

基于 LS-SVM 的测井物性参数的预测方法

陈 华,邓少贵,范宜仁

CHEN Hua,DENG Shao-gui,FAN Yi-ren

中国石油大学(华东),山东 东营 257061

University of Petroleum China,Dongying,Shandong 257061,China

E-mail:delaunay@163.com

CHEN Hua,DENG Shao-gui,FAN Yi-ren.Forecast method of logging physical property parameters based on LS-SVM. Computer Engineering and Applications,2007,43(23):208-210.

Abstract: Support Vector Machine(SVM) is a general machine learning method in recent years,by which good results have been obtained in fitting of small samples.Using new Support Vector Machine—Least Square Support Vector Machine(LS-SVM) to predict porosity,permeability,saturation is satisfied.The method is prone to use,it is seldom affected by uncertain factor and has powerful conformity skill of information and higher veracity in forecast.

Key words: Least Square Support Vector Regression;porosity;permeability;saturation;forecast

摘 要:支持向量机(SVM)是近年来发展起来的一种通用的机器学习方法,在小样本数据的拟合中已获得了很好的效果。采用新型的支持向量机——最小二乘支持向量机(LS-SVM)对孔隙度、渗透率和饱和度进行了预测,获得了满意的结果。该方法易于使用,很少受不确定性因素的影响,并具有较强的信息整合能力以及更高的预测准确性。

关键词:最小二乘支持向量回归机;孔隙度;渗透率;饱和度;预测

文章编号:1002-8331(2007)23-0208-03 文献标识码:A 中图分类号:TP391

1 引言

神经网络的理论基础是最小化经验误差,这种基于传统的渐进理论的学习方法,在训练样本点无穷多时是适用的^[1-3]。但是实际的情况是训练样本点数通常是有限的,有时甚至更少。支持向量机(SVM)是近年来发展起来的一种通用的机器学习方法,在小样本数据的拟合中已获得了很好的效果,在实际的工程应用中,支持向量机(SVM)^[4-6]作为一种新型的小样本建模分析工具越来越受到人们的重视,采用该方法在测井中对参数进行预测也开始得到应用^[7]。随着对支持向量机(SVM)研究的不断深入,发现在利用它进行非线性回归分析时,可比较完满地解决通用性和推广性问题。因此,深入研究 SVM 方法对于测井解释方法的研究具有重要意义。

由于支持向量机(SVM)具有良好的泛化能力,而最小二乘估计作为函数回归最基本的工具之一,在数据估计中占有举足轻重的地位,并得到了广泛应用。将最小二乘问题转化为支持向量机(SVM)形式问题加以解决,就可以保证得到的函数具有最小的预测风险,换句话说,具有最好的泛化能力。Suykens 首先提出了最小二乘支持向量机(LS-SVM)理论^[8,9],即通过将最小二乘线性系统引入支持向量机(SVM),代替传统的支持向量机(SVM)采用二次规划方法解决分类和函数估计问题。最小二乘支持向量机(LS-SVM)仍是基于多类核函数的机器学习,即

采用核函数,根据 mercer 条件,从原始空间中抽取特征,将原始空间中的样本映射为高维特征空间中的一个向量,以解决原始空间中线性不可分的问题。最小二乘支持向量机(LS-SVM)是传统支持向量机(SVM)的一种扩展,它是支持向量机(SVM)在二次损失函数下的一种形式。最小二乘支持向量机(LS-SVM)只求解线性方程,其求解速度快,在函数估计和逼近中得到了广泛应用。

2 最小二乘支持向量机(LS-SVM)的原理

Suykens 和 Vandewalle 对 Vapnik 的标准支持向量机(SVM)进行修改,提出了最小平方支持向量机(LS-SVM),其优化函数为:

$$\min_{w,b,e} J(w,b,e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^N e^2 \quad (1)$$

约束为:

$$y_k = w^T \phi(x_k) + b + e_k \quad k=1, \dots, N \quad (2)$$

相应的 Lagrange 函数表达式为:

$$L(w,b,e;\alpha) = J(w,b,e) - \sum_{k=1}^N \alpha_k \{w^T \phi(x_k) + b + e_k - y_k\} \quad (3)$$

其中, α_k 是 Lagrange 系数按照 KKT 条件:

基金项目:中国石油天然气集团公司石油科技中青年创新基金项目(No.06E1020)。

作者简介:陈华(1972-),男,讲师,博士生,主要研究方向为工业应用数学;邓少贵(1970-),副教授,博士,主要研究方向为应用地球物理;范宜仁(1962-),男,博士,教授,主要研究方向为应用地球物理。

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{k=1}^N \alpha_k \varphi(x_k) \\ \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{k=1}^N \alpha_k = 0 \\ \frac{\partial L}{\partial e_k} = 0 \rightarrow \alpha_k = \gamma e_k \\ \frac{\partial L}{\partial \alpha_k} = 0 \rightarrow w^T \varphi(x_k) + b + e_k - y_k = 0 \end{cases} \quad k=1, \dots, N$$

可以表示成下面的线性方程组:

$$\begin{pmatrix} I & 0 & 0 & -\phi(x) \\ 0 & 0 & 0 & -\mathbf{1}^T \\ 0 & 0 & \gamma I & -I \\ \phi(x)^T & \mathbf{1} & I & 0 \end{pmatrix} \begin{pmatrix} w \\ b \\ e \\ \alpha \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ y \end{pmatrix} \quad (4)$$

消除 w 和 e 后得到 $(N+1) \times (N+1)$ 的线性系统:

$$\begin{pmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & \Omega + \gamma^{-1} I \end{pmatrix} \begin{pmatrix} b \\ \alpha \end{pmatrix} = \begin{pmatrix} 0 \\ y \end{pmatrix} \quad (5)$$

其中:

$$\Omega_{k,j} = K(x_k, x_j) = \varphi(x_k)^T \varphi(x_j)$$

$$y^T = [y_1, \dots, y_N], \mathbf{1} = [1, 1, \dots, 1], \alpha^T = [\alpha_1, \dots, \alpha_N]$$

可以用不同的方法解方程(4), 后面作详细介绍。

由式(4)可得另一种方法, 即求逆方法:

$$\begin{aligned} b &= \frac{\mathbf{1}^T A^{-1} y}{\mathbf{1}^T A^{-1} \mathbf{1}} \\ \alpha &= A^{-1} (y - b \mathbf{1}) \end{aligned} \quad (6)$$

其中 $A = \Omega + \gamma^{-1} I$ 。

基于 LS-SVM 的回归预测函数变为:

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

式(7)中 α_k 和 b 由式(6)计算得到, 从而避免了二次规划问题。

3 最小二乘支持向量机 (LS-SVM) 实现算法

根据前面最小二乘支持向量回归机 (LS-SVM) 的原理和推

导过程, 其具体的非线性回归和预测实现算法如下:

(1) 把样本数据量化或标准化, 设处理后的样本集为:

$$\{(x_i, y_i) | x_i \in R^n, y_i \in R, i=1, 2, \dots, L\}$$

(2) 选择适当的正则化参数 γ 和核函数 $K(x, x')$ 。

(3) 求解线性方程组(5)。

(4) 构造回归预测函数(7)。

利用最小二乘支持向量回归机 (LS-SVM) 进行数据拟合时, 需要注意的事项主要有: 样本数据量化或标准化后, 根据实际经验或主成分分析选取合适的变量, 尽量使变量间独立性较大; 正则化参数 γ 可通过搜索法等方法自动获取^[10], 但这样做势必会增加计算量, 根据实验发现, 随着 γ 的增大, 精度越来越精确, γ 在达到某些值后, 比如下面例子中的 10 000, 误差已相对稳定, 因此, 只需把 γ 值设得稍大一点即可; LS-SVM 的计算速度和精度, 主要取决于方程组(5)的计算速度和精度, 采用合适的求解方法能提高计算速度和精度, 在下面测井物性参数的回归和预测中, 核函数采用径向基函数, 利用共轭梯度法求解效果较好。

4 基于最小二乘支持向量机 (LS-SVM) 的测井物性参数的回归和预测

储层孔隙度和渗透率是表征储油物性、建立各类地质模型的最重要参数, 在测井解释中被称为“四性”之一, 而饱和度则用来评价储集层的油气储量、含油性。对于孔隙度、渗透率和饱和度的预测方法, 从经验公式慢慢转向人工神经网络、支持向量机 (SVM) 等非线性非参数预测方法。本文尝试采用上述最小二乘支持向量回归机 (LS-SVM) 方法, 对“三低”储层中孔隙度、渗透率和饱和度进行回归预测, 获得较好效果。

为建立孔隙度、渗透率和饱和度预测模型, 选用大港油田某地区港 393 井资料进行研究, 提取该井 98 层资料中一部分可靠的、有代表性的数据作为样本进行训练和学习, 然后分别建立起测井变量到孔隙度、渗透率和饱和度值的映射关系模型, 用建立的模型对另一部分数据进行孔隙度、渗透率和饱和度预测, 检验实用效果。

表 1 训练样本数据

DEPTH	AC	CAL	CNL	DEN	GR	RLLS	RT	RXO	SP	PERM	POR	SW
3 298.275	253.938	20.731	1 330.217	2.388	65.546	5.892	6.114	5.553	48.835	8.834	0.086	0.911
3 298.400	254.404	20.707	1 298.631	2.375	64.452	6.511	6.771	5.621	48.487	10.724	0.091	0.810
3 298.650	253.295	20.627	1 223.636	2.354	61.208	7.575	7.797	6.822	47.010	13.711	0.099	0.679
3 298.775	251.562	20.620	1 189.533	2.352	59.952	7.867	7.998	7.292	46.290	13.760	0.100	0.664
3 299.025	248.226	20.640	1 137.920	2.365	59.900	7.739	7.701	7.148	44.650	11.736	0.093	0.728
3 299.150	247.124	20.656	1 134.304	2.375	60.920	7.500	7.433	7.071	43.443	9.929	0.088	0.790
3 299.400	245.647	20.641	1 166.281	2.387	62.701	7.084	7.019	7.150	42.174	8.823	0.080	0.899
3 299.525	245.301	20.626	1 183.983	2.389	62.229	7.015	6.959	7.153	41.283	8.887	0.081	0.898
3 299.775	244.338	20.610	1 190.502	2.388	57.243	7.011	6.923	7.160	38.703	10.972	0.092	0.791
3 299.900	243.485	20.612	1 182.408	2.387	54.409	6.962	6.827	7.170	37.585	12.375	0.098	0.749
3 300.150	242.193	20.625	1 177.899	2.382	50.955	6.724	6.506	7.204	34.029	14.611	0.105	0.718
3 300.275	241.925	20.630	1 188.39	2.380	49.836	6.558	6.310	7.140	32.898	15.837	0.108	0.713
3 300.525	242.446	20.618	1 228.078	2.374	48.637	6.275	5.993	6.811	30.748	18.416	0.112	0.704
3 300.650	243.064	20.581	1 251.477	2.371	48.262	6.242	5.959	6.790	29.270	19.712	0.115	0.692
3 300.900	244.173	20.534	1 274.994	2.368	48.141	6.304	6.015	6.966	28.261	20.745	0.117	0.674
3 301.025	244.121	20.522	1 267.607	2.370	48.192	6.364	6.067	7.245	28.106	20.179	0.117	0.673
3 301.275	242.362	20.540	1 219.728	2.379	48.520	6.419	6.112	7.857	29.169	17.202	0.112	0.699
3 301.400	240.584	20.562	1 186.133	2.385	49.445	6.431	6.136	7.800	30.260	15.123	0.106	0.736
3 301.650	236.461	20.682	1 097.449	2.400	52.872	6.511	6.291	7.486	35.990	10.342	0.089	0.865

表3 预测渗透率

PERM(实测值)	PERM(预测值)	绝对误差
12.416	12.603 0	0.187 0
12.943	13.153 2	0.210 2
9.054	9.152 8	0.098 8
9.699	9.656 6	0.042 4
13.559	13.531 1	0.027 9
17.110	17.162 5	0.052 5
20.599	20.480 9	0.118 1
18.992	19.118 9	0.126 9
13.038	12.830 8	0.207 2

表4 预测孔隙度

POR(实测值)	POR(预测值)	绝对误差(10 ⁻³ *)
0.096	0.095 7	0.272 6
0.098	0.097 5	0.468 7
0.083	0.082 9	0.080 7
0.085	0.085 5	0.482 6
0.102	0.101 8	0.167 8
0.110	0.110 1	0.056 4
0.116	0.116 4	0.422 9
0.115	0.115 7	0.736 2
0.098	0.098 3	0.309 2

表5 预测饱和度

SW(实测值)	SW(预测值)	绝对误差
0.731	0.726 2	0.004 8
0.684	0.683 2	0.000 8
0.856	0.856 6	0.000 6
0.853	0.851 2	0.001 8
0.729	0.729 3	0.000 3
0.710	0.710 2	0.000 2
0.681	0.682 1	0.001 1
0.680	0.678 4	0.001 6
0.790	0.787 0	0.003 0

由于孔隙度、渗透率和饱和度与测井值间的关系复杂,需要更多的附加物理测量值来综合表征,根据以往的研究资料,最终选择的测井变量主要有声波时差 AC、井径 CAL、补偿中子 CNL、密度 DEN、自然伽玛 GR、浅侧向 RLLS、深电阻率 RT、冲洗带电阻率 RXO、自然电位 SP 等,把它们作为输入值,孔隙度 POR、渗透率 PERM 和含水饱和度 SW 分别作输出值。在该层共有 28 个数据,其中 19 个作学习训练样本(表 1),9 个作预测样本(表 2)。在采用最小二乘支持向量机(LS-SVM)方法进行数据的回归和预测时,核函数采用径向基函数,径向基函数中的 δ 为 50,另外 γ 取值为 10 000。从表 3、表 4 和表 5 可以看出:用最小二乘支持向量机(LS-SVM)方法回归和预测孔隙度、渗透率和饱和度效果较好。

表2 测试样本数据

DEPTH	AC	CAL	CNL	DEN	GR	RLLS	RT	RXO	SP
3 298.525	254.235	20.672	1 260.437	2.362	62.934	7.096	7.360	6.106	47.987
3 298.900	249.687	20.626	1 158.730	2.356	59.540	7.902	7.930	7.315	45.526
3 299.275	246.246	20.652	1 146.688	2.382	62.100	7.261	7.188	7.104	42.821
3 299.650	244.957	20.611	1 192.792	2.389	60.302	7.016	6.956	7.148	39.734
3 300.025	242.713	20.616	1 176.406	2.385	52.449	6.862	6.679	7.187	36.187
3 300.400	242.016	20.63	1 205.883	2.377	49.092	6.388	6.117	6.961	31.842
3 300.775	243.714	20.555	1 269.283	2.369	48.099	6.253	5.968	6.833	28.677
3 301.150	243.454	20.525	1 248.519	2.373	48.203	6.401	6.097	7.613	28.471
3 301.525	238.393	20.599	1 147.338	2.391	50.853	6.461	6.196	7.598	32.037

5 结论

本文基于结构风险最小化原理建立了最小二乘支持向量机(LS-SVM)的孔隙度、渗透率和饱和度回归预测模型,并对大港油田某地区井目的层段孔隙度、渗透率和饱和度进行了预

(上接 207 页)

表2 采用本文提出的方法检测切变和渐变的结果

视频文件名	20041107_130001_				20041113_110001_				20041122_150000_			
	LBC_LBCNEWS2_				MSNBC_MSNBC				CNN_LIVEFROM_			
类型	切变	溶解	FOI	擦变	切变	溶解	FOI	擦变	切变	擦变	切变	擦变
ARB.mpg												
变化数	39	23	3	4	56	6	5	3	17	4		
正确数	39	21	3	4	55	5	4	3	17	4		
误检数	1	4	1	1	2	1	1	1	1	1		
漏检数	0	2	0	0	1	1	1	0	0	0		
PR	0.98	0.84	0.75	0.80	0.96	0.83	0.80	0.75	0.94	0.80		
RR	1.00	0.91	1.00	1.00	0.98	0.83	0.80	1.00	1.00	1.00		

测效果。为了增强算法的鲁棒性,可以在渐变检测中加入渐变位置线性校正,并在各种镜头变化检测的最后阶段加入相关度计算。(收稿日期:2007 年 1 月)

测,获得了满意的结果。该方法易于使用,很少受不确定性因素的影响,并具有较强的信息整合能力以及更高的预测准确性和可信度。另外,该方法对其它储集层参数的建模和预测以及岩性识别、储层分类等都具有较好的借鉴和应用价值,为非均质地层的储集层特性描述和预测提供了一条可行、有效的新途径。(收稿日期:2007 年 5 月)

参考文献:

- [1] 吕晓光,杜庆龙,曹维福.应用神经网络模型进行油层孔隙度、渗透率预测[J].大庆石油地质与开发,1996,15(3).
- [2] 郑艳辉,马力.神经网络预测木头油田储层孔隙度渗透率[J].河南石油,2002,16(3):15-17.
- [3] 张丽艳.砂砾岩储层孔隙度和渗透率预测方法[J].测井技术,2005,29(3):212-215.
- [4] 王才经.现代应用数学[M].东营:石油大学出版社,2004.
- [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.
- [6] Vapnik V N.Statistical learning theory[M].New York:John Wiley,1998.
- [7] 杨斌,匡立春,施泽进,等.一种基于核学习的储集层渗透率预测新方法[J].物探化探计算技术,2005,27(2):120-123.
- [8] Trejo R J.Kernel partial least squares regression in reproducing kernel Hilbert space[J].Machine Learning Research,2001,2:97-123.
- [9] 张晓晖,朱家元,张恒喜.基于 LS-SVM 的小样本费用智能预测[J].计算机工程与应用,2004,40(27):203-204.
- [10] 郭辉,刘贺平,王玲.最小二乘支持向量机参数选择方法及其应用研究[J].系统仿真学报,2006,18(7):2033-2036.

参考文献:

- [1] 庄越挺,潘云鹤,吴飞.网上多媒体信息分析与检索[M].北京:清华大学出版社,2002:73-75.
- [2] Bescós J, Menéndez J M, Cisneros G, et al. A unified approach to gradual shot transition detection [C]/Proc ICIP'2000 Vancouver, BC, Canada, Sep 2000:10-13.
- [3] Drew M S, Li Ze-nian, Zhong Xiang. Video dissolve and wipe detection via spatio-temporal images of chromatic histogram differences[C]/ICIP00, 2000, III:929-932.
- [4] Yuan Jin-hui, Zheng Wu-jie, Chen Le, et al. Tsinghua University at TRECVID 2004: shot boundary detection and high-level feature extraction. Beijing, P R China. 2004.
- [5] Wu M, Wolf W, Liu B. An algorithm for wipe detection [C]/IEEE Int Conf Image Processing, 1998:893-897.
- [6] Zhang H J, Kankanhalli A, Smoliar S W. Automatic partitioning of full-motion Video[J]. Multimedia Systems, 1993, 1:110-28.