多品种聚乙烯生产过程的在线检测

胡伟民　魏　雷（中海壳牌石油化工有限公司，广东　惠州　516000）

摘　要：在工业聚合反应器中，影响产品质量控制的一个主要困难是缺乏合适的聚合物性能在线测量方法，本文提出了一种预测聚乙烯共聚反应器熔体指数和密度的方法。从可用的在线温度和气体成分测量中导出了基于理论的模型来预测质量变量。这些模型中的可调参数是使用非经常性的实验室测量和递推参数估计技术在线更新的，并以一个工业反应堆的运行数据说明了这种方法的应用。结果表明，可以成功地预测熔体指数和密度，了解产品性能与预期目标的偏差，以便制造商采取纠正措施，减少所生产的劣质材料的数量，生产出一致的产品。

关键词：多品种；聚乙烯；生产过程；在线；检测

1 研究的目的

特别是在现代高密度聚乙烯（HDPE）市场上，对高质量、低价格的石油化工产品的需求越来越大。为了满足对高密度聚乙烯（HDPE）产品（如熔融指数）的各种严格要求，在等级转换操作期间保持 HDPE 性能的均匀性是很有必要的。但是，每级沉淀时间长，超调量大，以及大量不合规格的产品，阻碍了在实际生产中实现有效的级转换。在工厂运行过程中，在线测量 MI 是非常困难的。因此，通常的做法是根据可测量的关键变量来估计 Ml 值。在本研究中，我们建立了动态模型来评估高密度聚乙烯（HDPE）过程的累积 MIF。特别是，我们首先确定了四个或五个关键变量，然后推导了动态累积 EML 预测模型。特别是建立了 HDPE 装置各单元的最大似然预测模型。本文所建立的模型可以很容易地转化为瞬时最大似然预测模型。

2 现有的监测方案的优缺点

到目前为止，已经提出了大量的 MI 估计和相关方法。McAu-ley 和 MacGregor 提出的估算方法是基于各部件流量线性组合的对数。为了考虑未测量的杂质或干扰，他们采用了一个不断更新的常数项。在 Oh shima 提出的估算模型中，加入了一个表示温度的附加项，与 Mc Auley 和 Mac Gregor 的模型中采用线性组合的对数不同，它们对每个变量的对数采用线性组合。

3 基础理论

3.1 主成分分析

考虑数据矩阵 W ∈ $R^{m\times n}$ 有 m 行观测值和 n 列变量。将每列归一化为零均值和单位方差：x=（W-1 $\overline{W}$）S^{-1}，其中$\overline{W}$是平均向量，1 是元素值为 1 的列向量，S 是标准偏差的对角矩阵。协方差矩阵的伊根向量（P）可以从归一化数据集中获得。得分向量是数据矩阵 X 对每个特征向量的投影：

$$\mathbf{X}=\mathbf{t}_1\mathbf{p}_1^{\mathrm{T}}+\mathbf{t}_2\mathbf{p}_2^{\mathrm{T}}+\cdots+\mathbf{t}_K\mathbf{p}_K^{\mathrm{T}}+\mathbf{t}_{K+1}\mathbf{p}_{K+1}^{\mathrm{T}}+\cdots+\mathbf{t}_n\mathbf{p}_n^{\mathrm{T}}=\hat{\mathbf{X}}+\mathbf{E}, \tag{1}$$

其中 X 是数据矩阵 X 在由前 K 个特征向量构成的子空间中的投影，E 是与子空间正交的 X 的余数。

定义统计量 o 是为了检验 PCA 子空间要解释的新数据，Q 的置信限定义如下：

$$Q_\alpha=\Theta_1\left[\frac{c_\alpha\sqrt{2\Theta_2h_0^2}}{\Theta_1}+1+\frac{\Theta_2h_0(h_0-1)}{\Theta_1^2}\right]^{1/h_0} \tag{2}$$

$$h_0=1-\frac{2\Theta_1\Theta_3}{3\Theta_2^2},\quad \Theta_i=\sum_{j=K+1}^{n}\lambda_j^i,\ i=1,2,3$$

百分位 o 是在假设测试中出现 I 型错误的概率，cq 是对应于上百分位（1-α）的标准正态偏差。新数据与 pca 空间之差的一个热测量是统计 T2。置信限定义为：

$$T_\alpha^2=\frac{K(m-1)}{m-K}F_{K,m-1,\alpha} \tag{3}$$

其中 Fx，m-1，x 是自由度和 m-1 的 F 分布。只有当 Q ＜ o 和 l- ＜ lα 时，新数据才属于（1-z）置信限内的 PCA 子空间。

3.2 模糊 c- 均值聚类

模糊聚类的目的是将数据集 T 划分为具有模糊边界的 c 类，每个类应有自己的范数诱导矩阵，以便在一个数据集中反映不同几何形状的聚类。目标函数定义为：

$$\min_{\mathbf{U},\boldsymbol{\mu},\boldsymbol{\Sigma}} J(\mathbf{T};\mathbf{U},\boldsymbol{\mu},\boldsymbol{\Sigma})=\sum_{i=1}^{c}\sum_{j=1}^{m}(u_{ij})^mD_{ij,\boldsymbol{\Sigma}_i}^2 \tag{4}$$

$$u_{ij}\in[0,1],\quad 1\leqslant i\leqslant c,\ 1\leqslant j\leqslant m \tag{5}$$

$$\sum_{i=1}^{c}u_{ij}=1,\quad 1\leqslant j\leqslant m,\quad 0<\sum_{j=1}^{m}u_{ij}<m,\ 1\leqslant i\leqslant c$$

其中 ui 是与第 i 个集群共享的第 j 个数据点的成员值。分数向量代替原始数据用于模糊聚类。模糊化因子 q 决定了结果簇的模糊性。当它接近 1 时，簇的边界变得更清晰。利用拉格朗日乘子法，可以导出 Gustafson-Kessel 算法：

根据等式（5）中的约束条件，随机初始化隶属度。计算聚类中心和协方差：

$$\boldsymbol{\mu}_i^{(k)}=\frac{\sum_{j=1}^{m}[u_{ij}^{(k-1)}]^q\mathbf{t}_j}{\sum_{j=1}^{m}[u_{ij}^{(k-1)}]^q},\quad i=1\ldots c \tag{6}$$

$$\boldsymbol{\Sigma}_i^{(k)}=\frac{\sum_{j=1}^{m}[u_{ij}^{(k-1)}]^q(\mathbf{t}_j-\boldsymbol{\mu}_i^{(k)})^{\mathrm{T}}(\mathbf{t}_j-\boldsymbol{\mu}_i^{(k)})}{\sum_{j=1}^{m}[u_{ij}^{(k-1)}]^q} \tag{7}$$

计算距离：

$$D_{ij,\boldsymbol{\Sigma}_i}^{2(k)}=||\boldsymbol{\Sigma}_i^{(k)}||^{1/n}(\mathbf{t}_j-\boldsymbol{\mu}_i^{(k)})\boldsymbol{\Sigma}_i^{-1(k)}(\mathbf{t}_j-\boldsymbol{\mu}_i^{(k)})^{\mathrm{T}} \tag{8}$$

$$u_{ij}^{(k)}=1\Big/\sum_{l=1}^{c}(D_{ij,\boldsymbol{\Sigma}_i}^{2(k)}/D_{lj,\boldsymbol{\Sigma}_l}^{2(k)})^{2/(q-1)} \tag{9}$$

$$i=1\ldots c,\ j=1\ldots m.$$

如果成员的范数变化大于预定义公差（z），即：

$$\sum_{j=1}^{m}\sum_{i=1}^{c}||u_{ij}^{(k)}-u_{ij}^{(k-1)}||\geqslant\varepsilon,\quad k=k+1 \tag{10}$$

3.3 模糊 TS 建模

基于模糊规则的模型，将非线性系统分解为 c 个模糊规则。本文将从数据中提取的模糊聚类作为 FTS 模型的规则，每一规则都以得分向量和聚类中心的差值作为推理函数的输入。输出可以通过输入和回归矩阵 B 来评估：

R_i：If t_j is A_i then $y_{ij}=(t_j-\mu_i)B_i$ （11）

可通过解模糊获得输出：

$$\mathbf{y}_j=\sum_{i=1}^{c}u_{ij}\mathbf{y}_{ij}\Big/\sum_{i=1}^{c}u_{ij},\quad j=1\ldots m \tag{12}$$

由于 PCA 子空间的维数仅由输入决定，因此不考虑输入与输出之间的相关性。可能主成分的个数可以很好地解释输入，但忽略了 PCA 子空间以外与“噪声”有关的一些输出信息，因此本文采用交叉验证的方法来确定主成分的个数。数据被分成训练集和测试集，测试集的相对均方根误差是由训练集组成的回归模型中保留的 pcs 数的函数。

$$\mathrm{RRMSE}\equiv\left[\frac{1}{m_{test}}\sum_{i=1}^{m_{test}}\left(\frac{y_i-\hat{y}_i}{y_i}\right)^2\right]^{0.5}\times 100\% \tag{13}$$

最佳 PCs 数是产生测试集最小 RRMSEs 的 PCs 数。然而，这种方法将确定更多的 PCs 而不仅仅是使用输入数据。

4 示例

仿真数据集由四个输入变量和一个输出变量组成。输入变量（x1，x2）是自变量，其中输入数据由高斯分布生成。为了模拟多工况情况，独立变量具有不同平均值。输出变量（y）的数据由上述非线性函数生成。对于每个操作区域，生成 C1–C4120 个样本，其中 100 个样本用于训练模型，剩下的 20 个样本用作测试数据集，样本总数为 480 个。

为了说明时变的本质，第一个模型是用从 C1–C3 中平均抽取的 300 个样本建立的。线性和非线性 PLS 算法被应用于训练数据集。第一个潜在变量使用线性 PLS 的内在关系如图 1（a）所示，其中，开圆（O）表示输入 / 输出分数，实心圆（O）表示线性输入 / 输出关系。在图 1（b）中，使用基于误差的二次 PLS 提取非线性内部关系。基于误差的神经网络 PLS 算法也应用于训练数据集。回归结果如图 1（c）所示。内部关系为 1–4–1 网络结构，由交叉验证确定。模糊 PLS（Bang et al.，2003）的第一个内部关系有三个模糊规则，如下图所示。第 2（d）段。显然，建立一个全局模型来解释这三个区域的输入 / 输出数据是一个挑战，这三个区域具有高度非线性的性质。另一方面，为每个地区建立一个线性模型并总结为一个全局模型更有效。这方面的回归结果如图 1（e）和（f）所示。列出了这些方法与两个潜在变量的捕获方差。结果表明，更灵活的非线性方法（NNPLS 和 FPLS）可以改善响应变量的解释方差。然而，该方法更适合于解释数据的变化，因为通过将数据集分成若干组可以有效地缓解非线性。

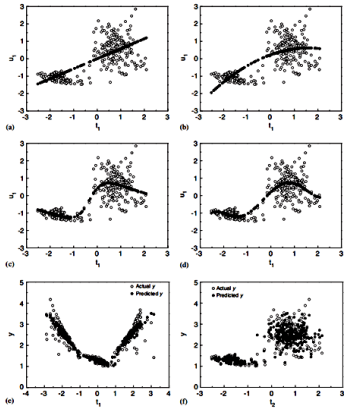

图 1

不同偏最小二乘法的第一个内在联系（a）（d），以及所提出的方法（e）和（f）的回归结果（b）二次 PLS（C）神经网络 PLS，（d）模糊 PLS，（e）第一个分数对输出，（f）第二个分数对输出。

第一个测试数据集的样本由训练数据集对应的三个不同区域组成，其中样本数为 60。NNPLS、FPLS 和所提出的方法在预测能力方面是相当的。结果表明，该方法能有效地处理输入输出数据的非线性特性，甚至优于非线性 PLS 方法。第二个训练数据集的样本是从样本数为 100 的 C4 区域抽取的，由于主成分分析（PCA）子空间不能解释新的事件，因此必须重建子空间来进行转换，训练数据的 FCM 参数可以从前一个子空间中用等式转换在新的子空间中，只需要对新事件的数据进行聚类，其中虚线和实线表示马氏距离的极限值等于 1，前一子空间上每个区域的线性模型也可以转移到新的子空间中。展示了两个训练数据集的回归结果，其中灰色圆表示使用 Eqs 从前一子空间转移的回归参数。因此，所提出的方法可以使模型适应于植物过程的时变特性，训练的数据不需要再用于建立新模型，与非线性 PLS 相比，PLS 算法必须使用所有的数据来适应全局模型。分块 RPLS 算法来适应 PLS 模型而不重用训练数据。第二个测试数据集的 RRMSE，其中样本数为 20。

5 结束语

本文提出了一种工业流化床聚乙烯反应器熔体指数和密度的在线预测方法。该方案由理论模型组成，这些模型将熔体指数和密度与反应器操作条件联系起来。当实验室分析结果可用时，模型中的可调参数通过递归预测误差方法在线更新。已经证明，这种技术能够成功地预测 MI 和密度。熔体指数和密度的在线预测为聚乙烯制造商提供了必要的信息，以便生产出更一致的聚合物。

参考文献：

[1] 赵炳阳 . 利用高密度聚乙烯装置生产线型低密度聚乙烯 [J]. 化工管理 ,2015,4(17):45.

[2] 王聪，王晶，李文光 . 相容共混物熔体的平衡转矩与组成关系的探讨 [J]. 塑料 ,2013,42(01):101-104.