

神经网络在车辆声学分类中的应用

刘壮明 管鲁阳[†] 鲍明 李晓东

(中国科学院声学研究所 北京 100080)

摘要 传统反向传播 (BP, Back-Propagation) 算法虽然解决了多层感知器的收敛问题, 但是训练时间长、收敛速度慢。本文针对训练样本分布状态未知的问题, 提出了一种有效的加速收敛方法, 即对不同的训练样本选择不同的学习率。将这种改进的 BP 算法应用到履带车与轮式车的声学分类中, 明显提高了算法的收敛速度、泛化能力及稳定性, 并可根据需要调整两种车辆的识别率。

关键词 多层感知器, 反向传播算法, 样本非均匀分布, 车辆声学分类

Application of Artificial Neural Networks in Acoustic Classification of Vehicles

LIU Zhuang-Ming GUAN Lu-Yang BAO Ming LI Xiao-Dong

(Institute of Acoustics, Chinese Academy of Sciences, Beijing 100080)

Abstract Basically, the conventional back-propagation (BP) algorithm overcomes the convergence problem of the multilayer perceptron, but slower convergence rate and longer training time are some disadvantages as compared with other competing techniques. In this paper, with the consideration of unknown data distribution, an efficient technique is proposed to improve the convergence rate, in which the learning rate coefficients are taken to be variable for different training samples. Based on the improved BP algorithm, experimental results in the application of acoustic classification for tracked and wheeled vehicles indicate a superior convergence rate, generalization capability and steadiness, and that the classification accuracy of vehicles could be adjusted according to request.

Key words Multilayer perceptron, Back-propagation (BP) algorithm, Imbalanced data distribution, Acoustic classification for vehicles

2006-06-05 收稿; 2007-03-18 定稿

作者简介: 刘壮明 (1981-), 男, 中国科学院硕士研究生, 研究方向: 信号处理、模式识别。

管鲁阳 (1979-), 男, 中国科学院声学研究所博士研究生。鲍明 (1973-), 男, 博士研究生。

李晓东 (1966-), 男, 研究员, 博士生导师。

[†] 通讯联系人 E-mail: guanluyang@mail.ioa.ac.cn

1 引言

随着科技快速发展与军事需要,无线传感器网络已经成为当前备受关注的、多学科交叉的新兴前沿研究热点。地面目标识别^[1]作为地面无无线传感系统的一个重要组成部分,其任务为:通过前端传感系统,检测到环境中的各类物理信号(包括声、振动、磁、图像等),并对所获得信号进行处理以确定环境中地面目标的信息(如类别、数量、移动方向、速度等)。本文选择基于传统反向传播(BP, Back-Propagation)算法的多层感知器作为分类器,研究并实现了地面目标识别中的一个典型问题——履带车与轮式车的声学分类。

传统BP算法^[2]是训练多层感知器最有效的方法之一,其主要思想是从后向前(反向)逐层传播输出层的误差,以间接计算出隐层误差。算法分为两个阶段:第一阶段(正向过程)输入信息从输入层经隐层逐层计算各单元的输值;第二阶段(反向传播过程)输出误差逐层向前计算出隐层各单元的误差,并用此误差修正前层权值。对于每个训练样本重复这一过程,直至误差收敛。

传统BP算法虽然解决了隐层权值修正问题,具有较强的并行处理能力和非线性功能,但也存在一些较明显的缺点。首先,BP算法是用梯度法求非线性函数极值,因而有可能陷入局部极小点,不能保证收敛到全局极小点。其次,BP算法逼近误差曲面^[3~5]的梯度变化是非均匀的,如果采用固定学习率,当学习率较小时,在误差曲面较平坦的区域,收敛较慢;当学习率较大时,又会在峡谷区域引起震荡。另外,BP算法的收敛速度,与学习率的选择以及权值的初始化也有很大关系。可见,选择固定的学习率,是导致BP算法收敛速度较慢的原因之一。针对这个问题,很多研究人员从不同角度提出了变学习率的BP改进算法^[6~8],其中有自适应变学习率及惯性项系数算法^[9](BPALM, Back-Propagation with Adap-

tive Learning rate and Momentum term),可以有效提高BP网络的收敛速度。

将BPALM算法用于履带车与轮式车的声学分类时,可以有效提高收敛速度,但是依然无法解决两种车辆识别率区别较大的问题。本文在BPALM算法的基础上,针对训练样本分布状态未知的情况,提出了应根据训练样本权重选择不同学习率的方法,不仅可以保证BP算法的收敛速度,而且能根据需要调整两种车辆的识别率。

本文第2节介绍了基于组样本训练的BPALM算法,并在此基础上提出了改进算法;第3节分析了履带车与轮式车的声音信号特征,介绍了改进的Mel频率倒谱系数(MFCC, Mel-Frequency Cepstrum Coefficient),然后利用BPALM改进算法实现了履带车与轮式车的声学分类;最后,在第4节中对全文得出结论。

2 BP算法介绍与改进

本文选择基于BP算法的双层感知器(即3层前馈神经网络,如图1所示)作为分类器,并使用组样本训练的方法。在BP算法中通常采用梯度法修正权值,为此要求输出函数可微,通常采用Sigmoid函数。如图2所示,不失其普遍性,我们研究处于某一层的第 j 个计算单元,脚标 i 代表其前层第 i 个单元,脚标 k 代表后层第 k 个单元, $O_j^{(p)}$ 代表样本 p 在本层的输出, ω_{ij} 是前层到本层的权值。

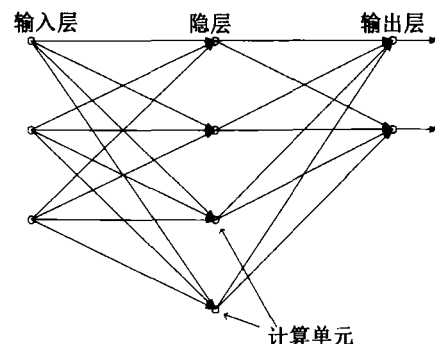


图1 双层感知器

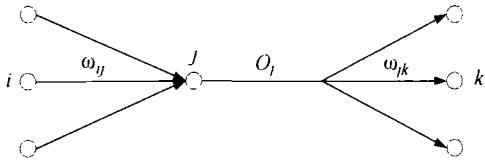


图 2 反向传播算法中的变量约定

假设共有 P 个训练样本， $y_j^{(p)}$ 代表每个节点的理想输出值，则 P 个样本的总均方误差为

$$E(n) = \frac{1}{N} \sum_{n=1}^N E^{(p)}(n) \tag{1}$$

$$= \frac{1}{2N} \sum_{p=1}^P \sum_j (y_j^{(p)} - O_j^{(p)})^2$$

其中 n 代表迭代次数。定义局部梯度

$$\delta_j^{(p)} = \frac{\partial E^{(p)}}{\partial (\sum_i \omega_{ij} O_i^{(p)})} \tag{2}$$

2.1 BPALM 算法

BPALM 算法^[9]的基本思想是：以连续两次观测的训练误差值为标准，若误差下降则增大学习率，加速收敛；若误差反弹则减小学习率，抑制振荡。定义误差改变率如下：

$$e_r(n) = \frac{\Delta E(n)}{E(n)} = \frac{E(n) - E(n-1)}{E(n)} \tag{3}$$

若 $e_r(n) < 0$ (有时取较小正数效果更好，如 10^{-2})

$$\eta(n+1) = \eta(n)[1 + ue^{-e_r(n)}], u \in (0, 1) \tag{4}$$

$$\alpha(n+1) = \alpha(n)[1 + ve^{-e_r(n)}], v \in (0, 1)$$

若 $e_r(n) > 0$ (或取 10^{-2})

$$\eta(n+1) = \eta(n)[1 - ue^{-e_r(n)}], u \in (0, 1)$$

$$\alpha(n+1) = \alpha(n)[1 - ve^{-e_r(n)}], v \in (0, 1) \tag{5}$$

则 BPALM 算法如下：

- (1) 创建 BP 网络，并选定权系数初始值；
- (2) 重复下述过程直至收敛。

从前向后各层计算各单元

$$O_j^{(p)} = \frac{1}{1 + e^{-\sum_i \omega_{ij} O_i^{(p)} + q_j(n)}}$$

对输出层计算 $\delta_j^{(p)} = (y_j^{(p)} - O_j^{(p)})O_j^{(p)}(1 - O_j^{(p)})$

从后向前计算各隐层 $\delta_j^{(p)} = O_j^{(p)}(1 - O_j^{(p)}) \sum_k \omega_{jk} \delta_k^{(p)}$

根据式 (3)~ 式 (5) 调整学习率 $\eta(n)$ 和惯性项系数 $\alpha(n)$

计算并保存各权值修正量

$$\Delta \omega_{ij}(n) = \alpha(n) \Delta \omega_{ij}(n-1) + \eta(n) \frac{\sum_{p=1}^P \delta_j^{(p)} O_i^{(p)}}{P} \tag{6}$$

$$\Delta q_j(n) = \alpha(n) \Delta q_j(n-1) + \eta(n) \frac{\sum_{p=1}^P \delta_j^{(p)}}{P} \tag{7}$$

修正权值 $\omega_{ij}(n+1) = \omega_{ij}(n) + \Delta \omega_{ij}(n)$

$q_j(n+1) = q_j(n) + \Delta q_j(n)$

2.2 BPALM 改进算法

在使用实际样本设计分类器的过程中，训练样本的具体分布状态未知，可能存在样本非均匀分布的问题^[10~12]，这可能会影响 BP 网络的收敛速度或实际分类效果。在不失一般性的前提下，一个两分类问题 (A 类和 B 类)，可能存在下面两种训练样本非均匀分布的情况：

(1) 属于 A 类的训练样本较多，而属于 B 类的训练样本较少，或相反情况。

(2) 高斯混合模型分类器^[13]，可以理解多个高斯概率密度函数的一个线性组合，只要有足够多数目的混合分量，就可以逼近任意一种概率密度函数。则任何一类样本，都可以通过聚类的方法细分为不同子类，而隶属于不同子类的样本个数可能存在较大差别。

基于以上考虑，假设 A 类共有 P_A 个训练样本，可细分为 M_A 个子类 (即聚类中心数)，每个子类的样本个数为 $P_{Ai} (i = 1, 2, \dots, M_A)$ ；B 类共有 P_B 个训练样本，可细分为 M_B 个子类，每个子类的样本个数为 $P_{Bi} (i = 1, 2, \dots, M_B)$ 。在使用 BPALM 算法时，计算权

值修正量时(如式(6))每个样本的学习率是相同的。在引入高斯混合模型的思想后,不同子类虽然样本个数不尽相同,但重要性是相同的,这就需要样本个数较少的子类学习率较大,样本个数较多的子类学习率较小。则将式(6)改写为如下形式:

$$\Delta\omega_{ij}(n) = \alpha(n)\Delta\omega_{ij}(n-1) + \eta(n) \sum_{p=1}^P \theta^{(p)} \delta_j^{(p)} O_i^{(p)} \quad (8)$$

其中, $\theta^{(p)}$ 用来调整样本 p 的学习率。则需要分别设定 A 类和 B 类样本的聚类中心数 M_A 和 M_B , 然后用 K-Means^[14] 方法对两类样本分别聚类, 得到 P_{Ai} 和 P_{Bi} , 进而由下式计算 $\theta^{(p)}$:

$$\theta^{(p)} = \begin{cases} \frac{P_A}{P_{Ai}(P_A M_A + P_B M_B)} & p \in A_i \subset A, \\ & i = 1, 2, \dots, M_A \\ \frac{P_B}{P_{Bi}(P_A M_A + P_B M_B)} & p \in B_i \subset B, \\ & i = 1, 2, \dots, M_B \end{cases} \quad (9)$$

需要特别说明的是, M_A 和 M_B 的选择并无理论方法可以指导, 需要根据训练样本的实际分布进行试探, 以得到较好的分类效果。

我们使用仿真样本对传统 BP 算法、BPALM 算法及 BPALM 改进算法进行了测试。结果是当训练样本均匀分布时, BPALM 算法及其改进算法的收敛速度均优于传统 BP 算法。当训练样本非均匀分布时, 传统 BP 算法及 BPALM 算法收敛速度较慢, 而 BPALM 改进算法可以明显提高收敛速度, 甚至接近训练样本均匀分布时的收敛速度。从而验证了 BPALM 改进算法的有效性, 下面将该算法应用到车辆声学分类中, 并与传统 BP 算法和 BPALM 算法进行比较。

3 履带车与轮式车的声学分类

MFCC 特征^[15,16] 已经在语音识别中得到广泛应用。在此基础上, 本文根据履带车与轮式车声信号的低频特性对其进行了一些优化^[17],

得到改进的 MFCC 特征。然后用该特征训练 BP 网络分类器, 包括传统 BP 算法、BPALM 算法及 BPALM 改进算法, 实现了履带车与轮式车的声学分类, 并详细讨论了这三种算法的收敛速度、泛化能力和稳定性。

3.1 履带车与轮式车的声信号特征提取

由于履带车与轮式车的主要能量集中在低频段(小于 300Hz), 所以设定采样率为 1000 Hz。对每个样本都进行去均值、归一化处理(FIR 滤波器系数为 [1, -0.95]), 以去掉直流偏置并补偿信道衰减。然后做短时傅里叶变换得到短时谱图(如图 3 图 4)。

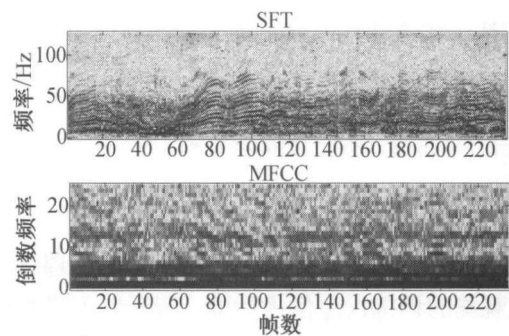


图 3 履带车声信号

(上: 傅里叶短时谱, 下: MFCC 短时谱)

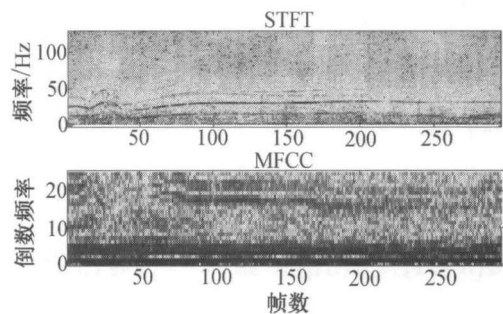


图 4 轮式车声信号

(上: 傅里叶短时谱, 下: MFCC 短时谱)

提取特征时, 借鉴并改进了语音识别中常用的 MFCC 特征^[17]。由于两类车辆的能量均集中在 50~250Hz, 其中履带车的谐频成分更丰富, 谱图结构更复杂, 因此采用倒谱分析的方法来凸现谐频中丰富的特性。频率域和 Mel 频率域的对对应关系(如图 5 所示)如下:

$$f_{Mel} = 900 \lg(1 + f/300) \quad (10)$$

然后参照 Mel 域均分原则, 构造 25 维三角窗滤波器组 (如图 6 所示), 则可得到改进的 MFCC 特征。最后, 得到 25 维改进的 MFCC 特征作为分类器的识别特征, 其中履带车样本 4550 个, 轮式车样本 3250 个。

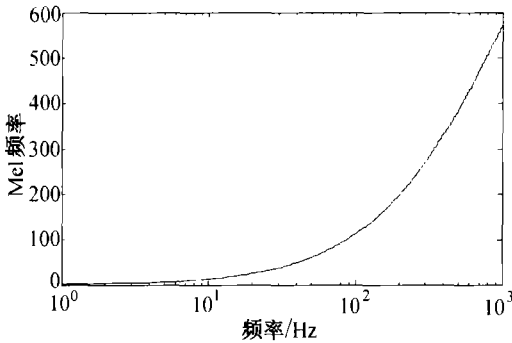


图 5 类 Mel 频率于 Nyquist 采样频率的对应关系

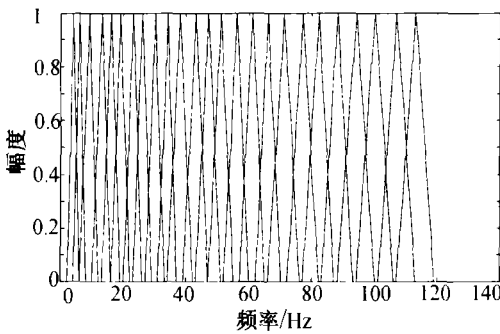


图 6 三角形的类 Mel 滤波器组

3.2 履带车与轮式车声学分类测试结果

采用 25 维改进的 MFCC 作为识别特征, 选择双层感知器作为分类器, 比较传统 BP 算法、BPALM 算法及 BPALM 改进算法。为了比较这三种 BP 算法的性能, 各 BP 网络的输入层、隐层和输出层节点相同, 分别为 25、16、2; 初始学习率 $\eta = 2.5$, 初始惯性项系数 $\alpha = 0.9$, 初始权系数均在 $[-1, 1]$ 之间随机生成, 且每次训练只迭代 300 次。

首先, 测试各种算法的收敛速度及泛化能力: 从 4550 个履带车样本中随机抽取 1500 个样本, 从 3250 个轮式车样本中抽取 1500 个样本, 将此 3000 个样本用于训练, 全部的 7800 个样本用于测试。然后, 测试各种算法的稳定

性: 连续训练 20 次计算平均正确率, 每次都随机抽取履带车和轮式车样本各 1500 个用于训练。最后, 尝试减少训练样本个数和训练迭代次数, 再次比较各算法的收敛速度、泛化能力及稳定性。

3.2.1 算法收敛速度与泛化能力

图 7~ 图 10 是各种算法训练迭代 300 次收敛曲线, 其中图 9 和图 10 中 BPALM 改进算法的聚类中心数不同。图 9 中履带车与轮式车的聚类中心数分别为 3 和 4, 图 10 中分别为 30 和 45。每种算法的收敛曲线均用上下两张图来表示, 其中上图为分类正确率, 实线代表训练样本, 虚线代表测试样本 (体现泛化能力); 下图为均方误差收敛曲线。比较这 4 张图可以看出, BPALM 算法与其改进算法的收敛速度大致相同, 均明显优于传统 BP 算法。

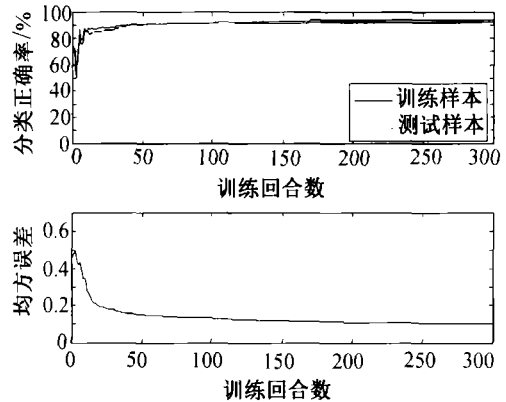


图 7 传统 BP 算法
(训练样本 93.5%, 测试样本 92.1%)

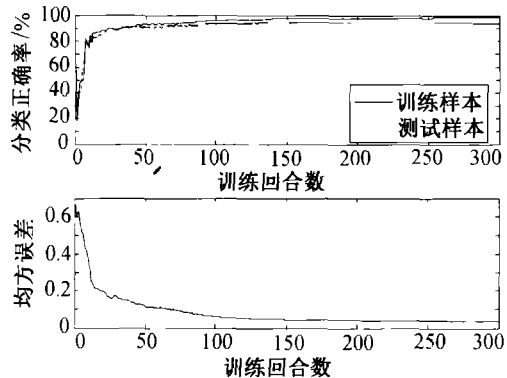


图 8 BPALM 算法
(训练样本 98.2%, 测试样本 94.2%)

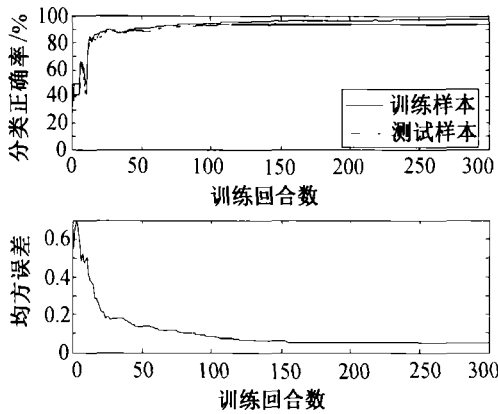


图 9 BPALM 改进算法 (3, 4)
(训练样本 97.3%, 测试样本 94.1%)

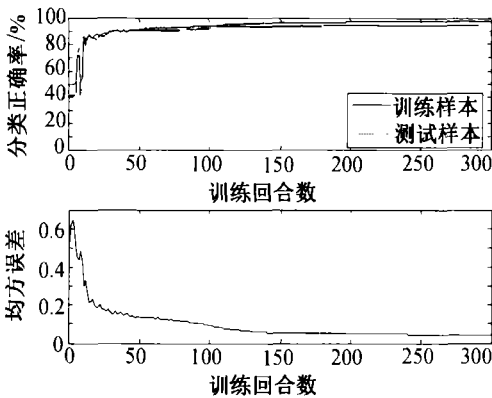


图 10 BPALM 改进算法 (30, 45)
(训练样本 97.6%, 测试样本 95.0%)

3.2.2 算法稳定性

将 3.2.1 节中测试的 3 种 BP 算法连续训练 20 次, 每次随机抽取履带车与轮式车各 1500 个训练样本, 然后训练迭代 300 次, 以测试各算法的稳定性。在所有训练过程中, 各算法全部收敛, 20 次训练后的正确率均值如图 11 所示。可明显看出, 传统 BP 算法收敛速度较慢, 而且履带车识别率明显高于轮式车。使用 BPALM 算法, 在迭代 300 次后, 训练样本分类正确率、测试样本分类正确率、履带车识别率及轮式车识别率均明显提高, 故训练收敛速度明显优于传统 BP 算法, 但是履带车识别率仍然明显高于轮式车。为解决该问题, 使用 BPALM 改进算法, 并使轮式车的聚类中心数大于履带车, 则可以在保证收敛速度与 BPALM 算法基本相同的情况下, 平衡两种车辆的识别率。

在上述测试的基础上, 减少履带车的训练样本个数, 即每次随机抽取履带车与轮式车训练样本数分别为 1000 和 1500, 连续测试 20 次, 每次训练迭代 200 次。在此情况下, 同时涉及训练样本非均匀分布中包含的两种情况, 以测试传统 BP 算法、BPALM 算法与 BPALM 改进算法 (履带车与轮式车的聚类中心数分别为 20 和 7) 的稳定性。从图 12 中可明显看出, BPALM 算法的收敛速度及泛化能力虽优于传统 BP 算法, 但轮式车的识别率仍然明显高于履带车; BPALM 改进算法在保证收敛速度与泛化能力的前提下, 不仅平衡了两种车辆的识别率, 而且泛化能力 (测试样本分类正确率) 也略优于 BPALM 算法。另外, 需要特别强调的是, 在使用 BPALM 算法的过程中, 有 3 次训练最终没有收敛, 图 12 中 BPALM 算法的结果为其余 17 次的平均值; 而使用 BPALM 改

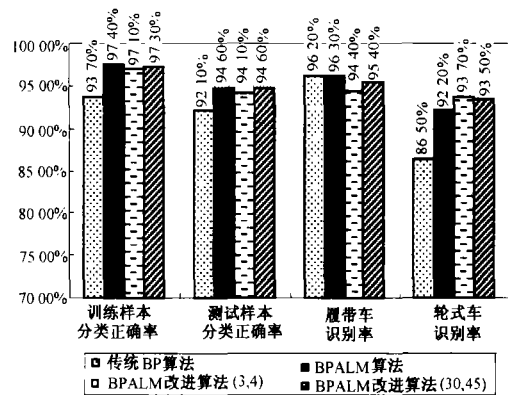


图 11 训练 20 次正确率均值 (300 次迭代)

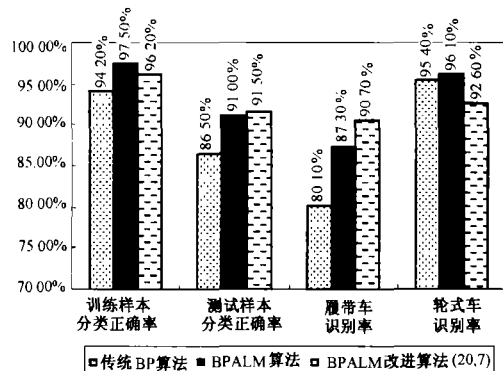


图 12 训练 20 次正确率均值 (200 次迭代)

进算法并没有出现不收敛的情况, 提高了算法的稳定性。

4 结论

本文首先介绍了 BPALM 算法, 在此基础上, 针对可能存在的训练样本非均匀分布问题, 提出了 BPALM 改进算法。然后根据履带车与轮式车声信号频谱的低频特性, 采用改进的 MFCC 系数作为识别特征。最后使用该特征对传统 BP 算法、BPALM 算法及 BPALM 改进算法分别进行了测试。

在测试过程中发现, BPALM 算法及其改进算法, 在收敛速度和泛化能力上比较接近, 且均明显优于传统 BP 算法。但由于训练样本分布状态未知, 可能存在非均匀分布的问题, BP 传统算法与 BPALM 算法均会出现一种车辆的识别率明显高于另一种车辆的现象, 甚至 BPALM 算法可能会不收敛。而 BPALM 改进算法可以有效地解决这两个问题, 不仅保证了收敛速度、泛化能力和稳定性, 而且可以通过调整聚类中心数平衡两种车辆的识别率, 甚至可以根据需要来增大或减小某一种车辆的识别率。

参 考 文 献

- [1] D Li, D W Kerry, Y H Hu, *et al.* Detection, classification and tracking of targets in distributed sensor networks. *IEEE Signal Processing Magazine*, 2002, 19(2): 17-29.
- [2] S Haykin. *Neural networks: a comprehensive foundation*. Second edition, Beijing: Tsinghua University Press, 2001. 156-255.
- [3] B Widrow, M A Lehr. 30 years of adaptive neural networks: perceptron, madaline, and backpropagation. *Proceedings of the IEEE*, 1990, 78(9): 1415-1442.
- [4] D R Hush, B G Horne. Progress in supervised neural networks. *IEEE Signal Processing Magazine*, 1993, 10(1): 8-39.
- [5] D R Hush, J M Salas, B Horne. Error surfaces for multi-layer perceptrons. *Proceedings of IJCNN*, June 1991, I: 759-764.
- [6] R A Jacobs. Increased rates of convergence through learning rate adaptation. *Neural Networks*, 1988, 1(4): 295-307.
- [7] L G Allred, G E Kelly. Supervised learning techniques for backpropagation networks. *Proceedings of IJCNN*, 1990, I: 721-728.
- [8] N B Karayiannis, A N Venetsanopoulos. Fast learning algorithm for neural networks. *IEEE Trans on Circuit and Systems*, 1992, 39(7): 453-474.
- [9] C C Yu, B D Liu. A backpropagation algorithm with adaptive learning rate and momentum coefficient. *Proceedings of IJCNN*, 2002, 2: 1218-1223.
- [10] K Z Huang, H Q Yang, I. King, *et al.* Learning classifiers from imbalanced data based on biased minimax probability machine. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004, 2: 558-563.
- [11] S Visa, A Ralescu. The effect of imbalanced data class distribution on fuzzy classifiers - experimental study. *Proceedings of 14th IEEE International Conference on Fuzzy Systems*, 2005: 749-754.
- [12] G Wu, E Y Chang. KBA: kernel boundary alignment considering imbalanced data distribution. *IEEE Trans on Knowledge and Data Engineering*, June 2005, 17(6): 786-795.
- [13] G R Xuan, W Zhang, P Q Chai. EM algorithm of gaussian mixture model and hidden markov model. *IEEE International Conference on Image Processing*, 2001, 1: 145-148.
- [14] R O Duda, P E Hart, D G Stork. *Pattern classification*. Second edition, New York: John Wiley & Sons, 2000.
- [15] S B Davies, P Mermelstein. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE trans on Acoustics, Speech, and Signal Processing*, 1980, 28(4): 357-366.
- [16] 甄斌, 吴玺宏, 刘志敏等. 语音识别和说话人识别中各倒谱分量的相对重要性. *北京大学学报, 自然科学版*, 2001, 37(3).
- [17] 鲍明, 管鲁阳, 苗浩等. 基于声信号改进 MFCC 特征的地面车辆分类. *声学技术*, 2005, 24: 406-409.