

利用遗传算法改进 SOM 网络初始权值的乐器分类^①

杨松, 于凤芹

(江南大学 物联网工程学院, 无锡 214122)

摘要: 针对 SOM 网络在分类中由于其初始权值的随机性而导致的训练次数过多且易陷入局部最小的问题, 提出了利用遗传算法改进网络初始权值的乐器分类。仿真实验提取 10 种乐器的 12 阶 MFCC 系数, 之后使用遗传算法计算出每种乐器各阶系数的适应度值, 并以此作为网络的初始权值, 之后使用已赋初值的 SOM 网络分类。仿真实验结果表明: 利用遗传算法改进 SOM 网络初始权值的乐器分类方法的分类正确率最高可达到 83.51%。

关键词: 乐器分类; 自组织特征映射网络; 遗传算法

Using Genetic Algorithms to Improve the Initial Weights of SOM Network in the Musical Instrument Classification

YANG Song, YU Feng-Qin

(School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China)

Abstract: For the problem of excessive training and easy to fall into local minimum in SOM network in the classification caused by the randomness of its initial weight, using genetic algorithm to improve network initial weights in instrument classification is proposed. Simulation experiments extract 12-order MFCC coefficients of 10 different kinds of musical instruments. Then use the genetic algorithm to calculate the fitness value of each order in each instrument, and use the fitness value as the network initial weights. Simulation results show that: the way of using genetic algorithms to improve the initial weights of SOM network in the musical instrument classification is effective and the classification accuracy can reach 83.51%.

Key words: musical instrument classification, SOM network, genetic algorithms

乐器分类就是通过分辨不同乐器所产生的声频信号从而将乐器区分开来, 广泛应用于信息检索及模式识别领域^[1]。传统的弦乐, 打击乐, 木管乐, 铜管乐, 键盘乐等 5 类乐器之间的大类区分由于其频率范围的差异性而显得相对简单。文献[2]仅利用频率作为特征对 5 类乐器区分, 分类正确率达到 72%。而同类乐器之间频率差异性相对较小, 因此不能单纯以频率范围作为分类的标准。由于 MFCC (Mel Frequency Cepstral Coefficients, MFCC) 充分模拟了人耳的听觉特性: 人耳听觉特性与频率成非线性对应关系, Mel 频率倒谱系数(MFCC)则是利用它们之间的这种关系, 计算得到的频谱特征, 而将频率范围按 Mel 频率划分成 12 部分

的 MFCC 称为 12 阶 MFCC 系数, 是目前常用的声学特征参数, 适用于乐器分类。文献[3]中使用 MFCC 作为特征对 6 种乐器实现分类, 正确率达到 92.5%。除声学特征外, 分类器性能同样至关重要。自组织特征映射网络(Self-Organizing Feature Map, SOM)是常应用于分类领域的神经网络。相对于需要大量训练样本的高斯混合模型等其他分类器而言, SOM 网络因其具有精确抽取输入模式特征的能力而能在样本数较少的情况下表现出良好的分类效果^[4]。文献[4]中使用 SOM 网络对行驶的车辆进行分类, 达到 90% 以上的分类正确率。但是人耳对各阶 MFCC 系数的敏感程度是随着乐器种类的变化而变化的^[5]。因此直接以 MFCC 为特征,

① 基金项目:国家自然科学基金(61075008)

收稿时间:2011-07-24;收到修改稿时间:2011-09-07

SOM 网络为分类器来实现乐器分类时,由于网络的初始权值为随机赋值,没有考虑到人耳对各阶 MFCC 系数的敏感度,而这种随机性容易导致网络训练次数较多且训练结果易陷入局部最小。遗传算法具有较强的全局搜索能力^[6],通过遗传算法得出每种乐器的各阶 MFCC 系数的适应度值,将其作为 SOM 网络训练的初始权值,可以达到减少训练次数提高分类正确率的目的,同时可以防止训练过程陷入局部最小。

本文实验对象为中提琴、大提琴、沙锤、小鼓、中国箫、笙、小号、大号、钢琴、手风琴等 10 种不同乐器。仿真实验首先提取每种乐器信号的 12 阶 MFCC 系数,之后采用遗传算法计算出每种乐器的各阶系数的适应度值,并将其作为 SOM 网络的初始权值进行分类。仿真实验结果表明,基于遗传算法改进的 SOM 网络分类正确率最高可提高 5.83%。

1 利用遗传算法改进 SOM 网络初始权值的乐器分类的原理

1.1 SOM 神经网络

设 $[x_0(t), x_1(t), \dots, x_{n-1}(t)]^T$ 表示输入节点 i 在时刻 t 的网络输入值。 $w_{ij}(t)$ 为神经元输入节点 i 与输出节点 j 在 t 时刻的连接权值,原始 SOM 网络中 $w_{ij}(0)$ 一般为随机数,同时设置初始节点半径 r , 记为 $R(0)$, 0 代表这是初始时刻的值。之后计算所有输入节点与输出节点 j 之间的距离:

$$d_j = \sum_{i=0}^{n-1} [x_i(t) - w_{ij}(t)x_i(t)]^2 \quad (1)$$

记具有最小距离的节点为 j^* 。接下来调整权值只对具有节点 j^* 及其在节点半径范围内的节点进行,新的权值为:

$$w_{ij}(t+1) = w_{ij}(t) + \eta(t)[x_i(t) - w_{ij}(t)x_i(t)] \quad (2)$$

式中, j 为在 t 时刻节点半径内的节点, i 为所有输入节点, $0 \leq \eta(t) \leq 1$ 为随时间减小的增益项,可以为任意形式的减函数。之后按所有节点的总数的 10% 速度减小节点半径并输入新的数据继续调整网络权值^[7]。

在 SOM 网络中可能存在多个局部最小值点,而 w_{ij} 的初始值的随机性极有可能导致网络陷入局部最小。且由 SOM 网络的实现过程可以看到,公式(2)对权值的调整是网络的核心。因此给 SOM 赋予有效的初始权值会使网络表现出更好的分类能力。

2 遗传算法

2.1 相关 Web 页面获取

遗传算法是一种借鉴生物界的进化规律演化而来的随机化搜索方法。其过程是在一个有一定规模的种群中,比较适应环境的个体,可以获得更多的繁殖机会,不适应环境的个体获得较少的繁殖机会。父代的繁殖通过交叉和变异等遗传操作传给下一代^[6]。

本文利用了遗传算法优秀的全局搜索能力,计算出每种乐器各阶 MFCC 系数的适应度值,对于第 j 种乐器的第 j 阶系数,其适应度函数形式设计如下:

$$Fitness_{ij} = \Gamma_{ij} - 10\Delta_{ij} \quad (3)$$

其中 Δ_{ij} 为第 i 种乐器第 j 阶 MFCC 系数聚类中心到本类所有数据点的欧氏距离之和, Γ_{ij} 为第 i 种乐器第 j 阶 MFCC 系数聚类中心到其他各类聚类中心的欧氏距离之和。适应度函数的计算结果作为 SOM 网络的初始权值。

公式(3)作为适应度函数体现了本类之间数据的密集程度以及类与类之间的离散程度,即本类之间数据越紧密,与其他类数据距离越大,适应度值就越大,说明越有利于分类。因此将其作为 SOM 网络的初始输入权值,避免了原始 SOM 网络由于其权值的随机性导致的训练次数较多,且易陷入局部最小的缺陷,提高了 SOM 网络分类的性能。

词库中的本体词。

3 算法实现

利用遗传算法改进 SOM 网络初始权值的乐器分类的步骤如下:

- 1) 读入乐器信号并使用汉明窗分帧,每帧 256 点,帧移 128 点;
- 2) 提取待分类的 10 种乐器信号的 12 阶 MFCC 系数;
- 3) 对第 i 种乐器的第 j 阶 MFCC 系数利用遗传算法求出其适应度值 $Fitness_{ij}$, 遗传算法的适应度函数如式(3)所示;
- 4) 将第 3) 步得到的适应度值作为 SOM 网络初始输入权值 w_{ij} , 获得改进后的 SOM 网络;
- 5) 使用第 4) 步得到的网络对待分类的 10 种乐器进行分类并记录分类的标记结果;
- 6) 将标记结果送入正确率判断函数中进行统计,得到最终的分正确率。

4 仿真实验及结果分析

仿真实验软件为 Matlab7.0, 实验样本来自于清华大学出版社 1997 年出版的震撼音乐软件。采样率统一为 11025Hz, 采用汉明窗分帧, 每帧 256 点, 帧移 128 点。共有乐器信号样本 200 个, 每种 20 个, 涵盖了所有的 5 大类乐器, 分别是: 中提琴、大提琴、沙锤、小鼓、中国箫、笙、小号、大号、钢琴、手风琴。

图 1 为大号和中提琴的 12 阶 MFCC 系数经过遗传算法计算后的每阶系数的适应度值:

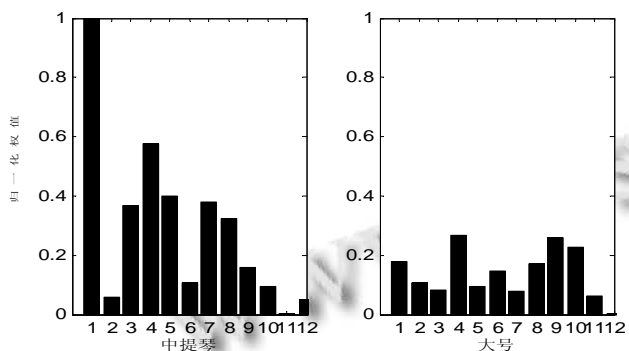


图 1 大号和中提琴的 12 阶 MFCC 系数归一化的每阶权值

由图 1 可以看出, 中提琴与大号在经过遗传算法计算得出的每阶权值不同, 这是有中提琴和大号本身的结构和材料决定的: 中提琴的音色柔美醇厚, 在其频率段上分布相对不均匀, 且在低频段的幅值较大; 而大号音色低沉笨重, 其频率段上的呈现相对均匀的分布, 且幅值变化不大。因此中提琴和大号的权值分布出现如图 1 的情况所示。

图 2 为利用遗传算法改进的 SOM 网络与原始 SOM 网络在相同训练次数下分类正确率的对比图:

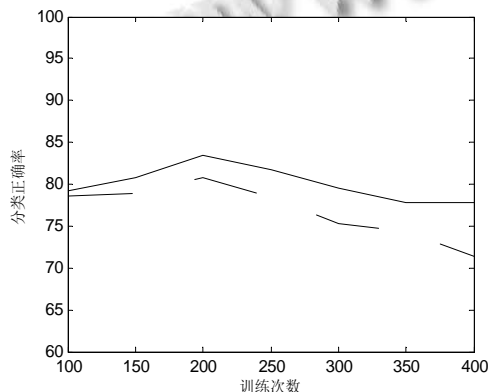


图 2 不同训练次数下的分类正确率

由图 2 可知, 基于遗传算法改进的 SOM 网络的分类正确率要高于原始 SOM 网络, 最高可提高 5.83%。在训练到 200 步时达到最高的分类正确率, 在继续训练会出现过训练的情况。由图 2 可以看出, 该方法是有用的。

5 结语

本文利用通过遗传的全局搜索能力, 将 SOM 网络和 MFCC 参数有机的结合在一起, 虽然付出了一定的时间代价, 但是在分类正确率上得到了提高且以此改善了 SOM 网络与 MFCC 直接结合所导致的训练次数过多且易陷入局部最小的缺陷, 达到了良好的分类效果。但是由于引入了遗传算法, 所以导致整体算法时间增长, 因此今后的研究中应注重开发更快速高效的算法。

参考文献

- 1 Mueller M, Ellis D, Klapuri A. Signal Processing for Music Analysis. IEEE Journal of Selected Topics in Signal Processing, 2011,1(99):1-24.
- 2 Eggink J, Brown G J. A missing feature approach to instrument identification in polyphonic music. 2003 IEEE International Conference on Acoustics, Speech and Signal Processing. Sheffield, 2003: 553-556.
- 3 Jie X, Jian W, Yan YC. SOM-based classification method for moving object in traffic video. 2010 Third International Symposium on Intelligent Information Technology and Security Informatics. Suzhou. 2010: 138-142.
- 4 张奇, 苏鸿根. 基于高斯混合模型的乐器识别方法. 计算机工程, 2004,30(18):133-134.
- 5 Roisim L, Jacqueline W, Michael ON. An Exploration of Genetic Algorithms for Efficient Musical Instrument Identification. Signals and Systems Conference. Ireland, 2010: 1-6.
- 6 Yan T, Xu JT, Jiang WG. A Load Distribution Optimization among Turbine-generators based on PSO-GA. 2011 International Conference on Intelligent Computation Technology and Automation. Nanchang, 2011: 15-18.
- 7 Shuai J, Yi L, Guang ML. SOM-based hand gesture recognition for virtual interactions. 2011 IEEE International Symposium on VR Innovation. 2011: 317-322.