

煤层气高产井和低产井预判方法研究

何金蔚*, 郭大立

西南石油大学理学院, 四川 成都

收稿日期: 2023年5月6日; 录用日期: 2023年5月26日; 发布日期: 2023年6月7日

摘要

煤层气单井产量低一直是制约煤层气开发利用的主要技术瓶颈。为实现煤层气效益开发和可持续发展, 急需对煤层气低产井、高产井进行准确预测, 从而进行科学决策和技术革新。根据单井产气量, 将煤层气井划分为高产井和低产井, 对于影响煤层产气量的地质工程参数, 先利用灰色关联分析法进行重要性排序, 再通过交叉验证法(CV)进行优选; 以典型的DJ区块为例, 影响煤层产井气量的主控因素为声波时差、体积密度、自然伽马、深侧向电阻率、总液量、总砂量和施工排量。在此基础上, 优选并引入AdaBoost算法, 提出高产井和低产井的预判方法, 预测符合率达到93.33%。最后, 对不同方法、不同样本个数、不同划分标准、不同区块进行讨论和对比分析, 结果表明AdaBoost算法效果明显优于BP神经网络等传统方法, 提出的预判方法适用于不同区块, 具有一般性, 对实现煤层气效益开发和可持续发展具有指导意义。

关键词

煤层气, 高产井, 低产井, 主控因素, Adaboost算法, 预判方法

A Pre-Judge Method for High Production Wells and Low Production Wells of Coalbed Methane

Jinwei He*, Dali Guo

School of Science, Southwest Petroleum University, Chengdu Sichuan

Received: May 6th, 2023; accepted: May 26th, 2023; published: Jun. 7th, 2023

Abstract

The low production of single well of coalbed methane (CBM) has always been the main limitation.

*第一作者。

Propose methods for predicting CBM high production wells and low production wells to achieve beneficial development and sustainable development of CBM, so as to carry out scientific decision-making and technological innovation. First of all, the article divided all wells into high production wells and low-production wells according to the average daily CBM production of each well. For the geological and engineering parameters affecting CBM production, the importance ranking was carried out by grey correlation analysis method, and then the main control factors were optimized by cross validation (CV). Taking the typical DJ block as an example, the main control factors affecting CBM production were AC, DEN, GA, RD, total fracturing fluid volume, total sand volume and construction displacement. On this basis, AdaBoost algorithm was optimized and introduced to put forward the pre-judge method of high-production wells and low-production wells, and the pre-judge coincidence rate reached 93.33%. Finally, different methods, different sample numbers, different division standards of production and different blocks were discussed and compared. The results showed that the effect of AdaBoost algorithm was obviously better than traditional methods such as BP neural network. The proposed prediction method is applicable to different blocks and has generality. It has guiding significance for realizing the benefit development and sustainable development of CBM.

Keywords

Coalbed Methane, High Production Wells, Low Production Wells, Main Controlling Factors, Adaboost Algorithm, Pre-Judge Method

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

1. 引言

非常规天然气肩负优化能源结构、补充低碳清洁能源、实现“双碳”目标的使命,成为近年来的研究热点[1]。截至2020年底,全国地面开发煤层气累计钻井21217口,投产井12880口,煤层气累计产气量484亿方,2020年产气量83亿方,主要来自沁水盆地和鄂尔多斯盆地[2];其中,直井单井日产气量870方,日产气量低于500方的低产井占比高达56%[3]。事实上,单井产量低、资源动用率低、产能转化率低[4]的问题一直是制约煤层气开发利用的主要技术瓶颈。为实现煤层气效益开发和可持续发展,急需对煤层气低产井、高产井进行准确预测[5],从而进行科学决策和技术革新。

目前,国内外学者对煤层气预测评价[6]逐步有了认识。Christian Oberwinkler等[7]利用神经网络模型进行了选井选层研究。Robert F.Shelley等针对较大规模的数据特征,认识到了影响压裂效果的特征参数优选更能够准确预测。Shahab Mohsagheh等运用虚拟智能系统来优选压裂改造技术,从而进行选井选层综合预测。虽然煤层气产量预测算法较多,但每个算法都有明显缺点。神经网络算法预测效果高,但容易出现过度拟合现象;支持向量机算法难以解决大规模样本问题和多分类问题;决策树算法对于大样本分类问题的预测精度低。

为提高煤层气产量预测精度,从而得到更优的煤层气高产井和低产井的预判方法,这里引入AdaBoost算法。通过灰色关联分析法和交叉验证法结合筛选出的主控因素,建立煤层产气量与自然伽马、体积密度和等主要参数的预测模型,能有效提高煤层气高低产井识别精度,进行高产井和低产井的预判,能在投产前给出生产意见,对实际生产具有指导意义。

2. 高产井和低产井主控因素

2.1. 数据来源

研究的煤层气数据来源于 DJ 区块, 该区块位于鄂尔多斯盆地晋西挠褶带南端与伊陕斜坡东南缘, 面积为 5784 km², 总体呈西倾单斜构造。区块主要煤层为山西组 5 号煤和太原组 8 号煤, 煤层气储层埋藏深度 300 m~2520 m, 煤层厚度 2.0 m~9.8 m, 平均渗透率为 0.40 mD, 区块煤岩层数多且分布稳定, 连续性较好, 总体具有较好的储集性能。

针对 DJ 区块的煤层气现场数据, 统计了连续生产一年的 127 井层的数据样本, 主要包含地质工程参数。由于数据存在异常和大量缺失, 因此对原始数据进行了插值和去异常值等数据预处理, 最终得到完整的 80 井层的数据样本(表 1), 其中产气量为生产一年的平均日产气量。影响煤层产量的因素多而复杂, 例如含气量和渗透率是影响煤层产气量的主要因素, 但仅有区块的范围和平均值, 而单井并没有具体的数值; 而且煤层渗透率是动态变化的, 无法取到明确的数值, 因此本研究不对含气量和渗透率做讨论。本研究利用常规测井参数刻画地质情况, 工程参数主要选取压裂参数和排采参数, 其中地质参数有 8 个, 工程参数有 6 个(图 1)。在排采参数的选取中井底压力、套压等参数为测量所得, 泵挂深度、泵径由地层条件决定, 而冲程为一定排量下抽油机地最大能力, 都属于不可控的, 唯一能够作为操作变量改变的只有冲次, 所以选择冲次作为排采参数。预处理后的数据能够全面准确地反映该区块的生产情况。

Table 1. DJ block average daily gas production and geological engineering parameters (part)
表 1. DJ 区块日均产气量及地质工程参数(部分)

井号	GA/API	DEN /(g/cm^3)	AC /($\mu\text{s}/\text{m}$)	RD /($\Omega\cdot\text{m}$)	CNL/%	SP/mV	...	总砂量 /(m^3/min)	总液量 /(m^3)	产量 /(m^3/d)
X1	145.76	2.67	178.94	528.73	21.29	56.19	...	39.27	1678.47	987.65
X2	169.37	1.74	369.27	360.78	18.48	55.17	...	38.51	1012.57	1067.42
X3	67.53		183.24		33.17	49.19	...	41.57	1345.67	2354.87
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
X79	48.84	1.27	297.05	271.37	33.47	37.46	...	39.97	1445.03	3082.54
X80	68.93	2.01	389.14	347.71	38.51	61.13	...	39.56	1850.74	1785.62

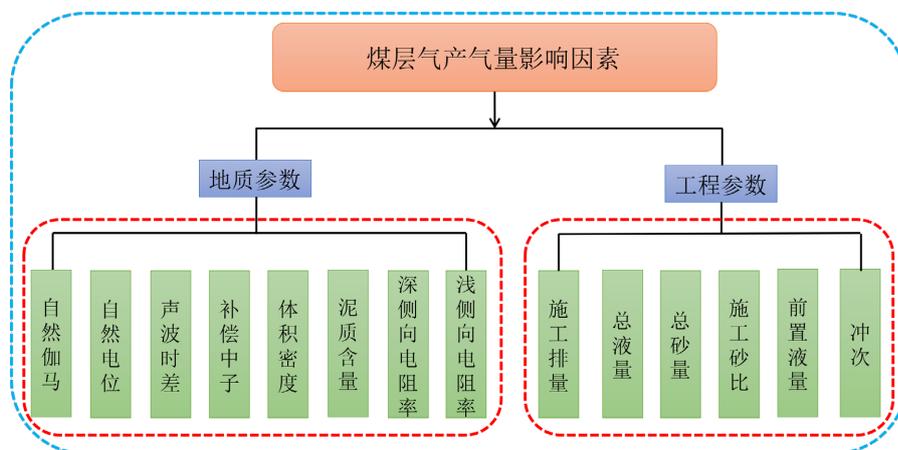


Figure 1. Influencing factors of CBM production
图 1. 煤层气产气量影响因素

2.2. 灰色关联分析

影响煤层气产量的因素很多且错综复杂, 不同的地质构造和施工方法是造成煤层产气量差异的主要原因。为了有效研究地质和工程参数对煤层产气量的影响, 首先通过灰色关联分析法[8]计算各地质工程参数与煤层产气量的灰色关联度。

灰色关联分析法能反应因素之间的趋势。以煤层产气量为母序列, 以煤层气井地质工程参数为子序列, 计算煤层产气量与各地质工程参数间的灰色关联度, 关联度越高, 变化趋势越接近。对灰色关联度进行排序得到结果, 见图 2。

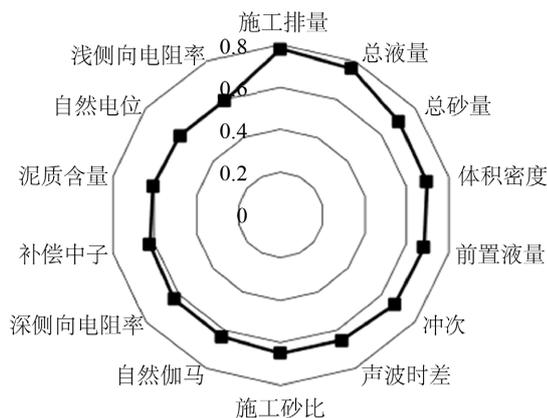


Figure 2. Grey Relational Degree Sorting Results
图 2. 灰色关联度排序结果

由图中的灰色关联度排序结果可知, 在排名靠前的因素中工程参数占比较大, 这是由于压裂技术和排采工艺本身是为煤层气井增产的改造技术, 因此关联度可能较高。但在本研究关于煤层气高产井和低产井的预判方法研究中, 需同时参考地质工程参数的影响, 所以还需要对初选后的参数进行优选。

2.3. 交叉验证

K 折交叉验证法[9]是常用的检验模型参数的方法, 将数据集 A 分割为 K 个互斥子样本, 保留一个子样本进行验证, 剩余 K-1 个子样本用来训练, 重复 K 次得到选择参数个数。

为解决灰色关联分析法得到的主控因素不全面的问题, 利用交叉验证法分别对地质参数和工程参数进行优选。

通过交叉验证的结果(图 3)得到当地质参数取前 4, 工程参数取前 3 时模型达到最优。因此, 最终得到的主控因素为: 体积密度、声波时差、自然伽马、深侧向电阻率、施工排量、总液量和加砂量。

3. 高产井和低产井预判方法

3.1. 原理和方法

以 DJ 区块的主控因素为基础, 研究煤层气高产井和低产井预判方法。在尝试对比多种方法后, 优选出 AdaBoost 算法, 具体流程见图 4。关于各种方法的讨论在后面阐述。

Boosting 算法[10]是freund 和 schapire 在 1990 年提出的算法。AdaBoost 算法在 Boosting 算法中具有代表性。其核心思想是提高错误率高的弱分类器的权重, 再通过不断迭代更新权值, 训练出 T 个弱分类器, 最后通过线性组合得到强分类器。AdaBoost 算法优点在于分类快且不出现过拟合。

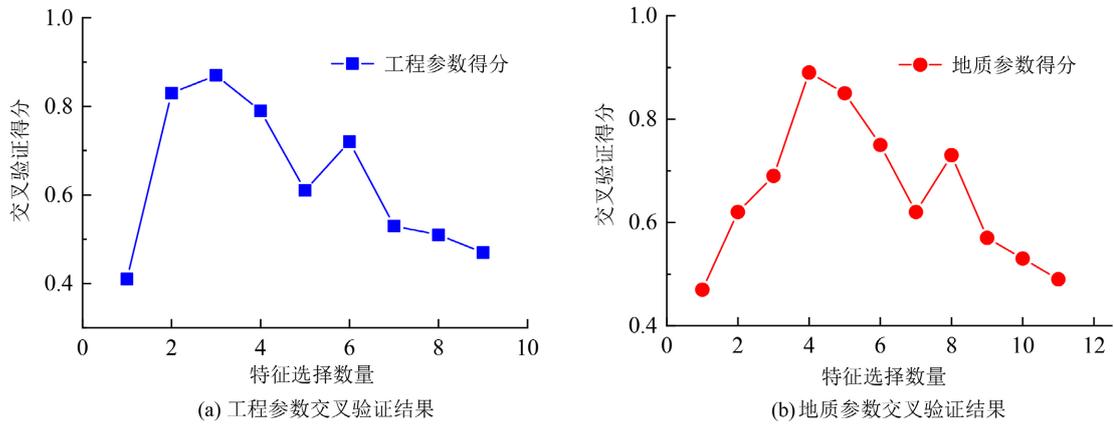


Figure 3. The result of CV
图 3. 交叉验证结果

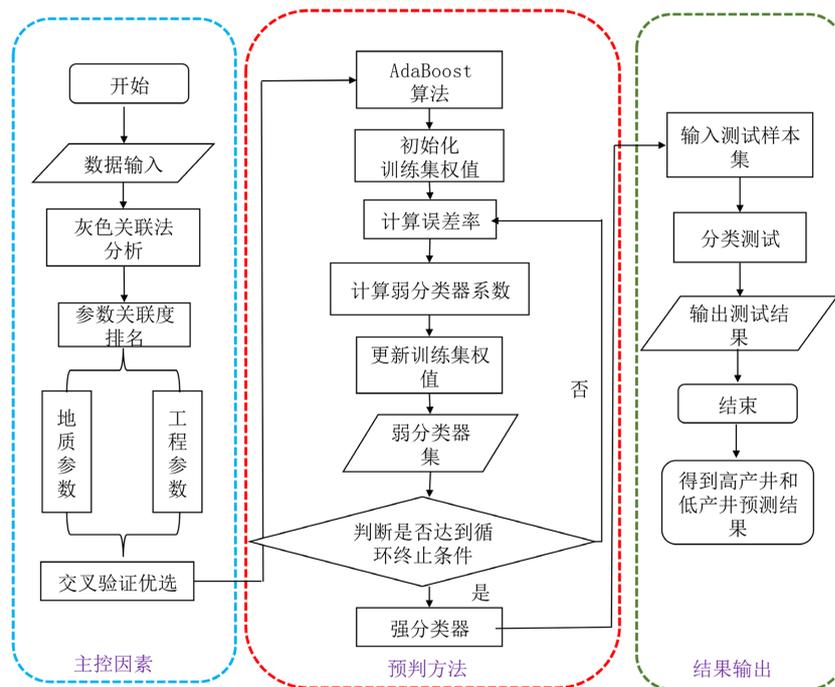


Figure 4. The flow chart of pre-judge methods for high production gas well and low production gas wells of CBM
图 4. 煤层气高产井和低产井预判方法研究流程图

基于 AdaBoost 算法建立高产井和低产井预测模型[11]的实现步骤为:

① 输入训练数据集

$$Z = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \quad (1)$$

式中, x_i 为训练数据集地质工程参数, y_i 为样本数据集标签类别。其中, $x_i \in X \subset R_n$, $y \in Y = \{-1, 1\}$, T 为迭代次数。

② 训练集权值初始化

$$D_1 = (\omega_{11}, \omega_{12}, \dots, \omega_{1N}) \quad (2)$$

式中, $\omega_i = \frac{1}{N}$, $i=1,2,\dots,N$, 训练数据集权值分布为 D_t , 有弱分类器 $H_t(x): X \rightarrow \{-1,1\}, t=1,2,\dots,T$ 。

③ 弱分类器 $H_t(x)$ 在 D_t 分布下的分类误差率

$$e_t = \sum_{i=1}^N P(H_t(x_i) \neq y_i) = \sum_{i=1}^N \omega_i I(H_t(x_i) \neq y_i) \quad (3)$$

④ 计算弱分类器 $H_t(x)$ 的系数

$$\alpha_t = \frac{1}{2} \log \frac{1-e_t}{e_t} \quad (4)$$

⑤ 更新权值

$$D_{t+1} = (\omega_{t+1,1}, \omega_{t+1,2}, \dots, \omega_{t+1,N})$$

$$\omega_{t+1,i} = \frac{\omega_i \exp(-\alpha_t y_i H_t(x_i))}{Z_t} \quad (5)$$

式中, Z_t 为规范化因子。

⑥ 得到强分类器

$$H(x) = \text{sign} \left(\sum_{t=1}^T \alpha_t H_t(x) \right) \quad (6)$$

3.2. 应用及检验

通过前文对高产井和低产井主控因素的研究, 把体积密度、声波时差、自然伽马、补偿中子、施工排量、总液量和加砂量这八个地质工程参数作为输入, 利用 AdaBoost 算法对模型进行训练。AdaBoost 算法一般用于二分类问题, 为了方便研究, 本文按照 1000 为标准将煤层气产量划为两个级别(表 2), 在研究区块优选出完整的不同产量级别的 65 个样本数据进行 AdaBoost 算法的训练, 而对其他 15 个样本数据进行预测。

Table 2. High production gas well and low production gas well
表 2. 高产井和低产井 AdaBoost 算法标签

产气量等级	Adaboost 算法标签
高产井	1
低产井	-1

根据选取的主控因素自然伽马(GA)、声波时差(AC)、体积密度(DEN)、深侧向电阻率(RD)、施工排量、总液量和总液量作为初次输入的训练数据集, 针对 80 井层的煤层气数据样本, 以 8 个主控因素为输入, 以平均日产气量为输出, 随机选取 65 井层的数据样本作为训练集, 对 AdaBoost 算法建立的高产井和低产井预判模型进行训练, 剩余 15 井层的数据样本作为测试集, 用来验证预判模型的符合率。

更改 AdaBoost 算法的迭代次数进行试算, 如图 5 所示, 结果表明当迭代次数为 200 次时, 错误率基本稳定在 10%, 因此后续迭代次数均固定为 200 次。

利用基于 AdaBoost 算法的煤层气高产井和低产井预判模型对测试集中的 15 组样本进行预判, 在 15 组测试样本中, 得到的部分结果见表 3。正确预判的样本有 14 组, 符合率为 93.33%。

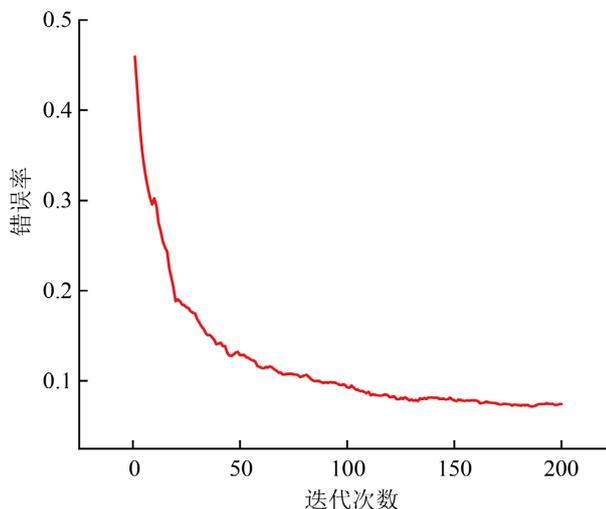


Figure 5. The relationship between the number of iterations and the error rate

图 5. 迭代次数与错误率的关系

Table 3. The result of AdaBoost algorithm predicts gas production (part)

表 3. AdaBoost 算法预测产气量结果(部分)

井号	体积密度 /(g/cm ³)	声波时差 /(μs/m)	自然伽马 /(API)	深侧向 电阻率 /(Ω·m)	施工排量 /(m ³ /min)	总液量/m ³	加砂量 /m+	产量 /(m ³ /d)	产量标签	预测标签
X1	2.68	178.94	53.64	528.37	9.62	1748.43	20.75	1607.82	1	1
X2	1.86	369.27	42.55	360.86	10.71	1011.51	34.61	966.53	-1	-1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
X14	1.49	402.84	96.43	82.05	9.84	1850.24	40.57	1732.25	1	1
X15	2.73	192.33	52.02	343.71	9.93	1258.45	39.63	1874.63	1	1

4. 推广及对比分析

下面对不同方法、不同样本个数、不同划分标准、不同区块进行讨论和对比分析,以验证基于 AdaBoost 算法的煤层气高产井和低产井预判方法的普适性。

4.1. 不同方法的对比分析

为验证本文建立的煤层气高产井和低产井预判方法的优越性以及主控因素选取的合理性,以主控因素为输入,分别利用 BP 神经网络算法[11]、SVM 和随机森林(RF)算法[12]进行对比,得到不同方法的预判结果(图 6)。

通过对比结果得到,AdaBoost 算法预测效果优于 BP 神经网络算法、SVM 和随机森林(RF)算法等传统算法。

4.2. 不同样本个数的对比分析

一般来说训练样本数越多,预测样本符合率越高。但在实际应用中,数据完整的井数不足,因此通过对对比不同训练集样本数对 AdaBoost 算法建立的高产井和低产井预判模型符合率的影响从而选择最优

样本个数。得到不同样本个数对比结果(表 4)。

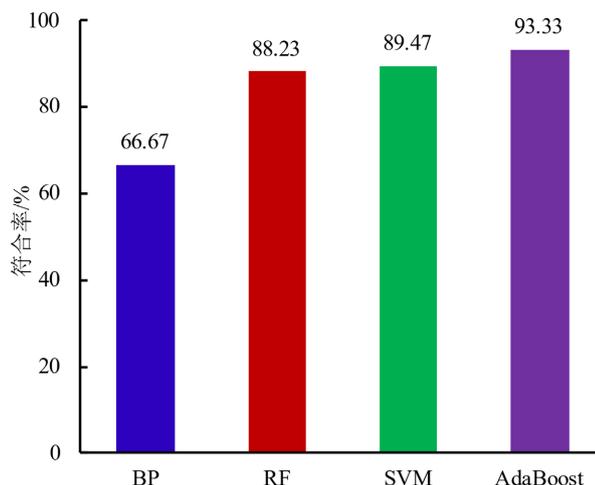


Figure 6. Comparison of coincidence rates of different methods

图 6. 不同方法符合率对比

Table 4. The impact of different number of training samples on the coincidence rate

表 4. 不同训练样本数对符合率的影响

训练样本个数/个	符合率/%
55	88.00
60	85.00
65	93.33
70	90.00

通过表 4 结果结合前文研究结果进行对比分析, 可以看出当训练样本个数选择为 65 时 AdaBoost 算法建立的高产井和低产井预判模型符合率最高。

4.3. 不同产气量划分标准的对比分析

通常情况下, 当煤层气井产气量低于 $500 \text{ m}^3/\text{d}$ 认为是低产井。在实际生产中, 当产气量低于 $1000 \text{ m}^3/\text{d}$ 时, 经济效益难以弥补开发成本。目前对产量划分没有明确的界限, 因此对不同产量划分标准进行对比分析, 找到一个合适的产量分界线, 发挥煤层气高产井和低产井预判方法的最佳效果[13]。

从表 5 的结果可以看出, 煤层产气量划分标准为 $1000 \text{ m}^3/\text{d}$ 时得到的符合率最高, 因此分析研究均在该划分标准下进行。

Table 5. The influence of different gas production classification standards on the coincidence rate

表 5. 不同产气量划分标准对符合率的影响

划分标准/ (m^3/d)	符合率/%
500	86.67
750	86.67
1000	93.33
1500	86.67

4.4. 应用于不同区块的对比分析

HC 区块位于鄂尔多斯盆地的东南缘, 面积为 1532 km², 总体呈单斜构造, 地层倾向北西。该区块煤阶类型为中高阶型, 煤层气储层埋藏深度 350 m~1200 m, 煤层厚度 8 m~15 m, 储层渗透率 0.02 mD~2.00 mD, 为低渗透储层, 气井无自然产能, 须经过压裂改造才能产气[14]。

BD 区块位于鄂尔多斯盆地东缘晋西挠褶带北段, 面积为 476.46 km², 总体呈西倾单斜, 断层和褶皱不发育。煤阶类型为中低阶型。煤层气储层埋藏深度为 300 m~1000 m, 煤层厚度 10 m~25 m, 储层渗透率 2.50 mD~8.00 mD, 煤层结构多样, 体现为较强的非均质性。

在研究了鄂尔多斯盆地 DJ 区块后, 对 BD 区块、HC 区块进行对比分析。对比 DJ 区块的地质工程参数分析发现, 基于灰色关联分析法和交叉验证法得到的主控因素存在差异(表 6)。

Table 6. Comparison of the main control factors of the three blocks

表 6. 三个区块主控因素比较

区块	地质参数	工程参数
DJ 区块	体积密度 声波时差 自然伽马 深侧向电阻率	施工排量 总液量 总砂量
HC 区块	含气饱和度 自然伽马 体积密度	总砂量
BD 区块	体积密度 孔隙度 深侧向电阻率	总液量 平均砂比

将 HC、BD 区块的主控因素作为输入, 分别基于 AdaBoost 算法建立 HC、BD 区块煤层气高产井和低产井的预判方法得到预测符合率结果(表 7)。

Table 7. HC, BD block prediction results

表 7. HC、BD 区块预测结果

等级	产气量	HC 区块	BD 区块	预测符合率/%	
				HC 区块	BD 区块
高产井		32	43	90.62	90.69
低产井		56	54	91.07	88.89
总计		88	99	90.91	89.90

通过对 HC、BD 区块的预测结果分析发现, AdaBoost 算法对这两个区块煤层气产量的预测效果较好, 因此本研究提出煤层气高产井和低产井预判方法适用于三个区块。

5. 结论

a) 利用灰色关联分析法和交叉验证法对 DJ 区块高产井和低产井的地质工程参数进行优选, 得到煤层气产量的 7 个主控因素, 分别是声波时差、体积密度、自然伽马、深侧向电阻率、总液量、总砂量和

施工排量。

b) 将本研究建立的 AdaBoost 高产井和低产井预判方法与 BP 神经网络、随机森林(RF)、SVM 算法进行对比得到 AdaBoost 算法预测符合率更高。

c) 对比分析了不同方法、不同样本个数、不同产气量划分标准得到 DJ 区块 AdaBoost 算法在训练样本个数为 65、产气量划分标准为 1000 m³/d 时符合率最高。再对比分析了 DJ、HC 和 BD 区块基于 AdaBoost 算法建立的煤层气高产井和低产井预判方法适用于多个区块。

d) 创新提出了煤层气高产井和低产井的预判方法。该方法应用于在煤层气井投产前, 如果是高产井, 则建议投产; 如果是低产井, 则建议不投产或采用新技术进行投产。

项目基金

国家科技重大专项(2016ZX05042, 2017ZX05070)。

参考文献

- [1] 朱庆忠, 杨延辉, 左银卿, 等. 中国煤层气开发存在的问题及破解思路[J]. 天然气工业, 2018, 38(4): 96-100.
- [2] 龚杰立, 李国富, 李德慧, 等. 山西省煤成气勘查开发现状及探索[J]. 煤田地质与勘探, 2022, 50(2): 39-47.
- [3] 武强, 涂坤, 曾一凡, 等. 打造我国主体能源(煤炭)升级版面临的主要问题与对策探讨[J]. 煤炭学报, 2019, 44(6): 1625-1636.
- [4] 高玉巧, 李鑫, 何希鹏, 等. 延川南深部煤层气高产主控地质因素研究[J]. 煤田地质与勘探, 2021, 49(2): 21-27.
- [5] 陈刚, 胡宗全. 鄂尔多斯盆地东南缘延川南深层煤层气富集高产模式探讨[J]. 煤炭学报, 2018, 43(6): 1572-1579.
- [6] 陈涛, 张占松, 周雪晴, 等. 基于测井参数优选的煤层含气量预测模型[J]. 煤田地质与勘探, 2021, 49(3): 227-235+243.
- [7] Tan, P., Pang, H., Zhang, R., *et al.* (2020) Experimental Investigation into Hydraulic Fracture Geometry and Migration Characteristics for Southeastern Sichuan Deep Shale Reservoirs. *Journal Petroleum Science and Engineering*, **184**, Article 106517. <https://doi.org/10.1016/j.petrol.2019.106517>
- [8] 冯敏玉, 孔萍, 胡萍, 等. 基于花前物候利用灰色关联分析法建立油菜花期预报模型[J]. 中国农业气象, 2021, 42(11): 929-938.
- [9] 关小果, 柴洪洲, 韩世静, 等. 基于系统间交叉验证的海上精密单点定位质量检核方法[J]. 大地测量与地球动力学, 2021, 41(12): 1235-1240.
- [10] Yang, C., Liu, G., Yan, C., Jiang, C. (2021) A Clustering-Based Flexible Weighting Method in AdaBoost and Its Application to Transaction Fraud Detection. *Science China Information Sciences*, **64**, Article No. 222101. <https://doi.org/10.1007/s11432-019-2739-2>
- [11] 娄高中, 谭毅. 基于 PSO-BP 神经网络的导水裂隙带高度预测[J]. 煤田地质与勘探, 2021, 49(4): 198-204.
- [12] 庞河清, 匡建超, 王众, 等. 核主成分分析与支持向量机模型在储层识别中的应用[J]. 物探与化探, 2012, 36(6): 1001-1005+1013.
- [13] 贾志超, 黄华州, 黄绍博, 等. 基于 AHP-熵权法的采动区煤层气开发潜力评价[J]. 煤田地质与勘探, 2021, 49(2): 117-124.
- [14] 贾慧敏, 胡秋嘉, 祁空军, 等. 高阶煤煤层气直井低产原因分析及增产措施[J]. 煤田地质与勘探, 2019, 47(5): 104-110.