

# 基于 PCA 与支持向量回归的储层渗透率预测

潘华贤,程国建,蔡磊

PAN Hua-xian, CHENG Guo-jian, CAI Lei

西安石油大学 计算机学院, 西安 710065

School of Computer Science, Xi'an Shiyu University, Xi'an 710065, China

E-mail: panhuaxian@gmail.com

PAN Hua-xian, CHENG Guo-jian, CAI Lei. Prediction of reservoir permeability based on PCA and Support Vector Regression. Computer Engineering and Applications, 2009, 45(35): 223-225.

**Abstract:** By studying of the logging technology and basic characteristics of the reservoir together with the logging parameters and core parameters which related to reservoir permeability, a reservoir permeability predicting method is introduced based on the Principal Component Analysis (PCA) and Support Vector Regression (SVR) according to the traditional predicting methods of reservoir permeability. The PCA is used to reduce the dimensionality of the logging parameters and core parameters through which the optimizing relevant parameters of permeability are chosen. Then, the optimized relevant logging parameters and core parameters are imported to the SVR to predict reservoir permeability. The experimental results show that the extracted parameters through the PCA have a good relevance with the reservoir permeability, and the SVR has a high accuracy. It has been shown the strengths and practical application of the PCA and SVR used for the prediction of the reservoir permeability.

**Key words:** Principal Component Analysis (PCA); permeability; Support Vector Regression (SVR); logging parameters; core parameters

**摘要:**在测井技术与储层基本特征研究的基础上,对与渗透率相关的测井参数和岩心参数进行了分析,根据传统的储层渗透率预测方法,提出了一种基于主成分分析与支持向量回归的储层渗透率预测方法。应用主成分分析对测井参数和岩心参数进行数据降维,优选出与渗透率最相关的参数,将优选出的测井参数和岩心参数作为支持向量回归模型的输入参数进行渗透率预测。实验结果表明,利用主成分分析算法提取的特征参数与渗透率有较好的相关性,且支持向量回归具有较高的预测精度,显示出主成分分析和支持向量回归在储层渗透率预测中的优势与实际应用价值。

**关键词:**主成分分析;渗透率;支持向量回归;测井参数;岩心参数

**DOI:** 10.3778/j.issn.1002-8331.2009.35.067 **文章编号:** 1002-8331(2009)35-0223-03 **文献标识码:** A **中图分类号:** TP305

渗透率是表征油气储层特性的一个重要参数,它代表流体的渗透能力。因此,正确认识该参数是进行油藏描述以及油藏工程研究的重要内容之一,获得准确的渗透率对石油工程师进行储层评价、产量优化、预测最终采收率具有重要的意义。

通常,根据渗透率与孔隙度的指数关系来求得渗透率, Kozeny-Carmen 方程是最典型的表征渗透率与孔隙度关系的方程,它将复杂的多孔介质简单化,并且忽略了孔隙喉道中汇聚流体以及流体通道扩张的影响<sup>[1]</sup>。然而,对于某些油气藏,渗透率与孔隙度的相关系数仅达到 0.6<sup>[2]</sup>,两者的相关性较差。因此,国内外出现了多种利用测井参数建立渗透率的经验公式,这些经验公式大部分是基于 Kozeny-Carmen 方程,它们将渗透率与测井参数或者测井参数的导出参数相关联,以获得储层渗透率。但是在非均质地层条件下,由传统统计学方法得到的经验公式并不适用于渗透率的计算,并且由于大部分经验公式都

受到线性方程的约束,以致于求得的渗透率很难与测井参数建立良好的相关关系。

神经网络是一个解决非线性、多维数据拟合的强有力工具,具有高度的非线性映射能力以及较强的自适应和自学习能力,可以较准确地预测储层物性参数,较常规的统计学方法具有较高的预测精度<sup>[3]</sup>。但是,神经网络存在模型结构复杂、参数选择困难、容易出现过拟合现象等问题。而支持向量回归 (Support Vector Regression, SVR) 能够较好地解决小样本、非线性、高维数和局部极小点等实际问题,有效地克服神经网络出现的缺陷。

该文的目的是利用支持向量回归模型预测储层渗透率,而与渗透率相关的测井参数和岩心参数的选取对于利用支持向量回归模型预测渗透率的精度影响很大。因此,文章通过使用主成分分析抽取与渗透率较相关的测井参数以及岩心参数,将其作为支持向量回归模型的输入来预测渗透率。分析研究支持

**基金项目:**国家自然科学基金(the National Natural Science Foundation of China under Grant No.40872087)。

**作者简介:**潘华贤(1984-),女,硕士研究生,研究方向为智能油藏工程、神经网络、数据挖掘、机器学习、软计算方法;程国建(1964-),男,博士,教授,研究方向为计算智能、模式识别、智能油藏工程、生物特征识别、商务智能等;蔡磊(1984-),男,硕士研究生,研究方向为智能油藏工程、图像处理、机器学习、模式识别。

**收稿日期:** 2009-07-29 **修回日期:** 2009-09-18

向量回归预测渗透率的能力。实验结果表明,支持向量回归具有较好的预测效果以及较高的可靠性,适用于储层渗透率预测。

## 1 PCA 与支持向量回归的基本原理

### 1.1 PCA

主成分分析(Principal Component Analysis, PCA)是一个广泛应用于人脸检测与图像压缩领域中的典型统计分析方法,它的核心是通过分析高维空间变量的相关性,在数据信息损失最小的前提下,将高维变量转换为少数综合的变量,这些变量称为主成分(Principal Components, PCs),其中每个主成分代表高维变量的线性组合,且各主成分之间相互独立,避免了信息重叠,第一个主成分的方差和贡献率最大,拥有的样本信息量最大。因此,在解决复杂问题时,只需考虑少数方差和贡献率较大的主成分,使问题得到简化。

### 1.2 支持向量回归

设给定的训练样本为 $\{(x_1, y_1)(x_2, y_2)\cdots(x_M, y_M)\} \subset R^n \times R$ ,  $x_1, x_2, \dots, x_M$  为输入样本,  $y_1, y_2, \dots, y_M$  为输出样本,  $M$  为样本总数,则支持向量回归的目标是寻找一个函数  $f(x)$ , 使之通过样本训练后,对于样本以外的  $x$ , 通过  $f(x)$  求出对应的输出值  $y$ 。为了得到该函数,首先利用非线性映射  $\phi(\cdot)$  将训练数据集映射到一个高维特征空间,使输入空间中的非线性问题转化为高维特征空间的线性问题,见图 1。

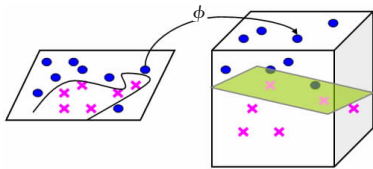


图1 低维输入空间到高维特征空间的转化

设高维空间的线性函数为<sup>[4]</sup>:

$$f(x) = \omega^T \phi(x) + b, \phi: \mathcal{X}^p \Rightarrow F; \omega \in F \quad (1)$$

其中,  $w$ 、 $b$  为回归参数,  $\phi(x)$  表示非线性函数, 它将输入空间  $x$  非线性映射到高维特征空间  $F$ 。非线性回归问题转化为在特征空间  $F$  中寻找最优回归函数。通过求解下列最小化问题来获得最优回归函数:

$$\begin{aligned} \min \Psi(\omega, \lambda) &= \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^M (\lambda_i^- + \lambda_i^+) \\ \text{s.t.} \begin{cases} (\omega \cdot x_i + b - y_i) + \lambda_i^+ \geq \varepsilon \\ (\omega \cdot x_i + b - y_i) - \lambda_i^- \leq -\varepsilon \end{cases} & \quad i=1, 2, \dots, M \end{aligned} \quad (2)$$

其中,  $C$  为惩罚系数, 用于控制产生误差时的惩罚程度。  $\lambda_i^+$ 、 $\lambda_i^-$  为松弛变量,  $\varepsilon$  为不敏感损失函数。若求得的误差值小于  $\varepsilon$ , 则表明所寻找的回归函数是最优的。

引入拉格朗日乘子, 将优化问题的求解转化为求解下列最小化问题:

$$\begin{aligned} \min \varepsilon \sum_{i=1}^M (a_i + a_i^*) + \frac{1}{2} \sum_{i,j=1}^M (a_i - a_i^*)(a_j - a_j^*)(x_i \cdot x_j) + \sum_{i=1}^M y_i (a_i - a_i^*) \\ \text{s.t.} \sum_{i=1}^M (a_i - a_i^*) = 0, 0 \leq a_i, a_i^* \leq C, a_i \cdot a_i^* = 0 \end{aligned} \quad (3)$$

其中,  $a_i$ 、 $a_i^*$  为拉格朗日乘子。引入核函数  $K(x_i, x_j)$ , 这样就得到一个等价的二次规划问题:

$$\min \varepsilon \sum_{i=1}^M (a_i + a_i^*) + \frac{1}{2} \sum_{i,j=1}^M (a_i - a_i^*)(a_j - a_j^*)K(x_i, x_j) + \sum_{i=1}^M y_i (a_i - a_i^*) \quad (4)$$

$$\text{s.t.} \sum_{i=1}^M (a_i - a_i^*) = 0, 0 \leq a_i, a_i^* \leq C, a_i \cdot a_i^* = 0$$

于是回归函数变为:

$$f(x, a_i, a_i^*) = \sum_{i=1}^M (a_i - a_i^*)K(x - x_i) + b \quad (5)$$

## 2 实验与分析

### 2.1 基于 PCA 的测井参数以及岩心参数提取

该文的实验数据采集深度为 900~1 387 m, 以 0.125 m 的间隔采集样本, 样本总数为 2 335。样本数据为测井参数以及岩心参数(包括 RILD、SP、R4\_0、GR、AC、PORT 等共 18 维)。基于 PCA 的测井参数以及岩心参数特征提取步骤如下:

(1) 设 PCA 的输入向量  $x_j \in X, i=1, 2, \dots, N(N=3 895), j=1, 2, \dots, M(M=18)$ 。

(2) 计算全体样本数据的均值  $\bar{x}_j$  以及标准差  $s_j$ , 将输入向量无量纲化  $y_j = \frac{x_j - \bar{x}_j}{s_j} \in Y$ 。

(3) 计算协方差矩阵  $C = \frac{1}{M} [Y - \bar{Y}][Y - \bar{Y}]^T$ , 其中,  $l=[1, 1, \dots, 1]_{1 \times N}, \bar{Y}=[\bar{y}_1, \bar{y}_2, \dots, \bar{y}_M]^T$ 。

(4) 根据特征方程  $(\lambda I - C)U = 0$  求解协方差矩阵  $C$  的特征值  $\lambda$  及与之对应的特征向量  $u$ 。

(5) 将协方差矩阵的特征值排序, 按照特征值的大小, 选取其对应的特征向量。

(6) 根据选取的特征向量求主成分  $T_j = y_{j1}u_{j1} + y_{j2}u_{j2} + \dots + y_{jm}u_{jm}$ ,  $j=1, 2, \dots, p, p \leq N$ 。

根据 PCA 的计算结果, 选取前 5 个主成分(自然电位 SP、自然伽玛 GR、声波时差 AC、孔隙度 POR、电阻率 RT), 因为前 5 个主成分包含了原 18 维数据信息量的 96.33%, 其方差及贡献率相对较高。因此, 将这 5 个主成分作为测井参数以及岩心参数新的综合特征向量, 将 18 维的测井参数以及岩心参数降维至 5 维。

### 2.2 利用支持向量回归预测储层渗透率

#### 2.2.1 模型的样本数据

为建立模型, 选用南 8-13 井的数据作为训练样本数据, 样本数据共 2 335 个, 分为学习样本和测试样本, 其中, 80% 的数据(1 868 个训练样本, 包括交叉验证样本)用于建模, 20% 的数据(467 个测试样本)用于模型测试。利用与南 8-13 井地质条件相同的南 9-162 井、南 10-154 井、南 11-136 井等 8 口井的部分井段样本数据验证所建模型的有效性。支持向量回归模型的输入参数为主成分数据(SP、GR、AC、POR、RT), 模型的输出参数为渗透率 PERM。

#### 2.2.2 核函数及其参数选择

核函数的选择及相关参数的确定是问题的关键所在, 不合适的参数将导致过拟合和欠拟合问题。研究发现, 高斯核函数相对于其他核函数表现出较好的性能。因此, 文章选取高斯核函数作为支持向量回归模型的核函数<sup>[5]</sup>。通过使用 LIBSVM<sup>[6]</sup>中的交叉验证功能, 以均方差和相关系数为标准, 反复测试, 最终选择惩罚系数  $C$  为 151,  $\gamma$  为 0.003 来训练回归模型。对于误差控制  $\varepsilon$ , 实验表明当  $\varepsilon=0.18$  时可满足预测需求。

表2 SVR 渗透率预测值与原始渗透率对比(部分数据)

井名	SVR 模型输入					原始渗透率/(%)	SVR 预测渗透率/(%)	绝对误差	相对误差/(%)
	SP/mV	GR/API	AC/(us·ft <sup>-1</sup> )	POR/(%)	RT/(Ω·m)				
南9-162井	50.278	92.896	252.433	20.847	14.143	15.515	17.706 10	2.191 10	14.122 460
	45.422	64.613	256.465	21.809	13.394	19.464	22.715 90	3.251 90	16.707 250
	42.250	60.175	247.994	19.789	11.497	13.027	12.866 74	0.160 26	1.2302 140
	40.416	49.182	246.072	19.330	10.281	18.498	20.953 70	2.455 70	13.275 490
	40.486	50.695	234.344	16.534	14.149	9.618 8	11.740 40	2.121 60	22.056 810
南10-154井	20.185	59.732	242.450	19.167	9.952	13.070	9.153 20	3.916 80	29.967 870
	26.269	55.972	238.749	16.547	19.658	7.765	6.606 65	1.158 35	14.917 580
	28.320	86.579	246.320	20.029	13.791	11.781	8.364 90	3.416 10	28.996 690
	27.288	47.302	267.250	24.555	9.042	61.071	67.805 20	6.734 20	11.026 840
	24.539	49.057	210.239	12.906	23.491	3.357	2.295 98	1.061 02	31.606 200
南11-136井	99.188	79.669	249.200	20.037	17.528	15.208	18.692 70	3.484 70	22.913 600
	71.062	47.903	253.100	20.267	13.156	24.445	21.149 10	3.295 90	13.482 920
	88.433	49.926	220.350	13.197	30.385	3.414	4.458 30	1.044 30	30.588 750
	59.708	41.161	251.300	20.577	16.392	35.173	42.941 77	7.768 77	22.087 310
	65.561	51.670	248.800	19.981	11.420	19.792	20.187 40	0.395 40	1.997 777

### 3 应用实例分析

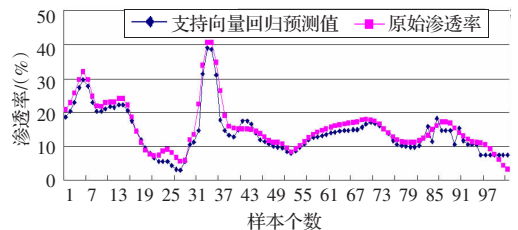
利用南8-13井的测试样本对所建立的支持向量回归模型进行测试,表1列出了训练样本与测试样本的误差对比,其中, $E_A$ 为相对误差的绝对平均值, $E_{\min}$ 为最小相对误差的绝对值, $E_{\max}$ 为最大相对误差的绝对值, $\gamma$ 为相关系数。由表1可知,测试样本的相对误差的绝对平均值仅16.9447%,并且相关系数达到了0.827175。结果表明,利用主成分分析提取的特征参数与渗透率具有较好的相关性,且支持向量回归对渗透率的预测精度高。南8-13井的支持向量回归预测值与原始渗透率对比参见图2,其中图2(a)为测试样本中前100个样本的预测值与原始渗透率曲线拟合图图2(b)为预测值与原始渗透率的交绘图。

表1 训练样本与测试样本误差对比

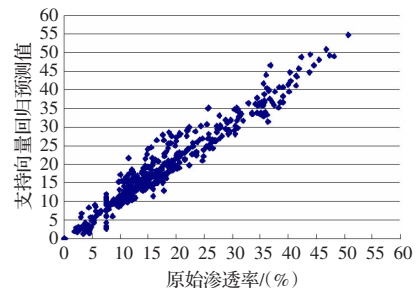
性能指标	训练样本	测试样本
样本总数	1 868	467
渗透率数据范围	0.01~58.347	0.01~54.717
$E_A$ (%)	11.949 59	16.944 7
$E_{\min}$ (%)	1.824 48e-06	0.040 7
$E_{\max}$ (%)	36.204 97	39.5132 45
均方差	21.087 9	34.165 6
$\gamma$	0.865 841	0.827 175

利用以上建立的SVR模型,对与南8-13井地质条件相同的南9-162井、南10-154井、南11-136井等8口井的部分井段样本数据进行渗透率预测。利用这些新样本进行渗透率预测来验证所建模型的有效性,实验结果表明,这8口井的预测结果与对应的原始渗透率基本一致。

表2为8口井中其中3口井的部分样本的渗透率预测结果,根据原始渗透率与SVR模型预测结果的绝对误差以及相对误差百分比两个评判标准评价模型预测渗透率的精确度,8口井对应的原始渗透率与SVR模型预测结果的绝对误差小于10,相对误差百分比在30%以内。测试结果中,虽然存在少量误差较大的情况,但是从总体上看,所建立的支持向量回归模型对渗透率的预测具有较好的预测效果,预测准确度高,说明了支持向量回归算法在储层渗透率预测上的有效性。据此,将主成分分析以及支持向量回归算法应用于储层渗透率的预测,具有很好的实际应用价值。



(a) 曲线拟合图



(b) 交绘图

图2 支持向量回归预测值与原始渗透率对比

### 4 结束语

(1)提出了一种基于主成分分析和支持向量回归的储层渗透率预测的方法。由于传统的经验公式受到孔隙度与渗透率的指数关系限制,在非均质地层条件下,无法准确确定渗透率,而该文提出的方法克服了传统经验公式的使用限制。

(2)利用PCA提取的自然电位、自然伽玛、声波时差、孔隙度、电阻率与渗透率具有良好的相关性,可以在非均质地层条件下的渗透率计算进行推广应用。

(3)通过使用支持向量回归,建立了渗透率的预测模型,将SVR的预测结果与原始渗透率进行对比。应用实例表明,SVR对渗透率具有较高的预测精度,充分发挥了SVR的优势。将支持向量回归应用于储层渗透率的预测具有实际应用价值。

### 参考文献:

- [1] 杨斌.神经网络及其在石油测井中的应用[M].北京:石油工业出版社,2005:77-83.