

# 模式识别技术在泥浆浓度反演中的应用

李德军, 吕艳华, 王润田

(中国科学院上海声学实验室, 上海 200032)

**[摘要]** 泥浆在建筑工程中使用非常普遍, 合理地控制泥浆的物理性能对于建筑工程施工及其质量控制非常重要, 通过声学方法可以有效地监测泥浆的体积浓度等物理参数。在通过声衰减和声速等介质的声学参数反演泥浆浓度的过程中, 数据拟合的好坏直接影响到反演的精确程度。通过模式识别技术, 利用聚类算法, 对数据进行分类、归类处理, 能有效的地提高反演的准确度。

**[关键词]** 模式识别; 最近邻法; 聚类算法; 泥浆浓度

**[中图分类号]** TB559 **[文献标识码]** A **[文章编号]** 1009-1742 (2007) 05-0081-04

泥浆在建筑、地质、电力、水利等部门的应用非常普遍, 对于建筑工程施工及其质量控制非常重要, 然而要用传统的取样法对泥浆的物理参数进行在线监测是比较困难的。借助超声学方法可以有效地监测泥浆的体积浓度等物理参数, 而且可以实现在线监测或实时监测。通常的做法是借助于超声波在泥浆介质中的传播, 通过超声波信号的变化, 反演泥浆浓度等介质的参数。由于建筑上使用的泥浆成分的多样性, 还很难找出一个准确的理论模型。因此必须借助一些先验知识并采取数据处理技术对泥浆浓度进行反演, 特别是对泥浆浓度超过20%的情况。

模式识别技术<sup>[1]</sup>在海洋遥感、计算机信息处理、自动控制等领域有广泛的应用前景。在基于声衰减和声速等介质的声学参数<sup>[2]</sup>的泥浆浓度反演的过程中, 将模式识别技术用于数据拟合、分类等处理过程, 并提出了一种合理的数据分析方法<sup>[3, 4]</sup>, 可以在高浓度的泥浆系统中使用。

## 1 最近邻归类法介绍

最近邻法是统计模式识别中比较直观、简单的

方法, 是常见的一种非参数分类方法。把每个训练样本作为该类的中心点, 对未知类别的模式  $X$  进行它与训练样本距离的计算, 并取最小值, 将该训练样本作为其归属类别。图1是一个二维二类的例子, 其中  $X_1, X_2$  为2个自由度,  $W_1, W_2$  为两类。两类的分界面是分段线性的。

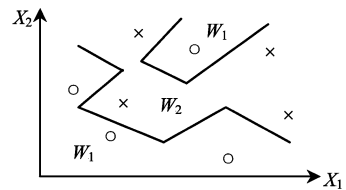


图1 最近邻法

Fig.1 Nearest neighbor

理论证明, 最近邻方法的平均错误率  $P$  与最小错误率贝叶斯分类器平均错误率  $P^*$  的关系为

$$P^* \leq P \leq P^* (2 - P^* M / (M - 1)),$$

其中  $M$  为分类数。当样本数趋于无穷时,  $P$  略大于  $P^*$ , 但小于  $2P^*$ 。最近邻方法得到这样的错误率, 是远远满足实际要求的。

最近邻方法计算量太大。在泥浆反演的实际工

[收稿日期] 2005-12-28; 修回日期 2006-04-10

[基金项目] 国家自然科学基金资助项目 (10574129)

[作者简介] 李德军 (1979-), 男, 山东莱阳市人, 中国科学院上海声学实验室硕士, 从事信号处理及声学理论的有关研究

作中, 笔者结合聚类算法, 对这种方法做了改进。

## 2 聚类算法介绍

聚类算法在统计理论、数据分析等方面应用极其广泛。聚类结果有一定的主观因素, 衡量结果是否合理应该从它是否符合实际情况来确定。

离差平方和准则是一种简单常用的聚类准则。设有  $N$  个样本, 分成  $K$  个群,  $\{X_1, X_2, \dots, X_k\}$ , 第  $j$  个群  $X_j$  有  $N_j$  个样本, 其中心点为

$$m_j = \frac{1}{N_j} \sum_{x \in X_j} x,$$

则  $X_j$  群内所有点和中心点  $m_j$  的距离平方和为

$$d_j = \sum_{x \in X_j} \|x - m_j\|^2.$$

$K$  个群的类内离差平方和为

$$J = \sum_{j=1}^k d_j.$$

如果分类恰当,  $d_j$  和  $J$  应当比较小。以  $J$  最小作为分群准则, 这样就应该把  $x$  分到离它最近的中心所代表的群中去。这种分类方法用在未知点“群”距离较大的场合尤其合适。

## 3 泥浆浓度反演的基本原理

假设泥浆系统是分层均匀的, 超声信号通过泥浆以后, 信号幅值和声速发生明显的变化<sup>[5,6]</sup>, 通过对声衰减和声速的变化来反演泥浆浓度。实验发现, 对小粒径泥浆粒子, 体积浓度低于 20% 时, 衰减与体积浓度近似成正比关系, 高于 20%, 二者没有明显的线性关系; 声速在低于 20% 浓度时, 声速随浓度增加而降低, 高于 20% 时, 声速随浓度增加而增加, 但都没有明显的正比关系。

在研究声衰减、声速和泥浆浓度的关系时, 由于存在体积浓度 20% 的分界点, 所以考虑分段研究。为了论述方便, 将用统一的函数来近似表示它们的关系。

假设体积浓度  $\phi$  与声衰减  $\alpha$  满足函数关系式为

$$\phi = f(\alpha) \quad (1)$$

体积浓度  $\phi$  与声速  $v$  满足关系式为

$$\phi = g(v) \quad (2)$$

令  $f(\alpha) = g(v)$ , 得到

$$K(\alpha, v) = 0 \quad (3)$$

得到的实验数据是离散值, 进行数据拟合, 用式 (1) 至式 (3) 表示它们的关系。假设函数分段连续。

由式 (1) 可以得到  $\phi$  随  $\alpha$  的变化率为

$$\theta_\alpha = \partial\phi/\partial\alpha \quad (4)$$

由式 (2), 可以得到  $\phi$  随  $v$  的变化率为

$$\theta_v = \partial\phi/\partial v \quad (5)$$

经过多次实验, 根据聚类准则对实验数据进行处理, 建立的数据库为  $(\alpha, v, \phi)$ , 通过  $(\alpha, v)$  反演  $\phi$ 。通过测定的  $(\alpha_x, v_x)$ , 将根据数据库中的训练样本求出  $\phi_x$  的值。实际上  $(\alpha_x, v_x)$  未必是建立样本库中的数据, 此时根据最小距离法, 求出样本库中离  $(\alpha_x, v_x)$  最近的点的体积浓度  $\phi$ , 再根据式 (4) 和式 (5) 进行修正, 将修正值作为  $\phi_x$  的反演值。

## 4 模式识别在反演中的具体应用

假设数据库中的  $\alpha$  和  $v$  的关系如图 2 所示 (对训练样本已经进行过聚类处理, 图 2 中 ‘ $\times$ ’ 代表训练样本)。将训练样本点记为  $\{A_1, A_2, \dots, A_j, \dots\}$ , 分别把样本库中的相邻 2 点作为一类 (同一个点属于 2 类), 即取  $\{A_j, A_{j+1}\}$ ,  $j \geq 1$  作为一类。对于一未知浓度的泥浆系统, 得到一组样本点 (去掉离大部分样本点距离较远的点), 如图 2 中的 ‘ $\circ$ ’ 表示。首先, 对样本点进行凝聚处理, 找到其中心点的位置, 假设中心点为  $O$ , 将未知样本归类的过程是计算  $O$  与每一类  $\{A_j, A_{j+1}\}$  的中心位置的距离, 找到最小值作为其归属类  $K: \{A_k, A_{k+1}\}$ ,  $K$  中的 2 个样本点记为  $M$  和  $N$ 。通过  $O$  向  $MN$  的连线作垂直线交于  $P$  点, 如图 3 所示。

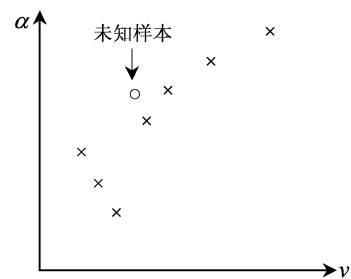


图 2 未知样本与数据库样本  
Fig.2 Unknown and database sample

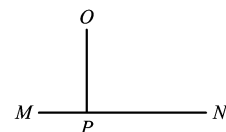


图 3 获得  $p$  点  
Fig.3 Access point  $p$

在  $P$  点, 由式 (4) 和式 (5) 求得  $(\theta_a, \theta_v)$ 。

$\theta_a > \theta_v$  表明, 此点声衰减随泥浆浓度变化更快, 此时, 在校正数据时  $P$  点应该向  $\Delta\alpha$  减少方向的直线上移动, 如图 4 所示。

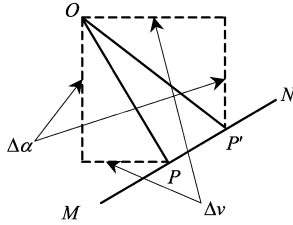


图 4 寻找最优值  $P'$

Fig.4 Search for the optimum point  $P'$

在  $P$  点处,  $\Delta\alpha > \Delta v$ , 所以将  $P$  移向  $P'$  点,

表 1 实验测得体积浓度与声衰减和声速的关系

Table 1 Experiment data about concentration, attenuation and velocity

$\phi/\%$	0.75	1.34	2.57	7.50	11.25	11.56	13.13	14.06	15.00	15.63	16.25	17.50	17.81	19.38	25.63
$\alpha/\text{dB}$	0.33	1.02	2.60	5.11	8.63	12.03	13.06	16.58	17.62	19.08	20.67	21.78	25.34	30.57	34.65
$v/\text{m}\cdot\text{s}^{-1}$	1 533	1 522	1 520	1 519	1 518	1 514	1 507	1 495	1 503	1 513	1 518	1 525	1 529	1 534	1 544

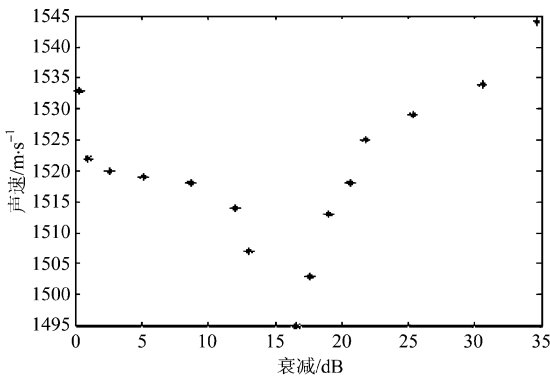


图 5 实验测得衰减和声速关系

Fig.5 Relation between attenuation and velocity in experiment

根据上述数据处理方法得到的结果和实测数据见表 2。理论计算值和实际测量值的对比曲线见图 6 和图 7。

### 5 误差分析

由表 2 的数据可以看出, 理论计算值和实际测量值之间有一定的误差, 一方面是实验过程中的存在测量和观测误差, 这主要是由于高浓度泥浆很容易沉淀, 影响观测数据的准确度; 另一方面, 由于

在  $P'$  点处,  $\Delta\alpha < \Delta v$ , 移动幅度可以近似使  $P'$  点处满足  $\Delta\alpha/\Delta v|_{P'} = \theta_v/\theta_a|_P$ 。

$\theta_v < \theta_a$  的情况下, 声速在该点对泥浆浓度的反演影响更大, 此时  $P$  点应该向  $\Delta v$  减少的方向移动。

$\theta_v = \theta_a$  时,  $P$  与  $P'$  重合。

设  $|MP'|/|P'N| = L$ , 令  $(\phi_1 - \phi_x)/(\phi_x - \phi_2) = L$ , 得到的反演数据为

$$\phi_x = (\phi_1 + L\phi_2)/(1 + L)$$

在一定条件下, 测得的泥浆体积浓度 (0.75~25.63)% 与相应的声衰减、声速的数据见表 1, 并建立其样本数据库。

由表 1 得到的衰减与声速的关系见图 5。

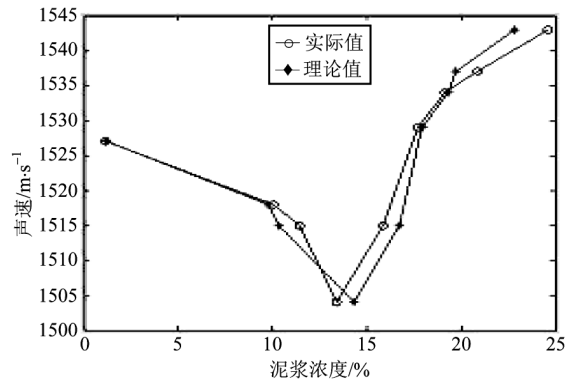


图 6 泥浆浓度与声速关系

Fig.6 Relation between concentration and velocity

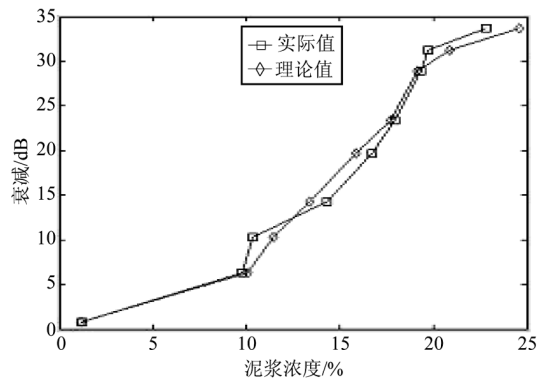


图 7 泥浆浓度与声衰减关系

Fig.7 Relation between concentration and attenuation

表2 误差分析  
Table 2 Error analysis

声速/ $\text{m}\cdot\text{s}^{-1}$	1 527	1 518	1 515	1 504	1 515	1 529	1 534	1 537	1 543
声衰减/dB	0.79	6.34	10.33	14.27	19.67	23.36	28.96	31.29	33.75
计算浓度/%	1.11	10.07	11.44	13.40	15.86	17.72	19.14	20.87	24.63
实际浓度/%	1.20	9.75	10.35	14.32	16.72	17.94	19.32	19.75	22.85
误差/%	8.7	3.37	10.53	6.42	5.14	1.21	0.93	5.67	7.79

测量过程中气泡的存在和悬浮质粒径大小分布的离散性以及温度、压力等都影响测量数据的准确性,这就要求获取比较多的样本,并在建库之前对数据进行合理的预处理。理论分析可知,反演泥浆浓度取自由度越多,反演结果就越准确,但在实验中,仅取两个自由度(声衰减、声速)来反演泥浆浓度,已经能够达到比较理想的效果。在实验过程中,注意减少误差的存在,在建库之前对数据进行合理的校正处理,反演结果会更加理想。

## 6 结语

影响泥浆中的声速的因素除了泥浆的浓度外,还有其他许多物理和环境等方面的因素,笔者主要从泥浆的声速以及对声能量的衰减量方面进行分析研究。通过对实验数据的凝聚处理和对样本库的数据分类处理,不仅简化了数据处理过程,而且在一定程度上降低了样本库的数据量,降低了建立数据库的工作量。所提出的凝聚求最小距离来确定归类

方法,在泥浆浓度反演的过程中,减少了实验误差,较好地满足了实际需要。

### 参考文献

- [1] 边肇琪. 模式识别[M]. 北京: 清华大学出版社, 2000
- [2] 王润田. 海底声学探测和底质识别技术的新进展[J]. 声学技术, 2002, 21(1): 96~98
- [3] 李世峰, 黄磊. 几种聚类方法的比较[A]. 第八届全国汉字识别学术会议论文集[C]. 北京: 中国中文信息学会, 2003
- [4] 钱同惠, 吴稼农. 加权接近度多特征模糊模式识别的应用[J]. 计算技术与自动化, 1998, 17(2): 51~53
- [5] Blue J E, McIeroy E G. Attenuation of sound in suspensions and gels[J]. JASA, 1968, 44(4): 1145~1148
- [6] Poverly M J W. Ultrasonic Techniques for Fluid Characterization [M]. Academic Press, USA, 1991

## Application of Fuzzy Pattern Recognition in the Measurement of Slurry Concentration

Li Dejun, Lü Yanhua, Wang Runtian

(Shanghai Acoustics Lab, Chinese Academy of Sciences, Shanghai 200032, China)

**[Abstract]** Slurry is widely used in construction projects, and it is important to control the slurry's physical characteristic properly. The acoustic method is used, which can effectively monitor the physical parameters of slurry, such as concentration. Data processing affects directly the precision in the measurement of slurry concentration by the sound attenuation and velocity. Based on the fuzzy pattern recognition, data are sorted and further classified, with cooperative clustering algorithm.

**[Key words]** fuzzy pattern recognition; nearest neighbor (NN); cooperative clustering algorithm (CCA); slurry concentration