

基于 HMM 与遗传神经网络的改进语音识别系统^①

吴延占

(西安工程大学 电子信息学院, 西安 710048)

摘要: 为了解决语音信号中帧与帧之间的重叠, 提高语音信号的自适应能力, 本文提出基于隐马尔可夫(HMM)与遗传算法神经网络改进的语音识别系统. 该改进方法主要利用小波神经网络对 Mel 频率倒谱系数(MFCC)进行训练, 然后利用 HMM 对语音信号进行时序建模, 计算出语音对 HMM 的输出概率的评分, 结果作为遗传神经网络的输入, 即得语音的分类识别信息. 实验结果表明, 改进的语音识别系统比单纯的 HMM 有更好的噪声鲁棒性, 提高了语音识别系统的性能.

关键词: 隐马尔可夫模型; 神经网络; 语音识别; 遗传算法

Improved Speech Recognition System Based on HMM and Genetic Neural Networks

WU Yan-Zhan

(School of Electronic Information, Xi'an Polytechnic University, Xi'an 710048, China)

Abstract: In order to solve the overlap between frames and improve the self-adaptability of the speech signal, an improved speech recognition system based on hidden Markov model (HMM) and genetic algorithm neural network is proposed in this paper. The major improvement is the adoption of wavelet neural networks in the training of Mel frequency cepstral coefficients (MFCC). And by using HMM models time series of speech signal, the speech's score on the output probability of HMM is calculated. The results will be used as the input of genetic neural network, the information of the speech recognition and classification can then be obtained. The experimental results show that, the improved system has better noise robustness than the pure HMM and the performance of the speech recognition system is also improved.

Key words: hidden markov model; neural network; speech recognition; genetic algorithm

随着语音识别技术的发展, 人们对语音识别的技术要求越来越高, 隐马尔可夫模型(Hidden Markov Model, 简称 HMM), 在语音识别中已经广泛得到应用. 但是, 其自适应能力差, 抗噪性也不是十分理想, 仅靠单一的 HMM 进行语音识别存在诸多困难^[1,2]. 而在广泛应用的人工神经网络(Artificial Neural Network, 简称 ANN)^[3,4], 在自适应、抗噪性方面具有良好的特性, 克服了 HMM 中存在的不足.

反向传播神经网络(BP)虽然已经是神经网络中前向神经网络的核心部分, 并且得到了广泛的应用^[5]. 然而, BP 神经网络也存在诸多缺点, 比如学习收敛速度太慢, 使其只能解决小规模的问题, 也不能保证收敛到全局最小点, 使得训练结果达不到全局最优性.

遗传算法优化后的反向传播神经网络(GA-BP)^[6-9]的混合模型语音识别方法, 有效地保证训练结果的全局最优性, 在语音识别的速度方面也有大幅度的提高, 但在噪声环境下, 遗传算法优化后的反向传播神经网络噪声鲁棒性并不是十分理想, 因此也对语音识别系统的性能带来了影响^[10].

本文运用小波神经网络结构简单、收敛速度快的优点, 对 MFCC 系数进行训练, 从而得到新的 MFCC 系数, 再进行特征提取后作为遗传算法优化后的输入. 然后利用优化后的遗传神经网络获得语音的分类识别信息进行语音识别. 实验结果表明, 基于 HMM 与遗传神经网络改进的语音识别系统进一步提高了语音识别系统的自适应性和噪声鲁棒性.

^① 收稿时间:2015-04-11;收到修改稿时间:2015-05-15

1 隐马尔可夫模型语音识别原理

隐马尔可夫模型是一种利用相关参数来表示,并用于描述随机过程中统计特性的概率模型^[11].它的本质是一种基于统计分布一致性的聚类分析,每个隐含的状态就是一个聚类,对 HMM 进行训练的过程就是寻找每个聚类之间的相关联的过程.它由两部分组成:一个是隐含的马尔可夫链,即为隐含层;另一个是实际的观测量,即为观测层. HMM 基于参数统计理论,利用概率密度函数计算出语音参数对模型的输出概率,找到最佳状态序列以后,用最大后验概率为准则进行识别.语音参数和隐马尔可夫模型关系如下图图 1 所示.

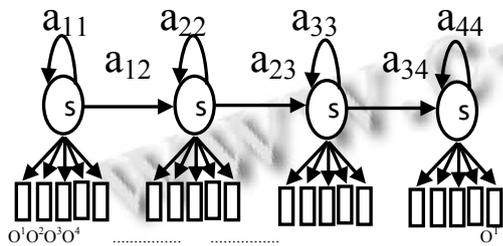


图 1 HMM 和语音参数的关系

对于整个 HMM 模型,利用 $\lambda = \{\pi, A, B\}$ 来描述,其中:

π 表示初始状态概率密度, $\Pi = \{\pi\} = (\pi_1, \pi_2, \dots, \pi_N)$, 其中 $\pi = P(q_1 = S_i), 1 \leq i \leq N$.

A 表示状态转移矩阵, $A = \{a_{ij}\}$, 其中 $a_{ij} = P(q_{t+1} = S_j | q_t = S_i), 1 \leq i, j \leq N$,

① B 表示观测层的概率分布矩阵, $B = \{b_j(q_t)\}, t=1,2,\dots,T\} b_j(q_t) = P(o_t = v_k | q_t = S_j)$, 其中 $1 \leq j \leq N, 1 \leq k \leq M, o_t$ 是在时间 t 的观测符号.

② S 表示隐含状态, $S = (s_1, s_2, \dots, s_N)$ 为 HMM 的状态集, N 表示隐含状态数,模型在时刻 t 的状态用 q_t 表示 ($q_t \in S, 1 \leq t \leq T$), 其中 T 表示观测序列的长度.

如下图图 2 是 HMM 语音识别系统框图,此算法框图模型是基于先训练后识别的基本模式进行的.

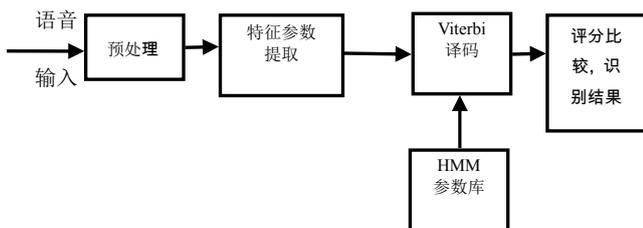


图 2 隐马尔可夫模型语音识别系统流程图

2 基于遗传算法的神经网络优化算法

BP 神经网络是人工神经网络中应用最为广泛的算法,但在广泛运用的同时,也逐渐出现一些不足之处,比如收敛速度慢、不能保证收敛到全局最小点等等.另外,网络结构,初始连接权值与阈值的选取对网络训练带来的影响也非常大,但是又无法准确获得.针对这些特点可以采用遗传算法对神经网络进行优化.

2.1 传统遗传算法

传统遗传算法实现步骤如下:

1) 随机产生一定数量的初始个体,这些随机产生的初始个体总体数目组成一个种群.

2) 用评价函数来评价每个个体的优劣,每个个体的适应程度(称为适应度)作为遗传操作的依据.

3) 从现有的种群中选取一定的个体作为新一代的个体,个体适应程度越高,被选择的机会越大.

4) 对于新生成的种群进行交叉、交异操作.

2.2 用遗传算法优化神经网络权值的学习过程

遗传算法(Genetic Algorithm,GA)是模拟著名天文学家达尔文的遗传选择和生物进化的计算模型,具有很强的宏观搜索能力和良好的全局优化性能^[12,13].因此采取遗传算法与 BP 神经网络相结合,训练时先用遗传算法对神经网络的权值进行寻找,将搜索范围缩小后,再利用 BP 网络来进行精确求解,可以达到全局寻找和快速高效的目的,并且避免局部最小问题.

遗传算法优化神经网络权值的学习过程,我们以 3 层 BP 神经网络为例,实现步骤如下:

1) 初始化种群 p.

2) 计算每一个个体的评价函数,并将其排序,可以按照下式概率值选择网络个体:

$$P_i = \frac{f_i}{\sum_{i=1}^n f_i} \tag{1}$$

式中, f_i 为个体 i 的适配值,可以用误差平方和 E 来衡量,其计算公式分别为:

$$f_i = \frac{1}{E(i)} \tag{2}$$

$$E(i) = \sum_{k=1}^m \sum_{o=1}^q (d_o - y_{o,i})^2 \tag{3}$$

式中 i 表示个体数, $i = 1, 2, \dots, n$. o 表示输出层节点数, $o = 1, 2, \dots, q$. k 表示学习样本数, $k = 1, 2, \dots, m$. y_o 表示实际网络输出, d 表示期望输出.

3) 以交叉概率 P_c 对个体 G_i 和 G_{i+1} 交叉操作,产生新个体,没有交叉操作的个体直接进行复制.

4)利用变异概率 P_m 变化产生 G_j 的新个体 G'_j .

5)把 4)得到新个体插入到种群 P 里面, 并通过计算得到新个体的评价函数.

6)判断算法是否结束. 若符合性能指标 ϵ_a , 则循环结束, 否则转入 3)进入下一轮运算.

算法结束后, 由群体中最优个体解码即可得到优化后的网络连接权值系数.

算法结束后, 由群体中最优个体解码即可得到优化后的网络连接权值系数.

3 基于HMM与遗传神经网络改进的语音识别系统设计

现有的语音识别系统只能处理平稳信号, 而人说话的语言频率不一样使得语音信号是一个准稳态信号, 这时就要把语音划分成若干帧以达到信号稳定的要求. 但这存在的不足之处就是并没有考虑到语音信号的动态特性, 根据神经网络在非线性映射方面比较好的效果, 同时神经网络也具有小波多分辨分析的性能, 从而可以从样本中提取出来新的特征信息.

本文采用基于 HMM 与遗传神经网络改进的语音识别系统, 对输入语音信号进行预处理后, 利用小波神经网络训练 MFCC 系数, 然后根据 HMM 参数库进行 Viterbi 译码, 归一化处理以后作为优化后遗传神经网络的输入, 即将 HMM 中全部状态累计概率作为优化后的遗传神经网络特征输入, 再根据神经网络的非线性映射能力对语音信号识别出所需要的结果. 改进后的语音识别系统流程图如图 3 所示.

基于 HMM 与遗传神经网络改进的语音识别系统算法实现:

1) 预处理. 包括预加重、分帧(把语音信号处理成可看作稳态的信号)、加窗函数, 预加重的滤波器设为:

$$H(z) = 1 - az^{-1} \quad (4)$$

2)小波神经网络训练 MFCC 系数. MFCC 和线性频率的对应关系如下

$$F_{mel} = 2595 \lg(1 + f/700) \quad (5)$$

式中, F_{mel} 是以美尔为单位的感知频率; f 是以 Hz 为单位的实际频率. 利用 MFCC 比较符合人耳听觉特性的基础, 用小波神经网络对 MFCC 进行训练, 从而获得新的 MFCC 系数作为特征提取的输入.

3)快速傅里叶变换与谱线能量的计算. 对每个滤波器取对数以后得到 MFCC 频率系数作为小波神经网络

的输入, 再通过小波神经网络的训练得到新的 MFCC 系数. 最后进行特征参数提取.

4)利用训练好的 HMM 参数库作为 Viterbi 译码的输入, 利用 Viterbi 算法得到相应语音信号的 Viterbi 评分, 在进行归一化处理.

5)由(4)得到的 Viterbi 评分, 利用优化后的遗传神经网络对其进行非线性, 获得分类的识别结果, 即通过上述识别过程共同识别作用, 最终得到语音识别结果.

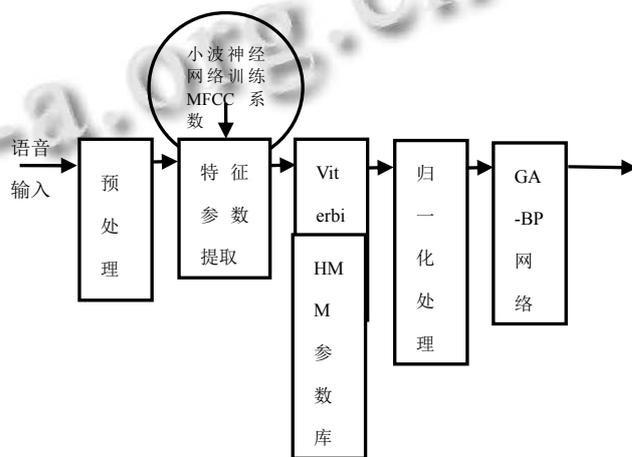


图 3 改进后的语音识别系统原理图

4 仿真实验及结果分析

实验语音文件从十个人中采集, 一个文件有中文数字 1-9 组成. 每个人录了四次, 其中三个用于培训和 一个用于测试. 记录格式的采样频率 8 kHz, 单通道 16 位采样点, 采取的帧长为 256 点. 记录后, 点检测去除 无声段, 其次是预加重. 然后, 语音段划分为 20 个帧, 从每一帧中提取的特征参数. 一个帧有 10 个特征. 因此, 每一个数字都会有 200 特点.

实验训练集采取 100 个不同人员的净语音样本; 在不同的信噪比下取 50 个不同人员的语音信息; 在不同的环境下, 采用的语音数据的信噪比分别为 40、35、30、25、20、15、5 和 0dB. 实验时, 为了验证改进的语音识别系统的有效性, 在 Matlab7.0 上分别对 HMM, HMM 与优化后的遗传神经网络和本文改进后的混合语音识别模型算法做对比. 实验分为两次实验过程, 第一次在加性高斯噪声下; 第二次在学校餐厅学生就餐时人声为噪声背景. 实验分别得出语音在加性高斯噪声下识别率如表 1 和图 4; 在学校餐厅时实验结果如

表 2 和图 5 所示.

表 1 数字识别率的比较(单位: %)

识别语音(信噪比 dB)	40 35 30 25 20 15 5 0
HMM	95 90 83 69 50 34 19 13
HMM 与优化神经网络	96 93 89 88 82 74 75 65
混合语音识别模型	96 95 90 89 88 82 75 65

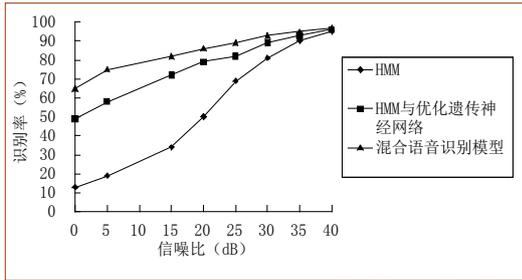


图 4 加性高斯噪声下语音识别性能比较

表 2 数字识别率的比较(单位: %)

识别语音(信噪比 dB)	40 35 30 25 20 15 5 0
HMM	96 92 88 79 50 34 23 16
HMM 与优化神经网络	98 94 92 84 79 74 61 49
混合语音识别模型	98 95 93 90 88 82 75 65

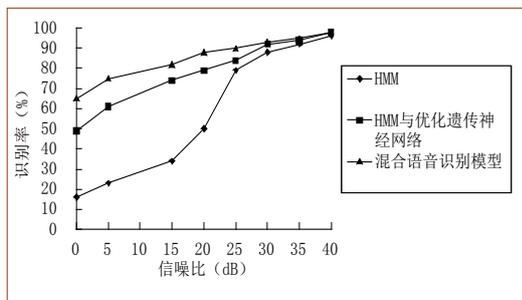


图 5 学校餐厅噪声下语音识别性能比较

由表 1 和表 2 中的实验数据可以看出, 改进以后的混合算法相比单一的 HMM 和优化的遗传神经网络相比具有更好的识别效果, 尤其对那些容易混淆的词语识别率也有所提高. 对于识别系统的信噪鲁棒性方面也有了明显的改变, 提高了语音识别系统的自适应能力. 神经网络的收敛速度是衡量语音识别的一个重要标准, 因此遗传神经优化算法与 BP 算法收敛速度作了如图 6、图 7 的比较, 实验结果表明, 优化算法收敛速度得到明显改善.

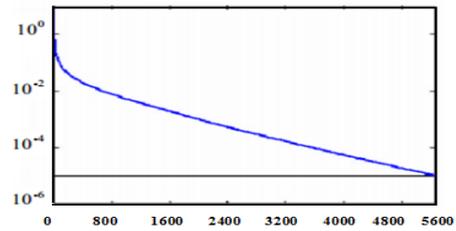


图 6 传统 BP 算法收敛速度

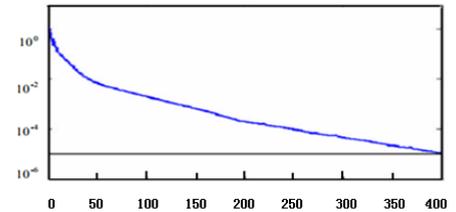


图 7 遗传网络优化算法收敛速度

5 结论

本文提出了基于隐马尔可夫 HMM 与遗传算法的神经网络改进的语音识别系统, 在隐马尔可夫模型与遗传网络算法优化的基础上, 引入小波神经网络训练 MFCC 系数, 充分利用了隐马尔可夫模型强大的时间建模能力与遗传算法优化后的收敛速度快、分类识别能力强的优点. 小波神经网络训练出的 MFCC 新系数应用到优化后的神经网络具有更高的识别率, 提高了识别系统的自适应能力.

参考文献

- 1 吕军,曹效英,徐宝国.基于语音识别的汉语发音自动评分系统的设计与实现.计算机工程与设计,2007,28(5):1232-1235.
- 2 郭超,张雪英,刘晓峰.支持向量机在低信噪比语音识别中的应用.计算机工程与应用,2013,49(5): 213-215.
- 3 Seman N, Bakar Z A, Bakar NA. The optimization of Artificial Neural Networks connection weights using genetic algorithms for isolated spoken Malay parliamentary speeches. 2010 International Conference on Computer and Information Application (ICCIA). IEEE. 2010. 162-166.
- 4 Lan ML, Pan ST, Lai CC. Using genetic algorithm to improve the performance of speech recognition based on artificial neural network. First International Conference on Innovative Computing, Information and Control(ICICIC'06). IEEE. 2006, 2. 527-530.
- 5 王晓东,薛宏智,马盈仓.基于自适应遗传算法的神经网络字符识别.西安工程大学学报,2008,22(2): 210-213.
- 6 钟林,刘润生.新神经网络结构及其在数码语音识别中的应

- 用.清华大学学报(自然科学版),2000,40(3):104-108.
- 7 包亚萍,郑骏,武晓光.基于 HMM 和遗传神经网络的语音识别系统.计算机工程与科学,2011,33(4):139-144.
- 8 冯宏伟,薛蕾.基于 HMM 和新型前馈型神经网络的语音识别研究.计算机工程与设计,2010,(24):5324-5327.
- 9 肖勇,覃爱娜.改进的 HMM 和小波神经网络的抗噪语音识别.计算机工程与应用,2010,(22):162-164.
- 10 Pan ST, Wu CH, Lai CC. The application of improved genetic algorithm on the training of neural network for speech recognition. Second International Conference on Innovative Computing, Information and Control(ICICIC'07). IEEE, 2007. 168-168.
- 11 Aggarwal RK, Dave M. Application of genetically optimized neural networks for Hindi speech recognition system. 2011 World Congress on Information and Communication Technologies (WICT). IEEE. 2011. 512-517.
- 12 An M, Yu Z, Guo J, et al. The teaching experiment of speech recognition based on HMM. The 26th Chinese Control and Decision Conference (2014 CCDC). IEEE. 2014. 2416-2420.
- 13 Silva WLS, de Oliveira Serra GL. A novel intelligent system for speech recognition. International Joint Conference on Neural Networks (IJCNN). IEEE. 2014. 3599-3604.