

支持向量机在重复压裂中的应用研究*

刘洪¹ 赵金洲² 胡永全² 林辉³ 洪志琼⁴ 冯万奎⁴ 陈玉飞⁴

(1. 重庆石油高等专科学校 2. “油气藏地质及开发工程”国家重点实验室·西南石油学院

3. 中国石化西南油气测试中心 4. 中国石油西南油气田分公司重庆气矿)

刘洪等. 支持向量机在重复压裂中的应用研究. 天然气工业, 2004; 24(3): 75~77

摘 要 井层优选和施工方案优化是油气田重复压裂技术的核心, 其“瓶颈”问题是“数据有限”和“模型与参数给不准”以及“许多问题的机理不清楚”, 无法获得问题的显式表达。文章尝试应用统计学来获取影响压裂效果的各项因素与压裂效果的关系模型和预测模型, 从而优选施工井层和优化施工方案。实践证明, 对于样本“数据有限”(小样本)的情况下, 支持向量机算法技术适应性强、精度高, 在重复压裂研究领域中具有广阔的应用前景。

主题词 重复压裂 选井(层) 方案优化 支持向量机 模式识别

支持向量机模式识别的基本理论

(1) 支持向量机的基本思想。支持向量机的基本思想是通过用核函数(内积函数)将样本空间映射到特征空间(其维数通常比样本空间高), 在特征空间中求出原样本集的最优分类面, 得到输入变量和输出变量之间的一种非线性关系, 即找出支持向量机来进行模式分类, 而计算的复杂性却基本不增加^[1], 其基本结构见图 1。支持向量机有着严格的理论基础, 采用结构风险最小化原则, 具有很好的推广

(2) 重复压裂有三种方式: 层内压新缝, 延伸原有裂缝和改向重复压裂。延伸原有裂缝用于施工失败和改造规模不够的井, 改向重复压裂适于低渗透中高含水期。前次形成的人工裂缝、地层流体压力变化、孔隙热弹性应力、邻井注水生产活动产生的诱导应力叠加结果决定了重复压裂裂缝延伸规律。

(3) 原有裂缝失效的原因是多方面的, 主要有结蜡结垢、微粒运移和裂缝闭合。裂缝失效是诸因素综合作用的结果。

(4) 重复压裂评估包括单井状况评估、前次压裂材料评估和工艺评估, 这是重复压裂选井选层的必需程序, 同时, 选井选层应遵守有关原则。

(5) 衰竭的低渗层重复压裂可能并不经济, 必须具备足够的剩余储量和地层能量以及足够的地层

能力; 支持向量机算法是一个凸二次优化问题, 保证找到的解是全局最优解; 能较好地解决小样本、非线性、高维数等实际问题, 因此, 应用 SVM 能够用尽可能少的样本得到最优分类面, 使通过该分类面进行预测得到的代价最小, 从而获得最优良的推广性。

(2) 支持向量机算法。SVM 是从线性可分情况下的最优分类面发展而来的, 基本思想可用图 2 的两维情况说明。图 2 中实心点和空心点代表两类样本, H 为分类线, H_1 、 H_2 分别为过各类中离分类线最近的样本且平行于分类线的直线, 它们之间的距离

系数。神经网络理论和遗传算法决策模型具有较强的适应性。

参 考 文 献

- 1 王志刚, 孙玉玲. 影响低渗透油田重复压裂效果的研究. 石油学报, 1990; 11(3)
- 2 胡永全, 赵金洲, 蒲万芬. 堵老缝压新缝重复压裂技术. 西南石油学院学报, 2001; 22(3)
- 3 Hu Yongquan *et al.* Refracturing improving oil productivity in Chinese oilfield. Oil & Gas Journal, Dec, 2002
- 4 胡永全, 任书泉. 水力压裂裂缝 P3D 延伸的数值模拟. 西南石油学院学报, 1992; (2)
- 5 胡永全, 任书泉. 非对称应力分布下裂缝几何尺寸计算. 西部探矿工程, 1999; (4)

(收稿日期 2003-06-30 编辑 钟水清)

* 本文系“十五”国家重点科技攻关项目研究成果之一(项目编号: 2001BA605A-05-02-03)。

作者简介: 刘洪, 1972 年生, 博士研究生; 毕业于西南石油学院; 从事油气藏开采和增产技术等方面的科研与教学工作。地址: (400042) 重庆市渝中区石油路 1 号。电话: (023) 89092047。

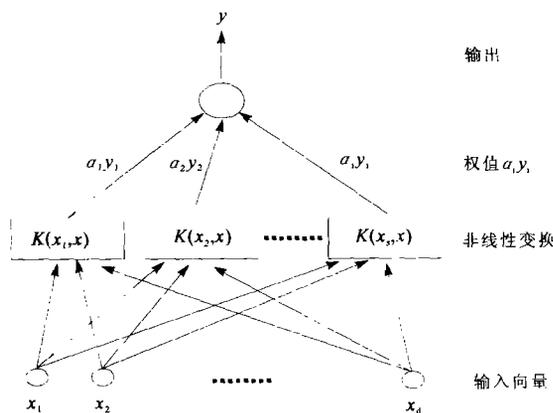


图1 支持向量机示意图

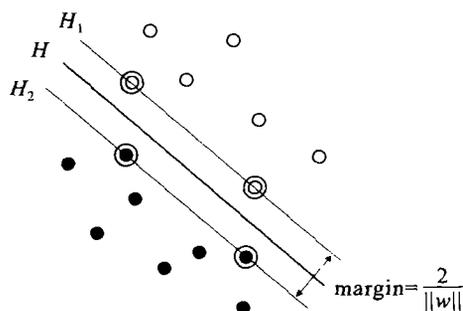


图2 最优分类面

离叫做分类间隔。所谓最优分类线就是要求分类线不但能将两类正确分开(训练错误率为0),而且使分类间隔最大,等于 $2/\|w\|$ (等价于使 $\|w\|^2$ 最小)。分类线方程为 $xw+b=0$, 对它进行归一化, 使得对线性可分的样本集 $(x_i, y_i), i=1, \dots, n, x \in R^d, y \in \{+1, -1\}$, 满足:

$$y_i[(wx_i) + b] - 1 \geq 0, \quad i = 1, \dots, n \quad (1)$$

H_1, H_2 上的训练样本点就称作支持向量。

对于非线性可分情况, 可将样本通过非线性函数映射到高维特征空间中, 使其线性可分, 再在该特征空间建立优化超平面: $w\Phi(x) + b = 0$

于是, 原样本空间的二元模式分类问题可以表示为:

$$y_i[(wx_i) + b] \geq 0, \quad i = 1, \dots, n \quad (2)$$

如果希望在经验风险和推广性能之间求得某种均衡, 可以通过引入正的松弛因子 ξ_i 来允许错分样本的存在。

$$y_i[(wx_i) + b] - 1 + \xi_i \geq 0, \quad i = 1, \dots, n \quad (3)$$

根据结构风险原则, 分类问题的最小风险界可

由下面的优化问题得到:

$$\min R(w, b, \xi) = \frac{1}{2}(w^T w) + C \sum_{i=1}^n \xi_i \quad (4)$$

通过引入 Lagrange 函数等一系列优化手段, 将上述问题转化成其对偶形式:

$$\max Q(\alpha) = -\frac{1}{2} \sum_{i,j=1}^n y_i y_j (\Phi(x_i) \Phi(x_j)) \alpha_i \alpha_j + \sum_{i=1}^n \alpha_i \quad (5)$$

$$\begin{cases} \sum_{i=1}^n y_i \alpha_i = 0 \\ 0 \leq \alpha_i \leq C \quad i = 1, 2, \dots, n \end{cases} \quad (6)$$

式中, $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$ 为 Lagrange 算子。则此为在不等式约束下的一个二次规划问题, 有唯一解。根据泛函理论, 存在内积函数 $K(x_i, x_j)$ 满足条件: $K(x_i, x_j) = \Phi(x_i) \Phi(x_j)$, 则 $K(x_i, x_j)$ 称为核函数。 α_i 即为下列满足式(6)的约束条件下的二次规划问题的解:

$$\begin{aligned} \max Q_i(\alpha, K(x_i, x_j)) = & -\frac{1}{2} \sum_{i,j=1}^n y_i y_j (K(x_i, x_j)) \alpha_i \alpha_j \\ & + \sum_{i=1}^n \alpha_i \end{aligned} \quad (7)$$

相应的最优分类决策函数为:

$$f(x) = \sum_{i=1}^k (\alpha_i - \alpha_i^*) K(x_i, x_j) + b \quad (8)$$

式中: k 是支持向量机数目, b 是为分类阈值。

重复压裂井层优选和优化施工方案设计

(1) 影响因素的确定。重复压裂井层优选和优化设计的决策参数数据的真实性、代表性和泛化性是决定评价效果的关键所在。研究参数既要包括油藏地质特征, 也要包括压裂施工、油井生产动态等方面的数据。通过综合地质、开发专家经验和反复测试分析, 我们选择下列参数: ① 候选井单井地质状况: 有效厚度、孔隙度、有效渗透率、含气饱和度、原始地层压力; ② 前次压裂施工参数: 排量、加砂量、含砂比、前置液量、携砂液量、泵压; ③ 裂缝目前状况诊断: 井底流压、支撑缝半长、裂缝导流能力; ④ 压裂井生产动态数据: 施工前产量, 施工后产量(日增气量增产有效期)。

(2) 建立决策参数数据库。重复压裂井层优选和施工方案优化设计决策数据库包括建模数据表、测试数据表和候选井数据表。所有数据表中的井均为研究区块或者研究层段上已经实施过重复压裂(或者压裂)作业的气井, 数据表中的参数为前面所

述的各项决策参数。①数据准备。根据该区块内压裂井(包括初次压裂)效果评价结果进行初步分类,通常需要考虑油气井的增产情况、增产有效期和经济评价结果来。类别的界限和数量根据油田的实际情况进行。本文在研究过程中采用日增产倍比作为重复压裂效果的评价指标,确定好评价参数和指标后,选择相应的比例来建立建模数据库。②建模数据表。建立建模数据表的关键在于选择进行建模研究井的泛化性和代表性。为了能够正确评价各个决策参数对前次压裂效果的作用效果,还必须考虑用尽可能多压裂井的各项数据。③测试数据表。测试数据表的井为该区块上的已经实施压裂作业的井,这些井可以是建模数据表中的井,也可以是进行研究的候选井,占研究总井数的15%。数据表的构成不包括模型压裂效果评价参数。④候选井数据表。候选井数据表的井,需要根据重复压裂选井选层的85/15法则,从研究的总井数中逐步筛选出15%左右最具潜能的候选井。

(3)决策参数预处理。对决策参数数据进行标准化处理,使影响资料质量的各种系统误差降到最低程度。常用的标准化处理采用归一化方法,将决策参数或者预测参数数据统一刻度在(0,1)之间,由此可以避免因各决策参数因量纲差异造成数值误差很大而给决策结果带来的负面影响。

(4)基于支持向量机的重复压裂决策。①决策模型的建立。根据建模数据, $(x_i, y_i) (i=1, 2, \dots, k)$, $x_i \in R^n$ 为影响重复压裂效果的因素, $y_i \in R$ 表示重复压裂效果。支持向量机重复压裂选井选层模型的建立,就是寻找 x_i, y_i 之间的关系:

$$f: R^n \rightarrow R; \quad y_i = f(x_i), \quad (i=1, 2, \dots, k)$$

式中: R^n 为影响重复压裂效果的因素; R 是重复压裂效果指标,即重复压裂效果。

根据支持向量机理论,重复压裂效果预测模型的建立,也即寻求表达式(8)成立,式中的 α_i, α_i^*, b 通过求解式(6)、(7)组成的二次规划问题获得。②决策模型参数的选择。学习样本集确定后,重复压裂井层优选和优化设计决策模型的建立,主要是选择相应的支持向量机参数:核函数和 C , 它们对预测结果的影响很大,它们的合理确定直接影响模型的精度和推广能力。本文通过样本数据和测试数据对

多项式核函数和径向基核函数(RBF)进行反复测试,最终确定决策模型中支持向量机参数:预测模型的核函数为径向基核函数,且 $d=3, C=1.5$ 。然后,根据方程(6)、(7)解得 α_i, α_i^*, b ,即可得到支持向量机重复压裂效果评价预测模型。③候选井预测结果。利用上面建立的模型对测试样本进行测试,根据测试结果确定模型的实用性,如果模型符合要求的精度,则表明建立的预测模型是合理的、可靠的。

结论与讨论

统计学理论是一种有限样本统计理论,是模式识别领域新近发展的一种新理论,着重研究在小样本情况下的统计规律及学习方法性质。为小样本机器学习问题建立了一个较好的理论框架,其通用的支持向量机学习算法,能够较好地解决小样本机器学习问题。

(1)对于已经有一定数量压裂井的油气田或者区块来说,利用统计学理论进行重复压裂井层优选和压裂方案优化设计,可以最大范围充分考虑各种影响压裂效果的因素,而且参数的选择和数量,可以根据油田的具体情况和研究任务的具体要求进行调节,从而使压裂施工设计更具科学性。

(2)统计学理论可以根据有限的学习样本,建立影响压裂效果的因素和压裂效果之间的非线性映射关系,并且这种映射关系具有很好的推广能力和较高的准确率。

(3)支持向量机的参数对模型的推广能力和准确率有很大的影响,而这些参数的选择主要是采用人工测试的办法获得,因此有必要和其他的搜索算法结合起来寻找最佳的参数。

参 考 文 献

- 1 Vapnik V N 著,张学工译.统计学理论的本质(第二版).北京:清华大学出版社,2000
- 2 张俊法,常剑,杨瑞敏.用综合判别方法优选压裂井层.断块油气田,2001;8(1):30~32
- 3 尚福华,唐国维,许少华.油井压裂效果预测模型的建立与应用.大庆石油地质与开发,1998;17(3):29~31

(收稿日期 2003-06-30 编辑 钟水清)