基于SOM神经网络和支持向量机的方言辨识

朱 颖,钱盛友,赵新民
ZHU Ying,QIAN Sheng-you,ZHAO Xin-min

湖南师范大学 物理与信息科学学院,长沙 410081
College of Physics and Information Science,Hunan Normal University,Changsha 410081,China
E-mail:shyqian@hunnu.edu.cn

ZHU Ying,QIAN Sheng-you,ZHAO Xin-min.Dialect identificatiom based on SOM and SVM.Computer Engineering and Applications,2009,45(22):200-201.

Abstract: A Chinese dialect identification system based on a mixed SOM neural network and SVM is proposed in this paper. Hunan dialects have been selected as the research object.SOM is applied to cluster for the MFCC of various dialects,and SVM is used as the final implement of decision and identification.The results show that this system has better real-time property and identification rate than the conventional methods especially at a low signal-to-noise ratio.
Key words: dialect identification;SOM neural network;Support Vector Machine(SVM)

摘 要:建立了一个基于SOM神经网络和支持向量机(SVM)的汉语方言辨识系统。该系统以湖南方言作为研究对象,借助SOM神经网络对不同方言的MFCC特征参量进行聚类,并用SVM作为最终的决策辨识器。实验结果表明:该系统与传统系统相比实时性和辨识率较好,特别适用于信噪比低的情况。
关键词:方言辨识;SOM神经网络;支持向量机
DOI:10.3778/j.issn.1002-8331.2009.22.064 **文章编号**:1002-8331(2009)22-0200-02 **文献标识码**:A **中图分类号**:TP391

1 引言

方言辨识作为语音识别中的一个新兴领域,对于语音识别技术的推广和应用有着重要的意义。这项研究已经得到了欧美各国的高度重视,已有不少新的技术成果出现,但在目前国内这方面的研究相对较少。中国作为一个多方言多民族语言的大国,完全有必要开展方言辨识的研究[1-3]。以湖南长沙、株洲、常德、衡阳四地方言为研究对象,在文献[1]研究的基础上,从语音环境的复杂性和方言辨识的实时性等方面考虑,提出了一种基于SOM神经网络和支持向量机的方言辨识方法。

2 基本理论

2.1 SOM神经网络

该网络由输入层和竞争层组成。两层之间的神经元通过权值相互联结在一起,对每个特征矢量序列,通过竞争层的竞争算法,在竞争层的某个神经元便会兴奋起来输出结果。

具体算法如下[4]:

(1)网络初始化:输入层和竞争层之间的权值初始值,用较小的随机数设定。

(2)输入向量的输入:首先,将各语音特征参量 $\boldsymbol{X}=[x_1,x_2,x_3,\cdots x_n]^{\mathrm{T}}$ 输入给输入层。

(3)在输出层计算各神经元的权值向量和各输入向量的距离。输出层第 j 个神经元和输入向量的距离,由下式给出:

$$d_j=\sqrt{\sum_{i=1}^{n}(x_i-w_{ji})^2} \tag{1}$$

式中,w_{ji} 为输入层的 i 神经元和竞争层的 j 神经元之间的权值。

(4)d_j 为最小,则将其称为胜出神经元,记为 j^*。

(5)权值的学习:正确神经元成为胜出神经元时,即

$$\Delta w_{ji}=\begin{cases}+\eta(x_i-w_{ji}):\text{正确识别时}\\-\eta(x_i-w_{ji}):\text{错误识别时}\end{cases} \tag{2}$$

式中,$\eta>0$ 为学习系数。

(6)重复第(2)步到第(5)步,进行权值学习。

2.2 支持向量机

支持向量机(SVM)是近来倍受关注的模式分类手段,是一种小样本学习方法[5]。其核心是"升维",即将样本映射到高维甚至无穷维空间,在高维空间采用处理线性问题的方法。支持向量机的计算量由支撑矢量决定,与样本维数几乎无关,从而可以避免"维数灾"。

假定训练样本数据 $(\boldsymbol{x}_i,y_i)$,$\boldsymbol{x}_i\in R^n$,$y_i\in\{-1,+1\}$可以被一个超平面 $(\boldsymbol{w}\cdot\boldsymbol{x})+b=0$ 分开,如果距离超平面最近的向量与超平面之间的距离是最大的,则判定这个向量被最优超平面分开,即

$$\begin{cases}(\boldsymbol{w}\cdot\boldsymbol{x}_i)+b\geqslant 1,\text{if } y_i=1\\(\boldsymbol{w}\cdot\boldsymbol{x}_i)+b\leqslant -1,\text{if } y_i=-1\end{cases} \tag{3}$$

基金项目:湖南省教育厅资助科研项目(No.06C517)。
作者简介:朱颖(1983-),男,硕士研究生,研究方向为语音信号处理;钱盛友(1965-),男,通讯作者,教授,博导,主要研究方向为信号检测与处理及智能仪器等。
收稿日期:2008-04-23 **修回日期**:2008-07-21

其中位于$(\boldsymbol{w}\cdot\boldsymbol{x}_i)+b=1$和$(\boldsymbol{w}\cdot\boldsymbol{x}_i)+b=-1$上的向量被称为支持向量。这样，SVM 问题可用下式带约束条件的优化问题来描述：

$$\begin{cases}\min\left(\frac{1}{2}\|\boldsymbol{w}\|+C\sum\xi_i\right)\\ y_i[\boldsymbol{w}^*\boldsymbol{x}_i+b]-1+\xi_i\geqslant 0\end{cases}\tag{4}$$

相应的决策函数为：

$$f(x)=\sum(a_i-a_j)K(x_i,x_j)+b\tag{5}$$

其中 $K(x_i,x_j)$为核函数。常用的核函数有线性核、多项式核、径向基核和多层感知器核等等，采用分类性能较好的高斯径向基核函数，惩罚系数 C 取 100，训练误差为 0.001。

通过以上理论分析知道：最优决策面仅由支撑向量决定，这样既降低了计算复杂度，又使分类器具有良好的鲁棒性，支持向量机能充分利用训练样本的分布特性，根据部分训练样本构建判别函数，不需要更多的先验信息，因此 SVM 具有较高的识别率和抗噪特性。

2.3　MFCC 参数

基于人耳听觉特性的美尔倒谱系数（MFCC）具有很好的辨识效果，在噪声环境中更能体现其优势[6]。文中就采用 MFCC 参数作为方言辨识中的特征参数，参数提取流程如图 1 所示。

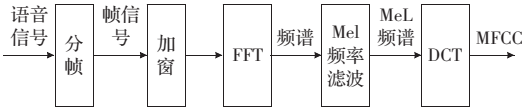

图 1　MFCC 参数提取流程图

3　方言辨识系统的构成

本实验采用的辨识系统如图 2 所示。它由 4 个部分构成：语音预处理、特征提取、SOM 神经网络、支持向量机。

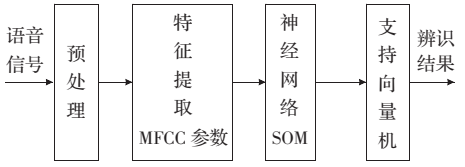

图 2　方言辨识实验系统

其中语音的预处理部分主要包括：语音抽样、A/D 转换、预加重、编码等，得到数字语音信号，然后用汉明窗进行分帧处理；特征提取部分主要是提取 16 阶 MFCC 参数作为方言特征参数。SOM 神经网络对特征参量数据进行聚类，将整个特征参量空间划分为若干聚类中心的子集；支持向量机作为识别器，对各子集进行分类识别，最后得到识别结果。

4　实验结果与分析

语音数据采用出生于湖南长沙、株洲、常德、衡阳的不同男女的发音，每人对表 1 中的 40 个单字用当地方言发音 3 遍，为便于对比另请普通话标准的男女试验者每人对表 1 中的各单字发音 3 遍，共选取 1 200 个样本。语音样本采用 WAV 格式，采样频率为 11 025 Hz，A/D 转换精度为 16 bit，采用 Hamming 窗进行加窗，窗长为 32 ms，帧移为 16 ms。取前两遍发音共 800 个样本的 MFCC 参数作为训练集，第三遍发音共 400 个样本的 MFCC 参数作为测试集。

表 1　实验所用单字表

	ma	po	ke	ti	fu	xū	you	bai	zhao	chuang
阴平	妈	坡	科	踢	夫	须	优	掰	朝	窗
阳平	麻	婆	咳	提	服	徐	油	白	着	床
上声	马	叵	可	体	抚	许	有	百	找	闯
去声	骂	破	客	替	负	序	又	败	照	创

采用 MATLAB 进行仿真，对 4 种湖南方言和标准普通话发音各 80 个样本，采用该方言辨识系统进行测试，辨识结果如表 2 所示。通过数据分析得到：长沙方言和株洲方言的相似度较高，而与衡阳方言相差较远；4 种湖南方言中常德方言与普通话较接近；采用同一辨识系统，针对不同的湖南方言，系统的辨识率各不相同，其中常德话的辨识率最高为 84.6%，长沙话的辨识率最低为 75.1%；与湖南方言相较，标准普通话更容易辨识，辨识率达到了 86.1%；系统的平均辨识率为 80.3%。

表 2　方言辨识结果统计

方言类别	识别结果					识别率/(%)
	长沙	株洲	常德	衡阳	普通话	
长沙	60	15	1	3	1	75.1
株洲	11	62	2	4	1	76.8
常德	0	1	68	1	10	84.6
衡阳	5	6	3	63	3	78.9
普通话	1	1	9	0	69	86.1

为进一步验证所提出方案的实时性和抗噪性，在训练样本中加入均匀的白噪声，测试样本仍然保持不变，重新进行上面的实验。另外分别采用 BP 神经网络、支持向量机进行两组对比实验。将采用不同辨识系统得到的 4 种湖南方言和标准普通话的平均辨识率、辨识时间进行比较，结果如表 3 所示。

由表 3 可知：随着信噪比的下降，三种方言辨识系统的辨识率都有所降低，辨识时间也都有不同程度的延长，其中以 BP 神经网络的变化最为明显，即在信噪比较低的复杂语音环境下，BP 神经网络辨识能力并不理想。相对而言支持向量机的辨识能力较强，但随着语言信号信噪比的不断降低，其学习时间增加明显，从而直接导致系统的辨识时间延长，即实时性不佳。所提出的采用 SOM 与 SVM 相结合的方言辨识系统，在信噪比变化的情况下，性能比较稳定，辨识率和实时性都有较好的改善。由此可见，基于 SOM 神经网络与支持向量机相结合的方言辨识系统是解决方言辨识问题的一条有效途径，尤其在信噪比低，实时性要求高的情况下，其优势更为明显。

表 3　三种不同辨识系统性能比较

信噪比	15 dB		10 dB		5 dB		3 dB	
辨识结果	平均辨识率/(%)	辩识时间/min	平均辨识率/(%)	辩识时间/min	平均辨识率/(%)	辩识时间/min	平均辨识率/(%)	辩识时间/min
BP 神经网络	76.4	5.4	75.1	7.8	71.2	9.7	61.8	12.0
SVM	78.3	2.3	77.1	3.4	76.7	4.6	75.6	6.1
SOM 与 SVM	79.6	1.5	78.4	2.0	78.4	2.4	77.7	3.2

（下转 205 页）