

应用于管道气体泄漏监测的 WSN 层级式数据融合方法<sup>①</sup>余 阳<sup>②</sup> 吴银锋 冯仁剑 万江文<sup>③</sup>

(北京航空航天大学仪器科学与光电工程学院 北京 100191)

**摘要** 为提高利用无线传感器网络(WSN)监测天然气管网泄漏的准确性和可靠性,提出一种小波支持向量机(SVM)和证据理论相结合的层级式数据融合算法。该算法利用小波变换方法对原始信号进行数据级消噪处理,并提取对泄漏敏感的特征参数;建立 SVM 多分类器模型,以特征参数作为输入向量在普通节点处进行泄漏检测特征级融合;采用改进的证据组合规则,在 Sink 节点处进行决策级证据组合,得到管网状态的最终决策。实验结果表明,该方法可有效地提高泄漏源位置检测的正确率,降低检测过程中的漏检率和误警率。

**关键词** 泄漏检测,无线传感器网络(WSN),支持向量机(SVM),证据理论

## 0 引言

天然气供给体系是城市的“生命线”之一,其主要输配方式为管道输送。随着管网越来越复杂,管线越来越长,重大的安全隐患问题也越来越突出。目前管网的安全检测方法,主要是人工使用泄漏检测仪器进行定期巡检,该方法效率低、实时性差,监测范围有限。对于覆盖面大的管网需要较多的人力资源,并且无法在线实施对整个网络进行全面监测。随着无线传感器网络(wireless sensor network, WSN)技术的不断发展,国外一些研究机构开始研究利用 WSN 的管网安全监测新方法。已有的研究包括:帝国理工大学和美国能源技术国家实验室利用 WSN 的地下管道监测系统 PipeNet 的联合开发研究<sup>[1]</sup>;Jawhar 等人开展的管道泄漏监测网络服务质量的研究,该研究针对管道网络特点,提出了一种线性无线网络结构的组网通信协议,保证了监测网络通信的健壮性、安全性<sup>[2-4]</sup>;Eydgahi 等人的用于管道泄漏检测的柔性 WSN 节点的研制<sup>[5]</sup>。上述研究主要集中在网络系统的结构设计和网络节点设计等方面,目前关于监测网络数据处理方面的研究还较少。

由于受到噪声和环境变化等因素的影响,网络

节点提供的检测信息具有不确定性和模糊性。对于同一泄漏事件,处于不同位置的节点可能做出不一致甚至是相互矛盾的诊断结果,致使系统难以做出正确决策。为此本文提出了一种用于管道气体泄漏检测的新的层级式数据融合方法,该方法能对网络中的多个节点的泄漏检测信息进行层级式的融合处理,以降低单节点泄漏识别过程中的不确定性,提高泄漏识别的准确率。

## 1 天然气管道监控 WSN 结构

天然气管道泄漏监控网络是由大量的普通传感器节点、Sink 节点和控制管理中心组成,如图 1 所示。为了提高网络的可扩展性和降低网络管理的复杂度,网络采用分簇的结构设计,分成为若干簇,每个簇包括一个 Sink 节点和若干个普通传感器节点。普通传感器节点安装在天然气管道上,负责信号的采集、数据的预处理以及将处理结果经过多跳发送至 Sink 节点。Sink 节点作为簇头节点,对普通传感器节点采集的数据进行决策级融合处理,判断管道是否发生泄漏,并以无线或有线的通信方式将检测结果发送至控制管理中心,同时接收控管中心的各种控制指令,实现防范和预警的功能。

① 863 计划(2009AA01Z201),国家自然科学基金(60974121)和北京市自然科学基金(8102025)资助项目。

② 男,1982 年生,博士生;研究方向:无线传感器网络;E-mail:sensory@buaa.edu.cn

③ 通讯作者,E-mail:sensory@buaa.edu.cn

(收稿日期:2010-09-02)

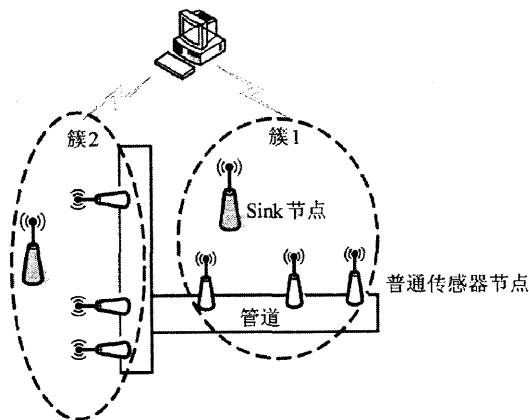


图1 天然气管道泄漏监控网络结构

## 2 层级式泄漏检测模型

在天然气管道泄漏监控网络中,由于受到测量噪声、传感器类型、节点数量和监测位置等因素的影响,网络中的检测信息表现出形式上的不确定性、多样性,数量上的巨大性,和关系上的复杂性。

首先,传感器采集的原始泄漏信号容易受到背景噪声和传播路径的干扰,导致检测信息具有很大的模糊性。为了提高检测精度,这里采用小波变换技术对原始泄漏信号进行去噪处理,并从去噪后的信号中提取对管道泄漏敏感的特征数据。其次,由于

每个节点上可能集成不同类型的传感器,因此需要对它们采集的信息进行综合处理,以提高网络中传感器节点的识别准确率。这里采用具有并行处理能力的支持向量机(support vector machine, SVM)多分类器方法来融合管道监测网络中的泄漏特征数据,提高网络的抗干扰能力,并同时准确地识别出管道的当前状态。最后,由于受到监测位置的影响,处于不同位置的节点可能得出不一致甚至相互矛盾的检测结果,致使系统难以做出正确的决策,并可能产生虚警或漏检问题。针对这一问题,将单节点的融合结果作为独立的证据,并在 Sink 节点处进行证据组合,从而提高证据集对正确命题的聚焦度。

基于上述分析,建立如图2所示的层级式数据融合泄漏检测模型:第一级融合为基于小波变换的数据级去噪,计算去噪后信号的时频域特征值,作为下一级融合的输入;第二级融合通过 SVM 多分类器方法获取各节点采集信号的判断结果;第三级融合利用改进的 Dempster-Shafer (D-S) 证据理论对多个节点的判断结果进行综合决策,实现对是否发生泄漏及泄漏量大小的判断。整个模型采用串行结构,前一级融合的结果作为下一级融合的输入,分别实现了数据级、特征级与决策级融合,保证输出结果的准确性和可靠性。

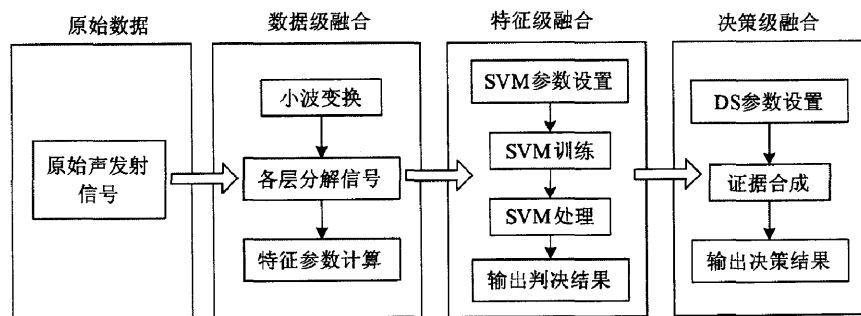


图2 融合检测模型

## 3 算法设计

### 3.1 基于小波变换的数据级融合

将小波变换方法应用于管道泄漏检测信号的预处理时,需要解决如下问题:小波基的选取、最大分解尺度的确定、小波阈值的选取和泄漏信号小波重构。最后,对降噪后的信号进行时域分析,提取出对泄漏敏感的特征参数。具体过程如下:

(1) 小波基的选取。这里采用 db 系列小波对

信号进行变换,该小波基对泄漏信号敏感,能够突出泄漏信号的奇异特征,减小信号的失真,保证信号的精确重构及良好的时频分析性能。

(2) 分解尺度的确定。对含噪泄漏信号进行离散小波变换,小波分解的尺度越大,越利于消除噪声,但尺度太大有时会丢失信号的某些重要局部奇异性。这里根据管道泄漏检测信号的频率特点,假设泄漏检测信号的采样频率是  $f_s$ , 最低识别频率为  $f_{\min}$ , 则最大分解尺度  $n$  应满足

$$\frac{f_s}{2^{n+1}} \geq f_{\min} \quad (1)$$

(3) 基于阈值决策的小波去噪。根据计算得到的最大分解尺度  $n$ , 对泄漏信号进行  $n$  层小波分解, 利用信号和噪声在各个尺度上的小波谱具有不同的表现特性, 选取一个适当的阈值, 令小于阈值的系数为 0, 大于阈值的点变为该点值与阈值的差值。本文采用启发式阈值法进行阈值的选择, 并对分解后的信号进行小波重构, 得到去噪后的泄漏信号。

(4) 对去噪后的信号进行时频域分析, 提取反映管道泄漏的特征参数, 包括峰值、平均幅值、方差和方根幅值等。

### 3.2 基于 SVM 多分类器的特征级融合

在特征级融合中采用 SVM 多分类器模型, 可检测出许多传统管道泄漏方法所不能检测到的小泄漏或缓慢泄漏, 并能克服传统方法的大样本要求。建立 SVM 多分类器模型包括输入向量的选择、分类原则以及内积函数和惩罚系数的确定等。网络中的传感器节点均采用相同的 SVM 多分类器结构, 如图 3 所示。其中,  $X_i (i = 1, 2, \dots, N)$  为输入的泄漏特征参数,  $Y_j (j = 1, 2, \dots, M)$  为 SVM 多分类器处理后的输出结果。

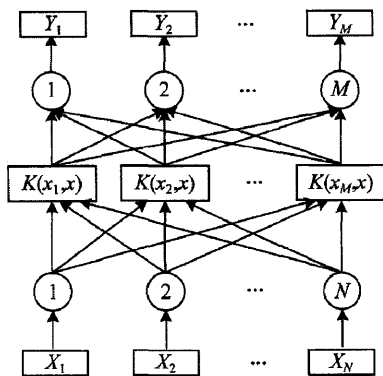


图 3 支持向量机分类器结构

#### 3.2.1 输入输出变量选取

研究表明, 信号的峰值、平均幅值、方差、均方根和方根幅值等参数可反映泄漏时引发应力波波动的大小或强弱; 峰值因子、脉冲因子和裕度因子是反映应力波信号剧变冲击的参数; 偏态因子和峭度因子能够反映泄漏引发应力波幅值分布情况的变化<sup>[6]</sup>。因此, 本文选取单模态声发射信号的峰值、平均幅值、方差、方根幅值、峰值因子、脉冲因子、裕度因子、偏态因子和峭度因子 9 个参数作为输入特征向量构建 SVM 多分类器。

由于单个 SVM 只能解决二类分类问题, 因此本文采用多个 SVM 联合构成的分类器来解决多管道状态的识别问题。根据泄漏检测的应用需求分析, 待识别的管道状态具有 3 种类型, 分别是正常、小泄漏和大泄漏, 故需要构造 3 个 SVM 与之对应, 其中第  $i$  个 SVM 对应第  $i$  个管道状态类型的判别输出为 +1, 而对其它状态类型的输出为 -1。即多分类器的输出节点数为 3, 当管道的状态分别为无泄漏、小泄漏和大泄漏时, 分类器的输出变量分别为 (+1, -1, -1)、(-1, +1, -1) 和 (-1, -1, +1)。

#### 3.2.2 SVM 分类器分类原则及参数优化

SVM 多分类器的分类原则是基于结构风险最小化原理和核函数方法<sup>[7]</sup>。其中, 判别函数为

$$f(x) = \text{sgn}\left(\sum_{i=1}^n \alpha_i y_i k(x_i, x) + b\right) \quad (2)$$

其中,  $x_i$  是训练样本, 个数为  $n$ ;  $x$  是实测样本。  $y_i = \{-1, 1\}$  为类别标号。  $\alpha$  为拉格朗日乘子系数向量,  $0 \leq \alpha_i \leq C$  且  $\sum_{i=1}^n \alpha_i y_i = 0, i = 1, 2, \dots, n$ 。  $b$  为分类阈值。  $k(x_i, x)$  为核函数。 本文采用的是径向基核函数, 其表达式为

$$k(x_i, x) = \exp[-\gamma \|x_i - x\|^2] \quad (3)$$

其中,  $\gamma$  为核函数参数。 实验表明, 该核函数的分类效果略优于其他核函数。

在参数选择方面, 惩罚因子  $C$  和核函数参数  $\gamma$  是必备的两个参数, 其取值好坏直接影响到分类精度。 这里采用双线性搜索法来求解最优参数, 其原理是利用不同的  $(C, \gamma)$  取值对应不同的性质, 以  $\log C, \log \gamma$  作为参数空间的坐标, 学习精度最高的参数组合  $(C, \gamma)$  将集中出现在直线  $\log \gamma = \log C - \log \bar{C}$  附近。 具体算法如下:

(1) 对线性 SVM 求解最佳参数  $C$ , 使得以之为参数的线性 SVM 学习精度最高, 称之为  $\bar{C}$ 。

(2) 对径向基核的 SVM, 固定  $\bar{C}$ , 对满足  $\log \gamma = \log C - \log \bar{C}$  的  $(C, \gamma)$ , 训练 SVM, 根据其对于学习精度的估算, 得到最优参数。

#### 3.2.3 SVM 多分类器的概率输出

标准 SVM 的判决输出属于硬判决输出, 这将导致在利用 SVM 多分类器进行数据融合时主要依赖于投票法<sup>[8]</sup>。 由于实际应用中需要一个具有软判决输出的 SVM, 因此这里采用 sigmoid 函数作为连接把 SVM 的输出  $f(x)$  映射到  $(0, 1)$  区间, 实现了二类 SVM 的后验概率输出, 其后验概率输出的形式如下:

$$P(y = 1 | f(x)) = \frac{1}{1 + \exp(Af(x) + B)} \quad (4)$$

参数  $A$  和  $B$  通过式

$$\min - \left( \sum_i t_i \log(p_i) + (1 - t_i) \log(1 - p_i) \right) \quad (5)$$

求得,其中  $p_i = \frac{1}{1 + \exp(Af(x_i) + B)}$ ,  $t_i = \frac{y_i + 1}{2}$ ,  $y_i$  为样本  $i$  的类别标签。

### 3.3 基于改进证据理论的决策级融合

根据管道泄漏检测系统的实际输出要求,构造如下的基于改进证据理论的决策级融合检测模型。

#### 3.3.1 识别框架

在证据理论中,由一组完备且互不相容的元素组成的集合称为识别框架,用符号  $\Theta$  来表示<sup>[9]</sup>。管道状态的识别框架为  $\Theta = \{A_1, A_2, A_3\}$ ,其中  $A_1$  表示“正常”, $A_2$  表示“小泄漏”, $A_3$  表示“大泄漏”。

#### 3.3.2 基本概率赋值(basic probability assignment, BPA)

根据 Platt 概率输出公式来对证据理论框架下的 BPA 函数进行赋值,其中如何对识别框架下  $\Theta$  分配 BPA 值是目前面临的主要问题。这里首先给出 SVM 识别误差上界的概念<sup>[10]</sup>:如果一组训练样本能够被一个最优分类面分开,那么对于测试样本分类错误率的上限是训练样本中平均的支持向量数占总训练样本数的比例,即

$$E(P(error)) \leq \frac{E(\text{支持向量数})}{\text{训练样本总数} - 1} \quad (6)$$

此式恰好反映了 SVM 对于样本  $x$  的不确定性,对应着输出识别框架中的  $\Theta$ 。因此,考虑采取如下方法来分配识别框架中各焦元  $A_j (j = 1, 2, \dots, k)$  所对应的 BPA 函数值。

对于样本集中的第  $i$  个样本点,应用训练好的 SVM 多分类器计算其识别结果的概率输出,识别结果中焦元  $A_j$  所对应的 BPA 函数值为

$$m_i(A_j) = \frac{1}{1 + \exp(Af(x_j) + B)} \frac{s_j}{l - 1} + \frac{1}{k - 1} \left[ \sum_{m=1, m \neq j}^3 \frac{\exp(Af(x_j) + B)}{1 + \exp(Af(x_j) + B)} \frac{s_m}{l - 1} \right] \quad (7)$$

$$m_i(\Theta) = \frac{1}{k - 1} \left( \sum_{m=1}^3 \frac{\exp(Af(x_j) + B)}{1 + \exp(Af(x_j) + B)} \frac{s_m}{l - 1} \right) \quad (8)$$

其中,  $\frac{1}{1 + \exp(Af(x_j) + B)}$  为由式(4)计算出的第

$j$  个二类 SVM 的后验概率输出,  $S_j$  为第  $j$  个 SVM 对应的支持向量个数,  $l$  为训练样本总数,  $\Theta$  为识别结果中的不确定性。对得到的 BPA 函数值再按式

$$m'_i(A_j) = m_i(A_j) / \sum_{j=1}^k m_i(A_j) \quad (9)$$

进行归一化处理,这样,就得到该 SVM 多分类器对应第  $i$  个检验样本点输出的 BPA 函数  $m'_i$ 。

在实际的管道泄漏检测过程中,检测节点距离泄漏点越近,提取到的泄漏特征信号相对越强,检测结果的可靠性越高;反之,距离越远,泄漏信号中的干扰成分越多,检测结果越不可靠<sup>[11]</sup>。为减小不可靠证据对组合结果的影响,提高证据集对正确命题的聚焦度,在决策级融合时,需要对 SVM 多分类器输出的 BPA 值加以预处理后再进行证据组合运算。这里采用的是基于节点可靠度的预处理方法:设传感器节点与泄漏源之间的距离为  $d_i$ ,距离泄漏源最近的节点的可靠度为 1,距离为  $d_{\min}$ ,则可通过式

$$\begin{cases} d_{\min} = \min(d_i) \\ Cred_i = T \sqrt{d_{\min}/d_i} \end{cases} \quad (10)$$

来计算证据的可靠度  $Cred_i$ 。其中,  $T$  为可靠度的影响因子,经过大量实验研究确定为 8。为了满足各命题的 BPA 函数之和为 1 的要求,经过可靠性处理后,各证据体的 BPA 函数可表示为

$$\begin{cases} m_i^*(A_j) = Cred_i m'_i(A_j), & A_j \neq \emptyset \\ m_i^*(\Theta) = 1 - \sum_{A_j \in \Theta} m_i^*(A_j) \end{cases} \quad (11)$$

#### 3.3.3 证据合成

设  $m_1$  和  $m_2$  分别为同一识别框架  $\Theta$  上的 BPA 函数,焦元分别为  $A_i (i = 1, 2, \dots, k)$  和  $B_j (j = 1, 2, \dots, l)$ ,则 2 条证据的组合公式<sup>[11]</sup>为

$$m_{1 \oplus 2}(A) = 0, \quad A = \emptyset$$

$$m_{1 \oplus 2}(A) = \frac{\sum_{A_i \cap A_j = A} m_1(A_i) m_2(B_j)}{1 - K} \quad (12)$$

其中,  $K = \sum_{A_i \cap A_j = \emptyset} m_1(A_i) m_2(B_j)$ ,表示证据冲突的程度; $\oplus$ 表示直和运算。

#### 3.3.4 最大信任度决策法

根据组合得到的具有证据结构特征的融合结果,可采用最大信任度值法做出决策判断。

$\exists A_1, A_2 \subset \Theta$ , 且  $m(A_1) = \max\{m(A_k), A_k \subset \Theta\}$ ,  $m(A_1) = \max\{m(A_k), A_k \subset \Theta$ , 且  $A_k \neq A_1\}$ , 若满足

$$\begin{cases} m(A_1) - m(A_2) > \varepsilon_1 \\ m(\Theta) < \varepsilon_2 \\ m(A_1) > m(\Theta) \end{cases} \quad (13)$$

则诊断结果为  $A_1$ 。其中,  $\varepsilon_1$ 、 $\varepsilon_2$  为预先设定的阈值,通过大量实验验证,确定  $\varepsilon_1$  为 0.15,  $\varepsilon_2$  为 0.1。

#### 4 实验分析

本实验通过管道泄漏模拟装置来模拟天然气管道泄漏。管道外径 160mm,壁厚 6mm,内部压力 0.2Mpa,使用泄漏阀模拟泄漏。在泄漏点附近布置 5 个传感器节点,各节点距泄漏源的距离依次为 50m、80m、100m、120m 和 150m。节点在每个采样过程中的采样点数为 2048,采样频率为 1MHz,采样持续时间为  $2048 \times 10^{-6}$ s,实测波速为 5410m/s。

图 4 给出了管道发生泄漏时,5 个传感器节点接收到的信号波形图。从图中可以看出,原始信号

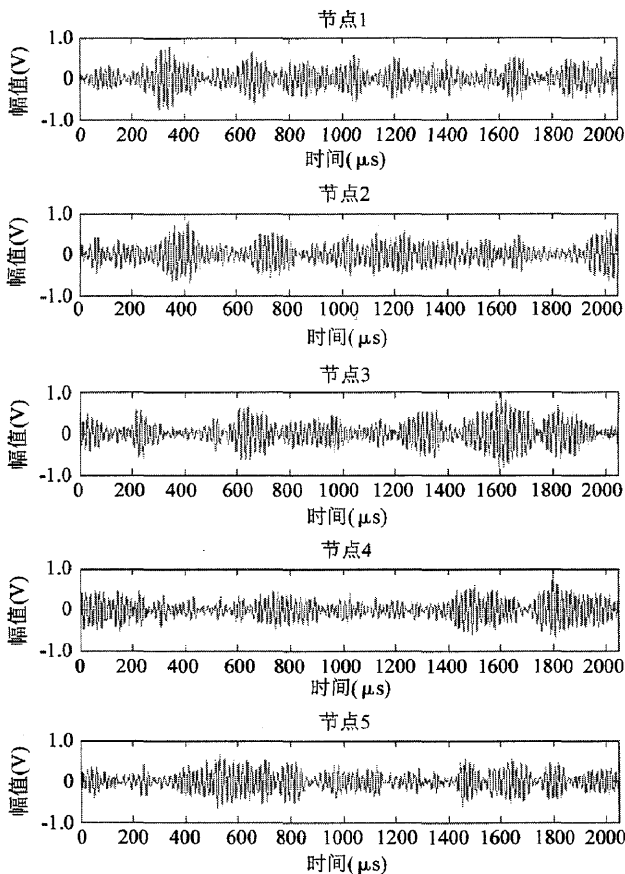


图 4 节点接收到的原始信号

为掺杂了大量噪声及干扰的连续型声发射信号,很难分辨出泄漏特征。对节点 1 信号进行 5 层小波分解,其结果如图 5 所示。可以看出,分解后的第 5 层信号具有明显的周期性冲击,反映了泄漏特征,并且第 5 层信号包含了原始信号的大部分能量,因此提取小波分解的第 5 层信号作为包含泄漏特征的单模态信号进行特征参数提取。

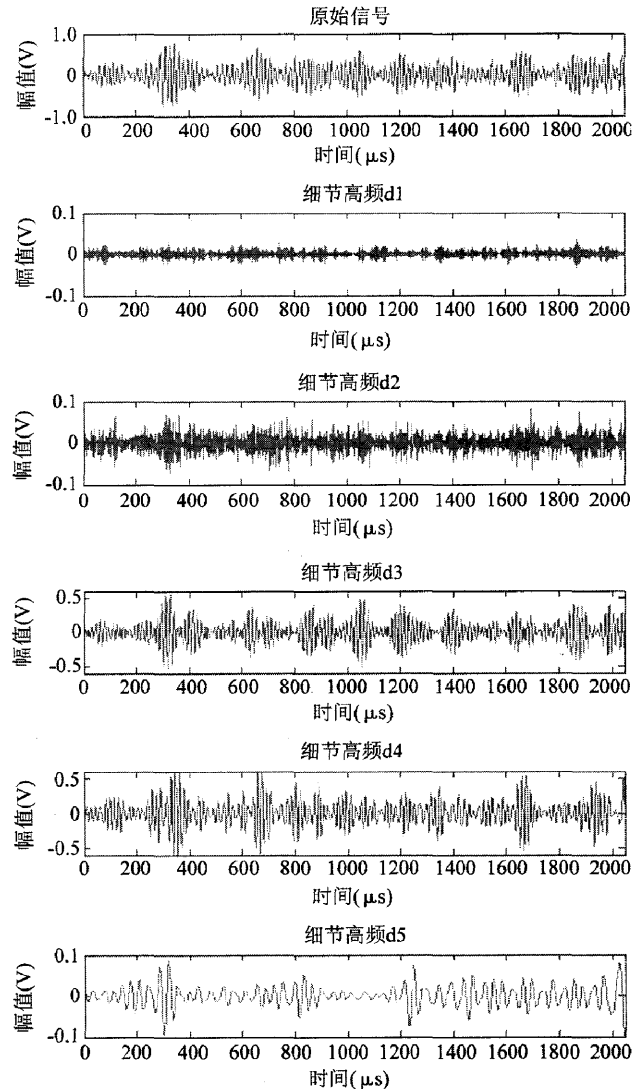


图 5 节点 1 信号小波分解

对于提取出的特征参数,作为输入特征向量送入 SVM 多分类器进行初级识别。本文独立选取 120 组数据(正常、小泄漏和大泄漏各 40 组)作为样本集,其中在每一种类型中随机选取 20 个样本(总共 60 个样本)组成训练样本集,其余的作为测试样本集。表 1 给出了 3 类信号的 2 个示例:1,2 为正常状态;3,4 为小泄漏状态;5,6 为大泄漏状态。利用 Matlab 中的 SVM 工具箱对分类器进行训练。训练结束后,选取测试样本输入到分类器中,并对输出结果作归一化处理,根据决策规则测试分类器的实际识别效果。

当管道某处发生小泄漏时,经 SVM 多分类器初级识别后所确定的基本概率赋值函数如表 2 所示。从表中可以看出,随着检测节点距泄漏点的距离不断增加,其检测的不确定性逐渐增大。另外,节点 4 的初级识别结果与其它 4 个节点不一致,因此不满

表 1 信号样本特征向量

峰值	平均幅值	方差	方根幅值	峰值因子	脉冲因子	裕度因子	偏态因子	峭度因子
0.04729	0.01272	0.00025	0.01077	2.96782	0.41931	3.71730	-1.4047	2.69521
0.04496	0.01395	0.00029	0.01195	2.62446	0.38066	3.22298	-1.4020	2.53662
0.07940	0.01381	0.00033	0.01126	4.33524	0.67553	5.74708	-1.3712	3.26362
0.07187	0.01295	0.00028	0.01077	4.28256	0.63152	5.54897	-1.4185	3.13221
0.11165	0.01607	0.00046	0.01320	5.17711	0.88072	6.94743	-1.2839	3.71906
0.12724	0.01649	0.00055	0.01307	5.41933	0.99074	7.71438	-1.5390	4.92900

表 2 支持向量机多分类器的初级识别结果

BPA	$A_1$	$A_2$	$A_3$	$\Theta$
$m_1$	0.2614	0.3400	0.2614	0.1372
$m_2$	0.2465	0.3210	0.2465	0.1860
$m_3$	0.2397	0.3118	0.2397	0.2088
$m_4$	0.2344	0.2344	0.3047	0.2265
$m_5$	0.2279	0.2964	0.2279	0.2478

足式(13),故无法进行最终决策。主要原因是:(1)检测节点自身出现问题导致 SVM 多分类器的输入特征向量存在一定的误差;(2)样本数目比较少,且不能覆盖所有可能的情况,导致分类器自身存在识别错误。

由表 3 可以看出,经过证据组合以后,证据集对命题  $A_2$  的支持度提高到 0.4638,与单节点的 SVM 识别结果相比,提高了 36.4% 以上。对不确定性命题的支持度降低到 0.044,显然满足式(13),其检测结果为“支持命题  $A_2$ ”,即发生了小泄漏。

表 3 证据合成结果

BPA	$A_1$	$A_2$	$A_3$	$\Theta$
$m_{1,2}$	0.2742	0.4040	0.2742	0.0476
$m_{1,2,3}$	0.2667	0.4469	0.2667	0.0197
$m_{1,2,3,4}$	0.2601	0.4296	0.3012	0.0091
$m_{1,2,3,4,5}$	0.2467	0.4638	0.2851	0.0044

通过 3 个层次的数据融合,解决了监测网络中存在的误识别问题,同时提高了证据集对正确命题的聚焦度,避免了初级识别结果存在的难以决策的问题。

## 5 结论

利用无线传感器对管道网气体泄漏进行检测时,数据处理采用层级式融合方法可以有效提高信息监测的准确性和可靠性。层级式融合方法的内容包括:数据级利用 db 小波变换提取包含泄漏特征的

单模态声发射信号,消除噪声干扰带来的影响;特征级建立支持向量机多分类器模型,并采用双线性搜索法选取分类器和核函数的参数,提高初级识别的准确率;决策级采用 sigmoid 函数将分类器初级识别结果转化为证据的基本概率赋值,实现赋值的客观化,改进的证据组合方法可解决不同节点之间的证据冲突问题。经过三级的数据融合处理,可有效提高管网泄漏检测精度,避免识别结果中存在的误识别和难以决策问题。

### 参考文献

- [ 1 ] Stoianov I, Nachman L, Madden S, et al. PIPENET: A wireless sensor network for pipeline monitoring. In: Proceedings of the 6th International Symposium on Information Processing in Sensor Networks. Massachusetts, USA, 2007. 264-273
- [ 2 ] Jawhar I, Mohamed N, Mohamed M, et al. A routing protocol and addressing scheme for oil, gas, and water pipeline monitoring using wireless sensor networks. In: Proceedings of the International Conference on Wireless and Optical Communications Networks, Singapore, 2008. 1-5
- [ 3 ] Mohamed N, Jawhar I. A fault tolerant wired/wireless sensor network architecture for monitoring pipeline infrastructures. In: Proceedings of the International Conference on Sensor Technologies and Applications, Cap Esterel, 2008. 179-184
- [ 4 ] Jawhar I, Mohamed N, Shuaib K. A framework for pipeline infrastructure monitoring using wireless sensor networks. In: Proceedings of the 2007 Wireless Telecommunications Symposium, Pomona, 2007. 1-7
- [ 5 ] Yuanwei Jin, Eydgahi A. Monitoring of distributed pipeline systems by wireless sensor networks. In: Proceedings of the 2008 IAJC-IJME International Conference, Nashville, USA, 2008. 1-10
- [ 6 ] Mitra P, Murthy C A, Pal S K. Unsupervised feature selection using feature similarity. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2002, 24 ( 3 ):

- 301-312
- [ 7 ] Kavzoglu T, Colkesen I. A kernel functions analysis for support vector machines for land cover classification. *Applied Earth Observation and Geoinformation*, 2009, 5 (11) : 352-359
- [ 8 ] Husan T L, Chih J L, Ruby C W. A note on platt's probabilistic outputs for support vector machines. *Machine Learning*, 2007, 68(3) :267-276
- [ 9 ] Florea M C, Jusselme A L, Bosse E et al. Robust combination rules for evidence theory. *Information Fusion*, 2009, 2(10) : 183-197
- [ 10 ] Platt J C. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. <http://research.microsoft.com/~jplatt>, 1999
- [ 11 ] 万江文, 陈斌, 冯仁剑等. 基于可靠度和一致强度的冲突证据组合算法. *高技术通讯*, 2009, 19(7) : 730-734

## A hierarchical data fusion method for detection of the leak of gas pipelines based on wireless sensor network

Yu Yang, Wu Yinfeng, Feng Renjian, Wan Jiangwen

(School of Instrumentation Science and Opto-electronics Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191)

### Abstract

To improve the accuracy and reliability of the leak monitoring of gas pipelines by using wireless sensor networks(WSN), this paper puts forward a hierarchical data fusion algorithm based on the combination of the wavelet support vector machine (SVM) method and the evidence theory. The algorithm is described below. In the signal level fusion, the noise elimination for primitive signals is conducted using the wavelet transform technology, and leak characteristic parameters are totally extracted as well. In the attribute fusion, a multi-classifier model based on SVM is constructed, and characteristic parameters as input vectors are sent to the multi-classifier for initial recognition. In the decision level fusion, the evidence combination is accomplished using the improved evidence combination methods at the sink node for final decision making. The experimental results show that the approach could improve the precision of the leak location detection and reduce the undetected rate as well as the false alarm rate.

**Key words:** leak detection, wireless sensor network (WSN), support vector machine (SVM), evidence theory