

## 基于模糊粗糙集和支持向量机的化工过程故障诊断

王鲜芳<sup>1,3</sup>, 王岁花<sup>1</sup>, 杜昊泽<sup>2</sup>, 王平<sup>1</sup>

(1. 河南师范大学 计算机与信息工程学院, 河南 新乡 453007; 2. 南京航空航天大学  
信息科学与技术学院, 南京 210016; 3. 密苏里大学哥伦比亚分校 工程学院, 密苏里 65211)

**摘要:** 为解决化工过程故障不易诊断的问题, 提出一种基于模糊粗糙集特征提取和支持向量机的故障诊断方法。首先, 利用模糊粗糙集对特征信息进行提取, 构筑相应的故障特征集合; 然后, 将故障特征集合对应的样本输入到支持向量机分类器, 实现对化工过程不同故障的识别。在 TEP 故障诊断中的应用表明了所提出方法的有效性。

**关键词:** 模糊粗糙集; 支持向量机; TEP 过程; 故障诊断

**中图分类号:** TP273

**文献标志码:** A

## Fault diagnosis of chemical industry process based on FRS and SVM

WANG Xian-fang<sup>1,3</sup>, WANG Sui-hua<sup>1</sup>, DU Hao-ze<sup>2</sup>, WANG Ping<sup>1</sup>

(1. School of Computer and Information Engineering, He'nan Normal University, Xinxiang 453007, China; 2. College of Information Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China; 3. College of Engineering, University of Missouri Columbia, MO 65211, USA. Correspondent: WANG Xian-fang, E-mail: xfw1969@hotmail.com)

**Abstract:** In order to solve the problem about fault diagnosis for the chemical industry process, a fault diagnosis approach is proposed based on feature extraction by using fuzzy rough sets and support vector machines. The feature information is extracted by utilizing fuzzy rough sets and the fault diagnosis sets is built firstly. Then, the samples corresponding to the fault diagnosis sets are input into the SVM multi-classifier to realize the identification of different fault diagnosis in the chemical industry process. Finally, the effectiveness of the proposed method is illustrated through fault diagnosis in TEP chemical industry process.

**Keywords:** fuzzy rough sets; support vector machines; TEP process; fault diagnosis

### 0 引言

化工过程反应复杂, 具有高度非线性和时变性, 一旦发生故障, 便不可逆转, 因此, 早期预警对化工过程监控起着非常重要的作用。如何从在线海量工业数据中挖掘出有用信息, 以便及时对化工过程进行故障诊断和决策是目前化工领域的一个研究热点。

针对化工过程的特殊性, 依斯曼化学品公司开发了一种高级仿真工业过程——田纳西-伊斯曼过程 (TEP)。近年来, 国内外一些学者对该过程的故障诊断进行了研究, 提出不少故障诊断方法, 如主元分析法 (PCA)<sup>[1]</sup>、神经网络方法<sup>[2]</sup>、符号有向图 (SDG)<sup>[3]</sup>等。这些方法各有特点, 但不同的方法适用于不同的故障模式, 对其他故障模式的误检率或漏检率较高。

支持向量机<sup>[4]</sup>(SVM) 是 Vapnik 等人提出的一种机器学习方法, 它以统计学理论为基础, 较好地解决了非线性、小样本和高维数等学习问题, 泛化能力强, 对故障诊断精度相对较高<sup>[5]</sup>。但是, 在使用支持向量机过程中, 过多的输入变量将直接影响诊断速度。

粗糙集理论 (RS)<sup>[6]</sup>是 Pawlak 提出的一种用于处理不确定信息的数据挖掘方法。该方法可以在不需要任何辅助信息的情况下, 通过直接分析不确定数据之间的关系提取有用特征, 进而得到简洁的知识表达形式。近年来, 粗糙集理论发展迅速, 已经在模式识别、决策分析、故障诊断等领域得到了应用<sup>[7]</sup>。

针对过多输入变量影响支持向量机故障诊断效果的问题, 本文在研究粗糙集理论和支持向量机的基

收稿日期: 2014-02-26; 修回日期: 2014-05-18.

基金项目: 国家自然科学基金项目(61173071); 国家公派高级研究学者及访问学者项目([2013]3018); 河南省高校创新人才支持计划项目(2012HASTIT011).

作者简介: 王鲜芳(1969—), 女, 教授, 从事生化过程建模、故障诊断、机器学习等研究; 王岁花(1963—), 女, 教授, 从事智能算法设计等研究.

础上,综合二者优势,提出一种基于模糊粗糙集特征提取和支持向量机的故障诊断方法.该方法先根据条件属性(集)对决策属性的正域大小选择属性,提取出对诊断故障贡献最大的最小故障特征集合,然后将最小故障特征集合对应的样本输入到支持向量机分类器,实现对化工过程不同故障的识别.在田纳西-伊斯曼过程(TEP)中的应用表明了所提出诊断策略的有效性.

## 1 基于模糊粗糙集和支持向量机的故障诊断基本原理

基于模糊粗糙集特征提取和支持向量机的故障诊断方法主要由数据采集与预处理、RS 数据处理和 SVM 故障诊断构成,如图 1 所示.

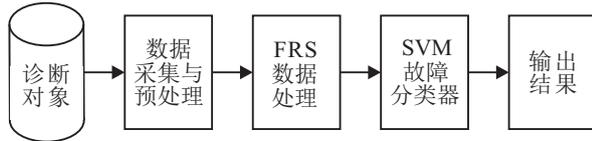


图 1 故障诊断方法流程

数据采集与预处理主要是对传感器采集到的原始信号进行归一化和故障特征提取,以得到原始的故障诊断样本,这些原始故障诊断样本往往含有冗余信息,需要经过 FRS 处理.经过 FRS 数据处理后,可以得到最小故障诊断特征的训练样本集和测试样本集,然后送至 SVM 故障分类器,得出诊断结果.

## 2 基于模糊粗糙集的 SVM 故障多分类器

### 2.1 数据约简<sup>[8]</sup>

模糊信息系统中条件属性和决策属性都可以是模糊的. $G = (U, A, F, D)$  称为模糊信息系统.其中

$$U = (x_1, x_2, \dots, x_n),$$

$$A = (a_1, a_2, \dots, a_m),$$

$$F = \{\tilde{A}_{jl} : U \rightarrow [0, 1] (j \leq m, l_j \leq r_j)\},$$

$$D = \{\tilde{D}_l : U \rightarrow [0, 1] (l \leq r)\}.$$

$x$  关于  $S$  相似的模糊等价类定义为

$$\mu_{[x]_S}(y) = \mu_S(x, y). \quad (1)$$

由文献 [8] 可得,对于  $x \in U$ ,属于模糊正域的隶属度为

$$\mu_{POS_P(Q)}(x) = \sup_{x \in U/Q} \mu_{PX}(x). \quad (2)$$

使用模糊正域的定义,新的模糊依赖度函数可以定义为

$$\gamma'_p(Q) = \frac{\mu_{POS_P(Q)}(x)}{|U|} = \frac{\sum_{x \in U} \mu_{POS_P(Q)}(x)}{|U|}. \quad (3)$$

### 2.2 模糊粗糙集的故障诊断特征提取

首先,建立原始故障诊断决策表  $A$ , 根据实际问

题的特点选择一种离散化方法,得到离散化的故障诊断决策表  $B$ ; 其次,初始化变量  $\delta$ 、模糊分隔数  $k$  及规则公式集合  $list\_B$ , 若  $list\_B \neq \emptyset$ , 则取  $\varphi \in list\_B$ , 若  $i < k$ , 则对  $d_j (j = (1, 2, \dots, m))$  计算  $\max_{s \leq r} D_B^s(l_j | a_j \in B)$ ; 再次,若  $\max_{s \leq r} D_B^s(l_j | a_j \in B) > \delta$ , 则产生规则  $N_B(l_j | a_j \in B)$ , 继续计算  $\max_{s \leq r} D_B^s(l_j | a_j \in B)$ , 直到小于最小可靠度为止; 最后,输出所有规则及其相应参数,输出故障特征子集.  $m, k$  分别为决策变量和模糊分类数目,  $\delta$  为最小可信度.

### 2.3 SVM 故障多类分类器

将由 2.2 节得到的故障特征子集作为多类 SVM 的输入,通过多类支持向量机的训练,可以判断故障的类型,进而达到故障预测的目的.

常用的多元分类 SVM 方法有 4 种: 一对多、一对一、导向无环图、二叉树. 一对一的分类方法虽计算复杂,但精度高,适合类别数目较少的情形. 本文的多类分类器是基于文献 [9] 的思想,采用“一对一”的方法构建  $k(k-1)/2$  个二元分类器,以对应  $k(k-1)/2$  个决策函数,  $k$  为所需分类问题的种类数. 测试集的分类标签由所有决策值和投票策略决定. 决策函数  $y^{s-t}(x)$  的表示式<sup>[10]</sup>为

$$y^{s-t}(x) = \text{sgn} \left( \sum_{i=1}^N a_i^{s-t} y_i K(x_i, x) + b^{s-t} \right), \quad (4)$$

其中  $s, t = 1, 2, \dots, k, t! = s$ . 如果对类  $c_s$  和  $c_t$  分类的决策函数  $y^{s-t}(x)$  的决策值显示样本  $x$  的预测类标签为  $c_s$ , 则类  $c_s$  的投票为

$$\arg \max_{s=1,2,\dots,k} \text{vote}_x(s), \quad (5)$$

其中  $\text{vote}_x(s) = \sum_{t=1, t! \neq s}^k v_x(s, t)$ ; 若  $y^{s-t}(x)$  得到的类  $x$  是  $c_s$ , 则  $v_x(s, t) = 1$ , 否则  $v_x(s, t) = 0$ .

## 3 TEP 化工过程故障诊断

### 3.1 TEP 化工过程故障

TEP 是由伊斯曼化学品公司开发的一种高级仿真工业过程,该仿真过程可以预设 21 个故障,其中 16 个故障是已知的,5 个故障是未知的,具体故障参见文献 [8].

### 3.2 数据获取

针对 TEP 过程故障存在的情况,设计 22 次不同的仿真实验,其中 1 次是正常情况,21 次对应 21 种不同的故障模式. 每次运行的仿真时间为 10h, 采样间隔为 2min, 共采集 600 个观测数据. 仿真开始时没有故障,仿真 1h 后引入故障,采集 570 个故障数据,用于知识表达系统的约简.

### 3.3 基于模糊粗糙集的 TEP 过程故障规则提取

选取有代表性的 4 种模式数据进行分析, 从每种故障类型的批次数据中分别选取有代表性的 200 个数据样本, 构成一个具有 800 个样本数据规模、33 个条件属性和 1 个决策属性的知识表达系统. 针对 33 个条件属性, 采用式 (3) 对数据进行约减, 得到约简后的 10 个条件属性. 这里的 1 个决策属性指特定的 4 种故障模式. 利用 2.2 节介绍的模糊粗糙集规则提取算法, 具体步骤如下.

Step 1: 利用约减后得到的 10 个条件属性构造新的故障数据集.

Step 2: 对数据集进行归一化  $[-1, 1]$  处理.

Step 3: 采用钟型函数对各个条件属性模糊化为 4 个模糊等价类.

Step 4: 初始化规则提取变量, 定义最小可信度  $\delta = 0.9$  及规则公式

$$\prod_{x \in U} (\tilde{B}/\tilde{A}) = \bigvee_{x \in U} (\tilde{A}(x) \wedge \tilde{B}(x)).$$

Step 5: 若  $i < k$ , 则对  $d_j (j = 1, 2, \dots, 4)$  计算  $\max_{s \leq r} D_B^s(l_j | a_j \in B)$ .

Step 6: 若  $\max_{s \leq r} D_B^s(l_j | a_j \in B) > \delta$ , 则产生规则  $N_B(l_j | a_j \in B)$ , 转入 Step 5, 否则转入 Step 7.

Step 7: 输出规则, 并得到其相应参数.

经过上述模糊粗糙集规则提取算法计算之后, 得到规则提取结果, 如表 1 所示.

表 1 TEP 过程故障规则分布表

故障类别	故障模式	规则数
0	a	8
1	b	12
4	c	15
5	d	10

表 1 中, a 表示无故障, b 表示故障 1, c 表示故障 4, d 表示故障 5, 共得到 45 条决策规则. 限于篇幅, 这里只给出判断故障 5 在 10 个约简条件属性下得出的 10 条规则, 如表 2 所示.

表 2 故障 5 的规则表

规则	属性										可信度
	C9	C10	C11	C12	C16	C17	C23	C24	C29	C30	
1	2	3	0	0	2	3	4	2	3	4	0.94
2	3	3	2	1	3	3		2	2	3	0.97
3	1	3			2		4		4	2	0.94
4	3	3	2		4	3		2	4	3	0.93
5	2				3		3	4		3	0.95
6	1	3		3	2		4	2	1		0.91
7	3		2	1		4	3		1	2	0.95
8	2	1		3	4		2	3	3	3	0.96
9	3	2		1		3	3		2		0.93
10		2	2	2			3		2		0.91

### 3.4 TEP 过程故障诊断

使用计算得到的约简数据集训练多类支持向量机, 核函数选择径向基核函数为

$$K(x_i, x) = \exp\left(-\frac{\|x_i - x\|^2}{2\sigma^2}\right). \quad (6)$$

训练后获得径向基函数的宽度  $\sigma = 0.5$ , 误差惩罚系数  $C = 100$ .

从故障模式样本中任取某种故障的 200 个数据进行测试, 利用构筑的多类支持向量机对 3 种类型故障模型分别进行判断, 得到诊断结果如表 3 所示, 然后依据得票多少即可判断故障模式.

表 3 模糊粗糙支持向量测试结果

	实际故障			
	0	1	4	5
0	9	1	2	1
1	0	9	0	0
4	0	0	7	0
5	0	0	0	8
未知	1	0	1	1
漏检率 %	1	0	3	2
误检率 %	0	1	8.5	11.3

### 3.5 几种诊断方法的比较

为了说明本文设计的故障诊断方法 (FR-SVM) 的效果, 分别采用文献 [8] 的模糊神经网络方法 (FR-RBF)、标准的 SVM、R-SVM 方法, 在同样的条件下进行故障诊断效果比较, 结果如表 4 和表 5 所示.

表 4 4 种故障诊断方法的精度比较 %

故障类型	方法			
	FR-RBP	SVM	R-SVM	FR-SVM
0	85.12	89.15	97.00	100.00
1	79.36	88.32	90.21	99.00
4	67.38	78.09	80.37	87.50
5	82.13	82.53	83.75	88.99

表 5 4 种故障诊断方法的时间比较 %

故障类型	方法			
	FR-RBP	SVM	R-SVM	FR-SVM
0	3.53	8.46	3.28	3.65
1	5.28	9.57	2.26	2.38
4	4.37	5.78	4.23	4.35
5	6.47	8.16	5.47	5.56

从表 4 可以看出, 相对于其他几种方法, 本文设计方法的故障诊断精度较高, FR-RBF 效果较差. 主要是由于样本数目有限, 预测模型泛化能力差. 从表 5 可以看出, 没有经过粗糙集处理的 SVM 故障诊断时间最长, FR-RBF 的耗时与 RF-SVM 相差不大, R-SVM 预测故障时间较短. 上述结果表明了本文所提出的设计方法在诊断精度和诊断时间上均有优势, 有利于对化工过程的实时监测.

## 4 结 论

针对 TEP 过程的特点,设计了基于模糊粗糙集和支持向量机的故障诊断方法.该方法在充分利用模糊粗糙集和 SVM 各自优势的基础上,通过简化模糊规则提高故障诊断速度和精度,达到了较好的诊断效果,能够为实际化工过程的实时故障诊断提供参考依据.

### 参考文献(References)

- [1] 王海清,宋执环,王慧. PCA 过程监测方法的故障检测行为分析[J]. 化工学报, 2002, 53(3): 297-301.  
(Wang H Q, Song Z H, Wang H. Fault detection behavior analysis of PCA-based process monitoring monitoring approach[J]. J of Chemical Industry and Engineering, 2002, 53(3): 297-301.)
- [2] 贺丁,赵劲松. 基于 Hopfield 网络的时滞分析故障诊断策略[J]. 化工学报, 2013, 64(2): 633-640.  
(He D, Zhao J S. Fault strategy of time delay analysis based on Hopfield network[J]. J of Chemical Industry and Engineering, 2013, 64(2): 633-640.)
- [3] 张建广,邱彤,赵劲松. 基于广义 SDG 的化工过程故障模拟分析[J]. 清华大学学报: 自然科学版, 2009, 28(9): 1561-1568.  
(Zhang J G, Qiu T, Zhao J S. Chemical process fault simulation analyses based on an extended SDG[J]. J of Tsinghua University: Science and Technology, 2009, 28(9): 1561-1568.)
- [4] Vapnik V, Golowich S, Smola A. Support vector machine for function approximation, regression estimation, and signal processing[C]. Neural Information Processing Systems. Cambridge: MIT Press, 1997: 281-287.
- [5] 曾辉. 改进加权支持向量机的研究及在故障诊断中的应用[D]. 广州: 华南理工大学自动化科学与工程学院, 2010: 3-10.  
(Zeng H. Research on improved weighted support vector machine and application in fault diagnosis method[D]. Guangzhou: School of Automation Science and Engineering, South China University of Technology, 2010: 3-10.)
- [6] Pawlak Z. Rough Sets[J]. Int J of Information and Computer Science. 1982, 11(5): 341-365.
- [7] 王国胤,姚一豫,于洪. 粗糙集理论与应用研究综述[J]. 计算机学报, 2009, 32(7): 1329-1342.  
(Wang G Y, Yao Y Y, Yu H. A Survey on Rough Set Theory and Applications[J]. Chinese J of Computers, 2009, 32(7): 1329-1342.)
- [8] 杨赛楠. 粗糙集神经网络故障诊断系统的方法研究及应用[D]. 无锡: 江南大学通信与控制工程学院, 2007: 31-35.  
(Yang S N. Research and application of fault diagnosis system based on rough sets-neural network[D]. Wuxi: School of Communication and Control Engineering, Jiangnan University, 2007: 31-35.)
- [9] 唐发明,王仲东,陈绵云. 支持向量机多类分类算法研究[J]. 控制与决策, 2005, 20(7): 746-754.  
(Tang F M, Wang Z D, Chen M Y. On multiclass classification methods for support vector machines[J]. Control and Decision, 2005, 20(7): 746-754.)
- [10] 吴瑞红,王亚丽,张环冲,等. 一种基于最小二乘支持向量机的葡萄酒品质评判模型[J]. 华侨大学学报: 自然科学版, 2013, 34(1): 31-35.  
(Wu R H, Wang Y L, Zhang H C, et al. An evaluation model of wine quality based on least square support vector machine[J]. J of Huaqiao University: Nature Science, 2013, 34(1): 31-35.)

(责任编辑: 齐 霖)