

文章编号:1671-251X(2020)07-0070-06

DOI:10.13272/j.issn.1671-251x.17605

煤矿瓦斯传感器人工调校噪声数据处理方法

胡峰, 叶福豪, 王国胤, 代劲

(重庆邮电大学 计算智能重庆市重点实验室, 重庆 400065)



扫码移动阅读

摘要:传统噪声数据处理方法对输入数据有一定的要求,且运行时间较长。而煤矿瓦斯传感器人工调校噪声数据存在数量少、质量差、时间不一致、易受环境影响等问题,采用传统噪声处理方法难以滤除该噪声数据。针对上述问题,提出了一种煤矿瓦斯传感器人工调校噪声数据处理方法。采用数据平均值填充煤矿瓦斯传感器浓度数据缺失值;采用多时间粒度构建煤矿瓦斯传感器浓度数据的特征集和样本集;采用高斯函数、混合高斯函数、二项式函数、三项式函数、分段二项式函数5种曲线拟合函数拟合人工调校噪声数据,并基于最小二乘法确定拟合函数参数,根据拟合效果得到最优的拟合函数;通过分析人工调校噪声数据得出该噪声数据与瓦斯浓度上升的斜率、峰值、调校前后浓度差等基本特征有关,根据这些基本特征识别出人工调校噪声数据并删除。实验结果验证了该方法的有效性。

关键词:瓦斯传感器;人工调校噪声数据;噪声数据识别;噪声滤除;多时间粒度;数据拟合
中图分类号:TD712 **文献标志码:**A

Manual adjustment noise data processing method for coal mine gas sensor

HU Feng, YE Fuhao, WANG Guoyin, DAI Jin

(Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts & Telecommunications, Chongqing 400065, China)

Abstract: Traditional noise data processing methods have certain requirements for input data, and have a long running time. However, there are some problems of manual adjustment noise data of coal mine gas sensor such as less quantity, poor quality, inconsistent time and easy to be affected by environment, it is difficult to filter the noise data by traditional noise processing methods. For the above problems, a manual adjustment noise data processing method for coal mine gas sensor is proposed. The data average value is used to fill the missing value of concentration data of coal mine gas sensor; the feature set and sample set of concentration data of coal mine gas sensor are constructed by using multi time granularity; five curve fitting functions, namely Gaussian function, mixed Gaussian function, binomial function, trinomial function and piecewise binomial function, are used to fit manual adjustment noise data, and parameters of the fitting function are determined based on the least square method, and the optimal fitting function is obtained according to fitting effect; through analysis of manual adjustment noise data, it is concluded that the noise data is related to the slope, peak and difference of gas concentration before and after adjustment, according to these basic characteristics, the manual adjustment noise data is identified and deleted. The experimental results verify effectiveness of the method.

Key words: gas sensor; manual adjustment noise data; noise data recognition; noise filtering; multi-

收稿日期:2020-05-11;**修回日期:**2020-06-11;**责任编辑:**李明,郑海霞。
基金项目:国家重点研发计划资助项目(2018YFC0808305)。
作者简介:胡峰(1978—),男,湖北天门人,教授,博士,主要从事数据挖掘、机器学习等方面的研究工作,E-mail:hufeng@cqupt.edu.cn。通信作者:叶福豪(1995—),男,河南洛阳人,硕士研究生,主要研究方向为数据挖掘、自然语言处理等,E-mail:512653694@qq.com。
引用格式:胡峰,叶福豪,王国胤,等.煤矿瓦斯传感器人工调校噪声数据处理方法[J].工矿自动化,2020,46(7):70-75.
HU Feng, YE Fuhao, WANG Guoyin, et al. Manual adjustment noise data processing method for coal mine gas sensor[J]. Industry and Mine Automation, 2020, 46(7): 70-75.

time granularity; data fitting

0 引言

为预防煤矿瓦斯灾害的发生,需采取有效的方式对矿井内瓦斯浓度进行监测^[1-3]。采用传感器监测瓦斯浓度时,需定期向传感器通入气体,观察其反应是否正常,这个过程称为人工调校。该过程中产生的数据称为人工调校噪声数据。在处理瓦斯浓度数据时需要滤除噪声数据,否则会对瓦斯浓度预测模型的精确度产生很大影响^[4]。

近年来,许多学者在噪声处理领域进行了深入的研究。秦鹏等^[5]提出了一种基于阈值分解的多级中值滤波方法以去除噪声。常威威等^[6]提出了一种基于小波变换和主成分分析的去噪方法。何清等^[7]提出了一种基于摄动的模糊聚类方法,用于滤除量测数据对中的噪声。杨朝辉等^[8]提出了一种基于支持向量机的噪声去除方法。Zhang Weiping^[9]提出了一种基于中值滤波和时域递归降噪的组合方法,用于去除噪声数据。Li Zhen 等^[10]提出了一种基于变分模式分解的时频峰值滤波方法来滤除随机噪声。J. A. Saez 等^[11]在多个数据集上证明了噪声过滤的重要性。S. Park 等^[12]提出了一种基于学习的自编码器去噪方法,该方法会校正输入传感器数据的不准确值。上述方法对输入数据有一定的要求,且运行时间较长。而煤矿瓦斯传感器人工调校噪声数据数量少、质量差、时间不一致、易受环境影响,难以用上述方法对其进行滤除。因此,本文结合多时间粒度、曲线拟合、最小二乘法等,提出了一种煤矿瓦斯传感器人工调校噪声数据处理方法,对人工调校过程中产生的噪声数据进行识别并删除。

1 人工调校噪声数据识别框架

1.1 人工调校噪声数据定义

为了保证传感器处于正常工作状态,检修人员会定期向其输送瓦斯气体,使传感器达到一定体积分数标准(1.5% CH₄~1.8% CH₄),停止向其输送气体后浓度恢复正常,恢复时间与传感器物理特性有关。噪声数据条件:1.5%<C<1.8%,C为调校时的最大瓦斯体积分数;H>60 s,H为调校持续时间。

以 2019-03-02 的瓦斯体积分数为例,传感器调校数据如图 1 所示。图 1 中第 4 个时间点—第 8 个时间点为人工调校过程,瓦斯体积分数先上升再下降,且人工调校前后时间点的瓦斯体积分数变化范围非常小。

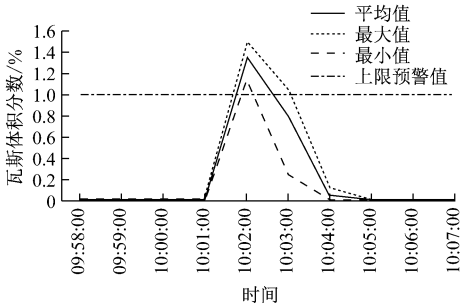


图 1 传感器调校数据

Fig. 1 Sensor adjusted data

1.2 人工调校噪声数据识别框架

传感器人工调校噪声数据识别框架如图 2 所示。

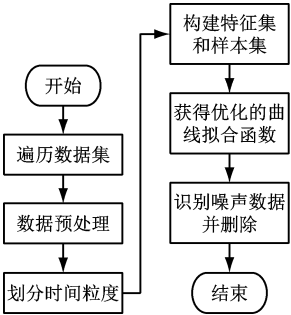


图 2 传感器人工调校噪声数据识别框架

Fig. 2 Sensor manual adjustment noise data recognition framework

1.2.1 数据预处理

煤矿传感器浓度数据集中最大的问题就是数据中含有缺失值及大量的噪声数据,在处理人工调校噪声数据前需对缺失值进行填充。取缺失值前 3 个时刻的数据平均值来填充,将小于 10 条记录的数据直接删除。

1.2.2 时间粒度划分

进行曲线拟合的一个重要前提是确定滑动窗口内时间粒度的大小。如果滑动窗口内时间粒度太小,会导致样本集数量过大,算法运行时间过长;如果滑动窗口内时间粒度太大,会导致实验精确度下降。因此,采用多时间粒度进行实验,设时间粒度分别为 1,5,10 min。

$$W = \{T_1, T_2, T_3\}$$
 (1)

式中:W 为时间窗口集合;T₁,T₂,T₃ 为时间粒度。

1.2.3 特征集和样本集构建

传感器监测数据量十分庞大且不同传感器监测时间间隔不同,因此需要对不同传感器的监测数据进行时间窗口统一处理。根据时间窗口生成一个样本,取该时间窗口内浓度最大值中的最大值、浓度最小值中的最小值、浓度平均值的均值分别作为新样本浓度最大值、最小值、平均值,使用处理后的浓度最大值作为实验的输入浓度值。时间窗口统一处理

后的样本特征为当前时刻、浓度上限预警值、传感器编号、浓度最大值、浓度最小值、浓度平均值。

1.2.4 优化的曲线拟合函数

通过对传感器人工调校噪声数据曲线分布进行观察分析,确定了一些符合其曲线分布的函数:高斯函数、混合高斯函数、二项式函数、三项式函数、分段二项式函数。

获得优化的曲线拟合函数步骤如下:首先,将原始传感器监测数据中的缺失值进行填充处理;其次,初始化时间窗口,构造特征集和样本集;然后,将人工调校噪声数据样本集中的当前时刻和浓度最大值输入到拟合函数中,使用最小二乘法求得拟合函数参数;最后,根据曲线拟合效果得到优化的拟合函数。

(1) 高斯函数:

$$f_1(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) \tag{2}$$

式中: $f_1(t)$ 为 t 时刻的高斯函数值; σ 为标准差; μ 为期望值。

(2) 混合高斯函数:

$$f_2(t) = \frac{k_1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(t-\mu_1)^2}{2\sigma_1^2}\right) + \frac{k_2}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{(t-\mu_2)^2}{2\sigma_2^2}\right) \tag{3}$$

式中: $f_2(t)$ 为 t 时刻的混合高斯函数值; k_1 和 k_2 为不同的高斯函数系数; σ_1, μ_1 为第1种高斯函数的标准差和期望值; σ_2, μ_2 为第2种高斯函数的标准差和期望值。

(3) 二项式函数:

$$f_3(t) = at^2 + bt + c \tag{4}$$

式中: $f_3(t)$ 为 t 时刻的二项式函数值; a 为二次项系数; b 为一次项系数; c 为截距。

(4) 三项式函数:

$$f_4(t) = dt^3 + at^2 + bt + c \tag{5}$$

式中: $f_4(t)$ 为 t 时刻的三项式函数值; d 为三次项系数。

(5) 分段二项式函数:

$$f_5(t) = \begin{cases} a_1t^2 + b_1t + c_1, & t < t_0 \\ a_2t^2 + b_2t + c_2, & t \geq t_0 \end{cases} \tag{6}$$

式中: $f_5(t)$ 为 t 时刻的分段二项式函数值; t_0 为一段调校数据中峰值所对应的时刻; a_1, b_1, c_1 为第1段函数的二次项系数、一次项系数和截距; a_2, b_2, c_2 为第2段函数的二次项系数、一次项系数和截距。

本文使用最小二乘法获得拟合函数的参数^[13-14],最小二乘法通过最小化真实值与预测值的误差平方和求得函数参数。设曲线方程为

$$f(t_i) = bt_i + c \tag{7}$$

式中: $f(t_i)$ 为拟合函数在 t_i 时刻的浓度预测值; t_i 为第 i 个样本对应的时刻。

$f(t_i)$ 的误差平方和为

$$J = \sum_{i=1}^m (f(t_i) - y_i)^2 \tag{8}$$

式中: m 为每个传感器处理后的样本总数; y_i 为 t_i 时刻的真实浓度。

对误差平方和求偏导并化简,得参数公式为

$$b = \frac{\sum_{i=1}^m (t_i - \bar{t})(y_i - \bar{y})}{\sum_{i=1}^m (t_i - \bar{t})^2} \tag{9}$$
$$c = \bar{y} - b\bar{t} \tag{10}$$

式中: $\bar{t}$ 为时刻均值; $\bar{y}$ 为真实浓度均值。

1.2.5 人工调校噪声数据处理

通过对人工调校噪声数据的分析,可知该噪声数据与瓦斯浓度上升的斜率、峰值、调校前后浓度差这些基本特征有关。

人工调校噪声数据处理过程如下:获得处理好的煤矿传感器样本数据集;得到浓度曲线的斜率 S 大于斜率阈值 V_a 的时刻;记录当前时刻为异常起始点,判断浓度上升峰值 $y_{\max}$ 是否大于传感器的预警阈值 V_b ;根据异常开始时刻 t_s 与瓦斯浓度差阈值 V_c 得到异常结束时刻 t_e ;根据拟合函数计算异常时间段的对应浓度值,判断其与真实浓度值 y_i 的均方误差 E 是否小于允许的误差阈值 V_d ;将符合要求的数据标记为人工调校噪声数据并删除。

2 方法描述

传感器人工调校噪声数据处理方法步骤如下。

(1) 构建数据。对 o 段人工调校数据集 $P = \{P_1, P_2, \dots, P_o\}$ 、 n 个传感器数据集 $D = \{D_1, D_2, \dots, D_n\}$ 进行缺失值填充,其中 P 中每个元素代表一段人工调校数据; D 中每个元素代表一个传感器监测数据。根据时间窗口构造样本特征集。处理后的人工调校样本集 Y 为 r 段人工调校噪声数据,即 $Y = \{Y_1, Y_2, \dots, Y_r\}$, Y 中每个元素代表经处理后的一段人工调校噪声数据。处理后的每个传感器浓度数据集 $U = \{(t_1, y_1), (t_2, y_1), \dots, (t_m, y_m)\}$, t_m 为第 m 个样本对应的时刻。

(2) 得到曲线拟合函数。将人工调校样本集 Y 分别输入到拟合函数(式(2)~式(6))中进行拟合,并使用最小二乘法得到拟合函数的参数,根据曲线拟合效果得到优化的拟合函数。

(3) 获取相关阈值。人工调校噪声浓度上升的

斜率阈值 V_a 、传感器的预警阈值 V_b 、调校开始与结束时的瓦斯浓度差阈值 V_c 分别由人工调校样本集对应的斜率、预警值、浓度差的平均值得到。误差阈值 V_d 的取值范围为 $0.01 \sim 0.50$, 统计样本真实值与预测值的均方误差 E 小于 V_d 的样本数, 取样本数最大时对应的最小 V_d 为最合适的误差阈值。

(4) 处理人工调校噪声数据。遍历数据集, 根据 $S = \frac{y_{i+1} - y_i}{t_{i+1} - t_i}$ 求得斜率。若 $S > V_a$ 成立, 则令 $t_s = t_i$; 对后面的数据进行比较, 直到达到首个浓度峰值 $y_{\max}$, 若 $y_{\max} \geq V_b$, 则令 $t_e = t_{i+1}$; 将 $|y_{t_e} - y_{t_s}| < V_c$ 成立的时刻 t_e 记为结束时刻, 否则将 t_e 置为下一个时刻, 继续判断该式; 计算 t_s 到 t_e 时间段的拟合函数值与真实浓度值的均方误差 E , 若 $E < V_d$ 成立, 则将这段时间的数据标记为人工调校噪声数据; 最后根据标记删除人工调校噪声数据。

3 实验分析

3.1 函数拟合效果

利用式(2)一式(6)拟合函数对数据进行拟合, 拟合效果如图3—图7所示。

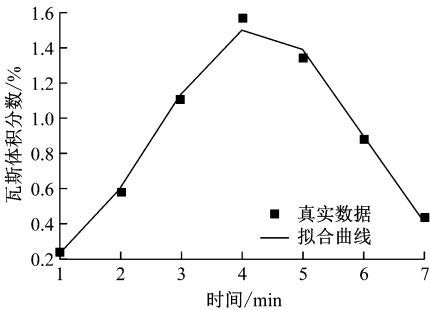


图3 高斯函数拟合效果

Fig.3 Gaussian function fitting effect

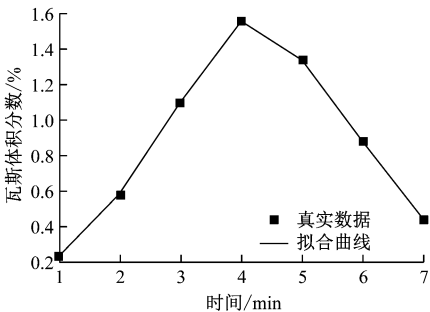


图4 混合高斯函数拟合效果

Fig.4 Mixed gaussian function fitting effect

为了反映拟合函数的效果, 本文使用均方误差^[15]指标对拟合结果进行对比。通过拟合函数预测值与真实值之差平方和的均值得到拟合函数的均方误差。拟合函数对应的均方误差见表1。可以看出, 分段二项式函数和混合高斯函数拟合较为准确, 混合高斯函数拟合效果最好。

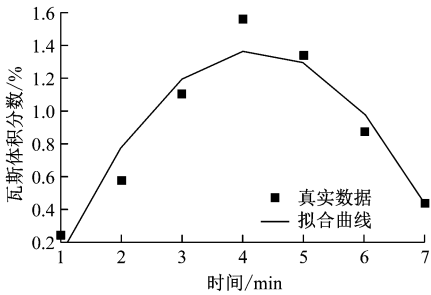


图5 二项式函数拟合效果

Fig.5 Binomial function fitting effect

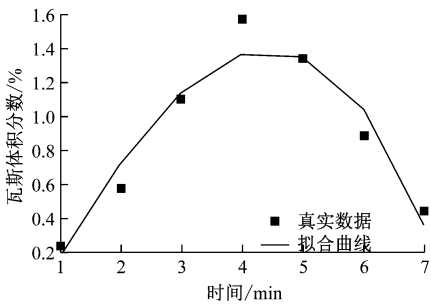


图6 三项式函数拟合效果

Fig.6 Trinomial function fitting effect

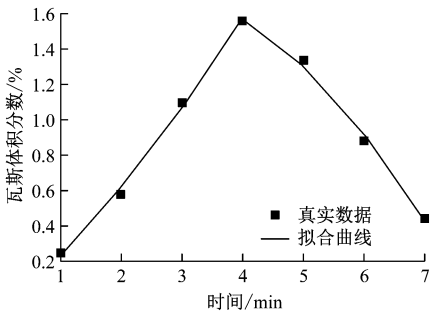


图7 分段二项式函数拟合效果

Fig.7 Piecewise binomial function fitting effect

表1 拟合函数的均方误差

Table 1 Mean squared error of the fitting function

拟合函数	均方误差
$f_1(t)$	0.001 299
$f_2(t)$	0.000 018
$f_3(t)$	0.016 087
$f_4(t)$	0.013 334
$f_5(t)$	0.000 870

3.2 拟合评价指标

混淆矩阵是精确度评价的一种标准格式, 用于比较分类结果和实际结果。混淆矩阵见表2, 其中, TP为真正例, FP为假正例, TN为真反例, FN为假反例。本实验中, 使用混淆矩阵的查准率作为评价指标, 查准率表示在所有被预测为正例的样本中实际为正例的样本概率, 即 $TP / (TP + FP)$ 。

表 2 混淆矩阵
Table 2 Confusion matrix

真实情况	预测结果	
	正例	反例
正例	TP	FN
反例	FP	TN

3.3 实验验证

本实验数据来源于重庆市某煤矿传感器监测到的瓦斯浓度数据,使用混合高斯函数进行实验。取 100 个传感器数据,共 344 个调校数据样本,其中训练集数据为 104 个,验证集数据为 240 个。

通过最小二乘法得到混合高斯函数的参数值 $k_1=9.505, \sigma_1=1.854, \mu_1=-0.424, k_2=-2.625, \sigma_2=0.440, \mu_2=-0.286$ 。根据参数阈值选取原则与实验性能指标,得到各参数值 $V_a=0.75, V_b=1, V_c=0.17, V_d=0.43$ 。

为验证本文提出方法的真实性能情况,采集 5 个瓦斯传感器从 2019-01-01—04-01 监测到的瓦斯浓度数据进行实验,得到人工调校噪声数据的查准率为 0.904 7。以 2019 年 1 月份的数据为例,传感器瓦斯体积分数部分数据如图 8 所示,传感器部分标记结果见表 3。

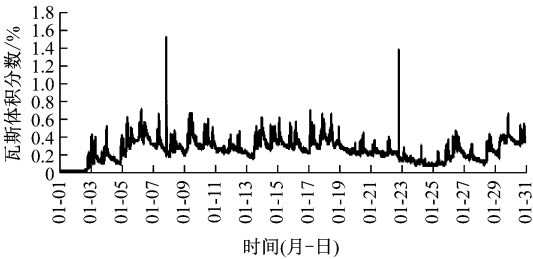


图 8 传感器瓦斯体积分数部分数据

Fig.8 Sensor gas concentration partial data

实验表明,在提出的拟合函数中,混合高斯函数拟合效果最好。人工调校噪声数据处理后得到的查准率为 0.904 7,本文方法可以有效识别煤矿瓦斯传感器人工调校噪声数据。

需要说明的是,由于人工调校噪声数据受传感器灵敏度、调校持续时间、传感器分布位置等多种未知因素影响,所以,在实验中难免会存在一些数据被误判或漏判,从而导致查准率下降。

4 结语

提出了一种煤矿瓦斯传感器人工调校噪声数据处理方法:首先,采用数据平均值填充煤矿瓦斯传感器浓度数据缺失值;其次,使用多时间粒度构建煤矿瓦斯传感器浓度数据的特征集和样本集;再次,运用多种曲线拟合函数拟合人工调校噪声数据,基于最

表 3 传感器部分标记数据
Table 3 Sensor partial labeled data

时间	瓦斯体积分数平均值	瓦斯体积分数最大值	瓦斯体积分数最小值	瓦斯体积分数上限预警值	标记
09:56:00	0.02	0.02	0.02	1	非调校数据
09:57:00	0.01	0.02	0.01	1	非调校数据
09:58:00	0.02	0.02	0.02	1	非调校数据
09:59:00	0.01	0.02	0.01	1	非调校数据
10:00:00	0.01	0.01	0.01	1	非调校数据
10:01:00	0.01	0.01	0.01	1	调校数据
10:02:00	1.2	1.5	0.98	1	调校数据
10:03:00	0.82	0.93	0.83	1	调校数据
10:04:00	0.14	0.21	0.01	1	调校数据
10:05:00	0.08	0.12	0.01	1	调校数据
10:06:00	0.01	0.01	0	1	调校数据
10:07:00	0	0	0	1	非调校数据
10:08:00	0	0	0	1	非调校数据
10:09:00	0	0	0	1	非调校数据
10:10:00	0	0	0	1	非调校数据
10:11:00	0	0	0	1	非调校数据

小二乘法确定拟合函数参数,根据拟合效果得到最优的拟合函数;最后,利用人工调校噪声数据的基本特征识别并删除人工调校噪声数据。煤矿瓦斯监测数据的实验结果表明,混合高斯函数对人工调校噪声数据的拟合效果较好,能有效识别煤矿瓦斯传感器人工调校噪声数据。优化的时间粒度与多模型融合的噪声滤除算法将是进一步的研究工作。

参考文献(References):

[1] 徐晓建. 煤矿安全风险防控及预警系统设计[J]. 工矿自动化, 2020, 46(3): 105-108.
XU Xiaojian. Design of coal mine safety risk prevention and control and early warning system [J]. Industry and Mine Automation, 2020, 46 (3): 105-108.

[2] 张瑜, 冯仕民, 杨赛烽, 等. 矿工不安全行为影响因素本体构建与推理研究[J]. 煤矿安全, 2019, 50(5): 300-304.
ZHANG Yu, FENG Shimin, YANG Saifeng, et al. Ontology construction and reasoning research on influencing factors of miners' unsafe behavior [J]. Safety in Coal Mines, 2019, 50(5): 300-304.

[3] 赵芳, 程国强, 孙业强, 等. 巷道高浓度瓦斯运移扩散规律的研究[J]. 煤炭技术, 2016, 35(1): 193-196.
ZHAO Fang, CHENG Guoqiang, SUN Yeqiang, et al. Study on laws of high concentration gas

- movement and diffusion in roadway [J]. Coal Technology, 2016,35(1):193-196.
- [4] 尚宇炜,郭剑波,吴文传,等.数据-知识融合的机器学习(2):泛化风险[J].中国电机工程学报,2019,39(16):4641-4649.
- SHANG Yuwei, GUO Jianbo, WU Wenchuan, et al. Machine learning methods embedded with domain knowledge(Part II): generalization risk [J]. Proceedings of the CSEE, 2019,39(16):4641-4649.
- [5] 秦鹏,丁润涛.基于阈值分解的多级中值滤波[J].微型机与应用,2003,22(12):50-52.
- QIN Peng, DING Runtao. Multistage median filtering based on threshold decomposition[J]. Microcomputer & Its Applications, 2003,22(12):50-52
- [6] 常威威,郭雷,刘坤,等.基于主分量分析的高光谱遥感数据噪声消除方法[J].计算机测量与控制,2009,17(6):1070-1072.
- CHANG Weiwei, GUO Lei, LIU Kun, et al. Denoising of hyperspectral data based on wavelet transform and principal component analysis [J]. Computer Measurement & Control, 2009, 17(6): 1070-1072.
- [7] 何清,徐树富,王加银,等.FCMBP聚类法在语音识别和模糊控制中的应用[J].系统工程学报,2001,16(6):430-437.
- HE Qing, XU Shufu, WANG Jiayin, et al. FCMBP clustering methods and its applications in speech recognition and fuzzy control[J]. Journal of Systems Engineering, 2001,16(6):430-437.
- [8] 杨朝辉,陈映鹰.基于支持向量机的椒盐噪声去除方法[J].计算机工程与应用,2009,45(22):150-152.
- YANG Zhaohui, CHEN Yingying. SVM-based approach for removing salt-pepper noise from images [J]. Computer Engineering and Applications, 2009, 45(22):150-152.
- [9] ZHANG Weipeng. Refuge chamber image processing algorithm based on combination of median filter and time domain recursive noise reduction [J]. Energy Education Science and Technology, 2012, 30(2): 241-248.
- [10] LI Zhen, GAO Jinghui, LIU Naihao, et al. Random noise suppression of seismic data by time-frequency peak filtering with variational mode decomposition [J]. Exploration Geophysics, 2019,50(6):634-644
- [11] SAEZ J A, KRAWCZYK B, WOZNIAK M. On the influence of class noise in medical data classification: treatment using noise filtering methods[J]. Applied Artificial Intelligence, 2016,30(4/5/6):590-609.
- [12] PARK S, GIL M S, IM H, et al. Measurement noise recommendation for efficient Kalman filtering over a large amount of sensor data[J]. Sensors, 2019,19(5):1168.
- [13] 汪奇生,杨德宏,杨根新.总体最小二乘求解线性回归模型的迭代算法[J].工程勘察,2014,42(9):85-88.
- WANG Qisheng, YANG Dehong, YANG Genxin. An iteration algorithm of total least squares for linear regression models[J]. Geotechnical Investigation & Surveying, 2014,42(9):85-88.
- [14] 陈永明,冯盛森,戴颖超.基于最小二乘法的拟合曲线CCD相机畸变校正[J].机电技术,2018,42(3):31-35.
- CHEN Yongming, FENG Shengmiao, DAI Yingchao. Distortion correction of fitting curve CCD camera based on least square method [J]. Mechanical & Electrical Technology, 2018,42(3):31-35.
- [15] 孙文胜,张冬,张公礼.均方误差短时反馈算法的残余频偏纠正及阈值选取的研究[J].电子器件,2009,32(4):792-796.
- SUN Wensheng, ZHANG Dong, ZHANG Gongli. Research of residual frequency offset correction based on the mean square error short-term feedback algorithm and decision of the threshold[J]. Chinese Journal of Electron Devices, 2009,32(4):792-796.