

文章编号: 1001-0920(2012)06-0953-04

## 基于相空间重构和支持向量机的多相催化剂失活预测

韩晓霞, 谢刚, 韩晓明, 谢克明

(太原理工大学信息工程学院, 太原 030024)

**摘要:** 针对多相催化剂在非定态下的复杂失活机理及活性受多种因素的影响, 获取催化剂失活过程的时间序列数据非常有限, 因而降低建模效率和预测精度的情况, 提出一种基于相空间重构和支持向量机结合的非线性时间序列预测方法. 将该方法应用于甲醇氧化羰基化反应中 Cu-Si-Al 碳酸二甲酯合成催化剂失活过程建模, 仿真结果表明预测误差在满意的范围之内, 所给出的碳酸二甲酯时空收率的预测值可以为反应器的正确设计和操作以及反应过程的优化提供有效信息.

**关键词:** 支持向量机; 相空间重构; 催化剂失活; 建模; 预测

**中图分类号:** TP18

**文献标识码:** A

## Application of phase space reconstruction and support vector regression for forecasting of catalyst deactivation

HAN Xiao-xia, XIE Gang, HAN Xiao-ming, XIE Ke-ming

(College of Information Engineering, Taiyuan University of Technology, Taiyuan 030024, China. Correspondent: HAN Xiao-xia, E-mail: hanxueeryang@163.com)

**Abstract:** The catalytic performance of heterogeneous catalysts is dependent on many factors and the mechanism of catalyst deactivation is very complicated. Moreover, the limitation of getting time series data during the deactivation process of catalysts reduces the modeling efficiency and prediction precision. Therefore, a method for nonlinear time series forecasting of the catalyst deactivation based on phase space reconstruction and support vector regression is presented. The method is applied to predict the deactivation process of the Cu-Si-Al based catalysts for the synthesis of dimethyl carbonate(DMC). The simulation results show that the prediction error of catalyst deactivation model is in a range of tolerance. The prediction space-time yield value of DMC can provide important information for the design and operation of reactors as well as the optimization of the reaction conditions.

**Key words:** support vector regression; phase space reconstruction; catalyst deactivation; modeling; forecasting

### 1 引言

近年来, 国内外学者热衷于碳酸二甲酯(DMC)工艺的开发与建设, 在众多合成方法中, 煤化路线——甲醇氧化羰基化法, 被认为是合成DMC最有前途的生产方法. 然而我国至今尚未实现DMC的生产工业化, 其中合成DMC催化剂快速失活是制约该反应工业化进程的关键因素. 通过各种方法计算并建立合成催化剂的失活模型具有重要的意义.

传统的催化剂失活模型主要是应用传统理论和数值模拟的方法, 如非线性参数估值法、最小二乘回归模型法<sup>[1]</sup>. 存在的主要缺点是: 1) 评价催化剂的寿命比较困难, 往往需要很长时间; 2) 由于催化剂失活

过程的复杂性和多样性以及数值方法的局限性, 其建模精度不能完全得到保证; 3) 依靠具有固定函数形式的失活速率方程建立催化剂的失活模型, 并不能完全反映不同催化剂情况下的失活规律.

近年来, 机器学习理论, 如神经网络、支持向量机(SVR)等方法, 不用通过假设催化剂失活模型的具体形式, 而是直接从数据中建立催化剂的动力学模型, 同时加以一定的先验知识来反映催化剂失活规律<sup>[2-3]</sup>. 然而, 该方法并没有考虑催化剂失活过程中数据之间的关联, 即历史信息与未来信息之间的自相似性特点. 对此, 本文引入相空间重构(PSR)技术, 提出一种基于相空间重构和支持向量机(PSR-SVR)结合的新策

收稿日期: 2011-10-09; 修回日期: 2012-03-19.

基金项目: 国家自然科学基金项目(60975032, 20606022); 山西省教育厅科技项目(2010107, 20110007).

作者简介: 韩晓霞(1976—), 女, 讲师, 博士, 从事软计算方法、非线性系统建模与优化的研究; 谢刚(1972—), 男, 教授, 博士生导师, 从事智能信息处理、故障诊断等研究.

略来对 Cu-Si-Al 合成 DMC 催化剂失活过程建模, 并对催化剂使用寿命进行预测. 本文以高维相空间重构的数据表示方法重构催化剂失活数据序列, 利用失活时间序列所计算出来的客观规律进行预测, 达到改写数据的规律性并探讨催化剂失活内在复杂本质特征的目的, 避免预测的人为主观性, 从而提高预测的精度和可信用度.

## 2 支持向量机和相空间重构

### 2.1 支持向量回归机

支持向量机<sup>[4]</sup>是一种基于统计学习理论的机器学习算法, 遵循结构风险最小化原理 (SRM), 通过核函数的映射, 将数据映射到高维特征空间进行线性处理, 在模型的复杂性和学习能力之间寻求最佳折衷, 以获得最好的推广能力, 较好地解决了小样本、高维、非线性和局部极小点等问题.

构造 SVR 回归估计函数

$$f(x) = \langle \omega \cdot \Phi(x) \rangle + b, \quad (1)$$

使得预测的期望风险函数最小, 即

$$\begin{cases} \min_{\omega, b, \xi} L = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*); \\ \text{s.t. } |y_i - \langle \omega \cdot \Phi(x) \rangle - b| \leq \varepsilon + \xi_i, \\ \xi_i \geq 0, \xi_i^* \geq 0, i = 1, 2, \dots, n. \end{cases} \quad (2)$$

其中松弛变量  $(\xi_i, \xi_i^*)$  和惩罚因子  $(C > 0)$  用于调节超出  $\varepsilon$  管道的样本点. 引入 Lagrange 乘子  $(\alpha_i, \alpha_i^*)$  将其转化成无约束二次规划问题进行求解, 其对偶问题形式为

$$\begin{cases} \max_{\alpha, \alpha^*} R(\alpha, \alpha^*) = \\ -\varepsilon \sum_{i=1}^n (\alpha_i - \alpha_i^*) + \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) - \\ \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n [(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j)]; \\ \text{s.t. } \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0, \\ 0 \leq \alpha_i, \alpha_i^* \leq C. \end{cases} \quad (3)$$

通过求解式 (3) 的二次规划问题可得到最优的 Lagrange 系数  $\alpha_i, \alpha_i^*$  以及阈值  $b$ . 只有一些系数  $(\alpha_i - \alpha_i^*)$  是非零的, 相应的输入向量  $x_i$  称为支持向量 (SVs), 支持向量可以视为压缩了的训练集里信息量最真实的数据点. 求解实际问题时只需对支持向量进行求解, 因此, 回归估计函数为

$$\begin{aligned} f(x) &= f(x, \omega) = f(x, \alpha_i, \alpha_i^*) = \\ & \sum_{i=1}^n (\alpha_i - \alpha_i^*) \langle \Phi(x_i) \cdot \Phi(x) \rangle + b = \\ & \sum_{x \in \text{SV}} (\alpha_i - \alpha_i^*) K(x_i, x) + b, \end{aligned} \quad (4)$$

其中  $K$  为核函数, 采用不同的核函数可以构建出不同的支持向量回归机. 本文采用径向基核 (Gaussian 核), 即  $K(x, x_i) = \exp[-(x - x_i)/2\sigma^2]$ , 它对于常见的模式识别问题都具有优异的泛化性能.

阈值  $b$  可通过下式计算:

$$b = \begin{cases} y_i - f(x_i)_{b=0-\varepsilon}, \alpha_i \in (0, C); \\ y_i - f(x_i)_{b=0+\varepsilon}, \alpha_i^* \in (0, C). \end{cases} \quad (5)$$

其中  $x_i$  和  $y_i$  分别为第  $i$  个支持向量以及与之对应的目标输出. 在 SVR 回归算法中, 惩罚因子  $C$  和管道因子  $\varepsilon$  是两个可以调节的自由参数,  $C$  表示模型的复杂性和逼近误差之间的平衡,  $\varepsilon$  表示训练数据合适的管道宽度, 文献 [5-6] 给出了  $C$  和  $\varepsilon$  的选取方法.

以上自由参数和核函数的选择共同决定了 SVR 回归估计的准确性和泛化能力.

### 2.2 相空间重构

实际的动态系统普遍存在一个问题——缺少关于系统内所有变量的信息, 相空间重构方法<sup>[7]</sup>为描述系统的动态特征提供了一条途径. PSR 的基本思想是通过考察系统的一个或多个分量, 将其在某些固定时间的延迟点上的观测量视为新的坐标, 由它们共同确定多维状态空间的一点, 重复这个过程, 即可构造出系统的近似状态空间, 相空间的维数即为时间序列延迟点的个数. 按照 Takens 定理<sup>[8]</sup>, 只要维数  $m \geq 2d + 1$ , 总可以在拓扑不变的意义上找到一个  $m$  维的嵌入相空间.

对于时间序列  $x_i (i = 1, 2, \dots, N)$ , 通过选定合适的嵌入维数 ( $m$ ) 和时间延迟 ( $\tau$ ), 便可构造一个  $m$  维的相空间

$$A_j = [a_j, a_{j+\tau}, \dots, a_{j+(m-1)\tau}]. \quad (6)$$

其中:  $j = 1, 2, \dots, N - (m-1)\tau$ ;  $N$  为时间序列长度.

有关嵌入空间维数  $m$  和时间延迟  $\tau$  的选取, 现在主要有两种观点. 本文采用 C-C 方法<sup>[9]</sup>来选取  $m$  和  $\tau$ . 该方法基于嵌入窗法的思想, 认为  $m$  与  $\tau$  是相关的, 两者的选取是相互依赖的, 应用关联积分能够同时估计出时间序列的延迟  $\tau_d (\tau_d = \tau \tau_k)$  和数据依赖的最大时间  $\tau_w$ . 时间延迟  $\tau_d$  确保  $x_i$  中各成分相互依赖, 但不依赖于  $m$ ; 而时间窗口  $\tau_w$  则依赖于  $m$ , 且  $\tau$  随  $m$  而变化.

取  $m = 2, 3, 4, 5, r_i = i \times 0.5\sigma, i = 1, 2, 3, 4$ , 计算

$$\bar{S}(t) = \frac{1}{16} \sum_{m=2}^5 \sum_{i=1}^4 S(m, r_i, t), \quad (7)$$

$$\Delta \bar{S}(t) = \frac{1}{4} \sum_{m=2}^5 \Delta S(m, t), \quad (8)$$

$$S_{\text{cor}}(t) = \Delta \bar{S}(t) + |\bar{S}(t)|, \quad (9)$$

其中  $\bar{S}(t)$  和  $\Delta \bar{S}(t)$  均反映了原时间序列的自相关特

性. 考虑到  $\bar{S}(t)$  值可以是正数或负数,  $\Delta\bar{S}(t)$  值总为正数, 因此, 寻找  $\bar{S}(t)$  的第1次过零点, 或  $\Delta\bar{S}(t)$  的第1个局部极小值点所对应的  $t$  即为最佳时延. 时间延迟  $\tau_d = t\tau_s$ , 对应着时间序列独立的第1个局部最大时间  $t$ . 同时, 综合考虑  $\bar{S}(t)$  和  $\Delta\bar{S}(t)$ , 寻找  $S_{\text{cor}}(t)$  的全局最小值所对应的  $t$  即可获得最佳嵌入窗  $\tau_w = t\tau_s$ , 对应着时间序列独立的第1个整体最大值时间窗口.

C-C方法是利用统计结果得到的, 在实际计算中运行良好, 容易操作, 计算量小, 具有较强的抗噪声能力.

### 3 催化剂失活模型建立及预测

#### 3.1 过程描述

催化剂的失活原因一般分为中毒、烧结和热失活、结焦和堵塞三大类, 需要对催化剂进行加速寿命试验, 即不停地改变温度、压力和进料比等主要因素进行烧结. 在甲醇氧化羰基化法合成DME工艺中, 由于反应过程的复杂性和各反应之间的强烈耦合作用, 系统具有非线性、不确定性、强相关性、混沌、耦合等特点, 对其建立合成DMC多相催化剂失活机理的动力学模型非常困难.

本文以某国家重点实验室研制的Cu-Si-Al合成DMC催化剂为研究对象, 简称Cu(I)Y型分子筛. 研究显示, Cu(I)Y型分子筛的失活过程与Si/Al<sub>2</sub>比、交换度、Cu<sup>+</sup>含量和热处理加热时间有最直接的关系, 对于Cu(I)Y型分子筛的活性能力影响较大.

采用ANN和SVR, 将影响活性的多种相关因素(如催化剂使用时间、交换条件、加热温度、活性组分含量等)分别关联在失活模型中, 以建立催化剂失活的灰箱模型. 在催化剂寿命试验中, 按照失活规律的初期、中期、末期3个阶段采集实验数据, 并在每个阶段根据相关因素选取尽可能多的正交试验点进行测量. 但是, 该方法获取的实验数据有限, 仅适合于初选催化剂, 而对于处在同一失活阶段的数据几乎没有效果, 模型的精度也不能完全得到保证, 进而无法预测催化剂的使用寿命.

实际上, 根据多相催化剂的失活过程包含化学混沌现象, 可以将Cu(I)Y型分子筛失活视为一个混沌动力学系统的演化问题. 通过催化剂失活过程整体行为的一维“投影”来“重构”复杂系统的整体行为. 根据混沌理论, 失活过程中产物DMC的时空收率(STY)的混沌时间序列是由与之相互作用的其他分量(包括Si/Al<sub>2</sub>比、交换度、Cu<sup>+</sup>含量和热处理加热时间等)共同决定的. 因此, 可以从STY的时间序列数据中提取和恢复出系统原来的规律, 这种规律是高维空间下的一种轨迹. 可以结合非线性学科的相空间重构技术

和支持向量机建立演化模型, 以评估和预测催化剂失活的非线性行为.

#### 3.2 催化剂失活预测

将Cu(I)Y型分子筛失活测试得到的DMC的STY历史时间序列数据记为  $\{x(i)\}$ , 则基于PSR-SVR的非线性模型学习与预测算法示意如图1所示.

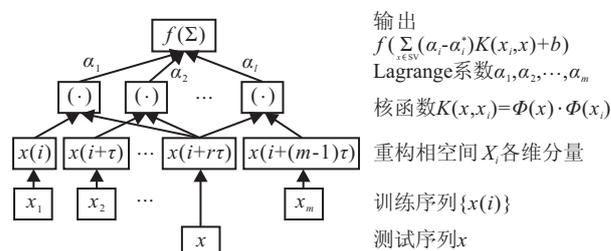


图1 基于PSR-SVR模型的学习与预测算法示意

图1中,  $s$ 是支持向量个数,  $a$ 是Lagrange系数,  $x(i), x(i+\tau), \dots, x(i+(m-1)\tau)$ 是重构相空间中向量  $X_i$  的各维分量,  $m$ 是重构相空间的维数,  $\tau$ 是时间延迟,  $i$ 是重构相空间中向量的序号. 采用C-C方法确定PSR中的  $m$ 和  $\tau$ .

根据求取的  $m$ 和  $\tau$ , 得到支持向量机学习和预测的样本集  $\{X_i\}$ , 它是重构相空间后的DMC时空收率时间序列, 即

$$X = \begin{bmatrix} x_1 & x_{1+\tau} & \cdots & x_{1+(m-1)\tau} \\ x_2 & x_{2+\tau} & \cdots & x_{2+(m-1)\tau} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n-(m-1)\tau} & x_{n-(m-1)\tau+1} & \cdots & x_{n-1} \end{bmatrix},$$

$$Y = \begin{bmatrix} x_{2+(m-1)\tau} \\ x_{3+(m-1)\tau} \\ \vdots \\ x_n \end{bmatrix}. \quad (10)$$

SVR对式(10)的高维样本集进行学习和训练, 以评估和预测催化剂失活过程的STY. 计算步骤和流程如下:

**Step 1:** 将DMC的STY历史时间序列数据  $\{x(i), i = 1, 2, \dots, N\}$  按照C-C方法进行相空间重构, 选取  $N = 3000$ , 计算给定其标准差  $\sigma$ .

**Step 2:** 按照式(7)~(9)计算  $\bar{S}(t)$ ,  $\Delta\bar{S}(t)$ 和  $S_{\text{cor}}(t)$ , 其中时间变量  $t$ 取小于等于200的自然数, 即  $\bar{S}(t) \xrightarrow{v \rightarrow s} t (0 \leq t \leq 200)$ ,  $\Delta\bar{S}(t) \xrightarrow{v \rightarrow s} t (0 \leq t \leq 200)$ ,  $S_{\text{cor}}(t) \xrightarrow{v \rightarrow s} t (0 \leq t \leq 200)$ .

**Step 3:** 依据3个统计量的计算结果画图, 确定相空间重构下的嵌入空间维数  $m$ 与时间延迟  $\tau$ .

**Step 4:** 将选定的嵌入维数  $m$ 与时间延迟  $\tau$ 代入式(6), 得到式(10)的高维样本集, 其中  $X$ 为SVR模

型的输入,  $Y$  为 SVR 模型的输出.

Step 5: 确定 SVR 模型的超参数最优值. 可采用文献 [6] 中混沌粒子群优化算法 (CPSO) 来优化 SVM 超参数 ( $C, \sigma, \varepsilon$ ).

Step 6: 用带有优化参数 ( $C^*, \sigma^*, \varepsilon^*$ ) 的 SVR 模型对测试集序列进行预测.

## 4 运行结果

实验选用 Cu(I)Y 分子筛中催化剂活性最高样品 HY56Cu75H9 进行失活测试. 表 1 列出了 Cu(I)Y 分子筛失活测试 100 个小时内的实验数据. 因前 16 个小时催化剂活性最好, 活性一直处于上升阶段; 反应 16 个小时后活性开始下降, 故表 1 从第 16 个小时开始记录 DMC 的时空收率值 (单位:  $\text{mg}/(\text{g} \cdot \text{h})$ ), 共 84 组数据.

表 1 Cu(I)Y 分子筛 100 个小时内的 STY 失活数据

时间/h	DMC 的时空收率 STY/(mg/(g·h))					
16-26	155.35	147.98	144.86	145.54	138.92	122.56
28-38	121.11	114.55	119.14	114.44	114.79	145.56
40-50	147.16	149.37	121.03	121.21	110.96	100.14
52-62	98.49	101.43	96.69	94.53	93.24	93.01
64-74	92.56	91.86	93.50	99.16	99.39	86.25
76-86	78.93	79.90	71.91	76.14	74.68	74.06
88-98	75.79	77.23	79.02	81.61	70.73	70.62
100	70.02	/	/	/	/	/

按照 84 组 Cu(I)Y 分子筛失活测试数据, 使用 C-C 方法进行相空间重构. 图 2 为 C-C 方法计算延迟时间  $\tau$ . 从图中可以看出,  $\Delta \bar{S}(t)$  的第 1 个近似极小值是 4,  $\bar{S}(t)$  的第 1 个近似零点也是 4, 于是, 确定延迟时间  $\tau = 4$  作为 STY 时间序列的延迟时间.

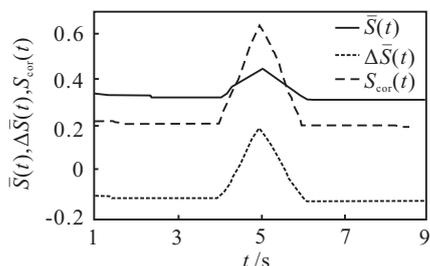


图 2 C-C 方法计算 STY 时间序列的延迟时间

图 3 给出了采用 C-C 方法计算 STY 时间序列的  $S_{\text{cor}}(t)$  曲线. 可以看出,  $S_{\text{cor}}(t)$  大约在  $t = 36$  时取得全局最小点, 依据  $\tau_w = (m - 1)\tau$ , 因而取  $m = 10$ .

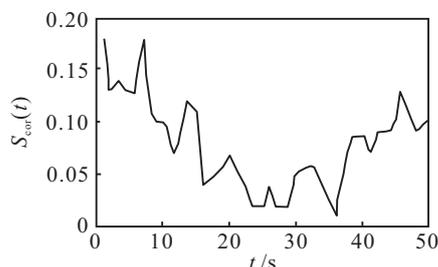


图 3 C-C 方法计算 STY 时间序列的  $S_{\text{cor}}(t)$  曲线

根据嵌入空间维数  $m = 10$ , 时间延迟  $\tau = 4$ , 重构相空间后 DMC 的时空收率 STY 时间序列  $X_i$  为  $n - (m - 1)\tau = 48$  组数据. 将所得 48 组数据分成两部分, 前 42 组数据作为训练集, 后 6 组数据作为测试集. 由重构相空间嵌入相点在相空间的轨迹, 用支持向量机进行学习训练, 构造出一个失活模型, 并根据 SVR 模型对 STY 的后 6 个数据进行预测. 图 4 为基于 PSR-SVR 预测算法的 Cu(I)Y 分子筛失活预测曲线.

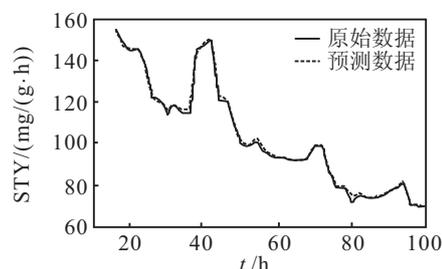


图 4 基于 PSR-SVR 的 Cu(I)Y 分子筛失活预测曲线

从图 4 可以看出, 基于 PSR-SVR 预测算法能较准确地预测 Cu(I)Y 分子筛失活过程的 STY 值, 最大误差约  $0.35 \text{ mg}/(\text{g} \cdot \text{h})$ , 在 Cu(I)Y 分子筛失活过程预测中已经达到非常高的精度.

## 5 结 论

与传统的催化剂失活模型相比, 基于 PSR-SVR 方法建模, 无需通过假设催化剂失活模型的具体形式, 避开了催化剂失活中多种因素难以取舍的缺陷, 直接从时空收率时间序列数据中得到催化剂的失活规律. 与通过失活率常数来刻画催化剂的失活过程相比, 本文方法能够更准确地反应催化剂失活过程的规律, 为进一步对连续反应装置的选取与优化, 及其反应放大后的生产提供了科学依据.

## 参考文献(References)

- [1] Anderson A A. Investigation of the oxidative carbonylation of methanol to dimethyl carbonate over copper zeolite catalysts[D]. Madison: Department of Chemical and Biological Engineering, University of Wisconsin-Madison, 2004.
- [2] Kito S, Satsuma A, Ishikura T, et al. Application of neural network to estimation of catalyst deactivation in methanol conversion[J]. Catalysis Today, 2004, 97(1): 41-47.
- [3] Han X X, Xie J, Ren J, et al. Modeling and optimization strategy for heterogeneous catalysis based on support vector regression and genetic algorithm[C]. Proc of the 2nd Int Conf on Intelligent Computation Technology and Automation. Changsha, 2009: 94-98.
- [4] Vapnik V. The nature of statistical learning theory[M]. New York: Springer, 1995: 4-80.