

# 基于遗传小波神经网络的语音情感识别

韩志艳, 伦淑娴, 王 健

(渤海大学 工学院, 辽宁 锦州 121000)

**摘要:**情感识别是人机交互领域中必须解决的关键问题。针对语音情感的识别问题,文中把遗传算法和小波神经网络算法相结合,即利用遗传算法具有的高度并行、随机、自适应搜索性能来选取初值进行训练,用小波神经网络来完成给定精度的学习。这样在解决复杂和非线性问题时,具有明显的优势。文中主要研究了四种基本的人类情感:喜悦、愤怒、悲伤和恐惧。并与BP算法和小波神经网络算法进行了比较,实验结果表明,该模型不但能够提高情感识别的正确率,缩短系统识别时间,而且为算法的实用性奠定了基础。

**关键词:**情感识别;神经网络;遗传算法;小波分析

中图分类号:TP391.9

文献标识码:A

文章编号:1673-629X(2013)01-0075-04

doi:10.3969/j.issn.1673-629X.2013.01.019

## Speech Emotion Recognition Based on Genetic Wavelet Neural Network

HAN Zhi-yan, LUN Shu-xian, WANG Jian

(College of Engineering, Bohai University, Jinzhou 121000, China)

**Abstract:** Emotion recognition is the key issue that must be solved in human-computer interaction. Aiming at the recognition problem of speech emotion, combined genetic algorithm and wavelet neural network algorithm, utilize the performance of height parallel, random and adaptive search to select initial values, and using wavelet neural network to finish the learning. This has obvious advantages of solving complex and nonlinear problem. Four basic human emotions including joy, anger, sadness and fear were studied, and compared with back propagation neural network (BPNN) and wavelet neural network. The experimental results indicate that this method effectively improves the correct rate of emotion recognition, shortens the system recognition time, and lays the foundation for algorithm practicality.

**Key words:** emotion recognition; neural network; genetic algorithm; wavelet transform

## 0 引言

随着计算机技术的发展,如何让计算机识别出人的情感,使之能像人一样进行自然生动的交互,是当前模式识别领域的一个新的研究热点。语音作为人类交流的最重要媒介之一,携带着丰富的情感信息。因此,可以考虑从语音信号中提取情感相关的特征,然后通过计算机来识别出说话者的情感状态。该项研究在机器人技术、新型人机交互技术等领域具有重要的应用价值<sup>[1~6]</sup>。

从20世纪90年代中后期开始,各国开始重视情感信息处理技术的研究,尤其是在情感特征的识别方面,取得了许多研究成果。比如基于Mahalanobis距离

分类方法、矢量量化法、神经网络法(ANN)、混合高斯模型法等。由于语音信号本身复杂性较高,由于人们对语音学的了解还不够深入,加上语音信号本身复杂性较高,导致目前语音情感识别方法在性能上难以令人满意<sup>[7]</sup>。神经网络是一种先进的智能识别方法,具有非线性映射、高度并行处理、自学习功能和较高的泛化能力。但是由于神经网络的学习效果过多的依赖于网络的初始参数,如果它们设置不好,神经网络易陷入局部极小点,同时收敛速度减慢甚至不收敛。在无先验知识的情况下,对神经网络的寻优过程主要通过反复修改各个参数值来使ANN具有优良的分类性能。但是由于神经网络的参数较多、取值范围较广,手动调整参数的方法难以奏效。而遗传算法的出现为解决此类复杂和非线性问题提供了较好的手段<sup>[8~11]</sup>。因此,文中结合遗传算法和神经网络算法的优点,先利用遗传算法在参数优化方面的优势来选取初值完成训练,再利用小波神经网络算法完成给定精度的学习,即形成一种遗传小波神经网络学习算法。它具有响应速度

收稿日期:2012-05-04;修回日期:2012-08-10

基金项目:国家自然科学基金资助项目(60974071);辽宁省自然科学基金(201102005)

作者简介:韩志艳(1982-),女,内蒙古赤峰人,讲师,博士,研究方向为语音情感信息处理。

快,识别精度高等优点。

## 1 小波神经网络的设计

相对于基于 Sigmoid 函数的 BP 网络,小波神经网络具有收敛速度快、收敛精度可控和结构可设计等优点<sup>[12-14]</sup>。小波神经网络一般采用单隐层结构,为了便于理解信号小波表达的物理意义,输出层激励函数一般采用线性函数,但文中为使网络具有更强的逼近能力,采用 3 层网络结构,并使用 Sigmoid 函数代替线性函数。其具体的过程如下。

### 1.1 小波基函数的构造

文中选用 morlet 小波作为母小波,通过将 morlet 小波进行伸缩和平移操作,得到一组连续小波基函数,即:

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \Psi\left(\frac{t-b}{a}\right), a, b \in R; a \neq 0 \quad (1)$$

其中:  $a$  为伸缩因子,  $b$  为平移因子, morlet 小波为:

$$\Psi(t) = \cos(rt) \exp(-t^2/2) \quad (2)$$

其中  $r$  取为 1.75。

### 1.2 小波神经网络模型设计

文中采用三层小波神经网络,其中由  $L$  个神经元构成输入层,并采用线性变换作为激励函数;由  $M$  个神经元构成隐含层,并采用小波函数作为激励函数;由  $S$  个神经元构成输出层,并采用 Sigmoid 函数作为激励函数。则所构造的小波网络模型为:

$$y_i^n = f\left[\sum_{j=0}^M v_{ji} \Psi\left[\frac{\sum_{i=0}^L w_{ij} x_i^n - b_j}{a_j}\right]\right] \quad (3)$$

其中:  $w_{ij}$  和  $v_{ji}$  分别表示输入层与隐含层及隐含层与输出层之间的连接权值;  $w_{0j}$  和  $v_{0i}$  分别表示隐含层和输出层的阈值;  $x_i^n$  ( $i = 1, 2, \dots, L, n = 1, 2, \dots, N$ ,  $N$  为样本总数)表示第  $n$  个样本的输入值;  $y_i^n$  ( $t = 1, 2, \dots, S$ ) 为网络的输出值,对应的目标输出为  $d_i^n$ ;  $a_j$  为第  $j$  个隐含层节点的伸缩系数;  $b_j$  为第  $j$  个隐含层节点的平移系数。

### 1.3 误差能量函数的设计

误差能量函数定义为实际输出与期望输出之间的差值,具体可用下式表示:

$$E = \frac{1}{2} \sum_{t=1}^S (y_t^n - d_t^n)^2 \quad (4)$$

### 1.4 误差偏导数的计算

通过求出第  $n$  个样本时误差能量函数  $E$  对每个网络参数的偏导数,得到满足  $E$  最小的网络参数。

$$\frac{\partial E}{\partial v_{ji}} = (y_i^n - d_i^n) f(z) (1 - f(z)) \Psi(T) \quad (5)$$

$$\frac{\partial E}{\partial w_{ij}} = \sum_{t=1}^S (y_t^n - d_t^n) f(z) (1 - f(z)) v_{ji} \frac{\partial \Psi(T)}{\partial T} \frac{x_i^n}{a_j} \quad (6)$$

$$\frac{\partial E}{\partial a_j} = \sum_{t=1}^S (y_t^n - d_t^n) f(z) (1 - f(z)) v_{ji} \frac{\partial \Psi(T)}{\partial T} \frac{(-T)}{a_j} \quad (7)$$

$$\frac{\partial E}{\partial b_j} = \sum_{t=1}^S (y_t^n - d_t^n) f(z) (1 - f(z)) v_{ji} \frac{\partial \Psi(T)}{\partial T} \frac{(-1)}{a_j} \quad (8)$$

其中:  $T$  和  $z$  分别表示隐含层及输出层神经元的输入值。

$$T = \frac{\sum_{i=0}^L w_{ij} x_i^n - b_j}{a_j} \quad (9)$$

$$z = \sum_{j=0}^M v_{ji} \Psi\left[\frac{\sum_{i=0}^L w_{ij} x_i^n - b_j}{a_j}\right] \quad (10)$$

$$\frac{\partial \Psi(t)}{\partial t} = -r \sin rt \exp(-t^2/2) - t \cos rt \exp(-t^2/2) \quad (11)$$

输出层神经元激励函数为:

$$f(t) = \frac{1}{1 + \exp(-t)} \quad (12)$$

$$\frac{\partial f(t)}{\partial t} = \frac{\exp(-t)}{(1 + \exp(-t))^2} = f(t) (1 - f(t)) \quad (13)$$

### 1.5 网络参数的调整

参考常规的 BP 算法,网络各参数的调整通过引入动量因子  $\mu$  来实现,具体如下:

$$v_{ji}(N+1) = v_{ji}(N) - \eta \frac{\partial E}{\partial v_{ji}} + \mu \Delta v_{ji}(N) \quad (14)$$

$$w_{ij}(N+1) = w_{ij}(N) - \eta \frac{\partial E}{\partial w_{ij}} + \mu \Delta w_{ij}(N) \quad (15)$$

$$a_j(N+1) = a_j(N) - \eta \frac{\partial E}{\partial a_j} + \mu \Delta a_j(N) \quad (16)$$

$$b_j(N+1) = b_j(N) - \eta \frac{\partial E}{\partial b_j} + \mu \Delta b_j(N) \quad (17)$$

其中:  $\Delta v_{ji}(N)$  和  $\Delta v_{ji}(N+1)$  分别为上一次和本次调整后的权值差值;  $\mu$  为动态参量;  $\eta$  为学习率,它的选取很重要,如果选的太小容易陷入局部极小值点,如果选的太大又可能产生震荡而导致不收敛。

## 2 优化方法

### 2.1 基本思想

首先用遗传算法对网络初始参数进行优化,在解空间中定位出较好的搜索空间,然后用小波神经网络在这些小的解空间中再次寻优,直到搜索出最优解。

### 2.2 优化算法具体实施步骤

第一步:随机产生一组实值串,并将网络的初始参

数作为一组染色体,而且每一个个体编码由七部分构成,如图1所示。

由于文中自变量数目较多,如果采用二进制编码,就会导致染色体长度偏长、搜索空间偏大以及效率较低等问题的出现。所以文中采用实值编码,这样不仅解决了上述问题,还避免了编码中一些附加问题的出现。

$M$	$w_{ij}$	$v_{\mu}$	$a_j$	$b_j$	$\eta$	$\mu$
-----	----------	-----------	-------	-------	--------	-------

图1 个体编码

第二步:解码实值串中的隐含层节点数  $M$ ,生成相应的神经网络结构;解码实值串中的其余部分,生成神经网络的初始权值(含阈値)、伸缩因子、平移因子、学习率和动量因子。

第三步:正向运行网络,根据下式计算适应度值:

$$f = \frac{1}{R + 1} \quad (18)$$

它反映的是网络的实际输出值与期望输出值之间的误差平方和,是衡量网络性能的主要指标,其值越小表示该网络性能越好,其中:  $R = \frac{1}{N} \sum_{n=1}^N \sum_{t=1}^S (y_t^n - d_t^n)^2$ 。

第四步:通过选择操作、交叉操作及变异操作等操作来产生下一代种群,生成下一代网络。

具体步骤如下:

●选择操作:文中采用轮盘赌选择法来进行选择操作,其个体被选中的概率如式(19)所示,并和式(18)中的  $f$  成正比例关系。

$$P_{si} = \frac{f[i]}{\sum_{j=1}^{pop\_size} f[j]} = \frac{f[i]}{\sum f} \quad (19)$$

其中:  $pop\_size$  为种群规模,  $f[i]$  为个体  $i$  的适应度值,  $\sum f$  为当前代种群的适应度和。

其具体的选择操作步骤如下:

1) 计算当代种群的适应度总和;

2) 产生一个随机数  $r_0 \in (0, 1)$ , 取  $r = r_0 * \sum f$ ,

并把满足式  $\sum_{i=0}^{j-1} f[i] < r$  且  $\sum_{i=0}^j f[i] \geq r$  的第  $j$  个个体选中进入交叉操作,重复  $pop\_size$  次操作,最终得到  $pop\_size$  个个体,一起进入交叉操作。

●交叉操作:交叉操作又叫基因重组,就是把两个父体部分结构加以替换,生成新的个体的操作。

其具体步骤如下:

1) 指定参数  $P_c$  作为交叉概率;

2) 生成随机数  $r \in [0, 1]$ , 如果  $r < P_c$ , 则选择  $v_i$  作为一个父代。如果用  $v'_1, v'_2, v'_3, \dots$  表示选择的父代,并把它们随机分成对  $(v'_1, v'_2), (v'_3, v'_4), \dots$  进行

交叉。以  $(v'_1, v'_2)$  为例进行的交叉操作如下:

$$\begin{cases} x = c \times v'_1 + (1 - c) \times v'_2 \\ y = (1 - c) \times v'_1 + c \times v'_2 \end{cases} \quad (20)$$

其中:  $c$  为  $(0, 1)$  之间的随机数,  $x$  和  $y$  均为后代,  $x$  和  $y$  的替换对象分别为两个父代  $v'_1$  和  $v'_2$ ;

3) 从  $i = 1$  到  $pop\_size$  重复步骤2)。

●变异操作:为防止产生空类,在变异操作前,先要为被变异的个体作备份。

其具体步骤如下:

1) 设定变异概率为  $P_m$ ;

2) 生成一个随机数  $r \in [0, 1]$ , 如果  $r < P_m$ , 则选择  $v_i$  作为一个父代。如果用  $v'_1, v'_2, v'_3, \dots$  表示上面选择的父代,以  $v'_1$  为例的变异操作为:

$$z = v'_1 + c \quad (21)$$

其中:  $v'_1$  为父代个体的变异基因,  $z$  为变异后子代个体的基因;

3) 从  $i = 1$  到  $pop\_size$  重复步骤2)。

第五步:重复第二步~第四步,直到达到进化代数  $g\_max$  或  $R \leq R_{max}$ , 这里  $R_{max}$  是遗传算法所要达到的性能指标。然后解码最终种群中的个体,从而得到经过遗传算法优化后的值,并将其作为小波神经网络使用的参数。

第六步:双向运行网络,不断调节网络参数,直到  $R \leq R_{min}$ , 这里  $R_{min}$  是小波神经网络要求达到的性能指标,保存各参数值,学习过程终止。

### 3 实验结果与分析

为了验证算法的优越性,将文中算法与 BP 算法和小波神经网络算法进行了比较分析。

实验选用 10 个汉语语句作为情感分析用的语音资料,并由 20 位(10 男 10 女)善于表演的说话者用喜悦、愤怒、悲伤和恐惧 4 种情感以及中性情感对每一个句子各发 5 遍,共采集到 1000 句实验数据,其中 800 句情感句,并将其分成两个集合,其中 400 句(10 句×4 种×10 人)为训练集,剩下的 400 句(10 句×4 种×10 人)为测试集。录音环境为安静的实验室,采样频率为 16kHz。

为了尽可能地利用语音信号中所包含的有关情感方面的信息,选取了语句发音持续时间与相应的平静语句持续时间的比值、基音频率平均值、基音频率最大值、以及与相应平静语句的基音频率平均值的差值、与相应平静语句的基音频率最大值的差值、振幅平均能量、振幅能量的动态范围、以及与相应平静语句的振幅平均能量的差值、与相应平静语句的振幅能量动态范围的差值、第一共振峰频率的平均值及前 3 个共振峰

峰值的平均值这 11 个情感特征作为情感识别用的特征参数。

由表 1,2,3 可以看出,采用 BP 神经网络进行识别平均识别率为 79%,采用小波神经网络进行识别平均识别率为 87.25%,而采用文中方法平均识别率为 91%,较其余两种方法大大提高了情感识别的准确率。

表 1 基于 BP 神经网络的识别结果

情感类别	识别结果				平均识别正确率%
	喜悦	愤怒	悲伤	恐惧	
喜悦	80	17	0	3	80
愤怒	14	75	6	5	75
悲伤	5	2	83	10	83
恐惧	2	4	16	78	78

表 2 基于小波神经网络的识别结果

情感类别	识别结果				平均识别正确率%
	喜悦	愤怒	悲伤	恐惧	
喜悦	87	8	1	4	87
愤怒	6	87	4	3	87
悲伤	4	1	92	3	92
恐惧	7	6	4	83	83

表 3 基于遗传小波神经网络的识别结果

情感类别	识别结果				平均识别正确率%
	喜悦	愤怒	悲伤	恐惧	
喜悦	92	4	2	2	92
愤怒	10	89	1	0	89
悲伤	2	1	96	1	96
恐惧	3	4	6	87	87

4 结束语

实验结果表明,文中提出的语音情感识别模型切合了情感语音研究的需要,能够很好地模拟人们的情感感知机制,并获得了很好的识别结果。但是文中只是针对特定文本的语音情感识别研究,所以要提高识别建模的自适应性,基于非特定文本的语音情感识别

将成为下一步研究的重要方向。

参考文献:

[1] 张石清,赵知劲. 噪声背景下的语音情感识别[J]. 西南交通大学学报,2009,44(3):442-447.

[2] 章国宝,宋清华,费树岷,等. 语音情感识别研究[J]. 计算机技术与发展,2009,19(1):92-96.

[3] 赵力,王治平,卢韦,等. 全局和时序结构特征并用的语音信号情感特征识别方法[J]. 自动化学报,2004,30(3):423-429.

[4] Kim E H,Hyun K H,Kim S H,et al. Improved Emotion Recognition with a Novel Speaker-independent Feature[J]. IEEE Transactions on Mechatronics,2009,14(3):317-325.

[5] 石瑛,胡学钢,方磊. 基于决策树的多特征语音情感识别[J]. 计算机技术与发展,2009,19(1):147-149.

[6] Cowie R,Cornelius R R. Describing the Emotional States That Are Expressed in Speech[J]. Speech Communication,2003,40(1-2):5-32.

[7] 赵力. 语音信号处理[M]. 北京:机械工业出版社,2009:261-272.

[8] 李影,徐涛,邢伟. 基于进化遗传算法的神经网络优化[J]. 长春理工大学学报,2006,29(3):48-50.

[9] Pan S T. Evolutionary Computation on Programmable Robust IIR Filter Pole-placement Design[J]. IEEE Transactions on Instrumentation and Measurement,2011,60(4):1469-1479.

[10] 李敏强,寇纪淞,林丹. 遗传算法的基本理论与应用[M]. 北京:科学出版社,2002.

[11] 梁化楼,戴贵亮. 人工神经网络与遗传算法的结合:进展及展望[J]. 电子学报,1995,23(10):194-199.

[12] 林遂芳,潘永湘,孙旭霞. 基于 HMM 和小波网络模型的抗噪语音识别方法[J]. 系统仿真学报,2005,17(7):1720-1723.

[13] Yoshihiro Y,Nikiforuk P N. A New Supervised Learning Algorithm for Multilayered and Inter-connected Neural Networks[J]. IEEE Transactions on Speech and Neural Networks,2000,11(1):36-46.

[14] Zhang Q H,Benveniste A. Wavelet Networks[J]. IEEE Transactions on Speech and Neural Networks,1992,3(2):889-898.

(上接第 74 页)

[8] 崔晨旸,石教英,王东辉. 几何特征映射下的 3 维模型相似性匹配研究[J]. 中国图形图像学报,2006,11(5):664-665.

[9] 杨育彬,林琿,朱庆. 基于内容的三维模型检索综述[J]. 计算机学报,2004,27(10):1297-1310.

[10] 柳伟. 三维模型特征提取与检索[D]. 上海:上海交通大

学,2008.

[11] Ohbuch R,Hata Y. Combining multiresolution shape descriptors for 3D model retrieval[C]//Proc. of WSCG 2006. Plzen, Czech Republic:[s. n.],2006.

[12] de Berg M,Cheong O,van Kreveld M. 计算几何-算法与应用[M]. 北京:清华大学出版社,2005.

# 基于遗传小波神经网络的语音情感识别

作者：[韩志艳](#)，[伦淑娴](#)，[王健](#)  
作者单位：[渤海大学 工学院, 辽宁 锦州 121000](#)  
刊名：[计算机技术与发展](#)  
英文刊名：[Computer Technology and Development](#)  
年，卷(期)：2013(1)

本文链接：[http://d.g.wanfangdata.com.cn/Periodical\\_wjfz201301021.aspx](http://d.g.wanfangdata.com.cn/Periodical_wjfz201301021.aspx)