

改进 LSSVM 在水平井产能预测中的应用

肖京男 汪志明 魏建光 赵姗姗

(中国石油大学(北京)石油天然气工程学院,北京 昌平 102249)

摘要:在水平井开发可行性论证及水平井优化设计中,产能预测是重要依据,常规产能预测方法,由于样本较少和影响因素多,预测精度难以保证。为此,采用最小二乘支持向量机方法回归出预测模型进行水平井产能预测。最小二乘支持向量机对标准支持向量机进行了改进,把不等式约束改为等式约束,把误差平方和损失函数作为训练集的经验损失,把解二次规划问题转化为求解线性方程组问题,较好地解决了水平井产能预测样本少、影响因素多的问题。引入粒子群优化算法来优选最小二乘支持向量机中的参数组合,既克服了交叉验证法耗时长的缺点,又发挥了最小二乘支持向量机的小样本学习能力强和计算简单的特点。以大庆油田某一区块10口水平井的生产资料作为样本,采用最小二乘支持向量机方法回归出预测模型,对该区块两口水平井的产能进行了预测,结果表明,预测产能与实际产能的最大相对误差小于15%,能够满足工程需要。

关键词:水平井;产量预测;最小二乘支持向量机;粒子群优化;大庆油田

中图分类号:TE319 **文献标识码:**A **文章编号:**1001-0890(2010)06-0095-04

Application of LSSVM Improved Horizontal Well Productivity Prediction

Xiao Jingnan Wang Zhiming Wei Jianguang Zhao Shanshan

(College of Oil & Gas Engineering, China University of Petroleum (Beijing), Changping, Beijing, 102249, China)

Abstract: The horizontal well productivity prediction is an important basis for decision making in the feasibility study of horizontal well development and optimum design. However, it is difficult to ensure the prediction accuracy due to limited sample data and complicated influencing factors. The least squares support vector machine regression was used to predict horizontal well productivity. Least squares support vector machine improved standard SVMs in which the inequality constraints were changed to equality constraints. The square error and loss function were regarded as a training set of empirical loss. The solutions of quadratic programming problem is transformed into the problem of solving linear equations. All these solved the problems of limited sample data and the complicated impacting factors. The particle swarm optimization algorithm to optimize least square support vector machine parameters overcome the shortcomings of time-consuming in cross-validation method, but also shown the characteristics of learning ability of support vector machine's small samples and simple calculation. Using 10 horizontal wells from Daqing Oilfield as samples, the prediction model was obtained using least squares support vector machine regression, and the horizontal wells' productivity were predicted using the regressed model. The prediction results shown that the maximum relative error of prediction is less than 15% which meets the engineering requirements.

Key words: horizontal well; production forecast; least squares support vector machine; particle swarm optimization; Daqing Oilfield

水平井产能预测方法的研究始于20世纪50年代,由于影响产能的因素较多,且各种因素交互重叠,利用常规方法进行产能预测时,难以保证精度。近年来,神经网络方法^[1-2]在油气田开发中得到广泛应用。但是传统的模式识别或人工神经网络方法都要求有较多的训练样本,而实际情况中已知样本较少。支持向量机是由Vapnik等人提出的一种新型

收稿日期:2009-11-01;改回日期:2010-10-11

基金项目:国家科技重大专项“复杂结构井优化设计与控制关键技术”(编号:2009ZX05009-005)资助

作者简介:肖京男(1987—),男,山东邹城人,2007年毕业于成都理工大学石油工程专业,在读博士研究生,从事井筒复杂流动与控制、现代完井工程方面的研究。

联系方式:(010)89734958,xjn343@126.com

机器学习算法,它建立在统计学基础上,遵循结构风险最小化准则^[3],更适合小样本的预测,但采用该方法进行水平井产能预测时,存在产能预测不准、数据样本少、最小二乘支持向量机中参数难于确定等问题。为此,笔者引入粒子群优化方法对最小二乘支持向量机进行参数组合优化,建立了精度较高的回归预测模型,并对大庆油田两口水平井的初期产液量进行了预测。

1 PSO-LSSVM 回归理论

近年来,Suykens 等人^[4]提出最小二乘支持向量机方法(least squares support vector machines, LSSVM),该方法采用最小二乘线性系统作为损失函数支持向量,具有良好的泛化能力,并有效克服了神经网络的维数灾难及局部极小问题。

最小二乘支持向量机^[5-6]基于正则化理论对标准支持向量机进行了改进,将传统支持向量机中的不等式约束改为等式约束,且将误差平方和(sum squares error)损失函数作为训练集的经验损失,这样就把解二次规划问题转化为求解线性方程组问题,提高了求解问题的速度和收敛精度,较好解决了小样本、非线性、高维数、局部极小点等实际问题。

设样本为 n 维向量,训练样本集表示为: $D = \{(x_k, y_k) | k=1, 2, \dots, n\}$, 其中, $x_k \in \mathbf{R}^n$, $y_k \in \mathbf{R}$, x_k 是 n 维输入数据, y_k 是输出数据。根据非线性映射 φ 把训练样本集 φ 从原空间映射到特征空间,在这个高维特征空间中构造最优策函数:

$$y(x) = \omega \cdot \varphi(x) + b \quad (1)$$

这样就把非线性估计函数转化为高维特征空间的线性估计函数。利用结构风险最小化原则,寻找 ω, b 等价于最小化误差 R :

$$R = \frac{1}{2} \|\omega\|^2 + \frac{\gamma}{2} R_{\text{emp}} \quad (2)$$

式中: $\|\omega\|^2$ 为控制模型的复杂度; γ 为正则化参数,控制对超出误差样本的惩罚程度; R_{emp} 为误差控制函数,即 ϵ 不敏感损失函数。

常用的损失函数有线性 ϵ 损失函数、二次 ϵ 损失函数和 Huber 损失函数。选取了不同的损失函数,可构造不同形式的支持向量机。用最小二乘支持向量机将目标损失函数优化为误差 ϵ_i 的二次项,故优化问题为:

$$\min J(\omega, \epsilon) = \frac{1}{2} \omega^T \cdot \omega + \frac{\gamma}{2} \sum_{i=1}^n \epsilon_i^2 \quad (3)$$

约束条件为:

$$y_k = \omega^T \cdot \varphi(x_k) + b + \epsilon_k \quad k=1, 2, \dots, n \quad (4)$$

用拉格朗日法求解这个优化问题:

$$L(\omega, b, \epsilon, \alpha) = J(\omega, \epsilon) -$$

$$\sum_{i=1}^n \alpha_i [\omega^T \cdot \varphi(x_i) + b + \epsilon_i - y_i] \quad (5)$$

对式(5)进行优化,令 L 对 $\omega, b, \epsilon, \alpha$ 的偏导数等于零,可得以下矩阵:

$$\begin{bmatrix} 0 & \mathbf{I}_v^T \\ \mathbf{I} & \mathbf{K}(x_i, x) + \frac{2}{\gamma} \mathbf{I} \end{bmatrix} \begin{bmatrix} b \\ \epsilon \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix} \quad (6)$$

式中: $\mathbf{I} = [1, 1, \dots, 1]^T$, $\mathbf{K}(x_i, x) = \varphi(x_k) \cdot \varphi(x_i)$ 。

用最小二乘法求出 ϵ 与 b 。最小二乘支持向量机也由此得名。

为了避免高维特征空间中的“维数灾难问题”,用输入空间的一个核函数 $\mathbf{K}(x, x_i)$ 等效高维空间的内积形式解决高维计算问题,所以要求回归函数为:

$$y = \sum_{i=1}^n \alpha_i \mathbf{K}(x_i, x) + b \quad (7)$$

$\mathbf{K}(x_i, x) = \varphi(x_k) \cdot \varphi(x_i)$ 称为核函数,它是满足 Mercer 条件的任何对称核函数对应于特征空间的点积。核函数的形式包括线性核函数、径向基核函数、Sigmoid 核函数等。

从以上分析可以看出,最小二乘支持向量机仅有两个参数(正则化参数 γ 和核参数 σ),降低了算法的复杂程度。但到目前为止,如何确定最小支持向量机中正则化参数 γ 和核参数 σ 的较优组合,一直是提高最小二乘支持向量机学习和泛化能力的主要研究问题之一。在此笔者引入粒子群优化算法,对正则化参数 γ 和核参数 σ 进行优选。

粒子群算法^[7-9]是由 Kennedy 和 Eberhart 于 1995 年提出的一种进化算法,该算法主要模拟鸟群飞行觅食的行为,通过鸟之间的集体协作使群体达到目的,其算法简单,采用实数编码,不需要像遗传算法那样进行二进制编码且调节参数较少,在解决非线性、多变量、多目标约束优化问题上表现出了优越性。在粒子群优化过程中,每只鸟参考自己的飞行方向、所经历的最优方向和整个鸟群所公共认识到的最优方向来决定自己的飞行方向、速度和距离。

2 PSO-LSSVM 预测水平井产能模型

根据区块已投产油井及新投产油井的储层特性参数、流体物性参数、井筒几何参数、生产参数等,预

测新投产油井初期产液量。水平井初期产量是储层物性参数、流体特性参数和井筒几何参数等参数的函数,表示为:

$$q = f(K, h, \mu, \Delta p, \dots, L) \quad (8)$$

2.1 样本的预处理

为了消除量纲的影响,首先将样本数据和测试数据进行标准化处理。笔者利用方差和均值进行了标准化处理,使处理后的数据均值为0,标准差为1。

$$\text{std}(x) = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})^2} \quad (9)$$

$$x'_j = (x_j - \bar{x}) / \text{std}(x) \quad (10)$$

$$y'_j = (y_j - \bar{y}) / \text{std}(y) \quad (11)$$

2.2 选择模型核函数

核函数的选择对支持向量回归模型准确性的影响是十分大的。但到目前为止,选择核函数的理论还不十分完善。RBF核函数在处理非线性样本时优于多项式核函数,同时RBF核函数和多项式核函数相比具有参数少的优点,因此笔者在进行产能预测时,选用了RBF核函数。RBF核函数的表达式为:

$$K(x_i, x_j) = \exp \left\{ -\frac{\|x_i - x_j\|^2}{\sigma^2} \right\} \quad (12)$$

2.3 模型参数算法的选择

模型的主要参数是正则化参数 γ 和核参数 σ ,这两个参数在很大程度上决定了最小二乘支持向量机的学习能力和泛化能力。笔者通过粒子群寻优的方法来寻找较优的参数组合。在参数寻优过程中,关注更多的是训练模型的推广能力,为了获得满意的参数组合,以测试样本集相对误差的平均值作为依据来进行优选。

相对误差的平均值 R_{av} 为:

$$R_{\text{av}} = \frac{1}{n} \left| \frac{Y_{\text{tst}} - Y_{\text{t}}}{Y_{\text{t}}} \right| \quad (13)$$

2.4 计算步骤

1) 对群体规模 n 、迭代次数、每个粒子的位置 x_i 和速度 v_i 、最小二乘支持向量机的正则化参数 γ 和核参数 σ 进行初始化。

2) 定义适应度函数为测试样本集相对误差的平均值。

3) 根据适应度函数,调整各粒子的位置和速度。

4) 判断所有粒子最优位置的适应值或迭代次数是否满足要求,若满足,则结束计算,并保存此时粒子群的整体最优位置值。若不满足则返回2)继续计算。

5) 读取预估数据,将数据预处理后进行预测,将得到的结果进行反归一化处理后得出预测值:

$$y_j = y'_j \text{std}(y) + \bar{y} \quad (14)$$

值得注意的是,若 γ 过大,对训练样本的拟合效果很好,但是对新样本的泛化能力较差,此时会发生严重的“过学习”现象;若 γ 取值很小,会发生严重的“欠学习”现象。

3 应用实例

在大庆油田某一区块选择10口水平井生产资料(见表1)作为样本数据进行训练,再以该区块4口水平井的生产资料(见表2)作为测试样本集进行评价。以4口水平井产量的相对误差结合粒子群优化算法求出最佳参数组合,将新投产肇18-平62井和肇9-平31井的数据(见表3)输入训练好的模型进行产量预测,结果见表4。

表1 10口水平井样本数据

Table 1 Sample data of ten horizontal wells

井号	初期产液量/ $\text{m}^3 \cdot \text{d}^{-1}$	渗透率/ $10^{-3} \mu\text{m}^2$	有效厚度/ m	黏度/ $\text{mPa} \cdot \text{s}$	地层压力/ MPa	水平段长度/ m	射开长度/ m	动液面/ m
肇60-平54	10.1	16.1026	2.53	35.5	12.8	698.3	318	1 097
肇48-平33	13.5	17.2852	6.01	39.1	13.57	784.6	496	700
肇53-平37	3.4	28.6993	1.17	43.3	13.00	707.7	103	1 209
肇51-平38	34.3	48.4885	6.85	36.0	12.66	520.0	335	1 033
肇68-平21	9.1	23.4716	5.19	42.5	13.30	623.0	344	742
肇27-平36	12.6	31.3885	2.45	31.7	13.38	411.0	179	986
肇9-平27	14.3	43.8839	3.03	31.7	13.38	577.0	59	757
肇60-平33	9.1	20.6650	3.76	43.3	13.00	680.0	443	911
肇26-平26	12.0	22.7720	2.17	31.7	13.38	472.3	246	953
肇50-平31	10.6	17.8580	4.41	39.1	13.57	739.8	621	886

表2 测试样本井数据

Table 2 Tested data of four horizontal wells

井号	初期产液量/ $m^3 \cdot d^{-1}$	渗透率/ $10^{-3} \mu m^2$	有效厚度/ m	黏度/ $mPa \cdot s$	地层压力/ MPa	水平段长度/ m	射开长度/ m	动液面/ m
肇66-平21	15.0	19.572	4.865	42.5	13.30	786.00	369	752
肇24-平31	9.4	19.568	4.109	31.7	13.38	518.70	242	633
肇19-平31	12.4	14.326	3.788	31.7	13.38	499.70	215	980
肇22-平30	9.7	21.672	3.675	31.7	13.38	469.44	213	953

表3 新井待预测数据

Table 3 Data for prediction of two horizontal wells

井号	初期产液量/ $m^3 \cdot d^{-1}$	渗透率/ $10^{-3} \mu m^2$	有效厚度/ m	黏度/ $mPa \cdot s$	地层压力/ MPa	水平段长度/ m	射开长度/ m	动液面/ m
肇18-平62	12.6	5.822 6	5.67	33.9	14.06	567	238	985
肇9-平31	14.3	26.205 6	4.05	20.2	13.60	330	134	606

表4 预测结果对比

Table 4 The performance of LSSVM model

井号	产量/ $m^3 \cdot d^{-1}$		相对误差, %
	实际	预测	
肇18-平62	12.93	14.41	14.37
肇9-平31	14.30	14.12	1.20

从表4可以看出,利用粒子群优化算法对最小二乘支持向量机进行参数优化后,对新井产能预测的最大相对误差为14.37%,最小相对误差为1.20%,能够满足工程要求。

4 结论

1) 针对水平井产能预测中数据样本少的问题,引入了最小二乘支持向量机回归模型,可以充分利用有限的样本数据得到最优解。

2) 针对最小二乘支持向量机中参数优选困难的难题,引入粒子群优化算法对最小二乘支持向量机中的参数进行了优选,既克服了常用交叉验证法耗时长的缺点,又发挥了最小二乘支持向量机小样本学习能力强的特点。

3) 利用PSO-LSSVM训练好的模型,对大庆油田两口水平井的初期产液量进行了预测,结果表明,通过粒子群寻优后,提高了最小二乘支持向量机预测精度,最大相对误差小于15%,能够满足工程要求,为水平井产能预测提供了一种较为可靠的方法。

参 考 文 献

- [1] 许敏,赵粉霞. 神经网络法在气井产能预测中的应用[J]. 天然气工业, 2000, 20(3): 73-75.
Xu Min, Zhao Fenxia. Application of nerve network method to gas well productivity prediction [J]. Natural Gas Industry, 2000, 20(3): 73-75.

- [2] 田冷,何顺利,顾岱鸿,等. 改进BP神经网络模型在长庆气田产能预测中的应用[J]. 石油天然气学报, 2008, 30(5): 106-109.
Tian Leng, He Shunli, Gu Daihong, et al. Application of neural network technique for productivity evaluation in Changqing Gasfield [J]. Journal of Oil and Gas Technology, 2008, 30(5): 106-109.
- [3] 魏茂安,马海,孙正义,等. 基于支持向量机的虚拟测井声波速度重构技术研究[J]. 石油钻探技术, 2006, 34(6): 83-87.
Wei Maoan, Ma Hai, Sun Zhengyi, et al. The reconstruction of the virtual well-log acoustic velocity technology based on support vector machine[J]. Petroleum Drilling Techniques, 2006, 34(6): 83-87.
- [4] Suykens J A K, Vandewalle J. Recurrent least squares support vector machines[J]. IEEE Transactions on Circuits and Systems: I, 2000, 47(7): 1109-1114.
- [5] 邓乃扬,田英杰. 数据挖掘中的新方法:支持向量机[M]. 北京:科学出版社, 2004: 236-243.
Deng Naiyang, Tian Yingjie. Data mining in the new method: the study of support vector machine SVM[M]. Beijing: Science Press, 2004: 236-243.
- [6] 李方方,赵英凯,颜昕. 基于Matlab的最小二乘支持向量机的工具箱及其应用[J]. 计算机应用, 2006, 26(增刊2): 358-360.
Li Fangfang, Zhao Yingkai, Yan Xin. Application of least square support vector machine toolbox of Matlab[J]. Computer Applications, 2006, 26(supplement 2): 358-360.
- [7] 李炜,石连生. 基于PSO-LSSVM的研究法辛烷值预测建模[J]. 化工自动化及仪表, 2008, 35(2): 25-27.
Li Wei, Shi Liansheng. Forecasting model of research octane number based on PSO-LSSVM[J]. Control and Instruments in Chemical Industry, 2008, 35(2): 25-27.
- [8] 陈传亮,田英杰,别荣芳. 基于粒子群优化的SVR算法与BP网络的比较研究[J]. 北京师范大学学报:自然科学版, 2008, 44(5): 449-452.
Chen Chuanliang, Tian Yingjie, Bie Rongfang. Research of SVR optimized by PSO compared with BP network trained by PSO [J]. Journal of Beijing Normal University: Natural Science, 2008, 44(5): 449-452.
- [9] 熊伟丽,徐保国. 基于PSO的SVR参数优化选择方法研究[J]. 系统仿真学报, 2006, 18(9): 2442-2445.
Xiong Weili, Xu Baoguo. Study on optimization of SVR parameters selection based on PSO[J]. Journal of System Simulation, 2006, 18(9): 2442-2445.