和地质因素 2 个方面,工程因素包括射孔段厚度、压裂液参数、支撑剂参数和施工参数等,地质因素包含煤体结构、储层参数以及渗透率等。本文以日产气量作为煤层气井压裂效果的评价参数。

### 2.1 数据处理

数据源自于 X 区块,其中包含 200 口井的 30 个地质及工程因素数据,为确保研究结果的准确性,对数据进行预处理,其步骤包括:①缺失值处理,剔除全部数据缺失的样本井,对于样本井口数据部分缺失的现象,采用 KNN 算法进行填补。②异常值处理,绘制箱形图筛选出部分异于最小值—最大值区间的异常值,对其归零,采用处理缺失值的方法对异常值进行处理。③归约处理,删除工程因素中的支撑剂类型以及支撑剂设计总量等定量数据。④标准化处理,采用 Z-Score 值标准方法对数据进行标准化处理。综上,得到完整的 171 口井的 26 个地质及工程因素数据。

### 2.2 主控因素识别

针对众多的地质及工程因素,对其进行相关性分析<sup>[18]</sup>。利用 Pearson 相关系数评价因素之间的相关程度,因素  $f_i$  和  $f_j$  的 Pearson 相关系数表达式为:

$$r(f_i, f_j) = \frac{\text{Cov}(f_i, f_j)}{\sqrt{\text{Var}(f_i) \text{Var}(f_j)}} \quad (9)$$

依据相关性分析的基本理论,确定各因素之间的相关系数。由计算结果可知,压裂液参数中的前置液量与携砂液砂比、前置液比、压裂液施工总量以及最大施工排量的相关系数分别为 0.75, 0.88, 0.69 和 0.70,表明该因素与其他因素均存在很大相关性,为避免多重共线性问题,将该因素剔除。同时压裂液设计总量与压裂液施工总量的相关系数高达 0.86,需剔除两者之一,本文选择剔除压裂液设计总量,最终剩余 12 个工程因素和 12 个地质因素。

初步剔除共线性因素后,采用 CBFS 算法对剩余因素进行主控因素识别,首先对数据进行无监督的离散化,再计算地质及工程因素之间的 Copula 互信息,并构建核数为 3 的并行集合,在查找最大相关特征和最小冗余特征子集时实现对小块的并行计算,提高计算效率。综上所述,利用 CBFS 算法对 X 区块煤层气井压裂效果的影响因素排序如表 1 所示。

基于 CBFS 算法,以准确率为交叉验证算法的



表1 X区块煤层气井压裂效果影响因素排序  
Table1 Sequence of factors affecting fracturing performance in CBM wells in Block X

序号	因素	序号	因素
1	煤体结构	13	支撑剂施工总量
2	含气饱和度	14	最高砂比
3	含气量	15	平均砂比
4	临储比	16	最大水平主应力
5	最大施工排量	17	射孔段厚度
6	垂直应力	18	携砂液液量
7	最小水平主应力	19	储层压力
8	前置液比	20	渗透率
9	压裂液施工总量	21	携砂液砂比
10	破裂压力	22	加砂强度
11	水头标高	23	施工排量
12	顶板封盖能力	24	支撑剂设计总量

评判标准,确定地质及工程因素的最佳特征数目。由地质及工程因素的交叉验证结果(图1)可知,最佳特征数目为3~6时,压裂效果预测准确率均超过80%,为保证利用筛选出的主控因素对压裂效果预测的准确率以及避免非主控因素造成的误差,在CBFS算法中k取值为5,得到影响煤层气井压裂效果的主控因素依次为煤体结构、含气饱和度、含气量、临储比和最大施工排量。

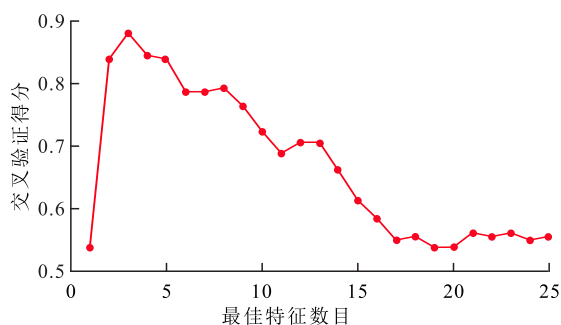


图1 地质及工程因素的交叉验证结果

Fig.1 Cross-validation results of geological and engineering factors

煤体结构受构造运动作用后,可压性和力学性质都会存在较大差异,煤体的破坏程度越大,压裂效果就越差<sup>[19]</sup>。煤体结构类型中,原生结构煤与日产气量呈正相关,碎裂煤和碎粒煤与日产气量呈负相关。

含气饱和度、含气量和临储比均为储层参数,其中含气量和含气饱和度为物性参数,临储比为动力参数,都是影响日产气量的重要因素。研究区块的含气量越大,煤层气井的日产气量越高,储层参数均与日产气量呈正相关。

最大施工排量为施工排量参数。在现场作业中,施工排量必须大于地层吸液能力<sup>[9]</sup>,如果施工排量过小,无法在煤层中压开裂缝或产生较长的裂缝,导致地层吸液能力变差,压裂效率降低;如果施工排量过大,会导致井口施工压力过高等不利影响,故必须将最大施工排量控制在一个合理的范围内,使得压裂效果更好。

### 3 模型验证及工程验证

从机器学习模型和工程实践2个方面验证CBFS-CV算法的准确性和有效性。

#### 3.1 机器学习模型验证

煤层气井产能预测的主要方法有数值模拟方法、数理统计方法、物质平衡法等,以上方法对数据要求较为严格且需要很强的数学理论知识,方法实用性不强且无法具体分析各种因素对产能的影响<sup>[20]</sup>。为验证CBFS-CV算法所选因素的准确性,利用机器学习算法中的随机森林回归<sup>[21]</sup>及梯度提升回归<sup>[22-23]</sup>模型对煤层气井的日产气量进行预测,并与实际值相比较,确定能够真实有效反映日产气量的回归模型。

采用评价指标中的均方误差(MSE)和决定系数(R<sup>2</sup>)对回归模型进行评价,表达式分别为:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (10)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (11)$$

选取20%的煤层气井地质及工程因素数据进行测试,由2种回归模型的实验结果(表2)和测试样本的预测值与实际值对比(图2)可知,梯度提升回归和随机森林回归模型的决定系数均达到0.8以上,且梯度提升回归模型决定系数高于随机森林回归模型,均方误差偏小,其拟合结果更贴近于真实值,表明依据煤体结构、储层参数以及最大施工排量3个主控因素对产能的解释程度可达88%,也证明了梯度提升回归模型具有良好的预测能力。

表2 2种回归模型的实验结果  
Table2 Experimental results of two regression models

回归模型	评价指标	
	R <sup>2</sup>	MSE
梯度提升回归	0.88	572.63
随机森林回归	0.84	818.07

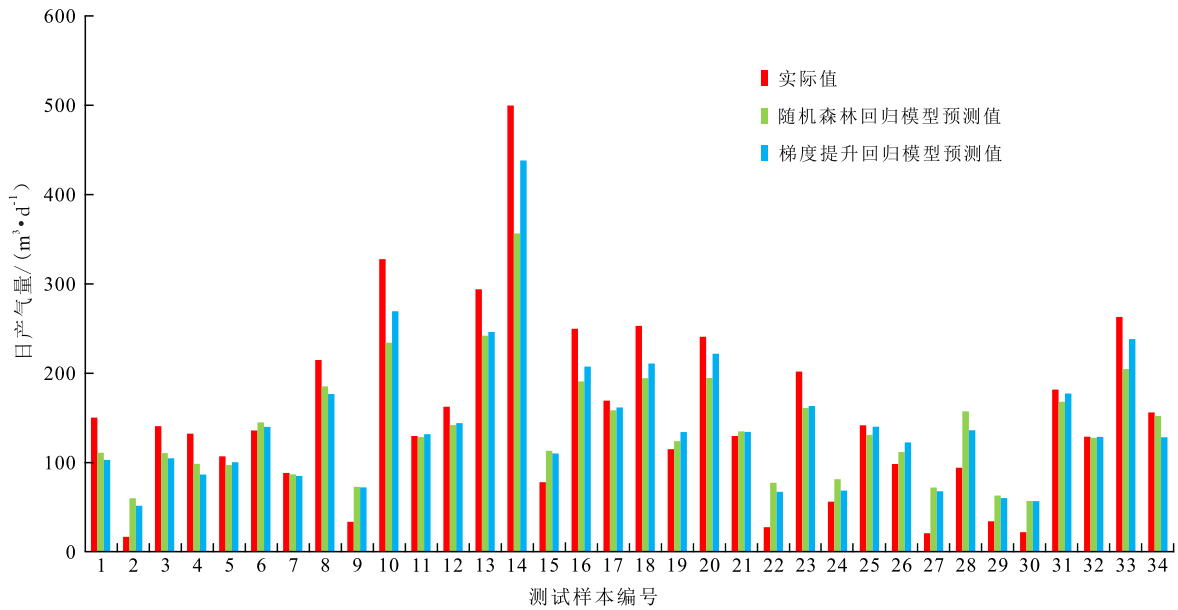


图2 2种回归模型预测值与实际值对比

Fig.2 Comparison of predicted values and actual values of two regression models

### 3.2 工程验证

选取研究区块某一典型井TS-X,该井初次压裂后,日产气量为 $288 \text{ m}^3/\text{d}$ ,日产水量为 $2.21 \text{ m}^3/\text{d}$ ,初次压裂效果不理想。依据研究结果,以煤体结构、储层参数和施工排量参数为出发点,结合压裂资料对该井进一步分析得出:①TS-X井煤体结构为碎粒煤且煤层厚度薄,含气量、含气饱和度和临储比均低于煤层气井压裂效果好的平均值。②目前,TS-X井储层日产气量小于 $500 \text{ m}^3/\text{d}$ ,但投产以来日产气量有大于 $500 \text{ m}^3/\text{d}$ 的情况,且其相邻井日产气量均大于 $500 \text{ m}^3/\text{d}$ ,说明该井有进一步进行重复压裂、提高产能的余地。③根据前期资料分析,研究区块的压裂最大施工排量控制在 $7.0\sim 8.5 \text{ m}^3/\text{min}$ 为宜,TS-X井初次压裂最大施工排量为 $8.03 \text{ m}^3/\text{min}$ ,还有进一步提升余地。④TS-X井压裂施工曲线形态为波动型,且该井日产气量和日产水量都较小,说明可能由于前期抽排速度过大或者中途停抽,导致煤粉

堵塞了渗流通道,满足煤层气井重复压裂的条件。

煤层含气量、含气饱和度及临储比属不可控因素,无有效改造手段,选择高含气量区域是获得高产气量的物质基础,且TS-X井的最大施工排量也在合理范围内,故需对煤体结构差、煤粉堵塞渗流通道等问题提出改造技术来进一步改善压裂效果。因此,综合分析结果,对TS-X井推荐采用氮气泡沫解堵工艺方案来进行重复压裂,该方案可以增加井眼内能量,干扰滞留煤粉,冲开堵塞分支从而提高日产气量。

采用改进方案后,TS-X井重复压裂施工曲线形态变为先升后降型(图3),表明地层吃砂能力强且加砂容易,更易形成高导流能力支撑裂缝,有利于后期排采,该井的含气量和含气饱和度均已达到了压裂效果好的煤层气井的平均值且见到了较好的增产效果,日产气量由 $288 \text{ m}^3/\text{d}$ 增至 $805 \text{ m}^3/\text{d}$ ,压裂效果明显改善,达到了工业化开采的条件。

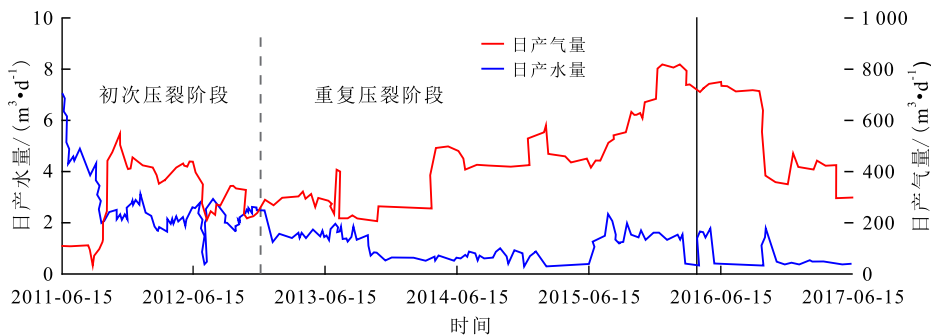


图3 TS-X井重复压裂前后的产气效果对比

Fig.3 Comparison of gas production performance before and after repeated fracturing in Well TS-X

## 4 结论

基于CBFS-CV算法,结合X区块的地质及工程数据,利用梯度提升回归模型进行预测检验,结果表明依据CBFS-CV算法筛选出的因素对产能的解释程度达到88%,并验证了该算法的可行性,为煤层气井压裂效果主控因素的识别提供了一种改进的算法。

煤体结构、储层参数和施工排量参数是影响煤层气井压裂效果的3个主控因素。储层参数和煤体结构类型中的原生结构煤与日产气量呈正相关;煤体结构类型中的碎裂煤和碎粒煤与日产气量呈负相关;最大施工排量控制在合理范围内,压裂效果更好。

通过工程验证方法,对初次压裂效果不好的典型井采用氮气泡沫解堵工艺方案进行重复压裂,解决了煤体结构差、煤粉堵塞渗流通道等问题,压裂效果明显改善,达到了工业化开采的条件。煤层气井的压裂效果受到多种因素共同作用,下一步将结合人工智能算法深入研究,形成一种适合多区块、多维度的煤层气井压裂效果影响因素识别的方法。

### 符号解释

$c$ ——Copula密度函数;  
 $C(\cdot)$ ——Copula函数;  
 $Cov(f_i, f_j)$ —— $f_i$ 和 $f_j$ 的协方差;  
 $D$ ——存储未被选特征与被选特征之间冗余信息值的集合;  
 $E$ ——存储对未被选特征的评价指标的集合;  
 $f_i, f_j$ ——未被选特征;  
 $f_1, f_2, \dots, f_s$ ——被选特征;  
 $f(x), f(y), f(x, y)$ ——随机变量 $X$ 和 $Y$ 的概率分布函数;  
 $F$ ——总特征集合;  
 $F(x_1, x_2, \dots, x_n)$ ——随机变量 $X$ 的联合分布函数;  
 $F_1(x_1), F_2(x_2), \dots, F_n(x_n)$ ——随机变量 $X$ 的边缘分布函数;  
 $F(y)$ ——随机变量 $Y$ 的分布函数;  
 $H(X)$ ——随机变量 $X$ 的信息熵;  
 $i$ ——特征编号;  
 $I(X, Y)$ ——随机变量 $X$ 和 $Y$ 的互信息;  
 $I(f_s, Y_{\text{label}} | \dot{S})$ ——在已知未被选特征集合 $\dot{S}$ 的情况下,类别变量 $Y_{\text{label}}$ 与被选特征 $f_s$ 之间的互信息;  
 $I_c(f_i; Y_{\text{label}})$ ——未被选特征 $f_i$ 和类别变量 $Y_{\text{label}}$ 之间基于

Copula的互信息;

$I_c(f_i; f_1; f_2; \dots; f_s)$ ——未被选特征 $f_i$ 和被选特征 $f_1; f_2; \dots; f_s$ 之间基于Copula的互信息;

$k$ ——特征数目;

$k_{\text{best}}$ ——最佳特征数目;

$MSE$ ——均方误差;

$n$ ——随机变量的维数;

$p(x_i)$ ——随机变量 $X$ 为 $x_i$ 的概率;

$P(f_i)$ ——未被选特征 $f_i$ 的边缘分布函数;

$r(f_i, f_j)$ —— $f_i$ 和 $f_j$ 的Pearson相关系数;

$R$ ——存储未被选特征与类别之间互信息的集合;

$R^2$ ——决定系数;

$s$ ——特征子集中元素个数;

$S$ ——特征子集;

$\dot{S}$ —— $S$ 的补集;

$Var(f_i), Var(f_j)$ —— $f_i, f_j$ 的方差;

$x_1, x_2, \dots, x_n$ ——随机变量 $X_1, X_2, \dots, X_n$ 对应分布函数和密度函数的自变量;

$X_1, X_2, \dots, X_n$ ——随机变量;

$y_i$ ——第 $i$ 口井的日产气量,  $\text{m}^3/\text{d}$ ;

$\bar{y}$ —— $n$ 口井的日产气量平均值,  $\text{m}^3/\text{d}$ ;

$\hat{y}_i$ —— $y_i$ 对应的预测日产气量,  $\text{m}^3/\text{d}$ ;

$Y_{\text{label}}$ ——类别变量。

### 参考文献

- [1] 郝宝华.我国煤层气储层特点及主控地质因素[J].山西煤炭管理干部学院学报,2010,23(1):112-113.  
 XI Baohua.Characteristics and main controlling geological factors of coalbed methane reservoir in China[J].Journal of Shanxi Coal-Mining Administrators College,2010,23(1):112-113.
- [2] 赵武鹏,刘春春,申兴伟,等.郑庄区块煤层气低产井增产技术研究[J].石油钻采工艺,2017,39(4):491-494.  
 ZHAO Wupeng, LIU Chunchun, SHEN Xingwei, et al.Study on the stimulation technologies for low-yield CBM wells in Zheng-zhuang Block[J].Oil Drilling & Production Technology,2017,39(4):491-494.
- [3] 倪小明,杨艳辉,叶建平.单一煤层重复水力压裂综合选井研究[J].煤矿安全,2016,47(2):170-174.  
 NI Xiaoming, YANG Yanhui, YE Jianping.Study on comprehensively selecting wells by repeated hydraulic fracturing in single coal seam[J].Safety in Coal Mines,2016,47(2):170-174.
- [4] 陆亚秋,王进,曹梦茜.基于改进的层次分析法的页岩气开发选区评价方法[J].油气藏评价与开发,2021,11(2):204-211.  
 LU Yaqiu, WANG Jin, CAO Mengxi.Evaluation method of shale gas development area selection based on improved analytic hierarchy process[J].Reservoir Evaluation and Development,2021,11(2):204-211.
- [5] ZHANG Jingchen.Numerical simulation of hydraulic fracturing coalbed methane reservoir[J].Fuel,2014,136:57-61.

- [6] ZHAO Jin, ZHAO Jinzhou, HU Yongquan, et al. Numerical simulation of multistage fracturing optimization and application in coalbed methane horizontal wells [J]. *Engineering Fracture Mechanics*, 2020, 223: 106738.
- [7] 杨焦生, 赵洋, 王玫珠, 等. 沁水盆地南部煤层气压裂、排采关键技术研究 [J]. *中国矿业大学学报*, 2017, 46(1): 131-138, 154.  
YANG Jiaosheng, ZHAO Yang, WANG Meizhu, et al. Study of key technologies on coalbed methane fracturing and drainage in the southern Qinshui basin [J]. *Journal of China University of Mining & Technology*, 2017, 46(1): 131-138, 154.
- [8] 冯绪兴, 谭学斌, 郝少伟, 等. 余吾矿煤层气井产能主控因素分析 [J]. *能源与环境*, 2020, 42(11): 77-81.  
FENG Xuxing, TAN Xuebin, HAO Shaowei, et al. Analysis of main controlling factors of CBM well productivity in Yuwu Coal Mine [J]. *China Energy and Environmental Protection*, 2020, 42(11): 77-81.
- [9] 杨兆中, 熊俊雅, 刘俊, 等. 基于 Apriori 关联分析的煤层气压裂效果主控因素识别 [J]. *油气藏评价与开发*, 2020, 10(4): 63-69.  
YANG Zhaozhong, XIONG Junya, LIU Jun, et al. Identification of main controlling factors on performance of CBM well fracturing based on Apriori association analysis [J]. *Reservoir Evaluation and Development*, 2020, 10(4): 63-69.
- [10] 杨显成, 蒋有录, 杨昕睿, 等. 煤层气含量的主控因素——以卡拉哈里盆地 XX 区块为例 [J]. *油气地质与采收率*, 2018, 25(1): 56-60.  
YANG Xiancheng, JIANG Youlu, YANG Xinrui, et al. Study on main controlling factors of the coalbed methane content: A case study of Block XX in Kgalahari Basin [J]. *Petroleum Geology and Recovery Efficiency*, 2018, 25(1): 56-60.
- [11] 李宇, 张亚飞, 刘广景, 等. 基于组合权重的煤层气井压裂后产量影响因素分析 [J]. *特种油气藏*, 2020, 27(3): 115-120.  
LI Yu, ZHANG Yafei, LIU Guangjing, et al. Production sensitivity analysis of fractured CBM well based on combination weight [J]. *Special Oil & Gas Reservoirs*, 2020, 27(3): 115-120.
- [12] 李郅琴, 杜建强, 聂斌, 等. 特征选择方法综述 [J]. *计算机工程与应用*, 2019, 55(24): 10-19.  
LI Zhiqin, DU Jianqiang, NIE Bin, et al. Summary of feature selection methods [J]. *Computer Engineering and Applications*, 2019, 55(24): 10-19.
- [13] 张军华, 任雄风, 赵杰, 等. 基于交叉验证支持向量机储层预测方法及应用 [J]. *科学技术与工程*, 2020, 20(13): 5 052-5 057.  
ZHANG Junhua, REN Xiongfeng, ZHAO Jie, et al. Reservoir prediction method and its application of support vector machine based on cross validation [J]. *Science Technology and Engineering*, 2020, 20(13): 5 052-5 057.
- [14] SNEHALIKA Lall, DEBAJYOTI Sinha, ABHIK Ghosh, et al. Stable feature selection using copula based mutual information [J]. *Pattern Recognition*, 2021, 112(1): 107697.
- [15] 赵梦婷. 高斯 Copula 过程及其应用 [D]. 武汉: 华中科技大学, 2016.  
ZHAO Mengting. Gaussian Copula processes and its applications [D]. Wuhan: Huazhong University of Science & Technology, 2016.
- [16] JAWORSKI P, HARDLE W K, RYCHLIK T, et al. Copula theory and its applications [M]. Berlin: Springer, 2010.
- [17] 陈哲, 王中鹏, 李尚儒. 萍乐坳陷曲页 1 井乐平组页岩气和煤层气地质条件分析 [J]. *油气地质与采收率*, 2019, 26(4): 50-55.  
CHEN Zhe, WANG Zhongpeng, LI Shangru. Geological condition analysis of shale gas and coalbed methane of Leping Formation in Well Quye1 of Pingle Depression [J]. *Petroleum Geology and Recovery Efficiency*, 2019, 26(4): 50-55.
- [18] 韩晟, 韩坚舟, 赵璇, 等. 距离权重改进的 Pearson 相关系数及应用 [J]. *石油地球物理勘探*, 2019, 54(6): 1 363-1 370.  
HAN Sheng, HAN Jianzhou, ZHAO Xuan, et al. A Pearson correlation coefficient improved by spatial weight [J]. *Oil Geophysical Prospecting*, 2019, 54(6): 1 363-1 370.
- [19] 周劼, 杨经栋, 汪勇, 等. 沁水盆地南部煤层气单井产量影响因素敏感性分析 [J]. *中国煤层气*, 2016, 13(5): 22-26.  
ZHOU Rui, YANG Jingdong, WANG Yong, et al. Sensitivity analysis on influencing factors of CBM single well production in south Qinshui Basin [J]. *China Coalbed Methane*, 2016, 13(5): 22-26.
- [20] 韩勇. 煤层气井产能预测及提高产能方法 [J]. *山西化工*, 2018, 38(6): 114-116.  
HAN Yong. Coalbed methane well productivity prediction and productivity enhancement method [J]. *Shanxi Chemical Industry*, 2018, 38(6): 114-116.
- [21] 朱庆忠, 胡秋嘉, 杜海为, 等. 基于随机森林算法的煤层气直井产气量模型 [J]. *煤炭学报*, 2020, 45(8): 2 846-2 855.  
ZHU Qingzhong, HU Qiujia, DU Haiwei, et al. A gas production model of vertical coalbed methane well based on random forest algorithm [J]. *Journal of China Coal Society*, 2020, 45(8): 2 846-2 855.
- [22] SAMADI Seyed Hashem, GHOBADIAN Barat, NOSRATI Mohsen. Prediction of higher heating value of biomass materials based on proximate analysis using gradient boosted regression trees method [J]. *Energy Sources, Part A: Recovery, Utilization and Environmental Effects*, 2021, 43(6): 672-681.
- [23] 侯春华. 基于长短期记忆神经网络的油田新井产油量预测方法 [J]. *油气地质与采收率*, 2019, 26(3): 105-110.  
HOU Chunhua. New well oil production forecast method based on long-term and short-term memory neural network [J]. *Petroleum Geology and Recovery Efficiency*, 2019, 26(3): 105-110.

编辑 何青芳