

doi: 10.3969/j.issn.1005-7854.2020.02.003

基于 K-means 算法识别瓦斯传感器软故障研究

胡 宇^{1,2}, 周代勇^{1,2}

(1. 瓦斯灾害监控与应急技术国家重点实验室, 重庆 400037;
2. 中煤科工集团重庆研究院有限公司, 重庆 400039)

摘 要: 针对井下瓦斯传感器设备出现的软故障如数据漂移、数据长期低于或高于正常值、数据周期性变动和数据出现大值等问题, 提出了一种基于轮廓系数自适应最佳聚类点的 K-means 算法识别瓦斯传感器出现软故障种类的方法。该方法是利用监控系统采集的瓦斯传感器软故障信号进行小包分解处理后, 结合 RBF 神经网络进行轮廓系数 K-means 自适应算法的软故障识别训练。K-means 自适应算法能够自适应优化聚类中心点, 利用聚类中心点的迭代循环计算出最优中心点, 选择最佳聚类点进行 K-means 聚类, 从而识别软故障信号的故障类型。实验证明, 自适应轮廓系数 K-means 算法能够有效地识别瓦斯传感器软故障类型, 提高了煤矿安全监控系统数据的准确性。

关键词: 监控系统; 故障识别; 神经网络; 聚类算法; 瓦斯传感器

中图分类号: TD712 **文献标志码:** A **文章编号:** 1005-7854 (2020) 02-0010-05

Identifying gas sensors soft fault based on K-means algorithm

HU Yu^{1,2}, ZHOU Dai-yong^{1,2}

(1. State Key Laboratory of Gas Disaster Monitoring and Emergency Technology, Chongqing 400037, China;
2. China Coal Technology Engineering Group Chongqing Research Institute, Chongqing 400039, China)

Abstract: In response to soft faults in underground gas sensor equipment, such as data drift, long-term data below or above normal values, cyclical data changes, and large data values, etc., the K-means algorithm based on adaptive optimal clustering points of contour coefficients is proposed to identify the types of soft faults in gas sensors. This method uses the soft fault signal of the gas sensor collected by the monitoring system to perform packet decomposition processing, and combines the RBF neural network to perform the soft fault recognition training of the contour coefficient K-means adaptive algorithm. The K-means adaptive algorithm can automatically optimize the cluster center point, use the iterative loop of the cluster center point to calculate the optimal center point, and select the best cluster point for K-means clustering, so that it can identify the fault type of the soft fault signal. The experiment proves that the adaptive contour coefficient K-means algorithm can effectively identify the soft fault types of gas sensors and improve the accuracy of coal mine safety monitoring system data.

Key words: monitoring system; fault identification; neural networks; clustering algorithm; gas sensor

煤矿安全监测监控系统中瓦斯传感器是对煤矿井下瓦斯浓度进行采集, 瓦斯传感器采集的数

据是否是煤矿井下真实的瓦斯浓度与煤矿的安全问题息息相关^[1], 因此, 对瓦斯传感器进行软故障的研究具有重大的意义。对于瓦斯传感器软故障的故障情况, 往往会出现瓦斯传感器漂移型(数据漂移)、偏置型(数据长期低于或高于正常值)、冲击型(数据出现大值)、周期型(数据周期性变动)等软故障^[2-3], 能够有效地识别瓦斯传感器的软故障类型是为煤矿安全生产提供了一

收稿日期: 2020-02-18

基金项目: 重庆市技术创新与应用示范项目 (cstc2018jscx-msybX0192); 重庆院自主创新引导科研项目 (2018YBXM10)

第一作者: 胡宇, 硕士, 主要从事煤矿安全监控系统、井下人员定位等研究。E-mail: 840110352@qq.com

套实用技术手段。

到目前为止，传感器软故障的检测技术有基于解析数学模型的方法：参数估计法，状态估计法和等价空间法^[4]，优点是模型机理清晰，结构简单，易实现，易分析，可实时诊断。缺点是计算量大，系统复杂，存在建模误差，模型的适应性差，可靠性差，容易出现误报、漏报等现象，外部扰动的鲁棒性，系统的噪声和干扰不敏感。不依赖于数学模型的软故障方法有基于数据驱动的方法，基于知识的方法和基于离散事件的方法，优点是不需要对象的准确模型，并且适应性强，其缺点是结构复杂，难于实现^[5-6]。针对煤矿井下瓦斯传感器设备安装位置、运行环境复杂多变等特点^[7]，以及煤矿安全监测监控系统中瓦斯传感器采集到数据记录和瓦斯传感器设备的运行情况。本文提出了一种轮廓系数自适应最佳聚类点的 K-means 算法检测瓦斯传感器设备软故障监测方法。该方法是利用聚类中心点的迭代循环计算出最优中心点，选择最佳聚类点再进行 K-means 聚类。同时结合 RBF 神经网络算法对故障瓦斯传感器特征能量谱作为输入，进行自适应最佳聚类点的 K-means 算法分类训练，就能够确定运行的瓦斯传感器出现软故障状态情况。

1 设备软故障信号小波包分析

1.1 软故障信号小波包分解

小波包分解是对信号进行有效的时频分解，其尺度函数是按二进制变化的，高频段进行指数等间隔划分，低频段做进一步分解，能够更加精细地表征低频信号。因此小波包分解能够自适应地对高频信号按照信号自身的频段进行细化信号的特征，细化后的信号频率完全与信号频率是相符合的，小波包分解是提升了信号的时频分辨率^[8-9]。小波包分解存在具体如下关系：

$$S = AAA3 + DAA3 + ADA3 + DDA3 + AAD3 + DAD3 + ADD3 + DDD3 \quad (1)$$

1.2 软故障信号特征量提取

小波包分解是一种信号频率越高，分解率较高的信号分解算法，其中有 $\varphi(t)$ 尺度函数， $\psi(t)$ 为小波函数，定义正交尺度函数确定函数簇，生成小波库。

$$\begin{cases} \varphi_{2m}(t) = \sum_k h_k \varphi_m(2t - k) \\ \psi_{2m+1}(t) = \sum_k g_k \varphi_m(2t - k) \end{cases} \quad (m \in Z) \quad (2)$$

h_k, g_k 分别为低通滤波器与高通滤波系数，小波包分解和重构递增公式为：

$$\begin{cases} d_k^{2m} = \frac{1}{2} \sum_L d_L^m h_{L-2k} \\ d_k^{2m+1} = \frac{1}{2} \sum_L d_L^m g_{L-2k} \end{cases} \quad d_k^m = \sum_L d_L^{2m} h_{k-2L} + \sum_L d_L^{2m+1} g_{k-2L} \quad (3)$$

信号高低频分解的同时，可以定义 $d_1 = \{d_k^1\}$ 为分解信号，分解和重构方式可达到完全离散化。但是小波包的分解存在严重的频率混迭现象，是因为小波包采样是隔点采样，分解一次，采样频率低一半。因此小波包分解时需要将故障信号先进行移频处理，使信号的最高频降低，将小波包分解和重构改为：

$$\begin{cases} d_k^{2m} = \frac{1}{2} \sum_L d_L^m h_{L-2k}^* \\ d_k^{2m+1} = \frac{1}{2} \sum_L d_L^m g_{L-2k}^* \end{cases} \quad d_k^m = \sum_L d_L^{2m} h_{k-2L}^* + \sum_L d_L^{2m+1} g_{k-2L}^* \quad (4)$$

h_{L-2k}^*, g_{L-2k}^* 分别是 h_{L-2k}, g_{L-2k} 的对偶算子。小波包对故障信号分解 N 层后，获得故障信号的特征信号 D_{Nj} ， $j = 1, 2, 3, \dots, 2^N$ 。同时提取各频带范围的信号，并计算各频带信号的能量为：

$$E_{Nj} = \int |D_{Nj}(t)|^2 dt = \sum_{k=1}^n |d_{jk}|^2 \quad (5)$$

式中： d_{jk} —重构信号 D_{Nj} 的第 k 个离散的幅值； n —重构信号 D_{Nj} 的离散点个数， $j = 0, 1, \dots, 2^N - 1, k = 1, 2, \dots, n$ 。将获得各频带信号能量的特征向量作为 RBF 神经网络的训练样本输入。

2 K-means 聚类算法的 RBF 神经网络训练

2.1 K-means 聚类算法

K-means 算法是利用质心与质心之间的距离迭代，选择最优质心距离来进行聚类划分的算法，当类间质心点的距离相互收敛时，算法质心的距离迭代结束，聚类计算分辨出具有共同特性的类别^[10]。优点是算法实现简单快捷，数据大小可伸缩，计算数据类高效。缺点是算法对初始距离质心点依赖性大，初始质心点的偏差可能导致产生的结果偏离实际的分类，算法在运行时需要不断地更改类质心距离，数据计算量比较大，算法开销的时间也比较大，聚类的种类是预先设定的，很难估计^[11]。

针对 K-means 算法存在的缺陷, 本文对 K-means 算法随机初始聚类质心点距离进行改进, 利用轮廓系数对个体样本进行聚类点距离缩小, 类质心点轮廓系数距离进行扩大, 样本轮廓系数和类质心点轮廓系数定义为 S_i 和 S , 其中与距离关系为:

$$S = \frac{1}{n} \sum_{i=1}^n S_i \quad (6)$$

$$S_i = \frac{x(i) - y(i)}{\max[x(i), y(i)]} = \begin{cases} 1 - \frac{x(i)}{y(i)}, & x(i) < y(i) \\ 0, & x(i) = y(i) \\ \frac{y(i)}{x(i)} - 1, & x(i) > y(i) \end{cases} \quad (7)$$

上述两公式中, 当样本 $i \in C$ 类时, $x(i)$ 表示个体样本 i 和同属 C 类其他个体样本到类质心点的平均距离, 当样本 $i \notin C$ 类时, $y(i)$ 表示个体样本 i 到其他非 C 类个体样本的距离最小值, 个体轮廓系数 S_i 是用于评价样本聚类到相应种类的适合度, 故 S_i 为 $[-1, 1]$, 当 S_i 越接近 1 时, 该样本到类质心距离就越小, 样本的归属于该类的可能性就越大, n 为样本总数, 核心指标为误差平方和 SSE:

$$SSE = \sum_{j=1}^k \sum_{p \in C_j} |p - m_j|^2 \quad (8)$$

式中: C_j — 第 j 个簇; p — C_j 的样本点; m_j — C_j 的质心; SSE — 所有样本的聚类误差代表聚类效果的好坏, 具体的算法步骤如下:

1) 设定参数 M , 初始化 $j = 1$, 当 $j \leq M$ 时, 完成下列循环:

a. 调用随机初始聚类的质心点完成初步 K-means 算法分类;

b. 初始聚类后, 计算各个样本到 C 类质心点的距离矩阵 $D(j)_c$ ($c = 1, 2, \dots, K$);

c. 调用个体轮廓系数公式计算相应的个体轮廓系数 $S_i(j)$;

2) 重新对上述求得结果进行重组排列为 $S_i(j)^m$, 对应的矩阵 $D(j)_c$ 表示为 $D(j)_c^m$, ($m = 1, 2, \dots, M$), ($c = 1, 2, \dots, K$), 用 $S_i(j)^m$ 和 $D(j)_c^m$ 综合判断是否是优秀样本。

3) 对优秀样本进一步求取均值, 利用求取后的均值作为聚类中心, 再进行 K-means 聚类。

2.2 基于 RBF 神经网络训练

RBF 神经网络训练算法是一种逼近任意非线性

性函数的训练方法, 具有很快的非线性学习收敛特点, 适用于故障诊断技术中^[12-15]。学习步骤如下:

1) 确定基函数中心 C_p

瓦斯传感器软故障时产生了故障能量特征值作为输入样本, 用上述优化 K-means 聚类算法求出输入样本的各类中心, 作为径向基函数的中心^[11]。

2) 径向基函数的宽度 δ_p

基函数宽度是采用 K-means 聚类算法聚类的结果, 即各个样本到聚类质心点的平均距离为基函数宽度, δ_i ($i = 1, 2, \dots, I$) 表示 I 个基函数的方差, 其宽度为:

$$\delta_1 = \delta_2 = \dots = \delta_I = \frac{d_{\max}}{\sqrt{2I}} \quad (9)$$

式中, $d_{\max}$ 表示选择的中心两两之间的距离最大值。

3) 权值 w_p

输出层权值线性方程表示为:

$$y(x_i) = w_1 \varphi_1(\|x_i - X_1\|) + w_2 \varphi_2(\|x_i - X_2\|) + \dots + w_p \varphi_p(\|x_i - X_p\|) \quad (10)$$

每个样本的期望输出为 d 数组:

$$d_i = [\varphi_1(\|x_i - X_1\|) \varphi_2(\|x_i - X_2\|) \dots \varphi_p(\|x_i - X_p\|)] \cdot [w_1 \ w_2 \ \dots \ w_p]^T \quad (11)$$

所用样本为:

$$[d_1 \ d_2 \ \dots \ d_N]^T = \begin{bmatrix} \varphi_1(\|x_i - X_1\|) \\ \varphi_2(\|x_i - X_2\|) \\ \dots \\ \varphi_p(\|x_i - X_p\|) \end{bmatrix} \cdot [w_1 \ w_2 \ \dots \ w_p]^T \quad (12)$$

$$\text{令 } \varphi = \begin{bmatrix} \varphi_1(\|x_i - X_1\|) \\ \varphi_2(\|x_i - X_2\|) \\ \dots \\ \varphi_p(\|x_i - X_p\|) \end{bmatrix}, \text{ 若 } \varphi^+ \text{ 表示 } \varphi \text{ 的}$$

伪逆矩阵, 权值可计算得:

$$[w_1 \ w_2 \ \dots \ w_p]^T = \varphi^+ [d_1 \ d_2 \ \dots \ d_N]^T \quad (13)$$

3 实验验证与分析

3.1 实验配置

实验以贵州土城煤矿为实验对象, 从土城煤矿

监控系统瓦斯数据库获取瓦斯传感器 5 种状态下的数据各 80 组。对各种状态的瓦斯数据进行小波包分解处理重构,利用小波包的相关算法提取信号的特征能量谱,组合成向量作为神经网络的输入样本,瓦斯传感器各种状态下的特征能量谱样本如图 1 所示。

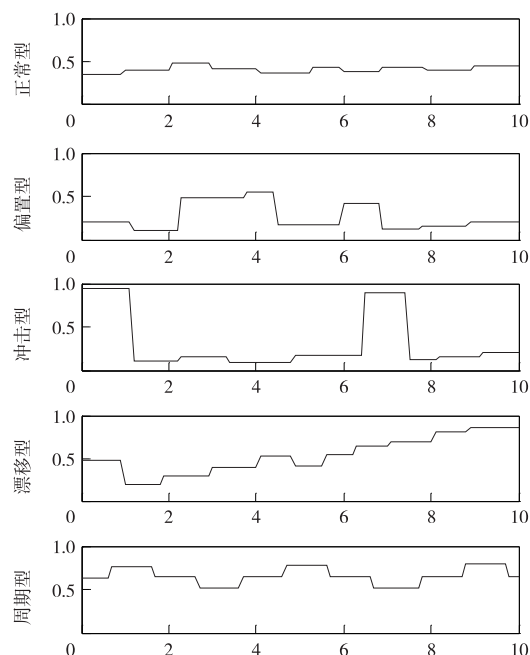


图 1 瓦斯传感器设备 5 种状态下的特征能量谱

Fig.1 Characteristic energy spectrum of the gas sensor device in five states

对自适应聚点 K-means 算法进行 RBF 神经网络训练。RBF 神经网络输出层节点为 3,是由于瓦斯传感器设备运行时运行状态细化为 5 种,映射关系具体对应关系如表 1。因此可以确定网络输出节点为 3。输入层节点为 8,是由于小波分解的特征能量谱向量来确定。隐含层的节点是自适应聚点 K-means 算法特点可确定为 5。RBF 神经网络采用

8-5-3 结构。通过提取特征向量组成样本,进行 RBF 神经网络训练。

表 1 瓦斯传感器运行状态与神经网络输出关系

Table 1 Relationship between the operating status of the gas sensor and the output of the neural network

瓦斯传感器运行状态	神经网络输出
I (正常型)	0 0 0
II (偏置型)	0 0 1
III (冲击型)	0 1 0
IV (漂移型)	0 1 1
V (周期型)	1 0 0

3.2 实验验证

从土城煤矿监控系统瓦斯数据库中取出软故障状态的数据各 20 组,分别对取出的数据进行测试。对神经网络算法训练的结果,进行实验验证如图 2 所示,可以从图中看出,100 组样本通过自适应聚点 K-means 算法识别的瓦斯传感器故障分为 5 类。表 2 中看出,故障状态传感器识别率高达 95%。因此,该自适应聚点 K-means 算法可以有效地识别瓦斯传感器软故障的类型,同时识别率能够达到预期的设计效果。

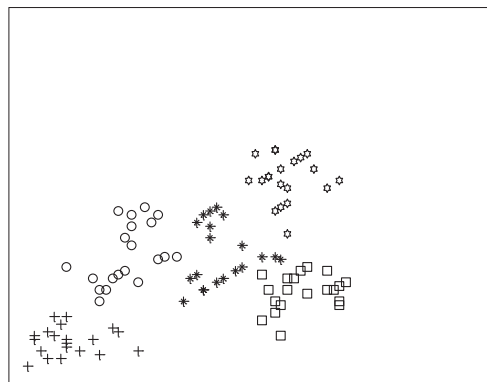


图 2 样本识别分类结果

Fig.2 Sample recognition classification results

表 2 实验输出结果

Table 2 Experimental output results

测试数据								网络输出			诊断结果	识别率/%
1	2	3	4	5	6	7	8	1	2	3		
0.261	0.342	0.392	0.423	0.361	0.342	0.413	0.372	0.001	0.002	0.001	I	95.2
0.113	0.091	0.493	0.612	0.172	0.253	0.094	0.165	0.004	0.008	0.994	II	96.3
0.952	0.123	0.085	0.054	0.145	0.096	0.065	0.153	0.003	0.997	0.004	III	95.7
0.425	0.166	0.281	0.322	0.562	0.331	0.522	0.442	0.002	0.998	0.996	IV	95.4
0.313	0.264	0.232	0.351	0.244	0.232	0.341	0.281	0.997	0.003	0.004	V	95.8

4 结论

1) 利用小波包对瓦斯传感器软故障信号进行

分解,分解算法降低了软故障信号的特征能量谱的向量维数,更好地体现出软故障时的特征信号,小波包分解是一种很实用的故障信号波形处理方法。

2) 加入轮廓系数的 K-means 聚类算法, 自适应地优化了聚类中心点, 使得聚类效果更加明显, 瓦斯传感器软故障识别率高达 95%。该算法运用到监控系统中不仅能够提高瓦斯传感器设备运行的可靠性, 而且也提高了煤矿安全监控系统的准确性。

参考文献

- [1] 董丁稳, 屈世甲, 王红刚. 矿井瓦斯监测数据特征分析及预处理[J]. 工矿自动化, 2015, 41 (6): 1-5.
DONG D W, QU S J, WANG H G. Characteristic analysis and preprocessing of mine gas monitoring data [J]. Industry and Mine Automation, 2015, 41 (6): 1-5.
- [2] 梁荣, 董丁稳. 矿井瓦斯监测数据消噪方法[J]. 工矿自动化, 2018, 44 (2): 18-22.
LIANG R, DONG D W. De-noising method of mine gas monitoring data [J]. Industry and Mine Automation, 2018, 44 (2): 18-22.
- [3] 张申, 刘鹏, 张彭. 感知矿山物联网云计算应用探索[J]. 煤炭科学技术, 2012, 40 (9): 72-75.
ZHANG S, LIU P, ZHANG P. Applied discovery on cloud computation of sensory mine Internet of things [J]. Coal Science and Technology, 2012, 40 (9): 72-75.
- [4] 杜运成. 基于等价空间方法的采样系统故障检测[D]. 上海: 东华大学, 2008: 68-70.
DU Y C. Fault detection of sample system based on parity space [D]. Shanghai: Donghua University, 2008: 68-70.
- [5] 王杰. 模糊综合评判在故障树分析法中的应用[J]. 电子设计工程, 2012, 20 (6): 41-44.
WANG J. Application of fuzzy comprehensive evaluation in the fault tree analysis [J]. Electronic Design Engineering, 2012; 20 (6): 41-44.
- [6] 高伟, 邢琰, 王南华. 基于定性模型的故障诊断方法[J]. 空间控制技术与应用, 2009, 35 (1): 25-29.
GAO W, XING Y, WANG N H. Fault diagnosis approach based on qualitative model [J]. Aerospace Control and Application, 2009, 35 (1): 25-29.
- [7] 孙继平. 安全高效矿井监控关键技术研究[J]. 工矿自动化, 2012, 38 (12): 1-5.
SUN J P. Research of key technologies for mine safety and efficiency monitoring [J]. Industry and Mine Automation, 2012, 38 (12): 1-5.
- [8] 郭瑞, 吴琼. 一种改进的小波算法在矿井漏电保护系统选线中的应用[J]. 工矿自动化, 2010, 36 (1): 29-32.
GUO R, WU Q. Application of an improved wavelet algorithm in fault line selection of leakage protection system of mine [J]. Industry and Mine Automation, 2010, 36 (1): 29-32.
- [9] 李杰, 刘贺平. 一种基于临界带宽的新小波包变换算法[J]. 计算机工程与应用, 2012, 48 (14): 5-7.
LI J, LIU H P. New wavelet packet transform algorithm based on critical bandwidth [J]. Computer Engineering and Applications, 2012, 48 (14): 5-7.
- [10] 刘鹏. 西安市 XX 线路公交行驶工况构建及混合动力公交动力系统参数匹配研究[D]. 西安: 长安大学, 2017: 32-33.
LIU P. Study on Xi'an xx road city bus driving cycle and parameters matching of hybrid electric city bus Transit Power systems [D]. Xi'an: Chang'an University, 2017: 32-33.
- [11] 傅德胜, 周辰. 基于密度的改进 K 均值算法及实现 [J]. 计算机应用, 2011, 31 (2): 432-434.
FU D S, ZHOU C. Improved K-means algorithm and its implementation based on density [J]. Journal of Computer Applications, 2011, 31 (2): 432-434.
- [12] 罗序荣, 任彬. 基于 RBF 神经网络的膝关节步态预测方法研究[J]. 工业控制计算机, 2019, 32 (8): 90-93.
LUO X R, REN B. Research on knee joint gait prediction method based on RBF neural network [J]. Industrial Control Computer, 2019, 32 (8): 90-93.
- [13] HE W M, WANG P L. Fault diagnosis of traction machine for lifts based on wavelet packet Algorithm and RBF neural Network [J]. Computational and Information Sciences, 2010, 10 (3): 372-375.
- [14] 郭伟. 基于互信息的 RBF 神经网络结构优化设计 [J]. 计算机科学, 2013, 40 (6): 252-255.
GUO W. Structural optimization algorithm for RBF neural network based on mutual information [J]. Computer Science, 2013, 40 (6): 252-255.
- [15] 刘思敏, 徐景田, 鞠博晓. 基于 EMD 和 RBF 神经网络的大坝形变预测[J]. 测绘通报, 2019 (8): 88-91.
LIU S M, XU J T, JU B X. Dam deformation prediction based on EMD and RBF neural networks [J]. Bulletin of Surveying and Mapping, 2019 (8): 88-91.