

doi: 10.11918/j.issn.0367-6234.2015.07.020

采用改进型 DENCLUE 和 SVM 的电子皮带秤故障诊断

朱 亮¹, 李东波¹, 何 非¹, 童一飞¹, 袁延强²

(1. 南京理工大学 机械工程学院, 210094 南京; 2. 南京三埃工控股份有限公司, 211100 南京)

摘要: 针对电子皮带秤的在线故障诊断问题, 提出一种基于改进型 DENCLUE 聚类分析和偏二叉树支持向量机(SVM)分类器的故障在线检测和诊断方法。由于故障数据随设备流量的变化而变化, 采用改进 DENCLUE 聚类算法对实时检测到的数据进行聚类分析, 分离出故障数据, 实现在线故障检测; 将 DENCLUE 算法中的密度估计方法引入到支持向量机中, 提出一种基于类内相似密度和类间相似密度构建可分性测度和二叉树结构的改进型 BTSVM, 结合标准数据集验证了改进型 BTSVM 的优越性, 并利用该分类器对检测出的故障进行故障模式在线识别诊断。对阵列式皮带秤进行试验, 结果表明, 提出的故障在线检测和诊断模型更适合散状物料连续称重系统的在线故障检测诊断。

关键词: DENCLUE; 二叉树; 支持向量机; 电子皮带秤; 在线故障诊断

中图分类号: TH165

文献标志码: A

文章编号: 0367-6234(2015)07-0122-07

Fault diagnosis of belt weigher using the improved DENCLUE and SVM

ZHU Liang¹, LI Dongbo¹, HE Fei¹, TONG Yifei¹, YUAN Yanqiang²

(1. School of Mechanical Engineering, Nanjing University of Science & Technology, 210094 Nanjing, China;

2. Nanjing Sanai Industrial Automation Co. Ltd. 211100 Nanjing, China)

Abstract: A method of on-line fault detection and diagnosis based on the modified DENCLUE clustering and partial binary tree support vector machine (SVM) is proposed for on-line fault diagnosis problem of bulk weighing equipment—electronic belt weigher. Firstly, in view of the fault data varying with equipment flow, a modified DENCLUE clustering algorithm is designed to realize the online fault detection by isolating the fault data after the clustering analysis of the real-time data. Secondly, the density estimation method in DENCLUE algorithm is introduced into the support vector machine, and then an improved BTSVM, in which the separability measure and binary tree structure is built based on the similar density within class and between class, is presented to recognize the detected fault on-line. The improved BTSVM is also verified the superiority by the standard dataset. Finally, the proposed online fault detection and diagnosis model is verified more suitable for the online fault detection and diagnosis of bulk weighing equipment by the array belt weigher experiments.

Keywords: DENCLUE; binary tree; support vector machine; belt weigher; online fault diagnosis

散状物料连续称重设备工作环境恶劣且长时间高荷载作业, 致使计量精度失准以及各种设备故障频繁发生, 直接造成大量贸易经济损失; 同时其故障维修服务也一直困扰着用户和厂家, 大部分故障都要派遣经验丰富的专业人员进行现场诊

断维修, 致使故障诊断维修成本一直居高不下; 此外, 许多故障都是在贸易运输过程中产生的, 需要及时检测、诊断及维修, 以避免更多经济损失^[1]。散状物料连续称重设备使用较多的有电子皮带秤、视觉秤、核子秤等, 虽然称重原理不同, 但设备的数据形态具有很大的相似性, 都会随着流量的变化而变化, 故本文以电子皮带秤为代表, 研究散状物料连续称重设备故障在线检测和诊断。随着计量精度要求的不断提高, 电子皮带秤已从单称重传感器发展到多称重传感器, 其故障在线检测和诊断的主要途径是对其多个称重传感器的数

收稿日期: 2014-06-14.

基金项目: 国家自然科学基金(51105157);

科技型中小企业技术创新基金(13C26213202062)。

作者简介: 朱 亮(1988—), 男, 博士研究生;

李东波(1957—), 男, 教授, 博士生导师。

通信作者: 李东波, lidongbo11@gmail.com.

据进行在线检测、分析、挖掘,以提取正常数据点,分离故障信息,最终识别故障数据,其实质是一个在线“聚类 & 多分类”的问题。

在故障信息分离阶段,应用聚类算法可降低故障数据维度,减少后续识别模型的训练时间,从而大大提高故障识别效率。聚类是无监督学习算法,算法简单且效率高,对随机性信号具有很强的鲁棒性。Khediri 等^[2]将核 k -means 聚类算法应用于故障数据前处理,Alaei 等^[3]提出一种基于 WFCM 聚类的在线故障检测方法,李亚敏等^[4]将相似性传播聚类算法引入到航空发动机突发故障诊断中。但文献[2-3]中需要提前指定聚类类数,文献[4-5]在处理含噪声数据时不太理想^[5]。DENCLUE(density-based clustering)算法是一种基于核密度估计(Kernel Density Estimation-based)的聚类算法^[6],无需提前指定聚类类数和聚类中心,而且对噪声具有很好的免疫效果^[7-8],但对密度不均匀的数据鲁棒性不好。

在状态识别的多分类阶段,机械设备故障样本的难以获取制约了人工神经网络等智能诊断方法的发展和应用^[9-10]。基于统计学习理论的支持向量机(support vector machine SVM)是一种针对有限样本的通用机器学习方法,能够避免过学习、局部极值等问题,有比神经网络更好的泛化性能^[11],通过引入再生核理论避免了维数灾难,成功解决了高维非线性问题,已广泛应用于故障诊断^[12-14]。故障诊断中常用的 SVM 多分类器主要有:“一对一”(1-a-1)、“一对其余”(1-a-r)、二叉树和决策导向无环图(decision directed acyclic graph DDAG)等等^[15]。本文采用二叉树支持向量机(Binary Tree SVM),BTSVM 与 1-a-1、1-a-r、DDAG 等组合策略相比,具有分类器数目少,不存在不可识别区域,分类速度快以及严格的泛化误差理论^[16-17]等优点,广泛应用于在线故障识别。

本文提出一种改进型 DENCLUE 聚类算法,将其应用于电子皮带秤在线故障检测,并提出一种二叉树设计方法,结合二叉树 SVM 对电子皮带秤进行在线故障识别。通过试验对本文诊断模型性能进行验证。

1 基于聚类和多分类的电子皮带秤故障在线检测和诊断

电子皮带秤称重传感器数据会随着其瞬时流量的变化而变化,故其未知的故障数据也随之变化,难以识别。正常数据点以及相同故障的数据点由于相似性极高会以集簇的形式出现,故本文基

于无监督学习思想,通过在数据空间中进行集簇聚类,分离正常数据点和故障数据点,实现在线故障检测。在线故障诊断即在检测出故障后,判别故障为哪一种已知的故障模式,是有监督学习的多分类问题。综上所述,电子皮带秤的故障在线检测和诊断是一个在线“聚类 & 多分类”的问题。

本文结合 DENCLUE 聚类与 BTSVM 分类器对电子皮带秤进行故障在线检测和诊断。在线故障检测时,采用基于 DENCLUE 的算法对实时采集到的称重传感器数据 $D' = \{x'_1, x'_2, \dots, x'_n\} \subset \mathbf{R}^m$ 进行在线聚类分析,提取正常传感器数据。得到正常数据 $D'_0 \subset D'$ 后,若正常数据数目等于传感器总数, $\#(D'_0) = \#(D')$, 则无故障;若 $\#(D'_0) < \#(D')$, 则存在故障。

在线故障诊断是在检测到故障后,采用已训练好的 BTSVM 分类器,以正常传感器数据的平均值、各现场环境数据以及原始数据 D'_0 为故障样本特征,对多类故障进行在线故障模式识别,具体流程如图 1。

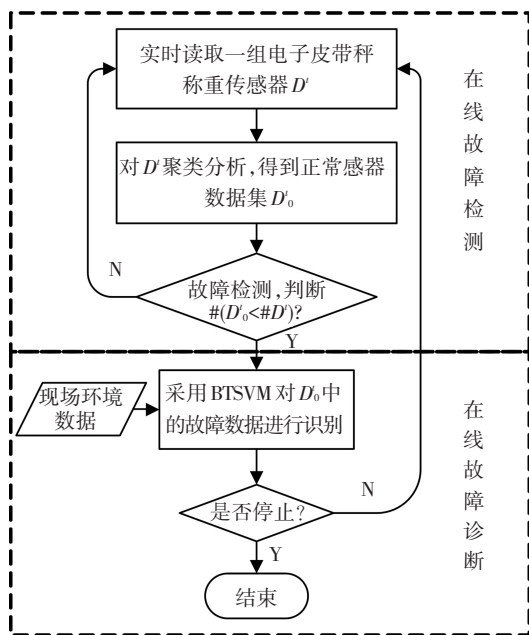


图 1 电子皮带秤故障在线检测和诊断流程

2 基于改进型 DENCLUE 算法的在线故障检测

2.1 DENCLUE 算法

DENCLUE 算法是一种基于密度的聚类算法,采用基于网格的方法提高了性能,其核心思想是采用核密度估计对数据空间中每点密度建模,并以自然的方式定义以密度吸引子(密度估计函数的局部极值点)为核心的集簇。

定义 1 若函数 $K: \mathbf{R}^m \rightarrow \mathbf{R}$ 满足 $\int_{\mathbf{R}^m} K(x) dx =$

1, 则 K 为一核函数

样本点 $x_i \in X \subset \mathbf{R}^m, m \in \mathbf{N}$ 对其周围点 $x \in X$ 的影响函数可用核函数表示, 若用高斯核函数, 有

$$K(x_i, x) = \frac{1}{\sigma^m (2\pi)^{\frac{m}{2}}} \exp\left(-\frac{d(x_i, x)^2}{2\sigma^2}\right) = \frac{1}{\sigma^m (2\pi)^{\frac{m}{2}}} \exp\left(-\frac{(x_i - x)^T(x_i - x)}{2\sigma^2}\right).$$

数据空间 $D = \{x_1, x_2, \dots, x_n\} \subset \mathbf{R}^m, m, n \in \mathbf{N}$ 中任一点 x 的密度估计为其他所有样本点对该点影响函数之和, 即

$$f^D(x) = \frac{1}{n} \sum_{i=1}^n K(x_i, x) = \frac{1}{n\sigma^m (2\pi)^{\frac{m}{2}}} \sum_{i=1}^n \exp\left(-\frac{d(x_i, x)^2}{2\sigma^2}\right). \quad (1)$$

定义 2 点 $x \in D$ 密度吸引于密度吸引子 x^* , 当且仅当开始于 x 的爬山算法最后收敛于 x^* . DENCLUE 算法中的多中心类定义见定义 3.

定义 3 集合 $C \subset D$ 为一关于参数 σ, ξ 的多中心类, 当且仅当其满足:

- 1) $\forall x \in C \exists x^* \in X: f^D(x^*) \geq \xi$, 且 x 密度吸引于 x^* , 其中 X 为密度吸引子集合;
- 2) $\forall x_1^*, x_2^* \in X$, 存在一条 x_1^* 到 x_2^* 的路径, 该路径上每点的密度值都不小于阈值.

在上述定义基础上, DENCLUE 聚类 (见图 2) 步骤如下:

- 1) 推导空间 D 中数据点的密度函数;
- 2) 对于任意密度吸引子, 将密度吸引于其的点定义为一簇;
- 3) 丢弃密度函数值小于噪声阈 ξ 值的簇及点;
- 4) 根据定义 3 合并相关簇.

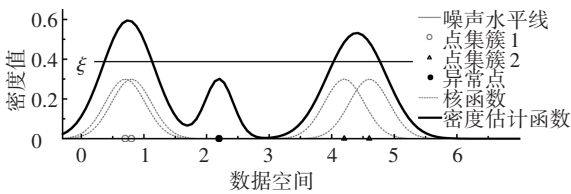


图 2 DENCLUE 聚类示意

虽然 DENCLUE 算法有诸如抗噪声能力强, 能发现任意形状簇等优点, 但同时也存在着严重的缺陷, 例如算法复杂度比其他基于密度的聚类算法要大, 处理密度不相同簇的数据时不具有很好鲁棒性等. 因此, 若要将其引入到电子皮带秤在线故障检测还有待改进.

2.2 基于改进型 DENCLUE 算法的电子皮带秤称重传感器在线故障检测

在实际情况中, 故障点相对于正常数据点是

少量的, 因此可作出以下合理假设:

若 D^t 为 t 时刻 N 个称重传感器、含有 k (未知) 类故障的样本数据, $D_0^t \subset D^t$ 为正常传感器数据点集合, $D_i^t \subseteq D^t, i = 1, 2, \dots, k$ 为第 i 类故障样本点集合, 则集合 D_0^t 中元素数目大于任一故障集合 D_i^t 中数目, 即 $\#(D_0^t) > \max_i(\#(D_i^t))$.

本文基于上述假设, 提出采用在线聚类分析提取正常传感器数据点, 分离故障数据以检测故障. 在线提取正常传感器数据点有两种思路: 一是采用诸如 DBSCAN、DENCLUE、模糊系统聚类无需提前指定类数 k 的聚类算法, 将 t 时刻数据点集合 D^t 聚类成 $\{C_1, C_2, \dots, C_k\}$, 然后得到 $D_0^t = \operatorname{argmax}_i(\#(D_i^t))$; 二是直接将正常数据从数据背景中提取出来. 后者比前者直接、简单、效率高, 更适合在线. 本文基于第二种思路, 对 DENCLUE 算法进行改进, 具体改进如下:

- 1) 采用相似系数 $r(x', x)$ 替代距离函数 $d(x', x)$.

电子皮带秤在不同输送流量时, 称重传感器数据点集合 D^t 内密度不相同, $d(x', x)$ 的分布也有较大变化, DENCLUE 需采用不同噪声阈值 ξ 和平滑因子 σ 才能将故障数据点与正常数据点分开, 即 DENCLUE 处理不同流量数据时不具有很好鲁棒性. 针对这一问题, 本文采用相似系数替代距离函数 $d(x', x)$ 、并重新构造密度估计函数以提高算法鲁棒性. 相似函数为

$$r(x_i, x) = \sqrt{\sum_{i=1}^m s(x_i(i), x(i))^2},$$

$$s(x_i(i), x(i)) = \begin{cases} 0, & x_i(i) = x(i) = 0; \\ \frac{|x_i(i) - x(i)|}{\max(|x_i(i)|, |x(i)|)}, & \text{else.} \end{cases}$$

- 2) 选取动态噪声阈值, 提取正常传感器数据点.

以全局极值点 $x^* = \operatorname{argmax}_x(f^D(x))$ 作为全局密度吸引子, 再选取动态噪声阈值 $\xi = \lambda \cdot f^D(x^*), \lambda \in (0, 1)$, 提取正常数据点集合 $D_0^t = \{x | f^D(x) > \xi, x \in D^t\}$, 余下的噪声信号即为故障数据. D_0^t 中可能有多个吸引子, 但符合定义 3.

通过改进 2), 算法只需找出全局密度吸引子、无需找全所有, 采用阈值过滤出正常数据点、取代了原算法中的 2)、3)、4) 步骤, 将原先的多类聚类算法化简为二类聚类, 大大降低了算法复杂度, 更适合在线. 改进后算法具体步骤如下:

Step 1 采用相似系数 $r(x', x)$ 构造相似核函数:

$$K(x_i, x) = \frac{1}{\sigma^m (2\pi)^{\frac{m}{2}}} \exp\left(-\frac{d(x_i, x)^2}{2\sigma^2}\right). \quad (2)$$

显然, 式(2)符合定义 1. 再将式(2)代入式(1)得相似密度估计函数:

$$f^D(x) = \frac{1}{n\sigma^m(2\pi)^{\frac{m}{2}}} \sum_{i=1}^n \exp\left(-\frac{r(x_i, x)^2}{2\sigma^2}\right).$$

Step 2 寻找全局密度吸引子 x^* , 提取正常数据点簇 $D'_0 = \{x | f^D(x) > \lambda \cdot f^D(x^*), x \in D'\}$.

采用改进型 DENCLUE 算法(见图 3)提取出正常数据点 D'_0 后, 若 $D'_0 = D'$, 即 $\#(D'_0) = \mathbf{N}$, 则无故障; 若 $\#(D'_0) < \mathbf{N}$, 则出现故障, 完成 t 时刻在线故障检测, 进入故障诊断阶段.

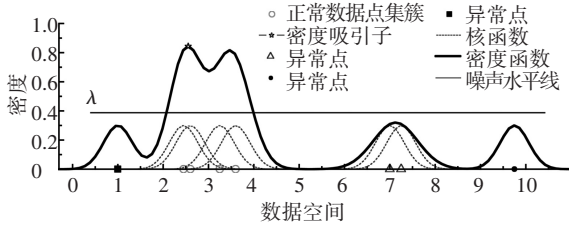


图 3 改进型 DENCLUE 聚类示意图

3 改进型偏二叉树支持向量机

3.1 二叉树支持向量机

对于 N 类问题, 二叉树 SVM 采用二叉树结构构造 $N - 1$ 个分类器. 二叉树的层次结构有两种: 一种是在每个内节点处, 由一个类与剩下的类构造分割面, 即偏二叉树; 另一种是在内节点处, 可以是多个类与多个类的分割^[18]. 本文采用偏二叉树结构 SVM 进行故障诊断.

BTSVM 整体分类性能主要与二叉树生成顺序、节点位置有关, 越是上层节点, SVM 的分类性能(分类间隔大)对 BTSVM 模型分类性能影响越大. 为了减少误差积累, 以及让分布广的类拥有更大的分割区域, 应最早分割可分性最大的类以提高分类推广能力^{[17][19]}. 本文以类内样本和类间样本的综合分布情况计算可分性测度.

3.2 类的可分性测度

为了更准确地估计各个样本类的分布情况, 寻找可分性最大的类以及构建最优性能的二叉树结构, 本文将 DENCLUE 聚类算法中的密度估计方法引入到二叉树支持向量机中, 以 DENCLUE 算法中的相似密度函数为基础建立可分性测度, 并以此对各个故障类进行排序. 由于各类故障数据真实分布无法获知, 可用有限样本数据的分布对真实分布进行近似估计, 本文分别定义类内平均相似密度、类间平均相似密度, 对类内、类间样本分布情况进行估计. 在含 k 类故障样本的数据空间 $\{D_0, D_1, D_2, \dots, D_k\}$ 中, 类内、类间平均相似密度定义如下:

第 i 类内平均相似密度为

$$\hat{F}_{D_i} = \frac{1}{N_{i x_p, x_q \in D_i}^2} \sum K(x_p, x_q),$$

$$N_i = \#(D_i), i = 0, 1, \dots, k;$$

第 i 类间平均相似密度为

$$\check{F}_{D_i} = \frac{1}{N_{i x_p \in D_i}} \sum_{t=1, t \neq i}^k \left(\frac{1}{N_{t x_q \in D_t}} \sum K(x_p, x_q) \right).$$

其中, 计算电子皮带秤各个故障类可分性测度时, $K(x_j, x)$ 为式(2)中所定义的核函数, 并且故障样本必须是同一流速情况下的故障样本, 否则无意义. 类内相似密度越小代表类内部分布越广, 类间相似密度越小代表该类与其他类差异越大、越可分. 根据类内分布越广、类间可分离度越大的原则^[20-21], 综合式(3)、(4)得第 i 类的可分性测度:

$$\mu_i = (\hat{F}_{D_i} + \alpha \cdot \check{F}_{D_i})^{-1}.$$

式中 α 为调节类内相似密度和类间相似密度的权重系数, 可通过参数优化确定. 由式(5)可看出, μ_i 值越大, 该类可分性越强.

在计算电子皮带秤 k 类故障的可分性测度时, 还需将样本按不同流量分为 $\{D^{0_1}, D^{0_2}, \dots, D^{0_s}\}$ s 个子样本集, 然后计算 k 类故障在每个流量样本 $D^{0_j}(j = 1, 2, \dots, s)$ 中的可分性测度 $\mu_i^{0_j}$, 最后取所有流量样本下的测度平均值作为每类故障的可分性测度:

$$\mu_i = \frac{1}{s} \sum_{j=1}^s \mu_i^{0_j}.$$

3.3 算法步骤

改进后的 BTSVM 算法步骤如下:

- 1) 确定权值, 根据式(5)或式(6)计算各类故障的可分性测度;
- 2) 将各类按其可分性测度由大到小排序(当存在两个或两个以上类别的可分性测度相同时, 依据人工经验自由排序), 得到类别排列 $c_1, c_2, \dots, c_k, c_i \in \{1, 2, \dots, k\}, i = 1, 2, \dots, k$, 即为各故障类的可分性级别;
- 3) 依据排列 c_1, c_2, \dots, c_k 确定偏二叉树各节点位置, 并采用二分类 C-SVM 训练算法构造各节点最优分类超平面, 构造流程与其他偏二叉树 SVM 几乎相同.

3.4 标准数据集试验验证

为验证改进型偏二叉树 SVM 算法的优越性, 本文以 UCI 机器学习数据库中 wine、iris、segment 和 vowel 数据集^[22] 分别与一对一 SVM、一对多 SVM 及改进 BTSVM 进行试验比较, 各算法均在 libsvm 基础上修改实现, 标准数据信息如表 1.

表 1 标准数据集信息

数据集	训练样本数	测试样本数	类别数	属性数
wine	90	88	3	13
iris	75	75	3	4
segment	210	2 100	7	19
vowel	528	462	11	10

为了方便后续数据处理,加快算法收敛速度,试验时对各样本先进行归一化处理,线性调整到 [0,1].算法中核函数统一采用高斯核,以网格搜索最优参数组合.改进 BTSVM 中的可分性测度权值 α 也采用网格搜索,通过搜索确定权值变化时对应的二叉树层次结构以及各个结构的 α 临界值.对每一种结构寻找其最优 SVM 参数组合,以获得最高识别准确率,并以此准确率为对应结构的识别准确率.试验结果如表 2,从表 2 中可看出改进后的 BTSVM 除了在 vowel 数据集中准确度稍逊于一对一算法,其他数据集中均高于其他算法,可见改进后 BTSVM 性能优越.

表 2 标准数据集试验结果

数据集	准确度/%			改进型 BTSVM 样本类别 可分性测度排列
	一对一	一对多	改进型 BTSVM	
Wine	97.73	95.45	97.73	1,3,2
Iris	95.08	94.67	97.33	1,2,3 GRASS,PATH, CEMENT,SKY, BRICKFACE, WINDOW, FOLIAGE
Segment	92.9	91.9	93.14	0,8,5,4,6,10,7,2, 9,3,1 或 0,8,5,4, 6,10,7,2,9,1,3
Vowel	63.42	62.2	62.34	

4 故障诊断模型的应用及试验验证

本文以南京三埃工控股份有限公司 QPS-皮带秤全性能试验中心的 3#阵列式皮带秤为故障诊断试验对象(图 4),皮带秤参数见表 3.该试验系统可循环走料.现场数据通过 RS485 总线传输,并采用单片机和周立功 RSM485CHT 转换器接收,上位机采用串口通信实时采集.改进型 DENCLUE 聚类算法是在 Matlab 环境下完成的,BTSVM 是在 libsvm 基础上修改,采用 Matlab 和 C/C++混合编程实现的.测试硬件环境为 Core i3-2.35 G 的 CPU,内存 6G,硬盘 500 G.

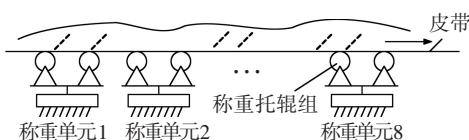


图 4 3#阵列式皮带秤示意

表 3 QPS-皮带秤全性能试验中心的 3#皮带秤参数

输送带宽度/mm	托辊间距/mm	称重单元数	输送带厚度/mm	托辊槽角/(°)	运行周期/s
1 000	1 200	8	12	30	49

4.1 阵列式皮带秤及其故障模式

3#阵列式皮带秤由 8 个称重单元组成,8 个单元都远离落料点,从而避免了落料冲击,尤其是突发的大块物料冲击对称重故障的干扰^[23].每个单元采样频率为 10 Hz,一圈运行周期约 49 s (2.21 m/s).每个时刻的每组采样数据由 8 个称重传感器实时数据及现场环境数据组成.试验过程中,在 3#阵列式皮带秤上分别模拟了表 4 中 6 类常见故障,每个称重单元的故障模式相同.由于实际情况中同一时刻最多只会 3 个称重单元区域出现故障,故本研究在试验时每组最多只模拟 4 个称重单元产生故障数据,通过在空载、200、500、800 t/h 不同流量下多次模拟试验,以验证本文故障在线检测和诊断模型的有效性.

表 4 阵列式皮带秤故障类别

故障码	故障类	故障码	故障类
F0	正常	F4	传感器接线不良
F1	传感器表面磨损	F5	托辊组托辊被卡
F2	称重架卡料	F6	传感器松动
F3	称重架松动		

4.2 基于聚类分析的在线故障检测试验及分析

电子皮带秤在线故障检测即对 8 个称重单元的检测数据进行聚类分析.试验分别采用 DENCLUE、DBSCAN、模糊层次聚类以及改进型 DENCLUE 对 8 个称重单元工作在空载、200、500、800 t/h 不同流量时的实时含噪声数据进行在线聚类分析,并对比各算法的准确度和实时性(见图 5).

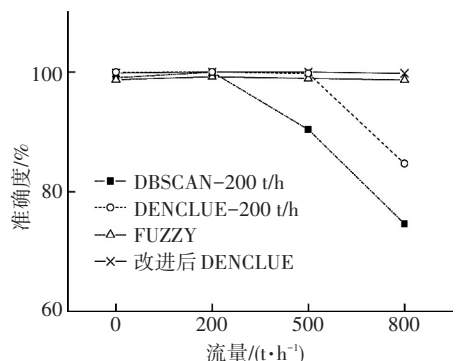
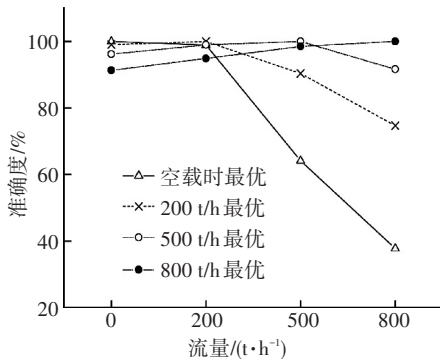


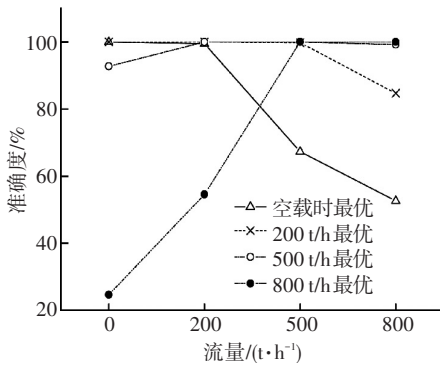
图 5 聚类算法精度比较

由图 5 可知,模糊层次聚类和改进型 DENCLUE 聚类算法的最优模型在任何流量下的聚类精度都较高且无明显变化,模糊层次聚类由

于噪声干扰, 聚类精度受到一定影响, 稍低于改进型 DENCLUE; 而采用距离函数的 DBSCAN、DENCLUE 算法在 200 t/h 得到最优聚类模型不具备很好的推广能力. 采用距离函数的 DBSCAN、DENCLUE 算法在各个流量时最优模型精度对比见图 6.



(a) DBSCAN 聚类效果图



(b) DENCLUE 聚类效果图

图 6 DBSCAN、DENCLUE 聚类效果图

由图 6 可知, 有必要使用相似函数替代距离函数, 以提高聚类算法对于不同流量数据的鲁棒性. 再次进行 4 种流量的在线聚类试验, 试验中 DENCLUE 和 DBSCAN 算法中均采用相似函数, 试验结果见表 5.

表 5 各聚类算法比较

聚类算法	平均聚类精度/%	平均每组时间/ms
模糊层次聚类	96.40	5.637
DBSCAN	99.70	2.113
DENCLUE	99.99	463.500
改进型 DENCLUE	99.99	0.722

从表 5 中可以看出, 改进型 DENCLUE 较之其他聚类算法拥有较高的聚类精度以及相当快的聚类速度, 更适合在线故障检测.

4.3 在线故障诊断试验及分析

从现场采集的空载、200、500、800 t/h 等 4 组故障数据中各抽选 250 个样本训练诊断模型, 比较各个诊断模型在散状物料连续称重设备在线故

障诊断中的性能. 为提高分类器在不同流量时的鲁棒性, 核函数均采用

$$K(x, y) = \exp(-r(x, y)^2 / (2\sigma^2)),$$

并通过修改 libsvm 中的高斯核实现; 最优参数是在网格 $C = [2^{-3}, 2^{-2}, \dots, 2^3]$, $\gamma = (2^{-1}, 2^0, \dots, 2^8)$ 范围内使用 5 折交叉验证方法搜索获得.

在使用数据集进行训练和预测时, 每个样本包括 6 个特征: 当前托辊传感器数据、同一时刻平均正常数据、同一时刻前相邻托辊称重数据(第 1 单元的前相邻称重数据取其本身)、同一时刻后相邻托辊数据(第 8 单元的后相邻称重数据取其本身)、现场温度、传感器灵敏度. 在线诊断测试时, 同一时刻平均正常数据为在线聚类分析所得. 在训练诊断模型前, 本文需要根据式(3)~(6)对正常类及各个故障类进行可分性测度估计(见表 6), 以建立偏二叉树结构.

由表 7 可看出, α 取任何值, F2 的可分性测度都是最大的, 故该类故障应最优先分离. 权值 α 通过以 1 为指数增量进行网格搜索优化, 研究发现, 当 $2^8 \leq \alpha \leq 2^{22}$ 时, 二叉树结构为固定的 8 种结构, 再以这 8 种结构构造 BTSVM, 对比 8 种结构的分类效果后, 得到 $2^{17} \leq \alpha \leq 2^{20}$ 为最优.

表 6 各故障类间、类内相似密度

故障码	类间平均相似密度	类内平均相似密度
F0	0.944 2	$1.982 7 \times 10^{-6}$
F1	0.998 4	$1.186 0 \times 10^{-6}$
F2	0.867 7	$4.169 4 \times 10^{-8}$
F3	0.977 3	$2.112 8 \times 10^{-7}$
F4	0.986 3	$2.112 8 \times 10^{-7}$
F5	1.000 0	$4.533 0 \times 10^{-7}$
F6	0.998 8	$2.035 0 \times 10^{-7}$

表 7 各权值 α 对应的二叉树结构

权值 α	可分性测度排序	最高识别准确率/%
2^8	F2>F0>F3>F4>F1>F6>F5	97.63
$2^9, 2^{11}$	F2>F0>F3>F4>F6>F1>F5	96.15
$2^{12} - 2^{14}$	F2>F0>F3>F4>F6>F5>F1	96.15
2^{15}	F2>F3>F4>F6>F0>F5>F1	98.02
2^{16}	F2>F3>F4>F6>F5>F0>F1	98.02
$2^{17} - 2^{20}$	F2>F3>F4>F6>F5>F1>F0	99.63
2^{21}	F2>F3>F6>F4>F5>F1>F0	99.20
2^{22}	F2>F6>F3>F4>F5>F1>F0	98.03

最后将二叉树结构 F2>F3>F4>F6>F5>F1>F0 的 BTSVM 与一对多算法以及一对一算法进行不同流量时的在线诊断试验, 将其诊断性能与其他分类器比较. 从表 8 可以看出“1 对 1”SVMs 平均测

试时间最长,而本文方法的测试时间明显最短,且精度最高,因此本文方法比较适合电子皮带秤的故障在线检测和诊断。

表 8 几种故障分类器性能比较

分类器	平均诊断精度/%	平均每组测试时间/ms
本文方法	99.46	0.531
1 对多	98.53	0.838
1 对 1	99.06	1.267

5 结 论

1) 采用相似密度函数替代 DENCLUE 聚类算法中的距离函数,解决了算法在处理密度不均匀数据时鲁棒性很差的问题,并采用动态阈值法替代爬山法,降低了算法复杂度,更适于在线聚类。

2) 引入 DENCLUE 中的密度估计方法,提出一种改进型偏二叉树 SVM 多分类器.该分类器利用类间、类内平均相似密度建立了可分性测度,再根据可分性测度构建偏二叉树结构.标准数据集试验表明:该分类器具有很好的优越性。

3) 结合改进型 DENCLUE 和改进型偏 BTSVM,提出了一种能准确对电子皮带秤常见故障进行在线快速检测和诊断的模型.最后通过阵列式皮带秤试验,验证了本文所提出电子皮带秤故障在线检测和诊断模型的优越性。

参 考 文 献

[1] 方原柏. 电子皮带秤[M]. 北京: 冶金工业出版社, 2008: 1-4.

[2] KHEDIRI I B, WEIHS C, LIMAM M. Kernel k -means clustering based local support vector domain description fault detection of multimodal processes [J]. Expert Systems with Applications, 2012, 39(2): 2166-2171.

[3] ALAEI H K, SALAHSHOOR K, ALAEI H K, et al. A new integrated on-line fuzzy clustering and segmentation methodology with adaptive PCA approach for process monitoring and fault detection and diagnosis [J]. Soft Comput, 2013, 17(3): 345-362.

[4] 李丽敏,王仲生,姜洪开. 基于相似性传播聚类的航空发动机突发故障诊断[J]. 振动与冲击, 2014, 33(1): 51-55.

[5] 杨小兵. 聚类分析中若干关键技术的研究[D]. 杭州: 浙江大学, 2005: 23-41.

[6] HINNEBURG A, KEIM D A. A general approach to clustering in large databases with noise [J]. Knowledge and Information Systems, 2003(5): 387-415.

[7] HINNEBURG A, GABRIEL H H. DENCLUE 2.0: Fast clustering based on kernel density estimation [J]. Lecture Notes in Computer Science, 2007, 4723: 70-80.

[8] 涂文燕,李德毅. 基于核密度估计的层次聚类算法 [J]. 系统仿真学报, 2004, 16(2): 302-309.

[9] 吴丹,金敏. 基于容错度自适应支持向量机的液压泵故障诊断[J].中国机械工程, 2011, 22(11): 2582-2587.

[10] 任能. 制冷系统故障检测、诊断及预测研究[D]. 上海: 上海交通大学, 2008: 3-4.

[11] LIU Zhiwen, CAO Hongrui, CHEN Xuefeng, et al. Multi-fault classification based on wavelet SVM with PSO algorithm to analyze vibration signals from rolling element bearings [J]. Neurocomputing, 2013, 99: 399-410.

[12] 苏小红,赵玲玲,谢琳,等. 阴影集的模糊支持向量机样本选择方法 [J]. 哈尔滨工业大学学报, 2012, 44(9): 78-84.

[13] OTHMAN N S, LAOUTI N, VALOUR J P, et al. Support vector machines combined to observers for fault diagnosis in chemical reactors [J]. The Canadian Journal of Chemical Engineering, 2014, 92(4): 685-695.

[14] 刘本德,胡昌华,蔡艳宁. 基于聚类和 SVM 多分类的容差模拟电路故障诊断 [J]. 系统仿真学报, 2009, 21(20): 6479-6482.

[15] 袁胜发,褚福磊. 支持向量机及其在机械故障诊断中的应用 [J]. 振动与冲击, 2007, 26(11): 29-34.

[16] FEI B, LIU J. Binary Tree of SVM: A new fast multiclass training and classification algorithm [J]. IEEE Trans on Neural Networks, 2006, 17(3): 696-703.

[17] BENNETT K P, CRISTIANINI N, SHAW T J, et al. Enlarging the margins in perceptron decision trees [J]. Machine Learning, 2000, 41(3): 295-313.

[18] 姚良,李艾华,王涛. 运用改进二叉树 SVM 算法的柴油机振动诊断 [J]. 振动、测试与诊断, 2010, 30(6): 689-693.

[19] 赵海洋,徐敏强,王金东. 改进二叉树支持向量机及其故障诊断方法研究 [J]. 振动工程学报, 2013, 26(5): 764-769.

[20] 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.

[21] 冯志刚,王祁,信太克规. 基于 EMD 和 SVM 的传感器故障诊断方法 [J]. 哈尔滨工业大学学报, 2009, 41(5): 59-63.

[22] UC Irvine Machine Learning Repository [DB/OL]. Irvine, California: University of California. 2014-05-03 [2014-05-10]. <http://archive.ics.uci.edu/ml/>.

[23] DONIS V K, RACHKOVSKII A E, SIN V M. How the conveyor belt length affects belt weigher accuracy [J]. Measurement Techniques, 2004, 47(2): 163-167.