

基于改进即时学习算法的动液面软测量建模

王 通¹, 高宪文¹, 刘文芳²

(1. 东北大学 信息科学与工程学院, 辽宁 沈阳 110819; 2. 沈阳工业大学 电气工程学院, 辽宁 沈阳 110870)

摘 要: 油田动液面参数软测量预测应用中, 软测量模型随生产的进行会逐步退化, 导致预测结果偏差较大, 无法在油田生产过程中加以使用. 对此, 提出采用基于子空间相似度的即时学习策略来对动液面预测模型进行自适应动态更新. 通过对生产阶段数据进行子空间的相似度计算, 提高建模样本选取的准确性. 设计两个记忆参数改变以往即时学习策略模型的更新方法, 在减少计算量的同时提高动液面的预测精度. 与以往即时学习算法进行实验对比, 结果表明, 改进算法对油田动液面测量精度高, 适应性强, 符合油田生产标准, 可以应用于油田实际生产.

关 键 词: 子空间; 即时学习; 模型更新; 相似度; 动液面

中图分类号: TP 29 文献标志码: A 文章编号: 1005-3026(2015)07-0918-05

Soft Sensor for Determination of Dynamic Fluid Levels Based on Enhanced Just-in-Time Learning Algorithm

WANG Tong¹, GAO Xian-wen¹, LIU Wen-fang²

(1. School of Information Science & Engineering, Northeastern University, Shenyang 110819, China; 2. School of Electrical Engineering, Shenyang University of Technology, Shenyang 110870, China. Corresponding author: WANG Tong, E-mail: tykj_wt@126.com)

Abstract: When soft sensor model is used to predict dynamic fluid levels in oil production, it will gradually degenerate during the process, resulting in the larger deviation of prediction results and the difficulties to be used in practical oilfield production. To solve this problem, a new just-in-time model based on the similarity of subspaces was proposed to realize adaptive dynamic updates for a prediction model of dynamic fluid level. According to the production data, the similarity of subspaces was calculated to improve the accuracy of selecting modeling samples. Two memory parameters were designed to change the update method in traditional just-in-time learning model, which could reduce the amount of calculation and improve the prediction accuracy of dynamic fluid level. Compared with the traditional just-in-time learning algorithm, the improved method has better measurement accuracy and adaptation for the prediction of dynamic fluid levels. The example showed that the proposed method was fitted in with the standard of oil production and could be applied to actual production.

Key words: subspace; just-in-time learning; model update; similarity; dynamic fluid level

软测量技术近年来在化工及石油应用领域快速发展, 取得了显著的效果. 以前人们往往利用神经网络、支持向量机、多元统计等方法进行软测量建模, 为了较好地体现实际生产过程的全局特性, 在建模过程中需收集大量的样本数据. 但在油田采油生产过程中不同时期的对象特性均不同, 采

用单模型建模不但导致模型学习时间过长、过程特性匹配不佳、精度和外推能力差等缺陷^[1], 亦会导致过于庞大的模型结构和冗长的训练时间. 因此采用单一模型对其进行描述已无法满足建模预测的要求. 文献[2]提出了一种多模型建模的方法, 通过选择更相似的样本集建立的即时模型

收稿日期: 2014-05-06

基金项目: 国家自然科学基金资助项目(61433004)。

作者简介: 王 通(1976-), 男, 辽宁沈阳人, 东北大学博士研究生; 高宪文(1955-), 男, 辽宁沈阳人, 东北大学教授, 博士生导师.

能更好地辨识当前的过程. 文献[3]采用了两个或多个模型的加权平均对旧数据进行增减, 利用更新局部模型的办法来替代单一的全局模型. 文献[4-5]则提出不同的动态在线部分更新模型方法, 替代全局静态模型. 其中, 即时(just-in-time, JIT)学习算法依据“相似输入产生相似输出”原理, 建立局部模型进行预测, 并取得了较好的效果. 基于JIT学习算法的模型预测精度与样本选取的质量密切相关, 传统样本选取方法易受到干扰及人为因素的影响, 误判率较高, 影响样本选取质量, 导致模型误差较大. 在一些生产过程中, 通常采用一段时间的监控数据来判断当前的工况, 且软测量建模过程中通常存在多率现象, 而采用多点输入模型结构的解决方法并不完备. 针对上述问题, 本文提出基于子空间特征相似度的即时学习策略来动态更新软测量模型. 根据软测量中的多率现象, 采用一段时间区间的辅助变量样本作为生产工况的依据进行子空间相似度计算, 选取子空间相似的样本进行预测模型的更新, 在此基础上提出相应过程执行算法, 并通过不同生产工况的实验验证该方法的有效性.

1 JIT学习算法的局部建模算法

LSSVM(least squares support vector machine)具有快速学习能力, 计算相对简单, 特别适合在线建模. 对于一组输入样本, LSSVM利用非线性映射将训练数据集非线性地映射到一个高维特征空间, 从而使非线性函数估计问题转化为高维特征空间中的线性函数估计问题^[6]. 设建模数据集 $\Omega = \{x_1, x_2, \dots, x_n\} = \{x | d(x_i, x_{\text{new}}) < L\}$, x_i 是测试样本 x_{new} 周围包含的 n 个历史样本, $d(\cdot, \cdot)$ 为两个样本之间的相似程度, 由建模样本数据集 Ω 利用 LSSVM 算法将数据集映射到高维线性特征空间, 并构建局部最优回归函数如下:

$$y = \sum_{i=1}^n a_i k(x_{\text{new}}, x_i) + b. \quad (1)$$

式中: a 为拉格朗日乘子; b 为偏置; x_{new} 为测试时刻样本; y 为输出; $k(x_{\text{new}}, x_i)$ 为核函数 $\exp(-\|x_{\text{new}} - x_i\|^2/2\sigma^2)$.

2 生产数据集样本相似度分析

在 JIT 学习算法的局部模型建立过程中, 数据集 Ω 的选取至关重要, 如果选取样本不具备测试样本特性, 不但会影响预测精度, 还会使模型频

繁更替增大计算量. 以往建模数据多通过基于距离的相关系数进行计算, 比较有代表性的是基于欧式距离的相关算法^[7]. Fujiwara^[8]从数据的相关性出发, 取得较好的结果, 但没有考虑随机数据对于模型整体统计特性的影响. 本文结合现场生产数据特性, 为提高模型的鲁棒性和自适应性, 在原有相关性计算的基础上针对适当连续时间内数据集进行子空间相似度计算, 将随机数据及数据误差的影响降到最低.

2.1 基于子空间相似度计算

子空间可以看作是向量的集合, 传统基于欧氏距离的计算并不适合子空间计算^[9].

对于 m 维线性空间 U 和 V , 定义空间的距离 $d(U, V)$ 为

$$d(U, V) = \sqrt{\sum_{i=1}^m d^2(u_i, V)}. \quad (2)$$

其中: $d(u_i, V)$ 为向量 u_i 到子空间 V 的距离.

$$d(u_i, V) = \min_{v \in V} \|u_i - v\|. \quad (3)$$

但仅根据样本空间距离进行相似度判断, 忽略数据之间的统计特性, 有可能造成相似样本判断错误^[8]. 本文采用 PCA 相关统计指标来评价数据的相似性^[10], 设建模样本数据集 X 为

$$X = [X_1, X_2, \dots, X_n]^T, X_i \in \mathbf{R}^{m \times n}.$$

其中: X_i 为第 i 时期样本集, $i = 1, 2, \dots, n$; 测试数据为 $X_K \in \mathbf{R}^{m \times n}$. 将 X_K 进行 PCA 分解得

$$X_K = T_K P_K^T + E_K. \quad (4)$$

其中: E_K 为测试数据的残差空间, 信息量较小, 取前 p 个主元进行后续运算.

计算 Q 统计量:

$$Q_j = \|(I - P_K P_K^T) x_{ij}\|^2. \quad (5)$$

计算 T^2 统计量:

$$T_j^2 = x_{ij}^T P_K \Lambda^{-1} P_K^T x_{ij}. \quad (6)$$

其中: x_{ij} 为 X_i 第 j 行; Λ 为数据矩阵 X_K 的协方差矩阵的前 p 个特征值组成的对角矩阵. 综合两个统计量的可用信息, 定义建模样本向量与测试样本空间的相关性系数为

$$J_j = \lambda \frac{T_j^2}{T_{j,\alpha}^2} + (1 - \lambda) \frac{Q_j}{Q_{j,\alpha}}. \quad (7)$$

其中, λ 是 0 到 1 之间的权重因子. 为权衡通过 PCA 分解得到的主元与残差在向量中互相影响的数据的量.

定义样本 X_K 与 X_i 相关度为

$$J_{K,i} = \frac{1}{m} \sum_{j=1}^m J_j. \quad (8)$$

在建立局部样本时, 对新输入样本通过式

(8) 计算与原建模样本数据集之间的指标 $J_{K,i}$. 通过 $J_{K,i}$ 与原建模数据集指标控制限 \bar{J} 判断后, 如果 $J_{K,i}$ 值小, 说明测试样本与建模样本相似度高, 则局部模型不进行更新, 如果 $J_{K,i}$ 超过 \bar{J} , 选取样本集重新计算 J , 在多个样本集中选取与测试样本相似度最高的重新建模. 而传统统计指标的相似度计算建模过程中, 通过选取连续固定长度的样本集与测试数据向量进行相似度计算, 在多个样本集中选取相似度最高的样本集建模. 但样本集的选取仅是在时间轴上连续选取, 对样本集内的数据无单独处理, 不能保证样本集内数据具有很好的相似度.

2.2 基于子空间相似度计算方法实验

将样本集中的各样本分别与测试样本按式(8)进行计算. 按设计要求取 J 值最小的 L 个样本进行建模. 为验证该相似度计算方法的有效性, 做如下实验.

选取套压、流量、泵效为即时学习策略建模的辅助变量, 辅助变量采样间隔不同, 以采集频率最低的流量为基准, 20 min 采集一次, 主导变量通常为 12 h 采集一次, 辅助变量 12 h 内 36 组数据为一个样本子空间对应一个主导变量动液面. 通过对比辅助变量样本子空间的相似度, 选取生产工况相近的样本进行建模. 建模样本选取不同时间条件下的生产数据, 部分数据描述见表 1.

表 1 辅助变量样本子空间部分数据样本
Table 1 Part samples of the auxiliary variables subspace

样本采集时间	流量 $t \cdot d^{-1}$	套压 MPa	泵效 %	动液面 m
2 月 24 日	76	0.25	95	118.5
	75.8	0.25	94.8	
	\vdots	\vdots	\vdots	
	75.7	0.25	94.6	
5 月 28 日	54.5	0.23	94	80.5
	55.1	0.23	95	
	\vdots	\vdots	\vdots	
	53.9	0.24	92.9	
7 月 21 日	53.6	0.25	92.4	124.42
	54.5	0.24	94	
	\vdots	\vdots	\vdots	
	54.4	0.24	92.9	

由表 1 数据中选择 3 个不同生产工况的样本数据各 30 组作为样本库, 随机选取不同生产时期样本作为测试样本, 通过相似度计算, 选取样本库中样本建模, 选取的正确结果百分比如下:

$$R = \frac{L}{K}.$$

式中: K 为选取样本数, 值为 20; L ($L \leq K$) 为相同类型的样本数. 通过与子空间距离相似度和子空间相关性方法比较证明方法的有效性. 结果对比见表 2.

表 2 不同方法的结果比较
Table 2 Comparison of different methods

方法	正确率/%		
	分组 1	分组 2	分组 3
基于子空间距离	66.7	70.0	76.7
基于子空间相关性	96.7	93.3	86.7

由表 2 可以看出, 基于子空间相关性的计算能够较为准确地辨识出相似的样本.

3 基于子空间计算的自适应 JIT 学习算法

JIT 学习算法解决了寻找全局泛化误差最小模型的难题, 提高了建模的灵活性, 但对于每个输入数据都需重新建立局部预估模型, 导致计算量增大, 模型更新频繁. 而采油过程为慢时变过程, 由随机事件导致的模型变更, 影响了预测的效果并且增大了系统的运行负担. 通过子空间相似度计算, 滤除随机事件的干扰, 在连续的生产过程中, 动液面不会发生频繁的大范围变化. 因此, 本文针对该工况, 对于预测输入数据并不实时更新模型, 而是在原 JIT 策略的基础上, 预先判断预测模型是否合适, 需要更新模型时, 采用 LSSVM 更新模型, 降低计算量.

具体实现过程中, 为提高算法效率, 判断模型更新标准, 设计采用两个记忆参数为①预测样本记忆参数 \bar{X}_K : 保留模型更新时的预测样本; ②相似度记忆参数 \bar{J} : 保留模型更新时最小相似度参数.

记忆参数设定的作用是在下一个测试样本 X_K 与 \bar{X}_K 计算相似度后, 通过与 \bar{J} 比较, 如果 $J_{K,i}$ 小于 \bar{J} 说明原预测模型满足当前预测需要, 无须重新建模, 否则需要更新模型.

实现过程分为两步, 具体如下:

步骤 1 建立初始模型

1) 设置建模参数(d 为建模样本个数) 及计算相关度参数(m, λ);

2) 读取测试样本辅助变量 $X_K \in \mathbf{R}^{m \times n}$;

- 3) 对 \mathbf{X}_K 进行 PCA 转换,求取载荷矩阵 \mathbf{P}_K , 设置 $i = 1$;
- 4) 读取建模样本 \mathbf{X}_i ,根据式(7)计算 $J_{K,i}$,判断 $i < d$ 成立, $i + 1$ 返回 4), 否则,执行下一步;
- 5) 根据 $J_{K,i}$ 的大小按序选取建模样本, $\bar{\mathbf{X}}_K = \mathbf{X}_K$,保留选取样本的 $J_{K,i}$ 最小值到 \bar{J} ;
- 6) 提取辅助变量子空间属性的趋势均值,使之与主导变量对应;
- 7) 采用 LSSVM 算法训练模型;
- 8) 读取 \mathbf{X}_K 子空间属性的趋势均值,预测动液面.

建立初始预测模型后,根据后续的输入测试数据进行预测,同时判断是否需要递推更新模型.

步骤2 判断更新模型

- 1) 读取测试样本辅助变量 $\mathbf{X}_K \in \mathbf{R}^{m \times n}$;
- 2) 对 \mathbf{X}_K 进行 PCA 转换,求取载荷矩阵 \mathbf{P}_K ;
- 3) 计算样本 $\bar{\mathbf{X}}_K$ 与测试样本相似度 $J_{K,K}$,如果 $\bar{J} \geq J_{K,K}$,利用原有模型预测测试样本,返回 1);如果 $\bar{J} < J_{K,K}$,执行 4);
- 4) 计算原建模样本与 \mathbf{X}_K 的相似度,计算建模样本集中样本与 \mathbf{X}_K 的相似度;
- 5) 确定替换样本, $\bar{\mathbf{X}}_K = \mathbf{X}_K$,保留选取样本的 $J_{K,i}$ 最小值到 \bar{J} ;
- 6) 计算辅助变量子空间属性的趋势均值,按递推 LSSVM 相应公式更新预测模型;
- 7) 预测动液面,返回 1).

4 实验研究

通常情况下异常数据会破坏整体数据的统计特性,但基于子空间相似度计算过程能够很好地抑制异常数据所导致的建模不准的问题.通过实验对基于子空间相似度的即时学习策略建模算法进行验证,并与传统基于相关性的即时学习策略算法进行对比,从而验证其在现场应用过程中性能的优劣.实验验证所使用的数据来自辽河油田某采油平台抽油机井两年内的部分监测数据,共 660 组数据,部分数据样本同表 1 所示.

实验一:选取连续平稳生产时期内的 600 组数据作为样本集,60 组数据作测试样本,传统基于相关性的相似度算法辅助样本选取主导变量采集时对应的一组辅助变量数据.分别进行实验,实验相关参数设置如表 3 所示,LSSVM 建模过程中相关参数采用交叉验证法确定.

预测结果如图 1 所示.

表 3 实验参数
Table 3 Experimental parameters

方法	d	m	λ
基于子空间相关性	30	36	0.6
基于向量相关性	30	1	0.6

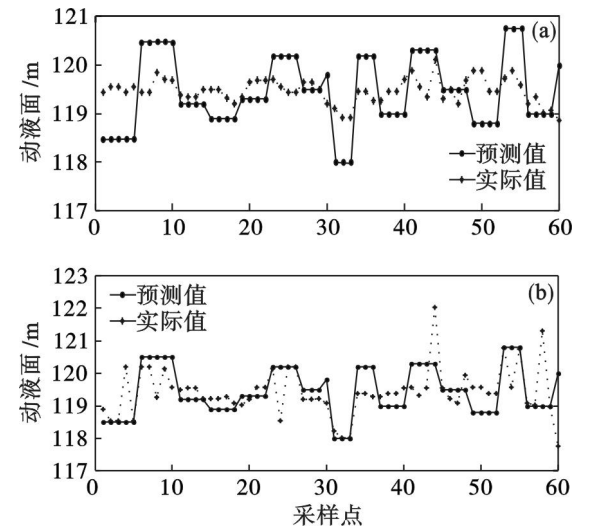


图 1 JIT 算法动液面预测
Fig. 1 JIT method prediction of dynamic fluid level
(a)—传统相关性相似度计算;
(b)—基于子空间相关性相似度计算.

从图 1a 和图 1b 对比可以看出,基于子空间相似度算法的预测结果较传统相关性的相似度算法预测结果有所改善,说明该算法的有效性.表 4 计算了两种算法的均方根误差和平均绝对误差.分析表 4 可以看出,基于子空间相似度的预测值误差减小,精度得到提高.

表 4 实验一的预测误差指标分析
Table 4 Analysis of prediction error indicators for experiment 1

指标	基于相关性的 JIT 学习算法	基于子空间相关性 的 JIT 学习算法
均方根误差	0.994 1	0.730 4
平均绝对误差	0.696 2	0.504 9

实验二:为长期稳产,在生产过程中改变采油的生产参数,选取该期间措施改变前后采油阶段的 600 组数据作样本集,60 组数据作测试样本,生产状态在措施前后发生波动,对两种方法分别进行实验,实验相关参数设置见表 3,预测结果见图 2.

从图 2a 和图 2b 对比可以看出,基于子空间相似度算法的预测结果较传统相关性的相似度算法预测结果在模型变化较大的工况下,效果提升明显,说明该算法的有效性.通过表 5 的对比可以

看出,基于子空间相似度的预测值误差减小,精度得到提高。

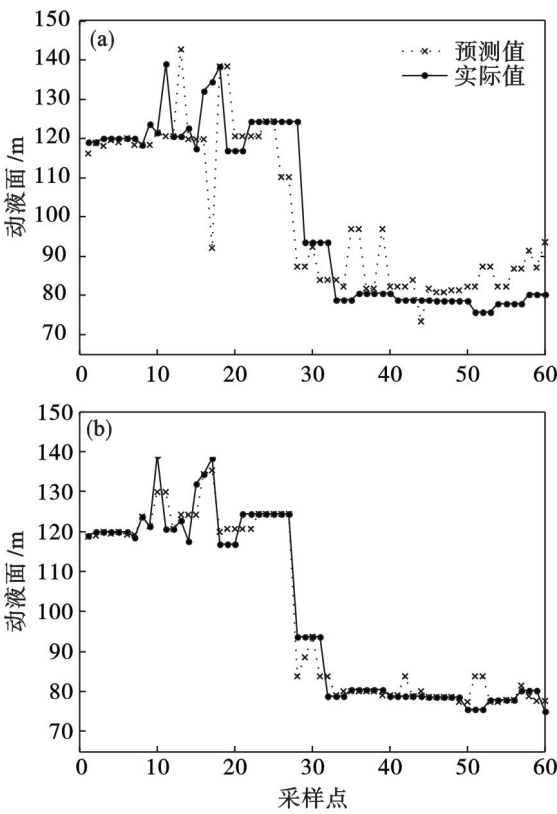


图 2 JIT 学习算法动液面预测
Fig. 2 JIT learning algorithm prediction of dynamic fluid level
(a)—传统相关性相似度计算;
(b)—基于子空间相关性相似度计算

表 5 实验二的预测误差指标分析
Table 5 Analysis of prediction error indicators for experiment 2

指标	基于相关性的 JIT 学习算法	基于子空间相关性 的 JIT 学习算法
均方根误差	11.157 2	3.586 3
平均绝对误差	7.304 9	2.132 0

5 结 论

本文采用基于子空间相似度的即时学习策略动态预测动液面数值,解决了软测量模型随生产的进行而逐步退化问题.在算法的设计中结合油田生产过程特性,运用生产阶段数据进行子空间的相似度计算,提高了建模样本选取的准确性;在解决传统 JIT 算法计算量偏大方面,提出采用两个记忆参数判断当前生产工况与模型代表工况的

相似程度,在工况发生变化时才进行模型的更新,提高动液面预测精度的同时降低了计算量.通过与传统即时学习算法的实验对比,结果表明该方法应用在油田动液面测量上精度较高,对数据波动的适应能力较强,更符合油田现场测试的要求,可以在油田动液面测量中加以应用。

参考文献:

[1] Kadlec P, Grbic R, Gabrys B. Review of adaptation mechanisms for data-driven soft sensors[J]. *Computers and Chemical Engineering*, 2011, 35(1): 1-24.

[2] 罗健旭,邵惠鹤.应用多神经网络建立动态软测量模型[J]. *化工学报*, 2003, 54(12): 170-177.
(Luo Jian-xu, Shao Hui-he. Developing dynamic soft sensors using multiple neural networks[J]. *Journal of Chemical Industry and Engineering*, 2003, 54(12): 170-177.)

[3] Du W L, Guan Z Q, Qian F. The time series soft-sensor modeling based on Adaboost LS-SVM[C]//The 8th World Congress on Intelligent Control and Automation(WCICA). Jinan, 2010: 1491-1495.

[4] Wang H Q, Li P, Gao F R, et al. Kernel classifier with adaptive structure and fixed memory for process diagnosis[J]. *AIChE Journal*, 2006, 52(10): 3515-3531.

[5] 刘毅,金福江,高增梁.时变过程在线辨识的即时递推核学习方法研究[J]. *自动化学报*, 2013, 39(5): 602-609.
(Liu Yi, Jin Fu-jiang, Gao Zeng-liang. Online identification of time-varying processes using just-in-time recursive kernel learning approach[J]. *Acta Automatica Sinica*, 2013, 39(5): 602-609.)

[6] 葛志强,刘毅,宋执环,等.一种基于局部模型的非线性多工况过程检测方法[J]. *自动化学报*, 2008, 34(7): 792-797.
(Ge Zhi-qiang, Liu Yi, Song Zhi-huan, et al. Local model based monitoring for nonlinear multiple mode process[J]. *Acta Automatica Sinica*, 2008, 34(7): 792-797.)

[7] Cheng C, Chiu M S. A new data-based methodology for nonlinear process modeling[J]. *Chemical Engineering Science*, 2004, 59(13): 2801-2810.

[8] Fujiwara K, Kano M, Hasebe S. Development of correlation-based pattern recognition algorithm and adaptive soft-sensor design[J]. *Control Engineering Practice*, 2010, 20(4): 371-378.

[9] Wang L W, Wang X, Feng J F. Subspace distance analysis with application to adaptive Bayesian algorithm for face recognition[J]. *Pattern Recognition*, 2006, 39(3): 456-464.

[10] Raich A, Cinar A. Statistical process monitoring and disturbance diagnosis in multivariable continuous processes[J]. *AIChE Journal*, 1994, 42(4): 995-1009.