

doi:10.3969/j.issn.1674-8530.13.0225

基于支持向量机的地面驱动螺杆泵井 工况诊断技术

刘广东

(胜利油田东胜精攻石油开发集团股份有限公司, 山东 东营 257000)



刘广东

摘要: 地面驱动螺杆泵井举升工艺的故障频发,限制了其进一步发展. 为了提高地面驱动螺杆泵井的经济效益和管理水平,提出了基于支持向量机的地面驱动螺杆泵井工况诊断技术. 选取产量、动液面等8个表征油井工作状态的变量作为输入参数,常见10种螺杆泵井工况作为输出参数,以东胜公司金家油田已存故障地面驱动螺杆泵井为基础,建立诊断样本集,采用投票法建立 $C_{10}^2 = 45$ 个子分类器,基于网格寻优、遗传算法和粒子群寻优算法对 C, g 进行优化计算,借助 Matlab 调用 Libsvm 工具箱对支持向量机模型进行训练,利用东胜公司金家油田15口地面驱动螺杆泵井进行诊断验证,并与人工神经网络方法进行了对比. 结果表明:采用支持向量机诊断方法诊断正确的有14口油井,诊断正确率为93.33%,与人工神经网络方法(88.90%)相比具有更高的精度,在小样本诊断问题中具有更强的优势,是一种切实可行的智能诊断方法.

关键词: 螺杆泵; 工况诊断; 支持向量机; Libsvm; 人工神经网络

中图分类号: S277.9; TE355.5 **文献标志码:** A **文章编号:** 1674-8530(2014)02-0125-05

刘广东. 基于支持向量机的地面驱动螺杆泵井工况诊断技术[J]. 排灌机械工程学报, 2014, 32(2): 125-129.

Working conditions diagnosis of surface-driving progressive cavity pump wells based on support vector machine

Liu Guangdong

(Dongsheng Jingtong Petroleum Development Co. Ltd., Shengli Oilfield, Dongying, Shandong 257000, China)

Abstract: The frequent faults of surface-driving progressive cavity pump (PCP) wells limit their further development. In order to improve the economic benefits and management level of surface-driving progressive cavity pump wells, a diagnosis method of working conditions of surface-driving PCP well based on support vector machine was proposed. The condition types of PCP wells were subdivided into 10 categories as outputs, and 8 variables were selected as inputs which can represent operation situation of oil wells. Sample sets were established based on the already existing failure oil wells of Jinjia oilfield in Dongsheng group company, and 45 classifiers were built using voting method. Best C and g were determined by three methods including grid optimization, genetic algorithm and particle swarm optimization. The Libsvm toolbox called by Matlab was used to establish and train the SVM model, and 15 PCP wells of Jinjia oilfield in Dongsheng group company were diagnosed to verify the SVM model, and the comparison between support vector machine and artificial neural network was conducted. The results show that the diagnosed working conditions of 14 PCP wells by the SVM method are in accord with their actual working conditions, with an accuracy of 93.33%. Compared to artificial neural network

收稿日期: 2013-11-11; 网络出版时间: 2014-01-18

网络出版地址: <http://www.cnki.net/kcms/doi/10.3969/j.issn.1674-8530.13.0225.html>

基金项目: 国家科技重大专项课题资助项目(2011ZX05011-003)

作者简介: 刘广东(1964—),男,山东德州人,高级工程师(liuguangdong.slyt@sinopec.com),主要从事采油工程工艺管理研究.

(88.9%), the SVM is more accurate, which is superior for small sample problems, and is a feasible diagnosis method for PCP wells.

Key words: progressive cavity pump; working conditions diagnosis; support vector machine; Libsvm; artificial neural network

螺杆泵井由于工况复杂、影响因素多导致其工况诊断较难实现. 目前, 国内常用的螺杆泵井诊断方法有生产参数经验诊断法、光杆受力测试诊断法和电参数测试诊断法^[1-4]. 随着人工智能技术的发展, 工况诊断也逐渐开始智能化, 它是一种全新的诊断方式. 目前神经网络方法普遍应用于机械故障诊断方面^[5], 但由于需要的样本数量多、计算易陷入局部极小等缺点, 在实际应用计算中还有一定不足. 支持向量机是最近兴起的一种人工智能方法, 在模式识别领域应用情况较好^[6-7], 它基于小样本集, 能有效解决局部极小等问题. 文中将支持向量机理论应用到地面驱动螺杆泵井的工况诊断中, 提出一种工况诊断新方法.

1 支持向量机

支持向量机是从统计学习理论的 VC 维理论和结构风险最小原则上发展起来的一种新型学习机器. 它可以在数目有限的样本数据里, 通过寻找合适的模型, 使模型的复杂程度和学习能力达到最优, 从而获得最好的学习机器推广能力. 它不仅可以处理线性情况下的二分类问题, 而且还可以通过引入核函数, 采用转换映射空间的方法, 有效地处理非线性问题, 这也使得支持向量机成为目前通用的分类器之一. 它的基本思想如图 1 所示.

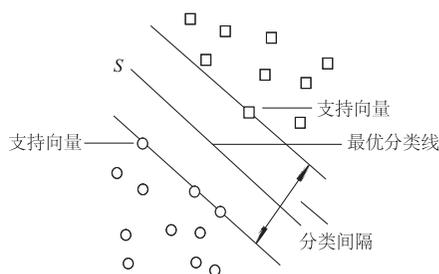


图 1 线性可分最优分类线示意图

Fig. 1 Diagram of optimal classification hyper-plane

支持向量机的主要思想是寻找 1 条最优分类线 S , 既能够正确分开两类数据样本, 又可以让两类样本之间的分类间隔最大. 在二维空间中为分类线, 当转化到高维空间时, 即为分类面.

假设有一样本数据集 (x_i, y_i) , $i = 1, \dots, n$, $x \in R^n$, $y \in \{+1, -1\}$, 用一个二次规划对偶问题来表示支持向量机分类的数学表达式:

$$\begin{cases} \max \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j), \\ \text{s. t. } \sum_{i=1}^n \alpha_i y_i = 0 \quad (\alpha_i \geq 0, i = 1, 2, \dots, n), \end{cases} \quad (1)$$

式中: n 为样本数; α_i 为拉格朗日乘子.

通过上述方程的求解, 可得到最优分类方程为

$$f(x) = \text{sgn} \left[\sum_{i=1}^n \alpha_i^* y_i (x_i \cdot x) + b^* \right], \quad (2)$$

式中: b^* 为常数; α_i^* 为拉格朗日乘子.

在实际的分类问题中, 通常是非线性不可分的, 因此, 引入核函数 $K(x_i, x_j)$, 将低维空间中的非线性问题转化成高维空间的线性可分问题. 另外, 为了控制误差, 引入惩罚因子 C . 最终, 目标函数 (1) 可以转化为

$$\begin{cases} \max \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j), \\ \text{s. t. } \sum_{i=1}^n \alpha_i y_i = 0 \quad (0 \leq \alpha_i \leq C, i = 1, 2, \dots, n), \end{cases} \quad (3)$$

相应的分类函数方程 (2) 变为

$$f(x) = \text{sgn} \left[\sum_{i=1}^n \alpha_i^* y_i K(x_i, x) + b^* \right]. \quad (4)$$

目前对于核函数的选择, 还没有理论原则可循. 对于同一个问题, 选用不同的核函数得到的效果可能差别很大. 但是一般而言, 径向基核函数不会出现太大偏差. 因此, 文中选取径向基核函数, 其表达式为

$$K(x, x_i) = \exp(-g \|x - x_i\|^2), \quad (5)$$

式中 g 为系数.

将式 (5) 带入式 (3) 后就得到了最终模型, 其中 C, g 需要进一步确定. 文中分别采用网格寻优、遗传算法寻优、粒子群算法寻优, 根据样本集进行计算, 选出分类结果正确率最高的 C, g 作为模型最终参数.

由于支持向量机只能区分二分类问题, 而文中

将地面驱动螺杆泵井的工况分为10类,属于多分类问题,因此需要进行处理.通常采用一对多、一对一(投票法)、决策导向无环图等方法^[8],文中采用投票法,即分别建立任意2种工况之间的支持向量分类机,可得到 $C_{10}^2=45$ 个分类器,判断某一生产状态属于哪种工况,则该工况得1票,最后得票最多者即为诊断得到的对应该生产状态的工况类型.

2 螺杆泵井工况诊断系统的构建

2.1 特征参数的选取

在对输入参数,即特征参数选择时,从螺杆泵井众多生产参数中找到那些能够直接表达其各组

成部分工作状况的特征,应遵循以下4个原则:①特征参数容易获取;②特征参数对输出影响大;③各特征参数之间应不相关或相关性很小;④特征参数集应尽可能包含整个油井的工作状况信息.

通过结合现场实测资料,选取产量 Q 、动液面 h_f 、测试扭矩 T_t 、测试轴向力 F_t 、是否出砂 S 、是否含蜡 W 、电流是否波动 E 和油套是否连通 TC 这8个特征参数,它们易获取,能够体现螺杆泵井工作时的状况.

对不同的工况类型进行分析,结果如表1所示.从表中可以看出,相邻2种工况之间的特征参数向量各不相同,因此理论上可以根据这8个输入参数将10种工况类型分开.

表1 工况类型分析表
Tab.1 Analysis table of working conditions types

工况类型	$Q/(t \cdot d^{-1})$	h_f/m	$T_t/(N \cdot m)$	F_t/kN	E	W	S	TC
正常	—	—	—	—	0	0	0	0
油管漏失	↓	↑	—	—	0	0	0	1
油管结蜡	—	—	↑	—	1	1	0	0
抽油杆断脱	↓	↑	↓	↓	0	0	0	0
泵漏失	↓	↑	—	↓	0	0	0	0
定子溶胀	—	—	↑	—	0	0	0	0
供液不足	↓	↓	—	—	0	0	0	0
砂卡	—	—	↑	—	1	0	1	0
油管脱落	↓	↑	↓	↓	0	0	0	1
定子脱胶	—	—	—	—	1	0	0	0

注:“↑”表示升高;“↓”表示降低;“—”表示没有变化或变化不大;“1”表示发生;“0”表示不发生

2.2 输入参数预处理

上述选取的8个特征参数只能表征某一状态下油井的生产情况是静态变量,不能反映该油井生产的动态变化,也就无法体现该油井目前的工作状况.因此需要对特征值进行预处理,从而提取油井当前生产时的工况信息.分别用 X_i 表示输入向量,预处理方法如下所示:

$$X_1 = \frac{Q}{Q_{\max}}, X_2 = \frac{h_f}{h_p}, X_3 = \frac{T_t}{T_c}, X_4 = \frac{F_t}{F_c},$$

$$X_5 = \begin{cases} 1 & \text{电流波动,} \\ 0 & \text{电流不波动,} \end{cases} X_6 = \begin{cases} 1 & \text{油井含蜡,} \\ 0 & \text{油井不含蜡,} \end{cases}$$

$$X_7 = \begin{cases} 1 & \text{油井出砂,} \\ 0 & \text{油井不出砂,} \end{cases} X_8 = \begin{cases} 1 & \text{油套连通,} \\ 0 & \text{油套不连通,} \end{cases}$$

式中: Q 为地面驱动螺杆泵井的实际产量, t/d ; Q_{\max} 为地面驱动螺杆泵井的最大产量, t/d ; h_f 为动液面, m ; h_p 为下泵深度, m ; T_t 为测试扭矩, $N \cdot m$; T_c 为计算扭矩, $N \cdot m$; F_t 为测试轴向力, kN ; F_c 为计算轴向力, kN .其中,计算扭矩和计算轴向力表征的是油井正常生产

时抽油杆的受力状态.并且,由于8个输入参数的量纲不同,上述变换也起到各参数归一化的作用,便于后期网络的计算.

2.3 输出参数预处理

根据地面驱动螺杆泵井工况分类,支持向量机的输出向量分别对应10种工况.对10种工况进行标签处理,依次分别为正常(1)、油管漏失(2)、油管结蜡(3)、抽油断脱(4)、泵漏失(5)、定子溶胀(6)、供液不足(7)、砂卡(8)、油管脱落(9)、定子脱胶(10).

3 诊断系统的实现

Libsvm^[9]工具箱是由台湾大学林智仁团队编制,可适用于多种语言调用,文中借助Matlab调用Libsvm工具箱,对建立的SVM模型进行训练、诊断,确定东胜公司地面驱动螺杆泵井工况类型,指导作业生产,提高螺杆泵井的管理水平,其计算流程如图2所示.

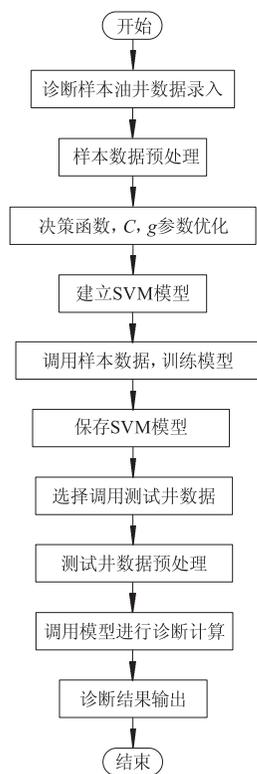


图 2 诊断设计流程图

Fig. 2 Flowchart of working condition diagnosis

4 实例计算分析

4.1 样本集建立

样本集是地面驱动螺杆泵井工况诊断的基础, 样本集的数量、范围直接影响到螺杆泵井的诊断精度. 对于样本集的选取应遵循原则: ① 应具有足够的数量; ② 应覆盖所有的工况类型.

近年来, 东胜公司金家油田进行了大量的螺杆泵转抽生产, 取得了良好的经济效益. 然而由于该区块原油黏度较大, 属于常规稠油, 导致螺杆泵运行出现了一些问题, 泵漏失、抽油杆断脱等现象比较普遍. 针对此类油井, 通过作业即可判断相应油井的故障类型. 因此文中以东胜公司金家油田早期故障螺杆泵井为基础, 建立诊断样本集, 根据支持向量机方法对目前生产螺杆泵井的工况进行诊断分析, 以便及时发现螺杆泵井潜在故障, 及时作业, 提高该区块地面驱动螺杆泵井的科学管理水平.

4.2 诊断分析

基于 Libsvm 工具箱对金家油田 J8 - X21 进行工况类型诊断分析, 该井基础数据: 油层静压为 7.14 MPa; 下泵深度为 690 m; 动液面为 634 m; 质量含水率为 50%, 生产气油比为 46 m³/m³; 采用 GLB120 - 27

螺杆泵进行生产, 转速为 30 r/min, 产液量为 1.0 t/d, 油压和套压分别为 0.1, 0.4 MPa. 2013 年 12 月油井测试数据: 测试扭矩为 271.6 N·m; 测试轴向力为 51.4 kN; 电流不波动, 油套不连通, 不含砂, 不含蜡.

根据前面支持向量机模型的设计及核函数的选取, 可以确定出最终决策函数为

$$f(x) = \text{sgn}\left[\sum_{i=1}^n \alpha_i y_i \exp(-g \|x_i - x\|^2) + b\right]. \quad (6)$$

输入训练样本后, 分别用网格寻优、遗传算法寻优和粒子群算法寻优 3 种方法确定合适的 C, g 值. 通常选取分类准确率最高的方法计算得出的 C, g 值, 当分类准确率相同时, 选取 C 值较小的一组, 以防过大的 C 值导致过拟合现象的发生, 使分类器的泛化能力降低. 因此, 通过计算选取 C, g 分别为 2.364 7, 0.982 57.

基本函数和参数 C, g 确定后, 输入训练样本即可训练得到 1 个模型. 将 J8 - X21 井的数据进行预处理, 结果如下所示:

$$\begin{aligned} & [X_1 \quad X_2 \quad X_3 \quad X_4 \quad X_5 \quad X_6 \quad X_7 \quad X_8] = \\ & [0.046 \ 9 \quad 0.972 \ 7 \quad 1.061 \ 0 \quad 1.049 \ 0 \quad 0 \quad 0 \quad 0 \\ & 0]. \end{aligned} \quad (7)$$

将预处理后的数据带入训练好的模型中, 通过 45 个二值子分类器的计算统计, 最终得到诊断仿真值为 7, 即 J8 - X21 井诊断故障为供液不足.

同理对东胜公司其他 14 口螺杆泵油井进行诊断分析, 结果如表 2 所示, 其诊断正确率为 93.33%.

表 2 东胜公司螺杆泵油井诊断结果表
Tab. 2 Working condition diagnosis results

序号	井号	诊断仿真值	诊断结果	实际工况
1	JJJ10 - 10 - X4	1	正常	正常
2	JJJ10 - 10 - X3	7	供液不足	供液不足
3	J8 - P9	1	正常	正常
4	J8 - P10	1	正常	正常
5	J8 - P7	1	正常	正常
6	J8 - X21	7	供液不足	供液不足
7	J7 - 12	1	正常	正常
8	J10 - 17X4	1	正常	正常
9	J8 - X11	2	油管漏失	泵漏失
10	JJJ10 - 10 - X5	7	供液不足	供液不足
11	GOGD9 - 908	1	正常	正常
12	GOGD7 - 906	1	供液不足	供液不足
13	GOGD14 - 905	1	正常	正常
14	GOGD9 - 31	1	正常	正常
15	GOGD9 - 19	5	泵漏失	泵漏失

注: 正常(1)、油管漏失(2)、油管结蜡(3)、抽油杆断脱(4)、泵漏失(5)、定子溶胀(6)、供液不足(7)、砂卡(8)、油管脱落(9)、定子脱胶(10)

由表2可以看出,测试的东胜公司15口生产井中,14口井诊断与实际工况相符,诊断正确率为93.33%,SVM诊断方法应用于螺杆泵井表现出了良好的精度和适应性。

基于相同输入参数和输出参数,借助人神经网络的方法对上述15口油井进行工况诊断^[10],其诊断正确率为88.90%,结果表明:文中提出的基于支持向量机的地面驱动螺杆泵井工况诊断方法较人工神经网络诊断方法具有更高的精度,在小样本诊断中具有更强的适应性,其为螺杆泵油井工况诊断提供了新的方法。

5 结 论

1) 针对传统的地面驱动螺杆泵井工况诊断方法精度低以及人工神经网络诊断方法需要大量的样本集等问题,提出了基于支持向量机的地面驱动螺杆泵井工况诊断方法,结合东胜公司15口地面驱动螺杆泵井进行了诊断分析,并与人工神经网络诊断结果进行了对比。

2) 支持向量机方法诊断正确率为93.33%,与人工神经网络方法(88.90%)相比,支持向量机方法由于其独特的统计学优势以及其在小样本诊断问题方面的适应性,具有更高的精度。

参考文献 (References)

- [1] 吴晓东,王世展,王智深. 地面驱动螺杆泵采油系统故障诊断方法[J]. 石油钻采工艺,2003,25(5):82-84.
Wu Xiaodong, Wang Shizhan, Wang Zhishen. Method to diagnose faults for oil extraction system with surface-driving progressive cavity pump used [J]. Oil Drilling & Production Technology, 2003,25(5):82-84. (in Chinese)
- [2] 王海文,陈镭,雷健. 螺杆泵井光杆受力法工况诊断技术[J]. 石油钻采工艺,2003,25(5):78-81.
Wang Haiwen, Chen Lei, Lei Jian. Diagnosis of the production system from the load borne by the polished rod [J]. Oil Drilling & Production Technology, 2003,25(5):78-81. (in Chinese)
- [3] 陈实,王海文. 螺杆泵井电参数法工况诊断技术[J]. 石油矿场机械,2007,36(2):53-55.
Chen Shi, Wang Haiwen. Diagnosis technology using electricity parameter method for production well adoption progressive cavity pump [J]. Oil Field Equipment,2007,36(2):53-55. (in Chinese)
- [4] 祖世强,白建梅,王燕嵩,等. 地面驱动螺杆泵井工况分析[J]. 石油钻采工艺,2006,28(S1):28-31.
Zu Shiqiang, Bai Jianmei, Wang Yansong, et al. Operational mode diagnosis of surface driven screw pump well [J]. Oil Drilling & Production Technology, 2006,28(S1):28-31. (in Chinese)
- [5] 何勇,张涌钢,黄德明. 基于神经网络的发动机实时检测与故障诊断系统[J]. 农业机械学报,2001,32(4):80-83.
He Yong, Zhang Yonggang, Huang Deming. Study on performance measurement and fault diagnosis of engine in real-time based on neural network [J]. Transactions of the Chinese Society for Agricultural Machinery, 2001,32(4):80-83. (in Chinese)
- [6] 张建华,冀荣华,袁雪,等. 基于径向基支持向量机的棉花虫害识别[J]. 农业机械学报,2011,42(8):178-183.
Zhang Jianhua, Ji Ronghua, Yuan Xue, et al. Recognition of pest damage for cotton leaf based on RBF-SVM algorithm [J]. Transactions of the Chinese Society for Agricultural Machinery, 2011,42(8):178-183. (in Chinese)
- [7] 刘路,王太勇,蒋永翔,等. 基于超球面支持向量机的刀具磨损状态识别[J]. 农业机械学报,2011,42(1):218-222.
Liu Lu, Wang Taiyong, Jiang Yongxiang, et al. Tool wear state recognition based on hyper-sphere support vector machine [J]. Transactions of the Chinese Society for Agricultural Machinery, 2011,42(1):218-222. (in Chinese)
- [8] Platt J C, Cristianini N, Shawe Taylor J. Large margin DAG's for multiclass classification [C]// Proceedings of Neural Information Processing Systems. Massachiest: MIT Press,2000:547-553.
- [9] Chang C C, Lin C J. LIBSVM: A library for support vector machines [J]. ACM Transactions on Intelligent Systems and Technology, 2011,2(3):27-36.
- [10] 薛建泉,李敏慧,张国栋,等. 基于BP神经网络的螺杆泵井故障诊断技术研究[J]. 西安石油大学学报:自然科学版,2013,28(3):74-77.
Xue Jianquan, Li Minhui, Zhang Guodong, et al. Fault diagnosis technique of screw pump wells based on BP neural network [J]. Journal of Xi'an Shiyou University: Natural Science Edition, 2013,28(3):74-77. (in Chinese)