文章编号:1673-1549(2011)05-0552-04

基于 RBF 神经网络的语音情感识别

张海燕, 唐建芳

(四川理工学院理学院, 四川　自贡　643000)

摘　要:介绍了径向基函数神经网络的原理、训练算法,并建立了 RBF 神经网络的语音情感识别的模型。在实验中比较了 BP 神经网络与 RBF 神经网络分别用于语音情感识别识别率,RBF 神经网络的平均识别率高于 BP 神经网络 3%。结果表明,基于 RBF 神经网络的语音情感识别方法的有效性。

关键词:径向基函数;RBF 神经网络;语音情感识别

中图分类号:TP391　　**文献标识码**:A

引　言

语音作为人类交流的重要媒介传递了丰富的情感信息,它除了包含实际发音内容外,还包含着说话人的喜、怒、哀、乐等丰富的情感信息。语音情感识别有着广泛的应用,如远程教学[1]、电子机器宠物[1-2]、辅助测谎[3]、司机在驾驶过程中的情感分析[4]以及临床医学[5]等。随着智能化人机交互计算机的高速发展,近年来,各个领域的研究者都十分关注如何从语音中自动识别说话人的情感状态,并使计算机作出更有针对性和人性化的响应。

目前语音情感识别的方法有很多,如隐马可夫模型(HMM)[6]、混合高斯模型(GMM)[7]、人工神经网络(ANN)、支持向量机(SVM)、线性判别分类器(LDC)、K 近邻法(KNN)和最大似然贝叶斯分类等,并取得了一定的效果。但由于研究对象(语种)各异,语料数据库没有统一的标准,造成识别结果相差悬殊,可比性差。总之,整个语音情感信息处理领域还处在一个较低的水平,发展前景还很广。而人类的情感具有很强的复杂性和不确定性的信息。LDC、KNN 和 SVM 等方法常用于确定性高的模型。而神经网络是典型的非确定模型,它具有 I/O 非线性映射特性、强大的普化能力及自学习、自组织、自适应能力,这决定了它处理这类不确定的、非线性映射问题具有独到的优势,它能探测并提取人类或者其它分类技术不能探测到的规律和趋势。在各种人工神经网络模型中,在模式识别中应用最多和最成功的是多层前馈网络,其中又以 BP 网络为代表。但 BP 网络容易陷入局部极小点的缺点容易导致识别错误。Hopfield 网络和随机型网络更适于解决联想记忆和优化计算问题,自组织竞争网络更适于训练期间不需指定期望输出的无导师学习式的模式自动聚类。而径向基函数(RBF)神经网络结构简单,既是最适于模式识别的多层前馈网,同时又避免了局部极小点问题,且学习速度也比 BP 网络大为加快,所以尝试选择 RBF 网络识别语音信号中的研究者普遍认同的高兴、愤怒、惊讶、害怕、悲伤、平静六种语音情感类型。

1　基于 RBF 网络的语音情感识别

1.1　基于 RBF 网络的语音识别原理

RBF 神经网络的基本思想[8]:把径向基函数作为隐单元的“基”,构成隐含层空间,隐含层对输入矢量进行变换,将低维的模式输入数据变换到高维空间内,使得在低维空间内的线性不可分问题在高维空间内线性可分。用于 RBF 网络的语音识别首先在数据库中每条语音信号提取相应的声学特征(韵律特征和音质特征),然后进行特征化简,将化简后的有效特征参数组成向量作为神经网络的输入。在训练时将训练样本对应的语音情感种类编号作为网络的期望输出,完成对网络的训

收稿日期:2011-08-25

基金项目:四川理工学院科研项目(2009XJKYL005)

作者简介:张海燕(1977-),女,四川乐至人,讲师,硕士,主要从事应用数学,粗集理论和人工智能方面的研究,(E-mail)zhang_petrel@163.com.cn

练。在识别阶段,测出一个未知情感的特征参数送入到神经网络,其对应的输出编号即表示未知语音情感所属的种类,从而完成未知情感的种类识别。其识别原理如图 1 所示。

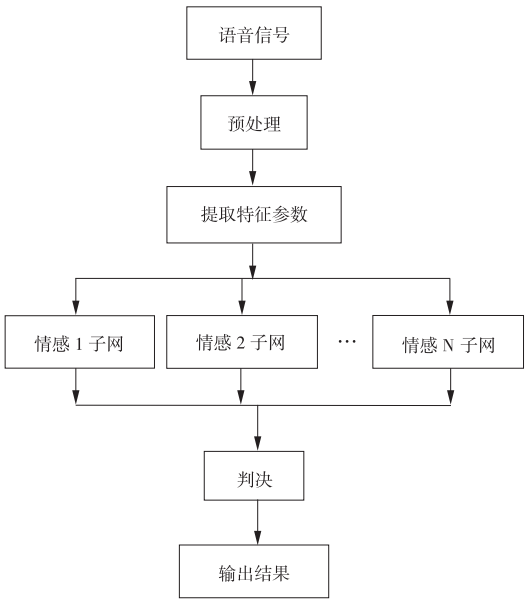

图 1　神经网络识别语音情感系统原理图

1.2　RBF 神经网络结构的确定[9]

神经网络用于模式识别一般有单输出方式、并列单输出方式和多输出方式。单输出方式多用于两种模式的识别分类,输出只取 0 和 1;当要识别多种模式时,可让输出取 0 ~ (N - 1)之间的值(N 为模式种类)。但这种方式中各种类别间耦合严重,影响了识别的效果。并列单输出方式为多个单输出方式神经网络的并列,网络个数等于模式种类数,每个网络只完成识别两类分类,即判断样本是否属于某个类别。这样可克服类别间的耦合,但连接权太多。多输出方式可以有 M 个输出节点,用它们的编码来代表 N 个类别,这样可以折衷前两种方式的优缺点。所以网络采用多输出方式,并用二进制方式对输出进行编码。按二进制编码原则,RBF 网络共需 6 个输出,设为 y1 ~ y6,可提供 16 种编码,六大类语音情感依次对应编码 0000H ~ 0101H,编码 0110H ~ 1111H 不用。隐层节点数与样本类别数和样本差异程度有关。隐层节点数太少必然使不同模式的输入样本归入同一节点代表的区域,造成模式划分的模糊;隐节点数太多则增大网络规模和数据处理量,同时给输出层训练时全局误差极小的获得带来困难。将隐层节点数初始设为 12,训练过程中根据训练的均方误差自动增减。RBF 函数选用高斯基函数。网络输入节点数选为特征参数的数目 12。按上述思想设计的 RBF 神经网络结构如图 2 所示。

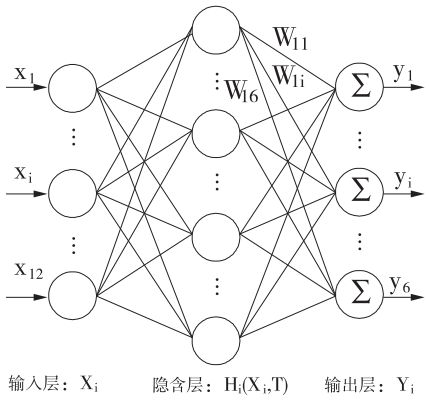

图 2　用于语音情感识别的 RBF 网络结构

1.3　RBF 网络训练算法[10]

在 RBF 中心初始化时,为了使聚类中心分布更合理化,依次从每一类语音情感类型样本中任意选出一个样本作为 RBF 中心初始值。若由于隐节点数大于种类数而使这样选出的中心数不够时,则再循环依次从每一类样本中任意选出另一样本作为聚类中心初值,直到初始的 RBF 中心数选足为止。

隐层节点的训练采用无监督的 K - 均值聚类法来完成,使聚类满足聚类集合中每一样本点到该类中心的距离平方和最小。输出层的训练采用有导师的线性最小二乘(LMS)学习算法来实现,调整输出层权矩阵 W,使网络的实际输出矢量所构成矩阵 Y 与对应的期望输出矢量所构成矩阵 Y_d 间的均方误差 $E = \|Y - Y_d\|^2$ 达到最小。这里 $Y = WH$, $H = \{h_{ij}\}$ 为隐层输出矢量构成的矩阵,

$$h_{ij} = \exp\left(-\frac{\|X_i - T_j\|}{2\sigma_j^2}\right)$$

$(i = 1,2,\cdots,N; j = 1,2,\cdots,M)$

N 为训练样本数, M 为隐节点数。连接权矩阵可由 $W = H^+ Y_d$ 获得, H^+ 为的伪逆,可由奇异值分解求得。

完成了隐层和输出层的训练后,再根据训练精度要求决定是否增加隐层节点。训练算法的实现如下:

(1)初始化聚类中心 $T_j (j = 1,2,\cdots,M)$ 。

(2)将所有训练样本按最邻近原则聚类,即按 $\theta_j = \{X_i \mid \|X_i - T_j\| \leqslant \|X_i - T_k\|\}$, $(k = 1,2,\cdots,M$ 且 $k \neq j)$ 的原则,将 X_i 归入第 j 个聚类 θ_j 中。

(3)计算 θ_j 中训练样本的平均值,即新的聚类中心:

$$T_j = \frac{1}{R_j}\sum_{X_i \in \theta_j} X_i$$

(4)判断本次 T_j 与前次 T_j 是否相等。若不相等,则转至(2)。

(5)按下式计算 RBF 的均方差(高斯半径)[9]:

$$\sigma_j^2 = \frac{1}{R_j}\sum_{X_i\in\theta_j}(X_i - T_j)^T(X_i - T_j)$$

由此得隐层高斯核函数为:

$$H_j(\|X - T_j\|^2) = \exp(-\frac{\|X - T_j\|^2}{2\sigma_j^2})$$

(6)根据 $W = H^+ Y_d$ 求输出层连接权。

(7)判断均方误差 $\|Y_d - WG\|^2$ 达到要求否。若达到,则完成网络学习,否则增加一隐层节点($M = M + 1$),转至(1)。

2 实验结果与分析

2.1 实验语料与特征提取

实验所用的 1 080 个样本数据库是按照文献[11]的方法和选取来自中国科学院 2005 年建立的汉语情感语料库的部分语料录制建立的。该数据库包括两名女性和三名男性(20 - 25 岁的在校本科生)的情感语句,包括高兴、愤怒、悲伤、害怕、惊讶、平静六种情感状态,每种情感以 50 句语义为中性的语料,可得 1 500 条样本。除此之外,还设计了 38 句语义自身与情感相关的语料,对上诉五名录音人员都以对应的情感朗读了一次,可得 190 条样本。这样就共获 1 690 条样本,去掉其中 610 条情感状态模糊不清的语句,最终形成了样本个数为 1 080 的语音库。

实验用 cool edit pro v2.1 和 praat 两种语音处理软件对每句语音信号通过分帧(每帧语音信号取 12ms)、加窗(hamming)等预处理后提取包含韵律类和音质类声学特征数据,形成了 30 维特征向量。然后利用粗集理论方法[12-14]进行特征化简,得到相关的 12 维特征:振幅峰值,RMS 激励最小值,RMS 激励最大值,RMS 激励平均化,最高分贝,最高分贝对应的频率,最低分贝,最低分贝对应的频率,最高强度,最低强度,第一共振峰带宽,第二共振峰带宽。

2.2 实验结果与分析

随机选取数据库中的 2/3 数据用以训练网络,1/3 数据用以测试网络。用训练好的网络对测试语音情感识别时,网络的输出向量实际不为二进制整数向量,其各个分量实际为靠近 0 或 1 的小数。这时将输出向量的各分量圆整到最近的整数即可形成二进制向量,按其编码即可对应查出被识别情感属于什么类。网络训练如图 3 所示,识别结果见表 1。

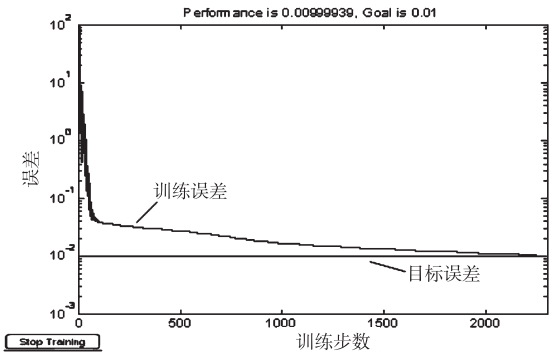

图 3 图语音情感识别的 RBF 网络训练图

表 1 基于 RBF 网络的语音测试集情感识别结果

样本种类	样本种类编号	样本数目	识别数目	识别率(%)	识别标准偏差
高兴	0000	86	54	63	0.025
愤怒	0001	100	58	58	0.205
悲伤	0010	80	51	64	0.135
害怕	0011	90	68	76	0.126
惊讶	0100	89	46	52	0.246
平静	0101	80	68	85	0.018

从表 1 识别的结果可以看出,RBF 神经网络的平均识别率为 66%,同时也用 BP 神经网络进行了网络训练,发现 RBF 网络的情感平均识别率高于 BP 神经网络 3 个百分点。从种类识别的分布标准差来看,个别类识别标准偏差偏大,表明分类不够集中,这可能是受语音信号特征提取,噪声以及算法的设计等影响。如果改善实验条件,使所得的数据库更具有针对性,同时改良算法、优化网络等等,都可以提高情感识别的效果。总的来看,实验结果表明了基于 RBF 神经网络的语音情感识别方法的有效性。

3 结束语

语音情感识别就是让计算机能够通过语音信号识别说话者的情感状态,是情感计算的重要组成部分。而由于情感信息的社会性、文化性以及语音信号自身的复杂性,语音情感识别中尚有许多问题需要解决,特别是符合人脑认知结构与认知心理学机理的情感信息处理算法需要得到进一步的研究。笔者利用 RBF 算法建立了有效的神经网络,并比较了 BP 神经网络识别,实验应用取得了较理想的语音情感识别效果。在今后的研究中,需要进一探讨算法与神经网络的结合,特别是优化神经网络的拓扑结构等来促进语音情感识别的发展研究。

参 考 文 献:

[1] Fragopanagos N,Taylor J G.Emotion recognition in human-computer Interaction[J].Neural Networks,2005,18(4):389-405.

[2] Bosch L E.Emotions,speech and the ASR framework[J].Speech Communication,2003,40(1-2):213-225.

[3] Cowie R,Douglas-cowie E N,Tsapatsoulis N,et al.Emotion recognition in human-computer interaction[J].IEEE Signal Processing Magazine,2001,18,(1):32-80.

[4] Malta Lucas,Chiyomi Miyajima,Norihide Kitaoka,et al.Analysis of Real-Word Driver's Frustration[J].IEEE Transactions on intelligent transportation systems,2011,1(12):109-118.

[5] France D J,Shiavi R G,Silverman S,et al.Acoustical properties of speech as indicators of depression and suicidal risk[J].IEEE,Trans on Biomed Engical Engineering,2000,47(7):829-837.

[6] Rabiner L R.A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition[J].Proceedings of the IEEE,1989,77(2):257-286.

[7] Vlassis N,Likas A.A greedy em algorithm for Gaussian mixture learning,Neural Process Lett,15(2002)77-87.

[8] 武开福,曹 伟.基于 RBF 神经网络的农田土壤含盐量的预测[J].节水灌溉,2011(1):18-20.

[9] 边肇其,张学工.模式识别[M].北京:清华大学出版社,2000.

[10] 王 雪.智能软计算及其应用[M].北京:清华大学出版社,2007.

[11] 谢 波,陈 岭,陈根才,等.普通话语音情感识别的特征选择技术[J].浙江大学学报,2007,41(11):1816-1822.

[12] 曾黄麟.智能计算[M].重庆:重庆大学出版社,2004.

[13] 唐建芳,曾黄麟,张海燕.基于粗集理论的语音情感识别研究[J].计算机科学,2009,36(8A):23-25.

[14] 曾光菊.基于粗神经网络的语音情感识别[J].四川理工学院学报:自然科学版,2011,24(4):472-476.

Speech Emotion Recognition Based on RBF Neural Network

ZHANG Hai-yan, *TANG Jian-fang*

(School of Science, Sichuan University of Science & Engineering, Zigong 643000, China)

Abstract: The principle of radial base function neural network and its train algorithm are introduced in this paper. Meanwhile, the model of speech emotion recognition based on RBF neural network is established. In the recognition experiments, BP neural network and RBF neural network are compared in the same testing environment. The recognition rate of RBF neural network is 3% more than BP neural network. The results show that the method based on RBF neural network speech emotion recognition is effective.

Key words: radial basis function; RBF neural network; speech emotion recognition