

基于大数据挖掘技术不停机间抽工作制度优化

——以大庆油田为例

高翔¹, 王云峰², 刘海波¹

(1. 中国石油大庆油田有限责任公司采油工程研究院, 黑龙江大庆 163453; 2. 中国石油大庆油田有限责任公司第九采油厂工程技术大队, 黑龙江大庆 163511)

摘要: 针对不停机间抽技术提高抽油机井泵效和降低能耗等方面效果显著, 但工作制度确定过程中易出现实际确定方法操作性不强, 主观因素影响大, 单井设计个性化弱等问题, 通过分析不停机间抽工作制度的影响因素, 在优选出9个相对独立因素的基础上, 利用大数据挖掘技术, 以不停机间抽正常运行时间和运行周期为分析挖掘对象, 对比分析了常用数据挖掘算法的适应性, 并优选出适应性最强的算法。结果表明, 回归计算中, BPNN算法要优于R-SVM和MRA算法; 分类算法中, C-SVM算法比BAYSD、NBAY等算法更优; 使用C-SVM—BPNN算法对不停机间抽工作制度优化后系统效率和泵效明显提高。研究结果对于不停机间抽井确定最优工作制度具有较好的指导作用。

关键词: 大数据; 数据挖掘; 不停机间抽; 工作制度

中图分类号: TE355

文献标识码: A

Optimization of non-stop intermittent pumping system based on big data mining technology —by taking Daqing oilfield as an example

GAO Xiang¹, WANG Yunfeng², LIU Haibo¹

(1. Oil Production Engineering Research Institute of Daqing Oilfield Co., Ltd., PetroChina, Daqing, Heilongjiang 163453, China; 2. Engineering and Technology Team of NO.9 Oil Production Plant of Daqing Oilfield Co., Ltd., PetroChina, Daqing, Heilongjiang 163511, China)

Abstract: The non-stop intermittent pumping wells are good to improve the pumping efficiency and reduce the energy consumption, while in the process of determining the working system, there are many problems, such as the low operability of the actual determination method, the great influence of subjective factors, and the weak individualization of single well design. By analyzing the influencing factors of the working system, based on the optimization of 9 relatively independent factors, by using big data mining technology, taking the normal running time and running period as the analysis mining object, the adaptability of common data mining algorithms is compared and analyzed, and the algorithm with the strongest adaptability is selected. The results show that BPNN algorithm is better than R-SVM and MRA algorithm in regression calculation, C-SVM algorithm is better than BAYSD and NBAY algorithm in classification algorithm, and C-SVM-BPNN algorithm is better than BAYSD and NBAY algorithm in system efficiency and pump efficiency. The research results have a good guiding role in determining the optimal working system of non-stop intermittent pumping wells.

Key words: big data; data mining; non-stop intermittent pumping; working system

随着大庆油田开发进入中后期, 低产液、低泵效、低流压井逐年上升, 这部分井存在着系统效率

收稿日期: 2020-03-24; 修订日期: 2020-04-13。

第一作者简介: 高翔(1988—), 男, 工程师, 现从事采油化学与采油工程系统理论研究工作。E-mail: cyygx@petrochina.com.cn。

基金项目: 中国石油天然气股份有限公司重大科技专项“油气开采降本增效技术与规模应用”(2016E-0212)。

低、能耗大、设备磨损严重等问题^[1-3]。为提高低效井的系统效率，达到节能降耗的目的，通常采用间抽采油技术^[4-5]。但在实际生产过程中，采用常规间抽技术往往因为停井时间长导致卡泵故障、井底流压上升，从而影响油井产量。相关研究表明，常规间抽技术平均影响油井产量 0.1 m³/d，部分井潜力未得到充分发挥^[6]。在综合考虑常规间抽技术缺点的基础上，相关工程技术人员提出一种抽油机不停机间抽技术。目前，大庆油田累计应用不停机间抽技术千余口井，与常规间抽技术相比，吨液百米耗电降低明显，系统效率显著提高，应用前景广阔^[7]。

不停机间抽技术应用过程中，工作制度的确定是最关键的部分^[8]。一般情况下，主要参考单井地质特征和供液能力，以油井产能为最大目标，结合 IPR（流入动态生产曲线）确定合理流压，通过持续观察动液面变化，最终确定最优的不停机间抽采油工作制度。但是该方法需要对每口井工作制度进行多次试验和反复验证，实际操作性不强，生产单位通常依据沉没度和产液情况，将运行工作制度定为若干种（表 1），从而大大降低了单井个性化设计强

度，也减弱了不停机间抽技术的应用效果。

近几年来，数字化、智能化、智慧化油田的建设将油田带到了“大数据、人工智能”时代^[9]，大数据挖掘技术作为机器学习、人工智能的基础，同样也开始应用到油气田勘探开发的各个领域^[10-13]。2012 年，石广仁教授介绍了八大类数据挖掘算法，并通过 34 个应用实例对算法的应用范围及条件、基本原理和完整计算方法进行了适用性比较^[14]；2015 年，檀朝东等人系统总结了大数据挖掘技术在石油工程的应用前景^[15]；2016 年，孙敬等人采用大数据挖掘技术建立的产能评价方程所预测的气井产量与实际产量相比，精度可以达到 90%^[16]；2018 年，李大伟等人开展了油气勘探开发常用数据挖掘算法优选工作，认为最优的回归算法是反向传播神经网络（BPNN），最优的分类算法是支持向量机分类（C-SVM）^[17]。本文则针对不停机间抽技术工作制度确定时出现的问题，通过对相关采油数据进行有效分析，明确影响工作制度确定因素的主次关系，并在对相关数据挖掘常用算法进行优选的基础上，构建单井个性化工作制度优化方法。

表 1 生产单位制定不停机间抽工作制度（运行周期 30 min）

产液量/(t·d ⁻¹)	沉没度小于 100 m	沉没度为 100~200 m	沉没度大于 200 m
Q≤1		摆动运行 25 min，整周运行 5 min	摆动运行 20 min，整周运行 10 min
1<Q≤3	摆动运行 25 min	摆动运行 20 min，整周运行 10 min	摆动运行 20 min，整周运行 10 min
3<Q≤5	整周运行 5 min	摆动运行 20 min，整周运行 10 min	摆动运行 15 min，整周运行 15 min
5<Q		摆动运行 15 min，整周运行 15 min	摆动运行 10 min，整周运行 20 min

注：整周运行是指抽油机像正常井一样运行；摆动运行是指抽油机在不停机条件下，依靠变频控制箱控制抽油机曲柄以“秋千式”低能耗小幅摆动运行；运行周期是指整周运行与摆动运行时间之和。

1 常用数据挖掘方法与模型建立

机器学习数据挖掘常用的算法可分为分类、回归、聚类等，其中分类和回归是最成熟、应用最广泛的算法。对具体的研究问题、研究对象和数据源，不同的分类和回归算法具有不同的适用性。常用的回归算法包括反向传播神经网络（BPNN）、多元回归分析（MRA）和支持向量机回归（R-SVM）等，常用的分类算法包括决策树（DTR）、支持向量机分类（C-SVM）、贝叶斯判别分析（BAYD）、朴素贝叶斯（NBAY）和贝叶斯逐步判别分析（BAYSD）等^[17]。由于 DTR 算法建立与应用非常复杂^[18]，BAYD 算法适应性弱于 BAYSD 算法，因此只针对其他几类常用的数据挖掘的回归和分类算法进行不停机间抽工作制度预测，并通过对比预测过程中产生的总平均相对误差绝对值来确定最佳的数据挖掘算法。

研究结果对于低产低效井应用不停机间抽技术

确定最优工作制度，从而达到节能最大化，具有较大的应用价值，也可以推广到其他数据挖掘案例，用于指导油气勘探开发、采油工作方案设计等数据挖掘工作。

1.1 模型与方法建立

假设有 n 个学习样本，每个样本有 $m+1$ 个参数 $(x_1, x_2, \dots, x_m, y_i^*)$ 的成组观察值 $(x_{i1}, x_{i2}, \dots, x_{im}, y_i^*)$ ($i=1, 2, \dots, n$)； $n > m$ ，且在实际应用中为了保证预测结果精度， n 一般要远远大于 m 。可将 m 个参数的 n 个学习样本定义为 n 个向量，学习样本的表达式为：

$$x_i = (x_{i1}, x_{i2}, \dots, x_{im}, y_i^*) \quad (i=1, 2, \dots, n) \quad (1)$$

式中： x_i 为学习样本； y_i^* 为学习样本的预测变量。

令 x_0 为 $(x_{i1}, x_{i2}, \dots, x_{im})$ 中向量的一般形式，且将 y 定义为一个单变量。

学习计算流程如下：①学习过程：使用学习样本得到拟合方程 $y = f(x_0)$ ；②误差检验：将学习样本代入拟合方程 $y = f(x_0)$ 中，得到预测值 (y_1, y_2, \dots, y_n) ，通过计算总平均相对误差绝对值 \bar{R}^*

检验该算法拟合度; ③回归(分类)预测: 将 k 个预测样本代入拟合方程 $y=f(x_0)$, 得到预测值 $(y_{n+1}, y_{n+2}, \dots, y_{n+k})$ 。其中, 预测样本为:

$$x_i^* = (x_{i1}, x_{i2}, \dots, x_{im}) \quad (i=n+1, n+2, \dots, n+k) \quad (2)$$

1.2 误差分析

为了表示预测样本和学习样本的预测变量 y 的结果精度, 通常采用相对误差绝对值 R_i 、平均相对误差绝对值 \bar{R} 和总平均相对误差绝对值 \bar{R}^* 来判断^[14]。

$$R_i = \left| \frac{y_i - y_i^*}{y_i^*} \right| \times 100\% \quad (3)$$

$$\bar{R} = \sum_{i=0}^{N_s} \frac{R_i}{N_s} \quad (4)$$

$$\bar{R}^* = \frac{R_i + \bar{R}}{2} \quad (5)$$

式中: N_s 为学习样本。

式(3)中的 y_i^* 值不能为 0, 因此, 回归学习过程中, 数据预处理时要删除 $y_i^* = 0$ 的样本; 而对于 \bar{R} , 式(4)中 N_s 对于学习样本为 n , 而对于预测样本为 k ; \bar{R}^* 是一个非常重要的指标, 用于确定预测结果是否可用, 根据相关研究结果, 如果 $\bar{R}^* < 10\%$, 结果可信, 否则需要对算法做进一步优选。

2 数据挖掘算法优选

2.1 学习样本获取

低产低效井产能的影响因素除了常规的地质特征, 还应包括油压、动液面以及生产制度等多方面的因素。通过对实际油井基础数据^[16]进行系统分析,

将影响因素归纳为 9 个独立变量(地层系数、日产量、孔隙度、含水率、油压、冲程、冲次、泵径和泵效)。

取大庆油田 A 采油厂的 15 个样本数据^[19-20], 其中, 14 个样本作为学习样本, 1 个样本作为预测样本, 每个样本都有 9 个独立变量数据^[21](表 2), 其中, 因变量 y^* 为工作制度运行时间。为了进一步准确获取 15 口样本井不停机间抽工作制度, 对样本井进行不同不停机间抽工作制度条件下动液面恢复试验。图 1 为 A1 井不同不停机间抽工作制度条件下动液面恢复情况, 从图中可以看出, 当该井的工作制度为正常运行时间 30 min, 摆动运行时间 30 min 时, 动液面稳定在 850 m 附近, 可以实现油井产能最大化, 所以认为该井不停机间抽最优工作制度为正常运行时间 30 min, 摆动运行时间 30 min。按照同样方法, 可以获得其他 14 个样本的最优工作制度(表 2)。

2.2 算法优选

2.2.1 输入参数

输入参数包括 14 个学习样本和 1 个预测样本的各个已知变量值 $x_i (i=1, 2, \dots, 9)$, 以及 14 个学习样本的预测变量 y^* 值。另外, 根据现场应用现状, 一般情况下, 不停机间抽运行周期可分为 30, 60, 90, 120 min, 所以对于回归计算, y^* 值为运行时间, 则摆动时间为运行周期时间减去运行时间; 而对于分类计算, y^* 值为运行周期, 表 3 为不同运行周期时间对应的运行周期类别。

2.2.2 学习过程

使用回归、分类算法对表 2 中 14 个学习样本进行学习计算, 分别建立 9 个独立变量 $x_i (i=1, 2, \dots, 9)$

表 2 不停机间抽工作制度分析预测基本数据

样本类型	样本编号	自变量 x_i									因变量 y^*	
		地层系数 $x_1 / (m \cdot \mu \cdot m^2)$	日产量 x_2 / t	孔隙度 $x_3 / \%$	含水率 $x_4 / \%$	油压 x_5 / Mpa	冲程 x_6 / m	冲次 $x_7 / (s^{-1})$	泵径 x_8 / mm	泵效 $x_9 / \%$	运行时 间/min	运行周 期分类
学习样本	A1	0.138	12.62	26.0	95.0	0.43	2.6	6.0	38	55.00	30	2
	A2	0.103	7.20	18.6	43.6	0.70	3.5	5.0	44	9.98	38	2
	A3	0.110	52.45	30.0	97.0	0.35	4.2	5.0	83	32.29	30	2
	A4	0.236	4.90	15.3	41.6	0.40	3.0	5.0	44	5.81	40	2
	A5	0.511	0.98	27.5	94.3	0.39	4.2	4.0	44	24.68	30	3
	A6	0.226	9.99	27.5	96.1	0.38	3.0	6.0	44	9.79	30	3
	A7	0.223	10.09	27.4	95.0	0.30	2.1	6.0	44	19.38	60	3
	A8	0.100	87.59	26.8	94.2	0.63	3.0	6.0	38	154.60	60	3
	A9	0.156	0.95	27.5	76.3	0.45	3.0	4.0	38	4.83	30	4
	A10	0.129	9.93	28.0	97.6	0.72	4.2	4.0	38	11.56	30	4
	A11	0.127	6.20	27.4	89.2	0.72	2.5	4.5	38	33.81	30	4
	A12	0.292	7.20	17.2	97.3	0.82	3.0	4.0	44	82.23	15	1
	A13	0.050	6.60	20.6	63.0	0.81	3.0	4.0	38	31.58	15	1
	A14	0.157	2.60	19.0	57.3	0.81	2.5	4.0	38	71.30	10	1
预测样本	A15	0.196	4.10	19.6	8.0	0.20	3.0	5.0	44	3.82	18	2

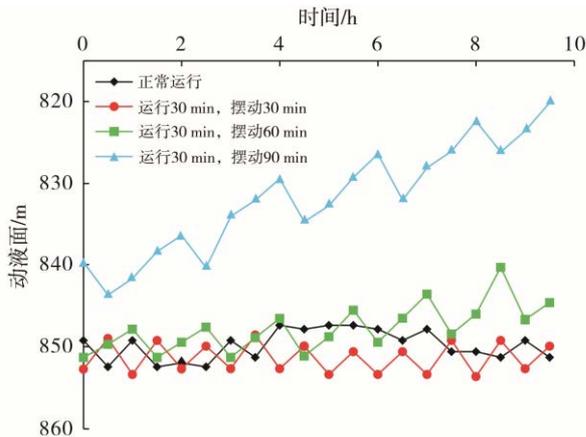


图1 A1井不同不停机间抽工作制度条件下动液面恢复

表3 不停机间抽工作制度运行周期分类

运行周期/min	30	60	90	120
运行周期分类	1	2	3	4

与运行时间(回归)或运行周期(分类)的工作制度的预测拟合方程 $y = f(x_0)$, 然后将表2中的14个学习样本和1个预测样本的 $x_i (i = 1, 2, \dots, 9)$ 值分别代入预测拟合方程 $y = f(x_0)$, 从而得到每个学习样本的运行时间 T (图2)和运行周期分类(图3), 并计算每种算法所得结果的误差(表4、表5)。

2.2.3 回归结果分析

由图2可知, BPNN、R-SVM和MRA三种回归算法最终预测的运行时间, 只有BPNN算法预测的结果与实际运行时间拟合度较高; 从表4也可以看出, 用R-SVM和MRA算法所得的结果精度较低, 而BPNN得到的结果精度则较高。虽然不停机间抽工作制度优化过程非线性关系很强, 但BPNN算法较适用。另外, MRA算法的 $\bar{R}^* = 53.27\%$, 所以预测得到的 y 值与相关的9个独立变量之间具有较强的非线性关系, MRA算法计算出的各独立变量的相关性由大到小排序为: $x_i (i = 5, 7, 6, 9, 8, 2, 1, 3, 4)$ 。由于用R-SVM和BPNN算法所得回归方程为非线性方程, 所以不能计算出预测运行时间 T 与各独立变量的相关性。

2.2.4 分类结果分析

图3为C-SVM、BAYSD和NBAY三种分类算法的最终预测运行周期分类, 从图中可以看出, C-SVM算法结果拟合度达到100%, 从表5也可以看出, C-SVM算法所得结果的精度非常高, R_f 、 \bar{R} 和 \bar{R}^* 均为0, BAYSD和NBAY两种算法的预测精度都很低。

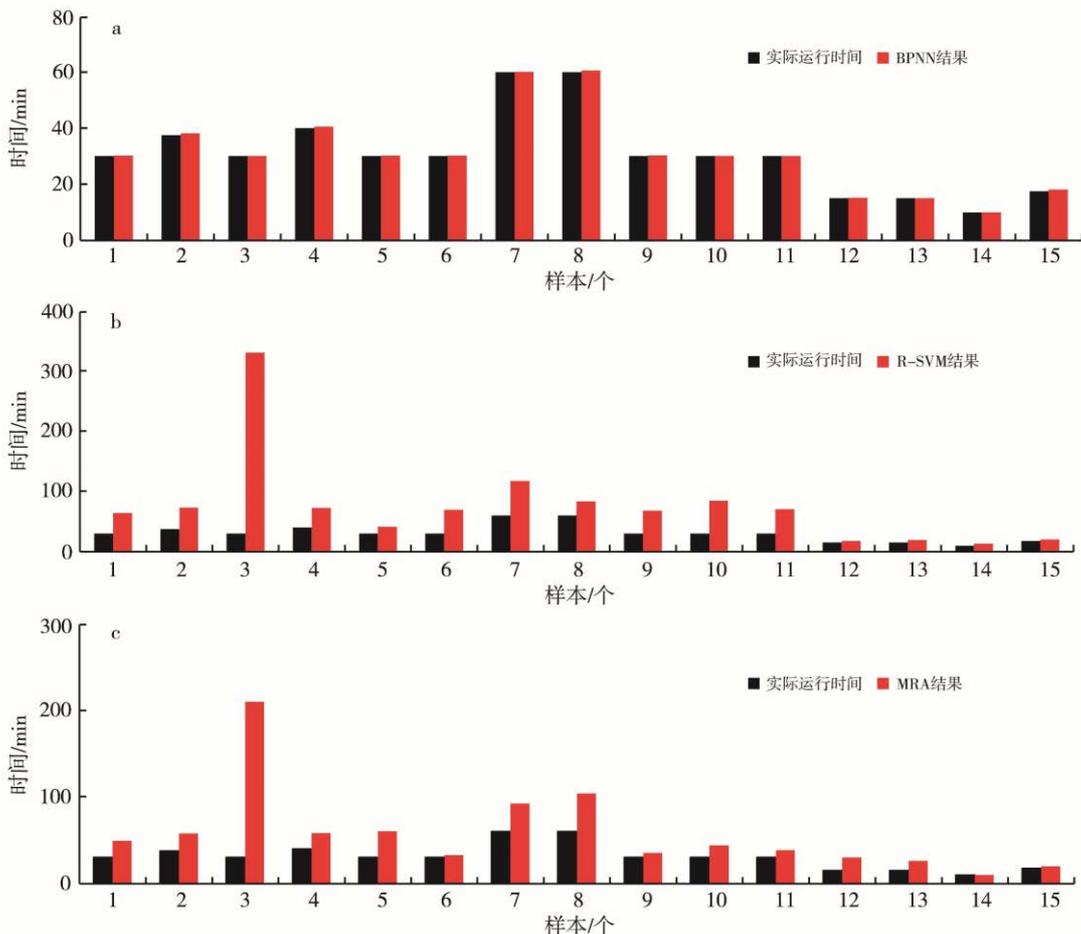


图2 不停机间抽运行时间预测结果

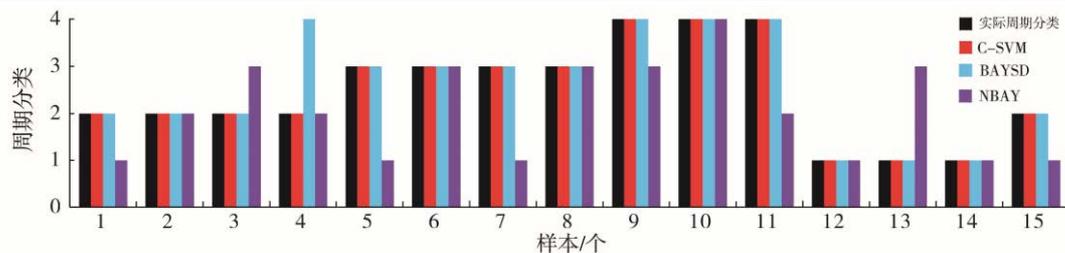


图3 不停机间抽运行周期分类结果

表4 三种回归算法计算不停机间抽运行时间结果误差 %

算法	回归预测方程	R_i	\bar{R}	\bar{R}^*	计算精度
BPNN	非线性	0.49	2.90	1.70	很高
R-SVM	非线性	150.39	14.82	82.61	中等
MRA	线性	91.14	15.4	53.27	中等

表5 三种分类算法计算不停机间抽运行时间结果误差 %

算法	回归预测方程	R_i	\bar{R}	\bar{R}^*	计算精度
C-SVM	非线性	0	0	0	很高
BAYSD	非线性	7.14	0	3.33	很低
NBAY	非线性	36.00	50	43.50	很低

综合数据挖掘的回归和分类算法, 优选得到 C-SVM—BPNN 算法。

3 C-SVM—BPNN 应用实例分析

通过 C-SVM—BPNN 算法对 A 采油厂所应用的不停机间抽技术进行工作制度优化, 由表 6 可知,

表6 部分井不停机间抽工作制度优化前后运行情况对比

井号	工况	工作制度			产液量/ ($t \cdot d^{-1}$)	泵效/%	系统效率/%
		周期分类	运行时间/min	摆动时间/min			
B1	优化前	2	15.0	45.0	3.1	11.51	15.09
	优化后	2	20.0	40.0	3.2	16.74	18.64
B2	优化前	2	10.0	50.0	1.7	3.81	16.75
	优化后	2	8.5	51.5	1.3	10.83	21.35
B3	优化前	2	20.0	40.0	1.4	4.80	7.41
	优化后	2	24.5	35.5	1.6	6.54	10.32
B4	优化前	2	15.0	45.0	1.6	3.39	20.26
	优化后	2	10.0	50.0	0.9	7.89	25.47
B5	优化前	2	27.5	32.5	1.3	7.41	10.71
	优化后	2	25.0	35.0	2.3	15.89	15.26

参考文献

- [1] 关宁, 欧阳华章, 李华. 抽油机低效间抽井产液变化规律[J]. 油气田地面工程, 2006, 25(2): 17-18.
- [2] 冯立, 王群崑, 吴虞. 降低采油工程投资的机采优化设计方法[J]. 大庆石油地质与开发, 2011, 30(3): 113-117.
- [3] 高翔, 吴虞, 王帅, 等. 抽油机井适应性性与节能设备匹配优选研究[J]. 中外能源, 2019, 24(1): 95-100.
- [4] 李军亮, 廖锐全, 陈晓春. 抽油井间抽周期的灰色预测

在产液量变化不大的情况下, 优化后泵效提高 2%~8%, 系统效率提高 3%~5%。

4 结论

(1) 常用数据挖掘算法用于优化不停机间抽工作制度, 最优的回归算法是 BPNN, 其次是 MRA 和 R-SVM; 最优的分类算法是 C-SVM, 其次为 NBAY 和 BAYSD。机器学习数据挖掘应用研究过程中, 对于具体的研究对象、研究问题和数据源, 不同的回归和分类算法具有不同的适用性, 所以针对不同问题要进行算法的优化选择。

(2) 利用 C-SVM—BPNN 算法对不停机间抽工作制度进行优化的实例表明, 优化后的不停机间抽井系统效率和泵效均有明显的提升, 说明 C-SVM—BPNN 算法具有较好的应用效果。

[J]. 断块油气田, 2012, 19(5): 634-637.

- [5] 苗国晶, 张传绪. 大庆外围低渗透油田间歇采油制度优化方法研究[J]. 石油钻采工艺, 2012, 34(2): 62-65.
- [6] 于小明, 何贯中, 金英兰. 抽油机井间抽制度合理性研究[J]. 大庆石油地质与开发, 2006, 25(4): 78-79.
- [7] 巩宏亮, 戚兴, 常瑞清, 等. 抽油机不停机间歇采油技术研究与应用[J]. 石油石化节能, 2017, 7(10): 3-6.
- [8] 于晶. 抽油机不停机间歇采油技术分析[J]. 化学工程与装备, 2019, 14(2): 53-54.

(下转第117页)

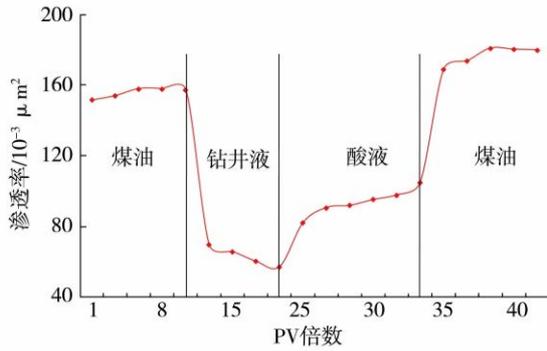


图7 3#岩心流动实验

降低酸液对储层的过度溶蚀，防止骨架破坏和微粒运移。

参考文献

- [1] 张荣军, 白云云, 秦文龙. 疏松砂岩油藏酸化机理研究[J]. 科技创新导报, 2010, 7(8): 1-3.
- [2] 张红军, 杜勇, 舒勇, 等. 疏松砂岩油藏酸化技术研究

- [J]. 钻采工艺, 2004, 27(3): 36-38.
- [3] 黄志宇, 何雁, 连西美. 酸化中酸渣沉淀的研究[J]. 西南石油学院学报, 1994, 16(3): 101-104.
- [4] 刘长松, 杜永慧, 李晓东, 等. 低渗多敏感性油藏酸化增注配方的研制[J]. 断块油气田, 2019, 26(1): 111-114.
- [5] 谢启超, 冯波, 宋鹏, 等. 合水地区长7致密砂岩储层微观孔喉结构分形特征[J]. 断块油气田, 2019, 26(2): 215-219.
- [6] 法鲁克·西维. 储层伤害原理、模拟、评价和防治[M]. 北京: 石油工业出版社, 2003.
- [7] 许正恩, 辛文明, 刘誉, 等. 高凝油油藏CO₂驱转水气交替驱动态及储层伤害特征[J]. 断块油气田, 2019, 26(5): 613-616.
- [8] 王鸿勋, 张琪. 采油工业原理[M]. 北京: 石油工业出版社, 1981.

(编辑 赵川喜)

(上接第 113 页)

- [9] 谭锋奇, 李洪奇, 孟照旭, 等. 数据挖掘方法在石油勘探开发中的应用研究[J]. 石油地球物理勘探, 2010, 45(1): 85-91.
- [10] 黄渊, 廖明光, 李斌, 等. 数据挖掘技术在碳酸盐岩储层评价中的应用[J]. 特种油气藏, 2014, 21(5): 37-42.
- [11] 吴树鹏, 熊华平. 油田开发数据挖掘技术的实现与应用[J]. 大庆石油地质与开发, 2002, 21(3): 49-51.
- [12] 李克文, 周广悦, 路慎强, 等. 一种基于机器学习的有利区评价新方法[J]. 特种油气藏, 2019, 26(3): 7-11.
- [13] 马文礼, 李治平, 孙玉平, 等. 基于机器学习的页岩气产能非确定性预测方法研究[J]. 特种油气藏, 2019, 26(2): 101-105.
- [14] 石广仁. 地学数据挖掘与知识发现[M]. 北京: 石油工业出版社, 2012.
- [15] 檀朝东, 陈见成, 刘志海, 等. 大数据挖掘技术在石油工程的应用前景展望[J]. 中国石油和化工, 2015, 3(1):

- 49-51.
- [16] 孙敬, 刘德华, 董春艳. 大数据挖掘技术评价致密气藏水平井产能[J]. 特种油气藏, 2016, 23(5): 74-77.
- [17] 李大伟, 石广仁. 油气勘探开发常用数据挖掘算法优选[J]. 石油学报, 2018, 39(2): 240-246.
- [18] 喻思羽, 李少华, 何幼斌, 等. 基于样式降维聚类的多点地质统计建模算法[J]. 石油学报, 2016, 37(11): 1403-1409.
- [19] 姜治生, 李长生. 不停井间抽技术在葡萄花油田的应用与认识[J]. 石油石化节能, 2018, 8(3): 12-15.
- [20] 于海山, 赵君峰. 低渗透油田机采举升方式优化评价[J]. 石油石化节能, 2018, 8(4): 7-10.
- [21] 邓吉彬. 抽油机井节能优化统一模型[J]. 石油钻采工艺, 2016, 38(6): 842-847.

(编辑 杜建波)