

基于机器学习的南海砂岩油藏合理井网密度预测方法

闫正和 李 伟 王亚会 缪 云 王丹翎

(中海石油(中国)有限公司深圳分公司, 广东 深圳 518000)

摘 要: 我国南海东部海相砂岩油田积累了大量的宝贵生产数据, 如何充分利用已开发油藏的井网部署经验确定未开发油藏的合理井网密度是当前开发决策人员所面临的主要问题之一。为此, 搜集了 101 个已开发油藏开发初期、中期和末期三个阶段的井网密度和预测采收率, 并通过最小二乘法回归得到了各油藏修正的谢尔卡乔夫公式, 建立了合理井网密度数据集, 并基于 XGBoost 机器学习方法得到了合理井网密度数据集与影响因素数据集相关关系的训练模型, 最后对模型进行了验证和矿场应用。

关键词: 南海砂岩油藏; 合理井网密度; 机器学习; XGBoost 方法

0 引言

南海东部海相砂岩油田由于天然能量充足, 无需进行注水, 因而人工可干预措施少, 合理的井网布置是影响开发效果的最主要因素^[1]。如何充分利用南海东部已开发油藏井网部署的成功经验对未开发油藏进行井网部署是当前开发决策人员所面临的主要问题之一^[2]。机器学习方法可通过对已有数据进行深层次数据挖掘从而建立目标函数与影响因素之间的训练模型并进行目标函数的预测。为此, 搜集了 101 个南海已开发油藏资料, 根据谢尔卡乔夫公式确定了合理井网密度, 基于 XGBoost 机器学习算法得到了南海东部砂岩油藏合理井网密度与影响因素的训练模型并进行了模型验证, 最后通过训练模型得到了南海 H 油藏的合理井网密度并进行了矿场应用, 取得了良好降水增油效果, 为我国海相砂岩油藏合理开发提供了技术支撑。

1 合理井网密度及影响因素数据集的建立

1.1 修正的谢尔卡乔夫公式

目前, 合理井网密度确定方法主要为谢尔卡乔夫方法, 其表达式为:

$$E_R = E_D e^{-a/S} \quad (1)$$

式中, E_R 为最终采收率; E_D 为水驱油效率; S 为井网密度, $\text{口}/\text{km}^2$; a 为井网指数。

中国石油勘探开发研究院统计了 144 个陆上油田资料, 得到了修正的谢尔卡乔夫公式, 即:

$$E_R = 0.3883(K_a/\mu_o)^{0.06971} e^{\frac{18.14}{(K_a/\mu_o)0.418S}} \quad (2)$$

式中, k_a 为渗透率, μm^2 ; μ_o 为原油粘度, $\text{MPa}\cdot\text{s}$ 。

中海油研究总院研究得到了适用于海上稠油油田的修正谢尔卡乔夫公式^[3], 即:

$$E_R = 1.7945\mu_o^{-0.1886} E_{D1.00} e^{\frac{18.14}{(K_a/\mu_o)0.418S}} \quad (3)$$

式中, $E_{D1.00}$ 为室内实验得到的含水率达到 100% 时的驱油效率。

对比三式可以看出, 式(2)、(3)均针对不同类型的油藏给出了式(1)中 E_D 和井网指数 a 的具体表达式。

南海东部海相砂岩油田储层连续性好、地层原油黏度低、天然水体能量充足等典型特点, 无法直接采用式(2)、(3)进行计算, 为此, 借鉴式(2)、(3)采用下式表征采收率与井网密度之间的关系:

$$E_R = m E_{D1.00} e^{\frac{a}{S}} \quad (4)$$

式中, m 为采收率修正系数, 取值范围为 0~1, 其数值越大, 表明矿场水驱油效率越接近室内实验所得到的含水率为 100% 时的驱油效率。

1.2 井网密度折算方法

谢尔卡乔夫公式中的井网密度一般指直井井网, 而南海东部海相砂岩油田普遍采用直井和水平井混合井网。水平井泄油面积远大于直井, 因而不能简单地采用直井和水平井加和的方法确定井网密度, 需要根据一定的准则将水平井数折算为直井数。本文采用等效泄油体积法^[4], 即水平井泄油区近似为椭圆形而直井近似为圆形, 同时由于南海东部海相砂岩油田储层较薄, 泄油区厚度均可假设为油藏厚度, 则可得到边水油藏直井泄油区体积:

$$V_v = \delta R^2 h \quad (5)$$

式中, V_v 为直井泄油区体积, m^3 ; R 为泄油半径, m ; h 为油藏厚度, m 。

水平井泄油区体积为:

$$V_h = \delta R \left(R + \frac{L}{2} \right) h \quad (6)$$

式中, V_h 为水平井泄油区体积, m^3 ; L 为水平井长度, m 。

则边水油藏水平井和直井折算比为:

$$n = \frac{V_h}{V_v} = 1 + \frac{L}{2R} \quad (7)$$

底水油藏直井泄油区可近似为一个锥体^[5], 其体积为:

$$V_v = \frac{1}{3} \pi r_1^2 \times Z \quad (8)$$

式中, r_1 为水锥下沿的半径, m ; Z 为水锥高度, m 。

水平井泄油区可近似为一个脊体，其体积为：

$$V_h = \frac{1}{3} \pi r_2^2 Z + r_2 ZL \quad (9)$$

式(9)中， r_2 为脊体下沿宽度半长， m 。

由此可得底水油藏水平井和直井折算比为：

$$n = \frac{V_h}{V_v} = \left(\frac{r_2}{r_1} \right)^2 + \frac{3r_2 L}{\pi r_1^2} \quad (10)$$

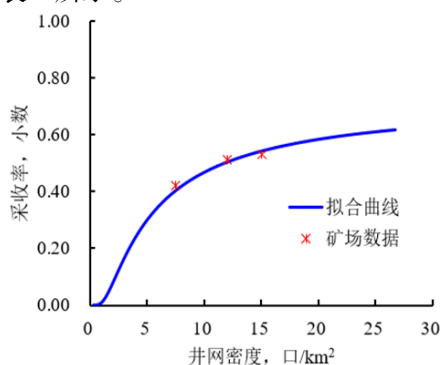
1.3 合理井网密度数据集的建立

南海东部海相砂岩油田已开发油藏总体上经历了总体开发方案时期(ODP时期)、开发中期、开发后期三个阶段。基于三个阶段开发历史数据，采用递减分析方法可以得到101个油藏对应井网密度下的采收率值。以P1、P2油藏为例，统计得到不同井网密度下的采收率值，如表1所示。

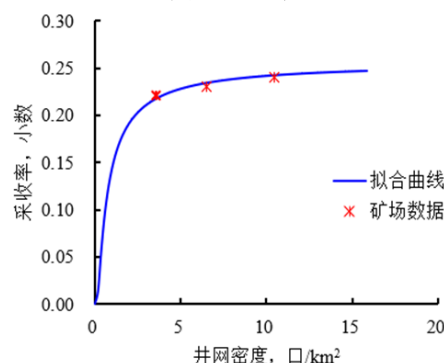
表1 P1、P2油藏不同时期井网密度及对应采收率

油藏	井网密度 / (口 · km ²)	对应采收率 / 小数
P1	2.00	0.480
	7.20	0.640
	14.40	0.690
P2	3.60	0.220
	6.50	0.230
	10.50	0.240

基于101个油藏三个阶段的井网密度和采收率数据，采用最小二乘法对式(4)进行回归可得到各油藏最优的 m 和 a 值，从而确定各油藏的井网密度与采收率的关系曲线，如图1所示。结合经济核算方法，可计算得到不同井网密度时的净现值，净现值最大时的井网密度即为合理井网密度，P1、P2油藏的 m 值、 a 值以及合理井网密度如表2所示。



(a) P1 油藏



(b) P2 油藏

图1 P1、P2油藏采收率与井网密度的关系曲线

表2 P1、P2油藏的 m 值、 a 值和合理井网密度

油藏	m	a	合理井网密度 / (口 · km ²)
P1	0.85	4.432	16.2
P2	0.3	0.587	11.4

1.4 影响因素数据集的建立

井网密度选择主要考虑储层基本物性、流体性质、井网特征、边底水能量大小以及海上油田特殊性：①储层物性，包括含油面积、油层厚度、孔隙度、平均渗透率、油藏中深、隔夹层密度；②流体性质，主要指地下原油粘度；③井网特征，包括水平井长度、底水油藏避水高度、边水油藏平面避水距离；④生产制度，主要为单井平均采液速度；⑤边底水能量，主要为水体倍数，即储层条件下边底水水体体积与孔隙体积比值。

2 合理井网密度机器学习模型建立

利用上述建立的合理井网密度及影响因素数据集，从数据挖掘角度建立合理井网密度机器学习预测模型^[6]。目前，机器学习的方法种类繁多，按照算法功能分类主要包括回归算法、决策树算法、贝叶斯算法、人工神经网络和集成算法等。XGBoost(eXtreme Gradient Boosting)算法在集成算法中具有高效率、高预测准确度等优点，故采用XGBoost算法建立合理井网密度机器学习模型。

2.1 模型训练及优化

选取南海砂岩油藏合理井网密度以及相关影响因素组成学习样本，用以XGBoost集成学习模型的训练。把学习样本的90%作为训练集(90个样本)，剩余的10%(11个样本)作为测试集，用以模型后续的验证和优化。

合理井网密度机器学习模型主要是调整决策树集提高合理井网密度预测值的准确性，因此提出合理井网密度准确率 r_s 作为评价指标^[7]。

$$r_s = 1 - \frac{\sum_{i=1}^n |S_{pi} - S_{ti}|}{n} \quad (11)$$

式中， n 为训练集样本总数， S_{pi} 为油藏 i 的合理井网密度预测值， S_{ti} 为油藏 i 的合理井网密度。

不断调整XGBoost学习模型的参数取值，根据实验结果确定模型参数的最优组合。修改模型参数为最优组合后，输入训练集数据重新训练学习，当准确率 r_s 达到98%作为训练结束条件。

2.2 模型测试与验证

选取南海砂岩油藏11个样本作为测试集，该部分油藏采用传统油藏工程方法拟合谢尔卡乔夫曲线得到合理井网密度，结合油田相关数据确定井网密度的影响因素作为测试集。

将训练集和测试集样本分别输入设置好最优参数组合的XGBoost机器学习模型，得到训练集和测试集对应油藏的合理井网密度值，以数据拟合得到的合理井网密度实际值为横坐标，以机器学习模型输出的合理井网密度值为纵坐标绘图进行对比，结果如图2和图3所示。

可以看到,预测值与实际值接近,测试集的合理井网密度预测值准确率高达 97.24%,平均误差为 2.76%,预测精度符合矿场需求,验证了 XGboost 机器学习预测模型的准确性。

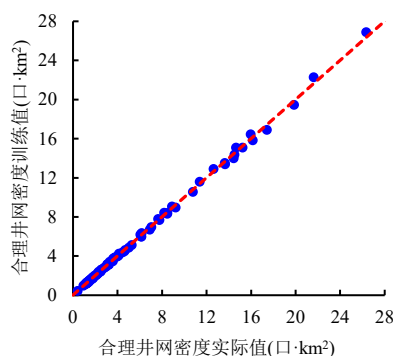


图 2 合理井网密度模型训练结果图

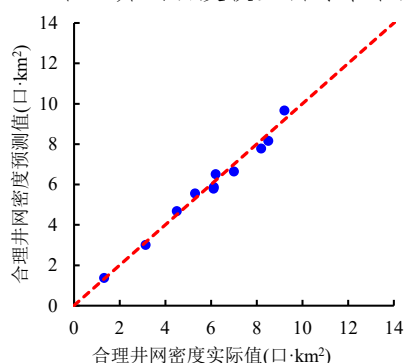


图 3 合理井网密度模型预测结果图

3 矿场应用

表 3 H 油藏合理井网密度影响因素参数

影响因素	取值	影响因素	取值
含油面积 / (km^2)	1.95	地下原油粘度 / ($\text{MPa} \cdot \text{s}$)	15
地质储量 / (10^4m^3)	134.44	单井控制储量 / (10^4m^3)	26.89
油层厚度 / (m)	4.5	水平井长度 / (m)	274
孔隙度 / (%)	27.4	边水避水距离 / (m)	-
渗透率 / ($10^{-3} \mu\text{m}^2$)	1047.3	底水避射高度 / (m)	4.85
油藏中深 / (m)	1577.55	单井平均采液速度 / ($10^4 \text{m}^3 \cdot \text{d}$)	14.66
隔夹层密度 / (个 $\cdot \text{km}^2$)	48.47	水体倍数 / 无因次	2.73

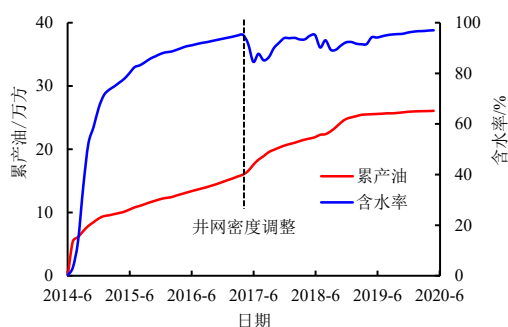


图 4 H 油藏累产油和含水率变化曲线

南海 H 油藏属于底水稀油油藏,合理井网密度影响因素如表 3 所示。井网密度调整前,有 1 口直井和 1 口水平井,折算井网密度为 $4.19 \text{口}/\text{km}^2$,含水率为 95.90%,采出程度为 35.55%。将该油藏数据导入上述基于 XGboost 机器学习模型的合理井网密度预测模型,预测合理井网密度为 $5.73 \text{口}/\text{km}^2$ 。2017 年 5 月开始对该油藏进行加密至合理井网密度 $5.73 \text{口}/\text{km}^2$ 。井网密度调整前后的 H 油藏累产油、含水率和月产油变化曲线如图 4 和图 5 所示。可以看出,井网密度由 $4.19 \text{口}/\text{km}^2$ 调整为合理井网密度 $5.73 \text{口}/\text{km}^2$,H 油藏含水率下降 10.6%,累产量增加 9.05 万方,采收率增加 6.73%,取得了良好的降水增油效果,验证了基于 XGboost 机器学习模型的合理井网密度预测方法的适用性和有效性。

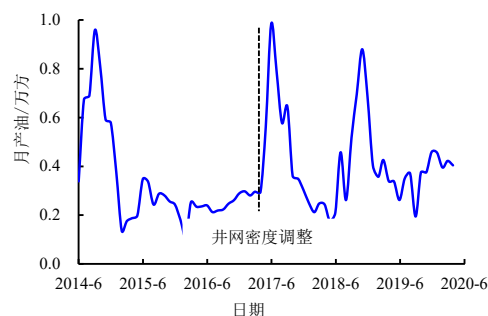


图 5 H 油藏月产油变化曲线

4 结论

提出了一种改进的谢尔卡乔夫公式,确定了合理井网密度及影响因素数据集以及数据处理、学习过程处理方法,建立了基于 XGboost 机器学习模型的合理井网密度预测方法。建立的合理井网密度机器学习模型可以很好地预测不同油藏的合理井网密度,预测准确率高达 97.24%,平均预测误差仅为 2.76%,预测精度符合矿场需求。矿场应用表明,根据合理井网密度预测结果进行井网调整,油藏采收率提高 6.73%,验证了基于 XGboost 机器学习模型的合理井网密度预测方法的适用性和有效性。

参考文献:

- [1] 张凤久,罗宪波,刘英宪,等.海上油田丛式井网整体加密调整技术研究[J].中国工程科学,2011,13(5):34-40.
- [2] 衣哲.多因素合理井网密度计算方法研究[D].青岛:中国石油大学(华东),2010.
- [3] 耿站立,安桂荣,周文胜,等.海上稠油油田井网密度与采收率关系研究[J].中国海上油气,2012,24(3):35-37.
- [4] 许家峰,倪学锋,甯波,等.海上油田合理油水井数比计算新方法[J].石油天然气学报,2012,34(11):109-113.
- [5] 侯君,程林松.常规底水油藏水锥高度计算方法研究[J].西安石油大学学报(自然科学版),2006,21(3):23-26.
- [6] 谷建伟,任燕龙,王依科,等.基于机器学习的平面剩余油分布预测方法[J].中国石油大学学报(自然科学版),2020,44(4):39-46.
- [7] 何龙.深入理解 XGBoost[M].北京:机械工业出版社,2020.