

多传感器融合技术在原油含水率测量中的应用

张冬至 夏伯锴 曾蕾蕾 任冬艳
中国石油大学信息与控制工程学院, 山东东营 257061
E-mail: sunny_doctor@126.com

摘要: 本文通过多传感器技术对原油含水率测量影响的多个参量进行测定, 提出基于多元非线性回归和神经网络的融合方法建立原油含水率预测模型, 并采用分段建模的方法进行改进。评价结果表明: 神经网络模型预测效果优于多元非线性回归模型, 原油含水率分段预测模型效果优于统一模型。尤其是改进的神经网络分段预测模型具有网络结构简化、收敛速度快, 泛化能力强的特点, 取得很好的拟合精度和预测效果。

关键词: 原油; 含水率; 多元非线性回归; 神经网络; 预测模型

1. 引言

原油含水率是原油生产中的重要数据, 研究油田开发状况的重要的参数指标。在原油生产和储运过程中, 原油含水率的检测十分重要, 历来倍受关注, 先后提出多种测量方法, 设计出不同形式的含水率测试仪表, 但由于原油含水率测量受到多种因素的影响, 且与其影响因素具有复杂的非线性关系, 始终缺乏一种既简便快捷、又准确实用的预测模型。本文采用多传感器技术对水分、温度、矿化度进行监测或标定, 研究了温度、矿化度等因素对原油含水率测量的影响, 对建立原油含水率预测模型进行探讨和研究。

2. 系统组成与测量原理

电磁波谐振技术原油含水率测量系统框图如图1所示。

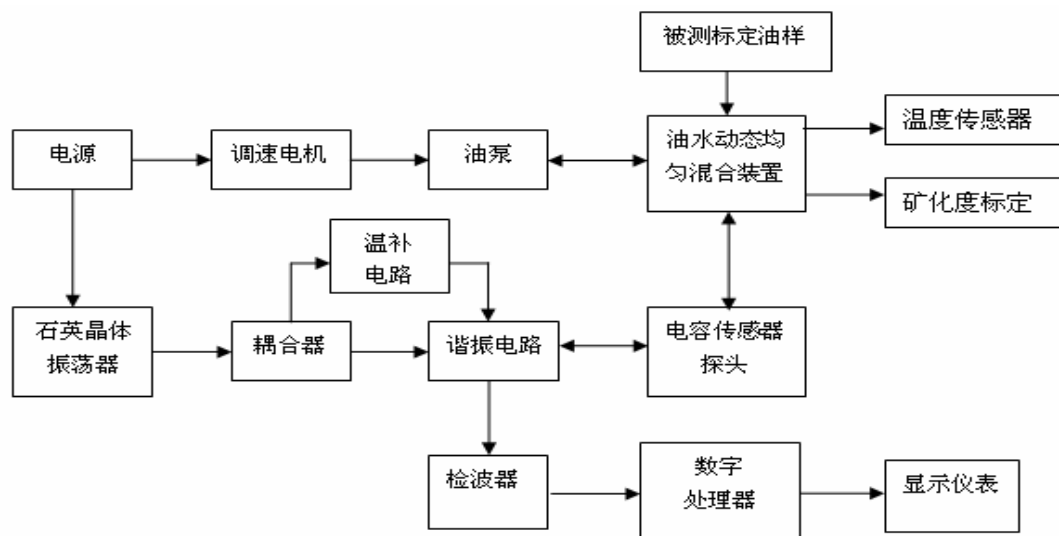


图1 电磁波谐振技术测量原油含水率原理框图

电容传感器利用油水介电常数差异较大的特性^[1]，在极化电场频率为 4×10^7 Hz下，能充分体现出水、水分子的取向极化对油水乳化液介电常数的贡献，对水分检测具有较高灵敏度。但在检测过程中，电容传感器输出电压还受到温度，矿化度，乳化液的结构、黏度，游离气等多种因素的影响。为提高水分测量精度，本文采用多传感器技术对水分、温度、矿化度进行了监测或标定。

3. 采样数据处理

用人工配比制取标定含水率的被测油样，用动态油水混合装置获得油水均匀混合的动态流体油样进行现场模拟试验。含水率配制范围0~100%，温度采集范围20℃~75℃，水矿化度（NaCl含量不同）标定范围0~50 g/L，采用多传感器技术^[2]进行参量监测，采集实验标定数据。

3.1 Grubbs 准则剔除异常数据

Grubbs准则既考虑了观测次数，又考虑了不同置信水平，鉴别能力较强，为保证测量的准确性，应用Grubbs准则对异常数据剔除。

第*i*个传感器测得数据 x_i 为一组测量列，若该测量列中的某个测量值 x_{im} 满足剔除准则

$$|x_{im} - \bar{x}_i| > g(n, 1-p)\sigma_{x_i} \quad (1)$$

则认为此测量值为异常数据，并予以剔除。这里 $g(n, 1-p)$ 中的*n*为测量数据的个数，而*P*为置信区间，本文中取 $p = 0.95$ ， $g(n, 1-p)$ 可查表得到。

其中 $\bar{x}_i = \frac{1}{n} \sum_{k=1}^n x_{ik}$ ， $\sigma_{x_i} = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (x_{ik} - \bar{x}_i)^2}$

3.2 Savitzky—Golay 滤波器平滑处理

从数据采集系统获得的分析信号，总叠加有随机噪声和干扰，在某种程度上存在着信号失真。Savitzky和Golay基于偏最小二乘原理提出的卷积平滑方式在保持分析信号有用信息方面是一种较好的卷积平滑算法^[3]。本文基于SG算法的思想设计了Savitzky—Golay滤波器，可以有效消除分析信号的高频噪声，很适合数据平滑。本文采用五点三次平滑方法，在不同温度标定点下，电容传感器测量信号平滑效果如图2所示。

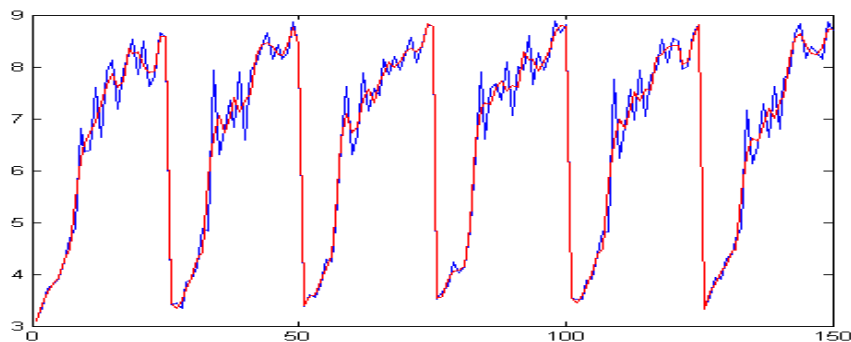


图2 电容传感器信号平滑效果

3.3 实验数据归一化处理

数据的量纲不同可能在很大程度上影响建模的效果,需对传感器测量数据进行归一化处理。测量数据的归一化处理采用式(2)处理。含水率作为目标参量,其值在0~1之间,不再进行处理,避免了反归一化转换。

$$x'_{ip} = \frac{x_{ip} - x_{i\min} + 0.1}{x_{i\max} - x_{i\min} + 0.1} \quad (2)$$

式中: x'_{ip} 为第 i 个传感器对第 p 组样本检测数据的归一化值; x_{ip} 为第 i 个传感器对第 p 组样本的检测数据或标定值。 $x_{i\max}$, $x_{i\min}$ 为第 i 个传感器输出或标定的最大、最小值。这样,对采样数据依次经过异常数据剔除,平滑及归一化处理后,建立了标准样本库。

4. 原油含水率预测模型的建立

4.1 基于多元非线性回归 (Multivariate nonlinear regression, 简称 MNLC) 的原油含水率预测模型

本文提出将非线性变换与最小二乘结合,建立基于样本矩阵非线性变换的多元回归模型,用于原油含水率预测。其基本思想是:先对自变量矩阵 X 作非线性变换,即采用自变量的非线性形式来建构拟合因变量 Y ,且待定参数为线性的,然后再用最小二乘方法处理,求解待定参数。

设多传感器测量矩阵 $X_{m \times n} = [x_1, x_2, \dots, x_n]$, 含水率标定矩阵 $Y_{m \times 1}$, 其中 m 为测量样本点个数, n 为传感器个数 (这里 $n = 3$), x_i ($i = 1, 2, \dots, n$) 为第 i 个传感器的测量向量。

本文采用多项式非线性变换的方法,在 X 矩阵中添加一个非线性部分,由各自变量的平方项和交积项来建构新的自变量矩阵 X^* ,引入待定系数矩阵 α 来拟合 X^* 与 Y 之间的数学关系,其矩阵形式为:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} & x_{11}^2 & x_{11}x_{12} & x_{12}^2 & x_{11}^2x_{13} & x_{12}x_{13} & x_{13}^2 \\ 1 & x_{21} & x_{22} & x_{23} & x_{21}^2 & x_{21}x_{22} & x_{22}^2 & x_{21}^2x_{23} & x_{22}x_{23} & x_{23}^2 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & x_{m1} & x_{m2} & x_{m3} & x_{m1}^2 & x_{m1}x_{m2} & x_{m2}^2 & x_{m1}^2x_{m3} & x_{m2}x_{m3} & x_{m3}^2 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_9 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix} \quad (3)$$

$$\text{记为: } Y = X^* \alpha + E \quad (4)$$

其中 E 为引进的残差矩阵。根据最小二乘法原理,由式(4)求得系数矩阵 α 的极小最小二乘解为: $\alpha = (X^{*T} X^*)^{-1} X^{*T} Y$ 。

利用标准样本库中的100组样本数据建立较为简洁健壮的MNLC模型,确定出 $\alpha = [0.1681 \ 0.1788 \ -0.3626 \ -0.4192 \ 0.8586 \ 0.1085 \ 0.3176 \ -0.5027 \ 0.4601 \ 0.3970]^T$ 50组样本数据用于模型预测,其拟合精度和预测效果明显优于多元线性回归(MLC)模型,即稳健性得到提高。

4.2 基于神经网络（ANN）的原油含水率预测模型

4.2.1 神经网络预测模型和算法

BP网络是目前应用最广泛的一种神经网络模型，它具有很强的非线性映射能力和泛化能力，但标准BP算法存在着收敛速度慢及容易陷入局部极小值等缺点。为此，本文采用一种收敛速度快、对初始值的设定鲁棒性强的LM算法。LM算法^[4]是一种基于标准数值优化技术的算法，是梯度下降法与高斯牛顿法的结合。在这种算法中，设 k 为迭代次数，定义误差能量函数为：

$$E(k) = \frac{1}{2} \sum_{i=1}^N e_i^2(k) = \frac{1}{2} e^T(k) e(k) \quad (6)$$

其中 $e(k) = [e_1(k), e_2(k), \dots, e_p(k)] = [e_{11}(k), e_{12}(k), \dots, e_{1p}(k), e_{21}(k), \dots, e_{pq}(k)]$ (7)

式中， p 为输出层节点的个数； q 为输入样本组数； $N = p \times q$ ； e 为网络训练误差矢量。每次迭代运算，各权值和偏置量按下式调整：

$$w(k+1) = w(k) - [J^T(w)J(w) + \mu I]^{-1} J^T(w)e(k) \quad (8)$$

其中 $J(w) = \begin{bmatrix} \frac{\partial e_1}{\partial w_1} & \frac{\partial e_1}{\partial w_2} & \dots & \frac{\partial e_1}{\partial w_m} \\ \frac{\partial e_2}{\partial w_1} & \frac{\partial e_2}{\partial w_2} & \dots & \frac{\partial e_2}{\partial w_m} \\ \dots & \dots & \dots & \dots \\ \frac{\partial e_N}{\partial w_1} & \frac{\partial e_N}{\partial w_2} & \dots & \frac{\partial e_N}{\partial w_m} \end{bmatrix}$ ，是Jacobi矩阵，它含有网络训练误差的一阶导数，

是权值和偏置值的函数； I 为单位矩阵； μ 为修正因子，在每次迭代过程中进行调整，从而使误差能量函数最终会减小到一定值，保证了该算法具有快速收敛和稳定性好的特点。

4.2.2 基于神经网络的原油含水率预测模型设计与建立

原油含水率神经网络预测模型框图如图3所示：

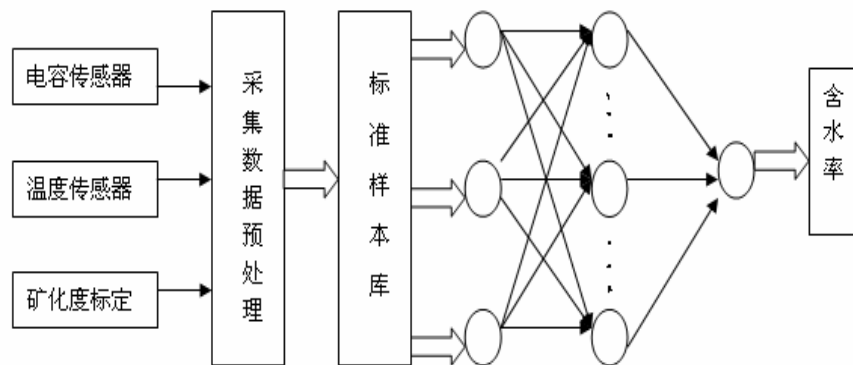


图3 原油含水率神经网络预测模型框图

在网络设计中，隐层神经元激励函数采用tan-sigmoid函数，表达式为式（9）；输出层采用sat-linear函数，表达式为式（10），防止了输出值溢出，限制在0~1之间。

$$\text{tan-sigmoid函数: } f(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \quad (9)$$

$$\text{sat-linear函数: } g(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases} \quad (10)$$

隐层节点的选择没有统一的规则可循，通过试错法确定网络拓扑结构为3-11-1。

在网络学习过程中，为解决过拟合^[5]问题，本文将样本数据分为三部分，一部分为训练集，其训练误差反向传输，用于调整权值；另一部分为监控集，不参与训练，其预测误差平方和（称为验证误差）用于控制网络的训练。当验证误差突然增大或其变化小于设定的终止阈值时，网络停止训练；其余的为测试集，用于网络模型测试。训练误差、验证误差曲线如图4所示。

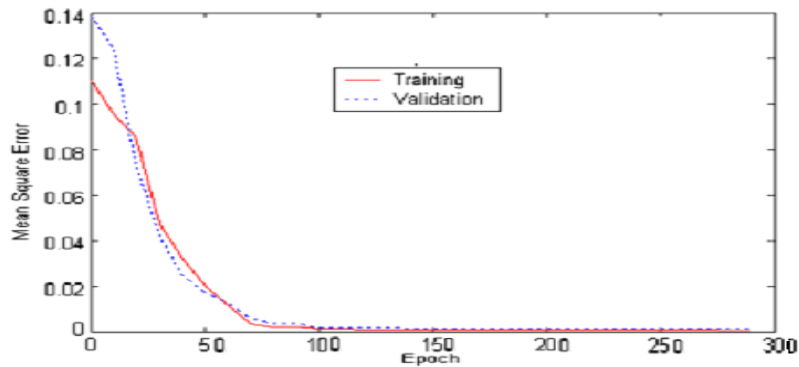


图4 神经网络模型训练误差、验证误差曲线

5. 原油含水率预测模型评价

基于多元非线性回归建立的原油含水率预测模型，避免了单传感器测量的局限性，使水分检测数据更趋于合理。该模型有利于研究温度、矿化度对含水率测量的影响关系，为克服温度、矿化度对系统的影响提供依据。

基于神经网络法的原油含水率预测模型具有较强的非线性映射能力，对多传感器数据融合处理效果良好，拟合精度和预测精度较高。表1列出了两种模型预测和标定之间的误差分析。由表1可以看出，ANN模型无论在平均相对误差（Mean RE）还是最大相对误差（Max RE）上都明显小于MNLC模型，而且误差分布也较为良好。

表1 MNLC模型和ANN模型的误差对比分析

项目	MNLC模型		ANN模型	
Mean RE	0.0681		0.0366	
Max RE	0.2162		0.1896	
RE分布比例	>10	23.3%	>10	4.8%
	>5	33.3%	>5	22.4%
	>3	18%	>3	16%
	≤3	25.4%	≤3	56.8%

6. 原油含水率分段改进预测模型

油水乳化液自身和含水率相关的物理特性（如油包水/水包油相转变，黏度等）也是原油含水率测量的影响因素，采用分段建模的方法可能会改善模型效果。本文对原油含水率的分段建模进行了探讨，根据含水率的不同，分为低含水段（ $\leq 30\%$ ）、中含水段（ $30\% \sim 70\%$ ）、高含水段（ $\geq 70\%$ ），分别采用多元非线性回归和神经网络的方法分段建模。表2列出了MNL C分段模型和ANN分段模型预测和标定之间的误差分析。

表2 MNL C分段模型和ANN分段模型的误差对比分析

含水段	MNL C模型		ANN模型	
低含水段	Max RE	0.0253	Max RE	0.00270
	Mean RE	0.0071	Mean RE	0.00053
中含水段	Max RE	0.0987	Max RE	0.0699
	Mean RE	0.0332	Mean RE	0.018
高含水段	Max RE	0.0790	Max RE	0.0281
	Mean RE	0.0354	Mean RE	0.0042

由表1和表2可以看出，MNL C分段模型和ANN分段模型和其统一预测模型相比，最大相对误差和平均相对误差均明显减小，预测精度有很大改善，可见分段预测模型在很大程度上改善了油水乳化液相转变、黏度等因素对原油含水率预测的影响。而且，改进的神经网络分段预测模型在很大程度上提高了拟合精度，并得到足够准确的预测结果，且具有网络结构简化、收敛速度快，泛化能力强的特点，是一种比较好的原油含水率预测模型。

7. 结论

- 1) 根据多传感器测量信号，采用Savitzky-Golay滤波器能达到较好的平滑滤波效果，且平滑点数的不同影响平滑效果的好坏。
- 2) 基于样本矩阵非线性变换与最小二乘结合，建立原油含水率多元非线性回归模型，稳健性明显优于多元线性回归模型。
- 3) 基于LM算法的原油含水率神经网络模型具有较强的非线性映射能力，对多传感器数据融合处理效果良好，拟合精度和预测精度较高，优于多元非线性回归模型。
- 4) 原油含水率分段预测模型，其效果优于统一模型。尤其是改进的神经网络分段预测模型具有很好的拟合精度和预测效果，且具有网络结构简化、收敛速度快，泛化能力强的特点，是一种比较好的原油含水率预测模型。

参考文献

- [1] 黄正华, 黄文霞. 石油产品含水率谐振技术取样测定法. 油气田地面工程, 2001, 20 (2): 45~46
- [2] 杨欣荣. 智能仪器原理、设计与发展. 长沙: 中南大学出版社, 2003
- [3] 梁逸曾, 俞汝勤. 化学计量学. 北京: 高等教育出版社, 2003
- [4] 董长虹. 神经网络与应用. 北京: 国防工业出版社, 2005
- [5] 李俭川, 秦国军, 温熙森等. 神经网络学习算法的过拟合问题及解决方法. 振动、测试与诊断, 2002, 22 (4): 260—264

Application of Multi-sensor Fusion Technology in Measuring of Water content ratio in Crude Oil

ZHANG Dong-zhi, XIA Bo-kai, ZENG Lei-lei, REN Dong-yan

College of Information and Control Engineering, China University of Petroleum, Shandong Donying, 257061

Abstract

Using multi-sensor technology, some parameters affecting the measurement of water content ratio of crude oil are measured, and prediction models of water content ratio of crude oil based on the methods of multivariate nonlinear regression and artificial neural networks are presented, and then being improved by subsection modeling. The assessed results show that the prediction effect of artificial neural networks model is better than that of multivariate nonlinear regression model, as well as the forecast effect of subsection model for water content ratio of crude oil is better than that of united model. In particular the improving neural network subsection prediction model can take advantage of simple network structure, fast convergence rate and strong generalization capability, and get a good modeling effect.

Key words: Crude oil; Water content ratio; Multivariate nonlinear regression; Neural networks; Prediction model

作者简介:

张冬至 (1981—), 男, 山东聊城人, 硕士研究生, 研究方向为工业过程建模仿真优化与控制, 多传感器信息融合及智能信息处理。

夏伯锴 (1956—), 男, 教授, 博士, 硕士生导师。