基于密度空间支持向量机的多工况过程故障检测

郭金玉 李 涛 李 元
(沈阳化工大学信息工程学院 辽宁 沈阳 110142)

摘 要 为了有效地对多工况数据进行检测,提出基于密度空间支持向量机(SVM)的多工况过程故障检测方法。运用局部概率密度方法对多工况数据进行预处理,消除多工况数据对过程故障检测特性的影响。在密度空间,运用正常数据和故障数据训练 SVM 模型获得权重向量和位移。把校验数据和测试故障数据作为 SVM 模型的输入,对其进行监视和检测。将该方法运用于田纳西—伊斯曼(Tennessee Esatman)多工况过程,仿真结果表明,对某些故障 PCA 和 KPCA 的检测效果较好,而对于某些故障 SVM 的检测效果较好。SVM 的平均故障检测率优于 PCA 和 KPCA。因此,不同的方法适用于不同类型的故障。

关键词 多工况过程 故障检测 局部概率密度 支持向量机

中图分类号 TP277 **文献标志码** A DOI:10.3969/j.issn.1000-386x.2022.07.006

MULTI-CONDITION FAULT DETECTION BASED ON SUPPORT VECTOR MACHINE IN DENSITY SPACE

Guo Jinyu Li Tao Li Yuan
(*College of Information Engineering, Shenyang University of Chemical Technology, Shenyang 110142, Liaoning, China*)

Abstract In order to effectively detect multi-condition data, a multi-condition fault detection method based on support vector machine(SVM) in density space is proposed. The local probability density method was used to preprocess the multi-condition data to eliminate the influence of multi-condition data on the fault detection. In the density space, the normal data and fault data were used to obtain the weight vector and displacement of SVM model. The validation data and test fault data were used as input of SVM to monitor and detect them. The proposed method was applied to the Tennessee-Eastman multi-condition process. The simulation results show that the fault detection effect of PCA and KPCA is the best for some faults, while the detection effect of SVM is the best for some faults. The average fault detection rate of SVM is better than that of PCA and KPCA. Therefore, different methods apply to different types of faults.

Keywords Multi-condition process Fault detection Local probability density Support vector machine(SVM)

0 引 言

在大数据背景下工业智能化时代已经到来,对于复杂工业过程而言,故障及时有效的诊断不仅关乎工业企业的经济利益,也关乎到工厂工人的生命安全,因此,对工业背景下的故障诊断也提出了更高的要求。在工业过程领域,故障诊断与检测是生产过程的重要环节,为了有效提高控制系统故障检测性能,基于数据驱动的检测方法被国内外学者深入研究和应用。由于该方法只需要在实际工业生产过程中获取历史数据,通过数据建立监测模型,从而得到了广泛的关注。

工业生产技术水平不断发展的今天,国内外学者对故障诊断和故障检测技术的研究越来越深入,多元统计分析以其独特的优势被广泛认可。主元分析(Principal Component Analysis,PCA)作为工业过程中对故障进行诊断和检测的最基础手段,一直发挥着重要作用,被广泛应用在多种场景,同时也衍生出多种新的故障检测方法[1-7]。PCA 作为多元统计分析方法的

收稿日期:2020-03-28。国家自然科学基金项目(61673279);辽宁省教育厅项目(LJ2019007)。**郭金玉**,副教授,主研领域:故障诊断,生物特征识别算法及应用。**李涛**,硕士生。**李元**,教授。

一种,处理的数据需要满足高斯、线性分布的前提假设。PCA 通过求得主元变量,将大量且复杂的数据投影到低维空间,保留主要数据,降低维数,方便计算,从而得到主元模型和统计控制限。由于算法自身在处理非线性时存在不足,导致检测结果不佳。为了改善 PCA 的不足,核主元分析(Kernel Principal Component Analysis,KPCA)[8] 被提出,在一定程度上扩大了 PCA 的使用范围。对非线性数据,KPCA 中核函数的优势就体现出来了,在低维空间中,样本分布呈现非线性,无法对其处理,需将数据映射到高维空间,去除数据非线性,然后运用 PCA 进行降维。由于 KPCA 鲁棒性较差,泛化能力不强,在解决多工况问题方面仍存在局限性。

支持向量机(Support Vector Machines,SVM)[9-10] 以其稳健的数学基础、强大的泛化能力和解决各种分类问题方面的诸多优势使其成为机器学习中的经典算法。早在 1999 年,支持向量机便由 Vapnik 提出[11],后因被国内外学者广泛研究学习而发展起来。SVM 在进行分类任务时具有独特的优势,这使其成为机器学习的主流技术。在处理多工况过程时,数据会出现非高斯性等问题,本文将多工况问题转换为单工况问题。为了使多工况数据变成单工况,并且使其近似服从高斯分布,达到满意的检测效果,本文结合局部概率密度方法,运用支持向量机算法对多工况过程进行过程监视。本文尝试将鲁棒性较好的 SVM 的分类特性用在故障诊断中,达到分离正常样本和故障样本的目的,从而提高过程监视性能的目的。

1　模　型

1.1　支持向量机

SVM 方法在解决数据集规模相对较小或样本非线性问题方面具有许多优点。SVM 算法在面对多工况过程时,往往会面临着众多非线性数据,由于非线性数据无法处理,需要将其投影到高维空间,去除数据非线性,建造最大分离超平面,使得数据能够进行有效分类。由于分离平面是基于支持向量构造的,所以 SVM 是解决高维问题的一种很好的解决方案。同时引入内核函数代替非线性映射,也避免了许多未解决的问题。

给定一个训练样本集,该训练样本集可表示为 $D=\{(x_1,y_1),(x_2,y_2),\cdots,(x_m,y_m)\}$,$y_i\in\{-1,+1\}$,在该样本训练集空间中找到一个最大分离超平面,把类别不同的样本有效分开,这是分类学习最基本的思想。分离给定数据分类的超平面如下:

$$\boldsymbol{w}^{\mathrm{T}}x+\boldsymbol{b}=0 \tag{1}$$

式中:$\boldsymbol{w}=(w_1;w_2;\cdots;w_d)$ 是权重向量;$\boldsymbol{b}$ 是位移项。

假如超平面$(\boldsymbol{w},\boldsymbol{b})$能将训练样本正确分类,那么对于$(x_i,y_i)\in D$,若 $y_i=+1$,则有 $\boldsymbol{w}^{\mathrm{T}}x_i+\boldsymbol{b}>0$;若 $y_i=-1$,则有 $\boldsymbol{w}^{\mathrm{T}}x_i+\boldsymbol{b}<0$。令

$$\begin{cases}\boldsymbol{w}^{\mathrm{T}}x_i+\boldsymbol{b}\geqslant+1 & y_i=+1\\ \boldsymbol{w}^{\mathrm{T}}x_i+\boldsymbol{b}\leqslant-1 & y_i=-1\end{cases} \tag{2}$$

欲找到有最大间隔的超平面,也就是要找到满足式(2)中的参数 w 和 b,使得间隔 γ 最大,即:

$$\min_{w,b}\frac{1}{2}\|\boldsymbol{w}\|^2 \tag{3}$$

$$\text{s.t. } y_i(\boldsymbol{w}^{\mathrm{T}}x_i+\boldsymbol{b})\geqslant1,i=1,2,\cdots,m$$

为了求解式(3),需要将其转化为“对偶问题”,用拉格朗日乘法求解,则该问题的拉格朗日函数为:

$$L(\boldsymbol{w},\boldsymbol{b},\boldsymbol{\alpha})=\frac{1}{2}\|\boldsymbol{w}\|^2+\sum_{i=1}^{m}\alpha_i(1-y_i(\boldsymbol{w}^{\mathrm{T}}x_i+\boldsymbol{b})) \tag{4}$$

式中:$\alpha_i\geqslant0$ 且 $\boldsymbol{\alpha}=(\alpha_1;\alpha_2;\cdots;\alpha_m)$。令 $L(\boldsymbol{w},\boldsymbol{b},\boldsymbol{\alpha})$ 对 $\boldsymbol{w}$ 和 $\boldsymbol{b}$ 的偏导为零可得:

$$\boldsymbol{w}=\sum_{i=1}^{m}\alpha_iy_ix_i \tag{5}$$

$$0=\sum_{i=1}^{m}\alpha_iy_i \tag{6}$$

通过求解,得到该模型为:

$$f(x)=\boldsymbol{w}^{\mathrm{T}}x+\boldsymbol{b}=\sum_{i=1}^{m}\alpha_iy_i\boldsymbol{x}_i^{\mathrm{T}}x+\boldsymbol{b} \tag{7}$$

假设样本出现了非线性数据,需要通过非线性映射 $\varphi(x)$ 投影到高维空间。分离超平面在高维空间中所对应的模型表示为:

$$f(x)=\boldsymbol{w}^{\mathrm{T}}\varphi(x)+\boldsymbol{b} \tag{8}$$

为了避免高维运算,引入核函数:

$$K(x_i,x_j)=(\varphi(x_i),\varphi(x_j))=\varphi(x_i)^{\mathrm{T}}\varphi(x_j) \tag{9}$$

通过核函数计算可得:

$$f(x)=\boldsymbol{w}^{\mathrm{T}}\varphi(x)+\boldsymbol{b}=\sum_{i=1}^{m}\alpha_iy_iK(x,x_i)+\boldsymbol{b} \tag{10}$$

因此,SVM 在对数据进行分类时,无论线性还是非线性,SVM 都可以将其有效转化,进而高效准确地对数据分类。

运用正常数据和故障对 SVM 进行训练,获得权重向量 $\boldsymbol{w}$ 和位移 $\boldsymbol{b}$。建立模型之后,SVM 能学习正常数据和故障数据的特性,从而将数据正确分类。将测试数据输入模型,通过超平面的划分,正常数据划分成一类,定义为标签 0;故障数据划分成另一类,定义为标签 1。

1.2 基于密度空间 SVM 的多工况过程故障检测

PCA、KPCA 和 SVM 算法适用于单工况的过程故障检测，然而工业过程通常包含多个工况，如果这些算法直接应用于多工况过程，其监视和检测性能就会下降。为了改进多工况过程故障检测性能，本文首先利用局部概率密度方法[12-13]将多工况数据处理成单工况数据，然后应用 SVM 进行过程监视与检测。基于密度空间 SVM 的多工况过程故障检测具体步骤如下：

（1）收集正常运行的历史数据集：

$$\boldsymbol{X}=[x_1,x_2,\cdots,x_m]^{\mathrm{T}}\in\mathbf{R}^{m\times n}$$

（2）对历史数据集 $\boldsymbol{X}$ 标准化后得到矩阵 $\boldsymbol{X}_1$。

（3）运用式(11)计算 $\boldsymbol{X}_1$ 的局部概率密度矩阵。

$$\hat{p}(x_i)=\frac{1}{k}\sum_{x_j\in kNN(x_i)}K\left\{\frac{d(x_i,x_j)}{d(x_j,x_j^k)}\right\} \tag{11}$$

式中：$K(\cdot)$为高斯核函数；x_i是 $\boldsymbol{X}_1$的样本；$kNN(x_i)=\{x_i^1,x_i^2,\cdots,x_i^k\}$是 x_i的 k 近邻域，x_i^k 表示 x_i 的第 k 个最近邻。

（4）运用正常和故障数据训练 SVM 模型获得权重向量和位移，然后把测试数据送入 SVM 进行分类。

2 仿真结果与分析

2.1 TE 多工况过程

Tennessee Esatman(TE)过程仿真平台已成为国际上通用的工业过程模型仿真平台[14-17]，在故障检测和诊断领域被国内外学者广泛使用。TE 过程变量非常多，其工业过程也很复杂，其中 2 个气液放热反应会产生 2 种主产品，此外，还与 5 个主要操作单元等共同组成 TE 过程。TE 过程模拟有 21 种预编程故障，丰富多样的故障类型能够真实反映实际工业工程中的众多问题。改变该过程中产物 G 和 H 的比例，可以对其进行各种操作模式。由于多工况过程具有不稳定性，受各种变量影响较大，所以采用多种控制策略来解决该问题。本文采用的是分散控制，由 Ricker 提出，可从文献[18]提供的网站上下载其仿真代码。本文只对该过程中的工况 1 和工况 3 进行研究。

2.2 仿真结果

PCA 和 KPCA 中的 1 200 个训练数据样本是从 TE 过程的工况 1 和工况 3 中选取的，此外，还需选取 400 个正常数据作为校验数据。由于 SVM 模型需要训练，故选取 800 个故障数据和 400 个正常数据作为 SVM 的训练数据。选取工况 1 和工况 3 中的故障 1 – 故障 5、故障 7 – 故障 10 为测试故障类型。测试故障数据集是在工况 1 和工况 3 中每个故障类型下各选取 400 个样本组成。对 TE 多工况过程的 9 个故障，运用局部概率密度进行预处理，并使用 PCA、KPCA 和 SVM 算法分别对测试数据进行故障检测。

表 1 表示 TE 多工况过程中影响每个故障的主要变量。使用 PCA 进行降维，通过 SPE 贡献图得出每个故障的主要影响变量。根据相关变量，可以分析每个故障产生的原因。

表 1 每个故障的主要影响变量

故障	变量
1	48
2	49
3	21
4	1
5	24
7	48
8	37
9	18
10	9

图 1 是正常数据和故障 1 中变量 48 的序列图。可以看出，变量 48 在测试数据集中有明显的阶跃型变化，对这种明显的故障变化，PCA 和 KPCA 很容易检测出来。图 2 为 3 种算法对校验数据和故障 1 的检测效果图，其中 PCA 和 KPCA 的故障检测率都为 100%，具有非常好的检测效果。采用 SVM 方法对校验和故障数据进行分类，若分类标签为 0，则表示正常数据；反之，若分类标签为 1，则为故障数据。就大多数故障数据而言，SVM 能够正确地对其进行分类且故障检测率为 98.12%，检测效果与 PCA 和 KPCA 相似。在故障 1 的检测中，PCA 和 KPCA 统计量比 SVM 更敏感。对于变量变化比较明显的故障，许多方法故障检测都很好。

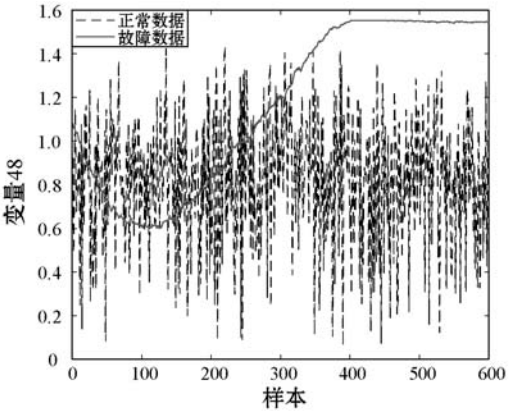
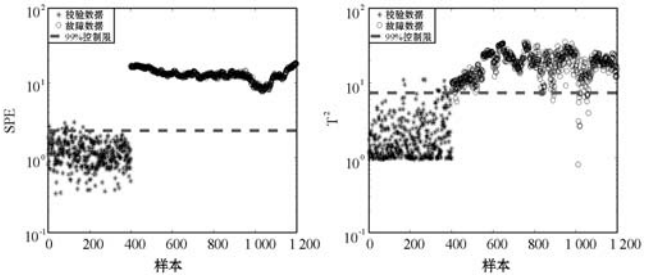
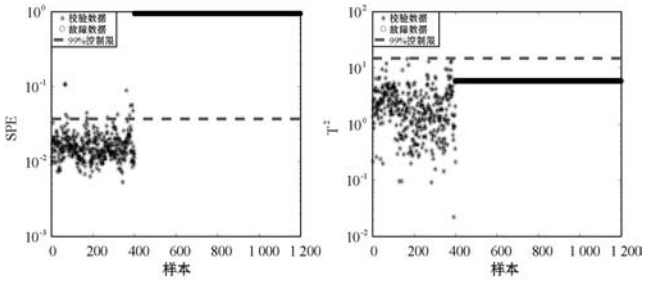


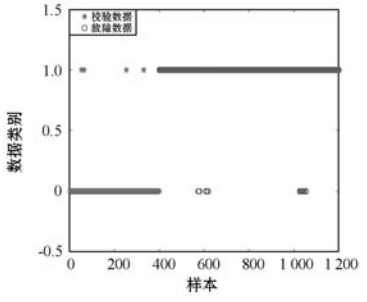
图 1 故障 1 中变量 48



(a) SPE-PCA　　(b) T^2-PCA



(c) SPE-KPCA　　(d) T^2-KPCA



(e) SVM

图 2　3 种方法对故障 1 的检测图

图 3 为正常数据和故障 4 中变量 1 的序列图,其图中故障数据的波动范围略大于正常数据,属于微小型故障。图 4 为 3 种方法对校验数据和故障 4 的检测图。由于故障数据波动范围略大于正常数据,PCA 和 KPCA 对该类故障检测不敏感,因此检测效果非常差,故障检测率也仅有 41.5% 和 11.63%。SVM 对该故障的故障检测率为 97.75%,由于训练数据集中包含故障样本,SVM 模型在训练时会学习到该种故障的变化并将其正确分类,因此,SVM 模型对故障样本的分类准确性高于 PCA 和 KPCA。

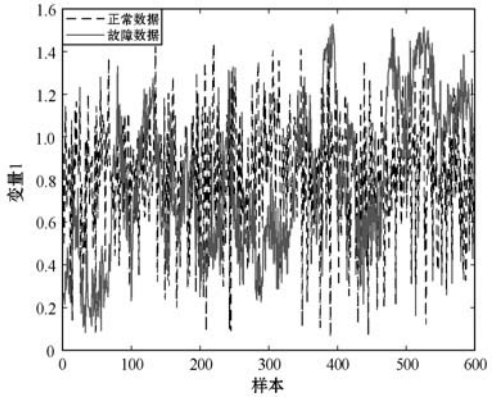
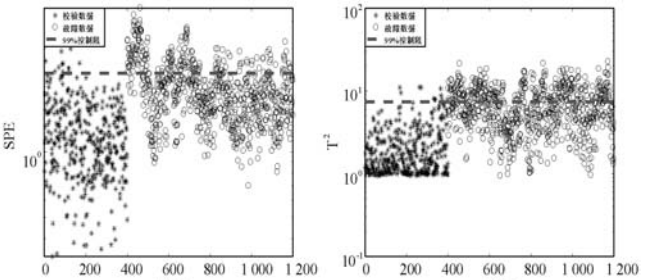
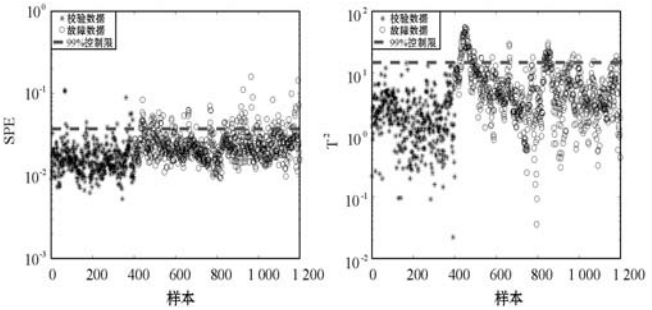


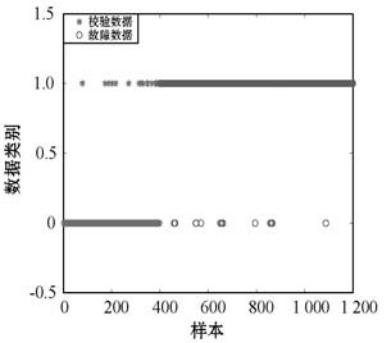
图 3　故障 4 中变量 1



(a) SPE-PCA　　(b) T^2-PCA



(c) SPE-KPCA　　(d) T^2-KPCA



(e) SVM

图 4　3 种方法对故障 4 的检测图

图 5 为正常数据和故障 9 中变量 18 的序列图。故障 9 的变量 18 属于脉冲型故障,波动范围明显大于正常数据集。正常数据的波动范围在 0.2 至 1.2 之间,而故障发生时数据波动范围在 0 至 1.6 之间。图 6 为 3 种方法对校验数据和故障 9 的检测图。分析可知,PCA 在检测故障时,SPE 统计量的波动幅度较大,部分数据未超出 99% 控制限,故障检测率为 68%,而 KPCA 的 SPE 检测指标的故障检测率为 100%,相比之下,KPCA 算法更适合于该故障的检测,检测效果更好。KPCA 通过核函数将将低维空间的数据投影到高维空间上,对这种变化的特征提取效果好。SVM 的故障检测率为 95.25%,对于大部分数据都能做到正确分类,而有些数据不能正确区分的主要原因是特征提取不够有效和合理,使得 SVM 分类不能很好地识别故障的特征。因此,与 KPCA 方法相比,SVM 的故障检测率较低。

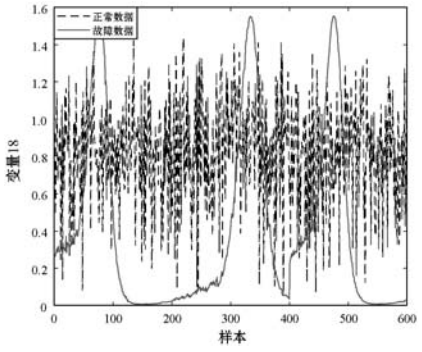
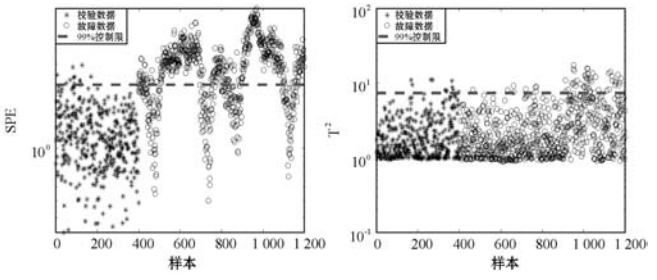
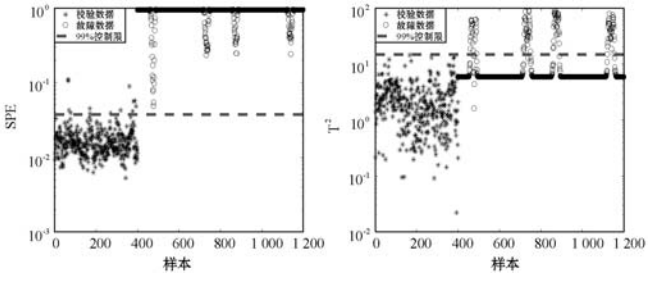


图 5 故障 9 中变量 18



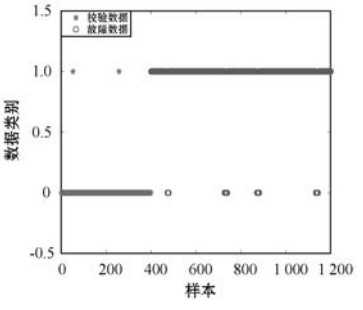
(a) SPE-PCA

(b) T^2-PCA



(c) SPE-KPCA

(d) T^2-KPCA



(e) SVM

图 6 3 种方法对故障 9 的检测图

对 TE 多工况过程的 9 种故障,运用误报率和故障检测率来衡量算法的优越性。表 2 是 3 种算法对校验数据的检测结果对比。可以看出,KPCA 的 T^2 指标的误报率最低,SVM 的误报率低于 PCA 和 KPCA 的 SPE 指标,高于 KPCA 的 T^2 指标,但是 SVM 的误报率是可以接受的。表 3 是 3 种算法对故障数据的检测率结果对比。通过表 3 可知,对故障 1、故障 2 和故障 7,PCA 的检测效果优于 KPCA 和 SVM。对故障 9 和故障 10,KPCA 的检测效果优于 PCA 和 SVM。SVM 对故障 3 – 故障 5 和故障 8 的故障检测率要明显高于 PCA 和 KPCA。SVM 的平均故障检测率最高,这表明 SVM 具有较强的鲁棒性和泛化能力。综上所述,不同算法适用于不同类型的故障。

表 2 三种算法的校验数据检测结果对比

检测类别	PCA		KPCA		SVM
	SPE	T^2	SPE	T^2	
误报率/%	2.75	2.75	3.25	0.5	2.25

表 3 三种算法的故障检测结果对比(%)

故障	PCA		KPCA		SVM
	SPE	T^2	SPE	T^2	
1	**100**	97.00	100	0	98.12
2	**100**	71.13	100	47.88	96.25
3	2.00	2.25	3.88	0.75	**43.00**
4	25.87	41.50	11.50	11.63	**97.75**
5	33.38	25.50	50.75	0.25	**64.88**
7	**100**	**99.38**	100	2.38	86.50
8	4.38	1.75	5.00	1.00	**34.38**
9	68.00	6.75	**100**	15.00	95.25
10	66.88	24.13	**91.88**	22.50	89.00
平均故障检测率	55.61	41.04	62.56	11.27	**78.35**

3 结 语

本文提出一种基于密度空间支持向量机的多工况过程故障检测方法。引入局部概率密度函数将多工况数据转化为单工况数据,消除多工况和非高斯特性。在此基础上运用 PCA、KPCA 和 SVM 分别进行故障检测。在实际的 TE 多工况工业数据中,运用本文方法对该过程进行监视和检测,由仿真结果可知,3 种算法分别适用于不同的故障类型,在平均故障检测率上 SVM 的检测效果更好,验证了 SVM 的有效性以及独特优越性。

参 考 文 献

[1] 周东华,李钢,李元. 数据驱动的工业过程故障检测与诊断技术[M]. 北京:科学出版社,2011:1 – 76.

[2] 张汉元,田学民. 基于异步 PCA 的故障识别方法[J]. 高校化学工程学报,2016,30(3):680 – 685.

[3] Gueddi I, Nasri O, Benothman K, et al. Fault detection and isolation of spacecraft thrusters using an extended principal component analysis to interval data[J]. International Journal of Control Automation & Systems,2017,15(2):1 – 14.

[4] Hamadache M, Lee D. Principal component analysis based signal-to-noise ratio improvement for inchoate faulty signals:

Application to ball bearing fault detection[J]. International Journal of Control Automation & Systems, 2017, 15(2): 1 – 12.

[5] Jaffel I, Taouali O, Harkat M F, et al. Kernel principal component analysis with reduced complexity for nonlinear dynamic process monitoring[J]. International Journal of Advanced Manufacturing Technology, 2016, 88(9 – 12): 1 – 15.

[6] Adedigba S A, Khan F, Yang M. Dynamic failure analysis of process systems using principal component analysis and bayesian network[J]. Industrial & Engineering Chemistry Research, 2017, 56(8): 2094 – 2106.

[7] Ge Z Q, Yang C J, Song Z H. Improved kernel PCA-based monitoring approach for nonlinear processes[J]. Chemical Engineering Science, 2009, 64(9): 2245 – 2255.

[8] Schölkopf B, Smola A, Müller K. Nonlinear component analysis as a kernel eigenvalue problem[J]. Neural Computation, 1998, 10(5): 1299 – 1319.

[9] Wu F, Yin S, Karimi H R. Fault detection and diagnosis in process data using support vector machines[J]. Journal of Applied Mathematics, 2014(8): 1 – 9.

[10] Shen L, Wang H, Xu L D, et al. Identity management based on PCA and SVM[J]. Information Systems Frontiers, 2016, 18(4): 711 – 716.

[11] Vapnik V N. The nature of statistical learning theory[M]. Springer, 1999.

[12] 郭金玉,刘玉超,李元. 一种基于改进局部熵 PCA 的工业过程故障检测方法[J]. 高校化学工程学报, 2019, 33(4): 922 – 932.

[13] Guo J, Wang X, Li Y. Fault detection based on improved local entropy locality preserving projections in multimodal processes[J]. Journal of Chemometrics, 2019, 33(3): 3116.

[14] Downs J J, Vogel E F. A plant-wide industrial process control problem[J]. Computers and Chemical Engineering, 1993, 17(3): 245 – 255.

[15] Mcavoy T J, Ye N. Base control for the Tennessee Eastman problem[J]. Computers & Chemical Engineering, 1994, 18(5): 383 – 413.

[16] Lee G, Han C, Yoon E S. Multiple-fault diagnosis of the Tennessee Eastman process based on system decomposition and dynamic PLS[J]. Industrial & Engineering Chemistry Research, 2004, 43(25): 8037 – 8048.

[17] Yin S, Ding S X, Haghani A, et al. A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process[J]. Journal of Process Control, 2012, 22(9): 1567 – 1581.

[18] Ma H, Hu Y, Shi H. Fault detection and identification based on the neighborhood standardized local outlier factor method[J]. Industrial & Engineering Chemistry Research, 2013, 52(6): 2389 – 2402.

(上接第 12 页)

[15] Duan Y, Wu O. Learning with auxiliary less-noisy labels[J]. IEEE Transactions on Neural Network and Learning Systems, 2017, 28(7): 1716 – 1721.

[16] Miao Q, Cao Y, Xia G, et al. RBoost: Label noise-robust boosting algorithm based on a nonconvex loss function and the numerically stable base learners[J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 27(11): 2216 – 2228.

[17] Varon C, Alzate C, Suykens J A K. Noise level estimation for model selection in Kernel PCA denoising[J]. IEEE Transactions on Neural Networks and Learning Systems, 2015, 26(11): 2650 – 2663.

[18] Sun J W, Zhao F Y, Wang C J, et al. Identifying and correcting mislabeled training instances[C]//Future Generation Communication and Networking, 2008: 244 – 250.

[19] Ekambaram R, Fefilatyev S, Shreve M, et al. Active cleaning of label noise[J]. Pattern Recognition, 2015, 51: 463 – 480.

[20] Malossini A, Blanzieri E, Ng R T. Detecting potential labeling errors in microarrays by data perturbation[J]. Bioinformatics, 2006, 22(17): 2114 – 2121.

[21] Zhang J, Sheng V S, Wu J, et al. Multi-Class ground truth inference in crowdsourcing with clustering[J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(4): 1080 – 1085.

[22] Li C, Sheng V S, Jiang L, et al. Noise filtering to improve data and model quality for crowdsourcing[J]. Knowledge-Based Systems, 2016, 107: 96 – 103.

[23] Nicholson B, Zhang J, Sheng V S, et al. Label noise correction methods[C]//2015 IEEE International Conference on Data Science and Advanced Analytics(DSAA), 2015: 1 – 9.

[24] Teng C M. Correcting noisy data[C]//16th International Conference on Machine Learning(ICML 1999), 1999: 239 – 241.

[25] Triguero I, Sáez J A, Luengo J, et al. On the characterization of noise filters for self-training semi-supervised in nearest neighbor classification[J]. Neurocomputing, 2014, 132: 30 – 41.

[26] Zhang J, Sheng V S, Wu J, et al. Improving label quality in crowdsourcing using noise correction[C]//24th ACM International on Conference on Information and Knowledge Management, 2015.

[27] Li C, Jiang L, Xu W. Noise correction to improve data and model quality for crowdsourcing[J]. Engineering Applications of Artificial Intelligence, 2019, 82: 184 – 191.

[28] Zhang J, Sheng V S, Nicholson B A, et al. CEKA: A tool for mining the wisdom of crowds[J]. Journal of Machine Learning Research, 2015, 16(1): 2853 – 2858.

[29] Witten I H, Frank E. 数据挖掘:实用机器学习工具与技术[M]. 北京:机械工业出版社, 2005.