

基于免疫聚类的 RBF 网络在说话人识别中的应用

周 燕^{1,2}, 胡志峰^{1,2}

(1. 江苏省现代企业信息化应用支撑软件工程技术研究开发中心, 江苏苏州 215104; 2. 苏州市职业大学电子信息工程系, 江苏苏州 215104)

摘要: 针对传统的基于 RBF(Radial Basis Function)网络的说话人识别系统中聚类中心的数量和位置难以确定的问题, 提出了一种基于人工免疫机制的 RBF 网络作为分类器的说话人识别系统。采用人工免疫机制可根据输入语音数据集自适应地确定 RBF 网络隐层中心的数量和初始位置。实际测试表明, 该系统具有快速学习网络权重的能力, 并且网络的全局寻优能力强, 识别率高, 是说话人识别的一种有效可行的新方法。

关键词: 说话人识别; RBF 网络; 免疫算法

中图分类号: TP18

文献标识码: A

文章编号: 1000-3630(2010)-02-0184-04

DOI 编码: 10.3969/j.issn1000-3630.2010.02.014

Application of immune algorithm based RBF network to human speaker recognition

ZHOU Yan^{1,2}, HU Zhi-feng^{1,2}

(1. Jiangsu Research & Development Centre for Modern Enterprise Information Software Engineering, Suzhou 215104, Jiangsu, China;

2. Department of Electronic and Information Engineering, Suzhou Vocational University Suzhou 215104, Jiangsu, China)

Abstract: When using traditional clustering algorithms based on RBF to recognize human speakers, it is hard to decide the number and locations of the cluster centers. To overcome these shortcomings, an RBF network based on artificial immune mechanism for human speaker recognition is proposed. The artificial immune mechanism can adaptively compute the number and initial locations of the centers in the hidden layer of the RBF network based on the audio sample data set. Experimental tests show that the system has a fast learning speed for network weights. The system is very good at searching for global optimum. It has a high recognition rate and is a new practical method for human speaker recognition.

Key words: human speaker recognition; RBF Network; immune algorithm

1 引言

说话人识别是从说话人的一段语音中提取说话人的个性特征, 通过对这些个人特征的分析 and 识别, 达到对说话人进行辨认或者确认的目的^[1,2]。由图 1 可见, 系统的任务主要有三项: 一是说话人语音采集与语音特征参数提取; 二是通过免疫算法计算得到说话人语音码书, 称为训练过程; 三是在说话人识别时将测试者的语音特征参数与已有的码书进行匹配并作出决策, 称为识别过程。

在说话人识别中, 运用 VQ(Vector Quantization)

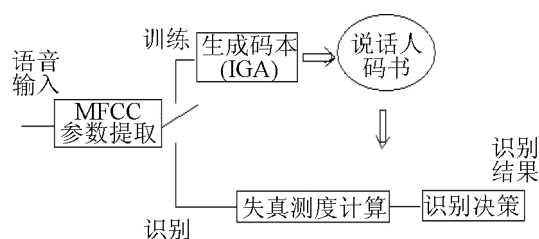


图 1 说话人识别系统的结构框图

Fig.1 Block diagram of human speaker recognition system

技术对说话人的语音特征参数进行聚类分析, 形成由各聚类中心构成的码本作为说话人的模型, 目前的聚类算法主要有 K 均值和 LBG 算法。文献[3]利用 K 均值聚类算法对语音数据聚类, 该算法具有简单、快速并且能够有效处理大数据的优点, 但是这种算法对于不同的初始值可能会导致不同的聚类结果, 而且, 由于该算法是基于目标函数的聚类算法, 因此极易陷入局部极小值。为了克服上述缺点, 利用遗传算法优化目标函数, 但是实验表明,

收稿日期: 2009-01-15; 修回日期: 2009-04-27

基金项目: 国家自然科学基金项目(60970058)、江苏省自然科学基金项目(BK2009131)苏州市职业大学校级基金项目(SZD09L26)

作者简介: 周燕(1980-), 女, 江苏苏州人, 讲师, 硕士, 研究方向为语音信号处理。

通讯作者: 周燕, E-mail: zhyan@jssvc.edu.cn

当样本数目、类别和维数较多时,基于遗传算法GA(Genetic Algorithm)的聚类方法常会产生早熟现象。近年兴起的人工免疫系统AIS(Artificial Immune System)^[4,5]研究发展迅速,成为智能系统中继模糊逻辑、神经网络、遗传算法之后的又一研究热点。基于生物免疫原理提出了多种人工免疫模型和算法,免疫计算的发展给解决以上的问题带来了新的思路。

本文提出的说话人识别系统由两个阶段组成:第一阶段采用人工免疫聚类算法(Artificial Immune Clustering Algorithm, AICA)来确定RBF网络隐层的聚类中心的位置和数量;第二阶段对RBF神经网络的输入样本集进行分析、学习、训练,求出输出层的权值。

2 系统描述

2.1 特征参数提取

特征提取是模式识别的关键问题,目前最为流行的特征参数是基于声道全极点模型的LPC倒谱系数(LPCC)^[6]和基于人耳听觉特性的MEL倒谱系数(MFCC)^[7,8],经研究表明,MFCC参数较之LPCC参数具有更好的识别效果,并且对噪声的影响不如LPCC敏感。因此MFCC的应用更为广泛,本文选取该特征参数。

根据MFCC参数和线性频率的转换关系:

$$Mel(f) = 2595 \lg(1 + \frac{f}{700}) \quad (1)$$

将实际频率尺度转换为Mel频率尺度,MFCC参数的计算过程如图2所示。

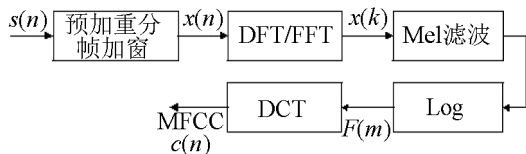


图2 Mel频率倒谱参数(MFCC)的提取
Fig.2 Extraction of Mel-Frequency Cepstral Coefficients(MFCC)

具体计算步骤为:

(1) 首先确定每一帧语音采样序列的点数,对每帧序列 $s(n)$ 进行预加重处理后再经过离散FFT变换,取模的平方得到离散功率谱 $|X(k)|^2$ 。

(2) 计算 $X(k)$ 通过 M 个 $H_m(n)$ 后得到的功率值,即计算 $X(k)$ 和 $H_m(n)$ 在各离散频率点上乘积之和,得到 M 个参数 $P_m, m=0, 1, \dots, M-1$ 。

(3) 计算 P_m 的自然对数,得 $L_m, m=0, 1, \dots, M-1$ 。

(4) 对 $L_0, L_1, \dots$,计算其离散余弦变换,得到 $D_m, m=0, 1, \dots, M-1$ 。舍去代表直流成分的 D_0 ,取 $D_1,$

$D_2, \dots, D_k$ 作为MFCC参数。MFCC参数的个数 k 通常取最低的12~16,本文选取 $k=12$ 。

2.2 RBF网络的结构

由于RBF网络结构简单,学习速度快,具有优异的函数逼近能力,因此在语音识别、系统辨识等方面得到了广泛应用。RBF网络是一种三层前馈神经网络,其结构如图3所示。

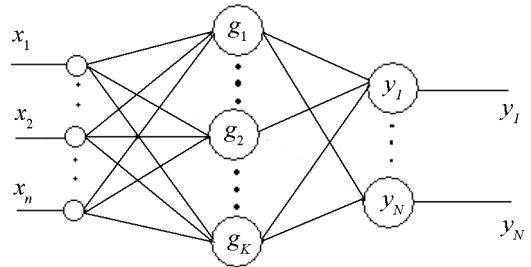


图3 RBF神经网络结构模型
Fig.3 Structure model of RBF neural network

第一层是输入层,这一层和其他的神经网络输入层一样,负责引入外部信息,自身无信息处理能力。第二层又被称为隐含层,是RBF神经网络的主要功能层,该层采用径向基函数作为传输函数,通常采用高斯函数。从输入层到隐含层的变换是非线性的,从隐含层到输出层的变换是线性的。假定输入样本为 x ,则隐含层第 i 个单元的输出为:

$$R_i(x) = \exp\{-|x - z_i|^2 / 2\sigma_i^2\}, i=1, 2, \dots, K \quad (2)$$

式中, K 为隐含层单元的个数, z_i 为基函数的中心, σ_i 是第 i 个隐含层单元的高斯函数的宽度,则输出层第 j 个节点的输出为:

$$Y_j(x) = \sum_{i=1}^K W_{ij} R_i(x), j=1, 2, \dots, N \quad (3)$$

式中, W_{ij} 是第 i 个隐含层节点到第 j 个输出节点的权值,输出节点共 N 个。

文中RBF网络结构与参数的训练过程如下:基函数的中心 z_i 、第 i 个隐含层单元的高斯函数的宽度 σ_i 通过人工免疫聚类算法(AICA)获得,从隐含层到输出层的变换权值 W_{ij} 通过遗传算法(GA)获得。重点是利用免疫算法确定网络的聚类中心 z_i ,聚类中心的个数即为RBF网络的隐层节点数。

3 确定RBF网络聚类中心的免疫算法

采用基于免疫原理的聚类方法,把语音数据对象视为抗原,把聚类中心看作是免疫系统中的抗体,语音数据对象的聚类过程就是免疫系统不断地产生抗体,识别抗体,最后产生出可以捕获抗原的最佳抗体过程。本文引入免疫机制聚类^[9],有效地

克服了一般 RBF 网络训练算法难以确定聚类中心的数量和位置的缺点,同时根据聚类问题的实际情况设计编码、交叉和变异算子,使得算法更快、更有效地收敛于全局最优解。

免疫聚类算法的基本步骤如下:

(1) 输入 n 个抗原,随机抽取 C 个初始抗体,组成抗体群;

(2) 每个抗体识别在其临近空间内的抗原,选择欧几里德距离为抗体与抗原的亲合力指标,按抗体与抗原的亲合力分组,分组的判定函数可定义为:

$$J(u, c) = \sum_{i=1}^c J_i = \left(\sum_{i=1}^c \sum_{j=1}^n (u_{ij}) \|X_i - C_j\|^2 \right) \quad (4)$$

这里 u_{ij} 是向量 X_i 属于组 C_j ($i=1,2,3,\dots,c$) 的程度,它的值介于 0 和 1 之间,每一次分组时, n 个数据对象按亲和力的大小,分属于不同的 C_i ($i=1,2,3,\dots,c$), 其中:

$$u_{ij} = \begin{cases} 1, & \text{对于每个 } k=i, \text{ 若 } \|X_j - C_i\|^2 \leq \|X_j - C_k\|^2 \\ 0, & \text{其他} \end{cases} \quad (5)$$

(3) 抗体的优化采用 *aiNet* 免疫网中有向搜索来寻找优化抗体的方法,利用公式: $C = C - \alpha(C - X)$, C 为形态空间中的抗体, X 为抗原, α 是学习率,对每一个 C_i , $i=1,2,3,\dots,c$, 在本分组内利用克隆并进化产生 C_{is} ($s=1,2,3,\dots,n$; $0 < s \leq n$), 然后计算每组中新产生的 s 个抗体与组内 s 个抗原的亲合力的大小,按亲合力最大原则选择新的最佳抗体(聚类中心);

(4) 根据抗体抑制原理,除保留 C 个分组中的最佳抗体之外,清除其余的所有抗体;

(5) 重复步骤(2)、(3)、(4)直到满足抗体最优条件:

$$\min J(u, c) = \sum_{i=1}^c J_i = \left(\sum_{i=1}^c \sum_{j=1}^n (u_{ij}) \|X_i - C_j\|^2 \right) \quad (6)$$

算法结束,输出聚类分析结果 Z_i ($i=1,2,3,\dots,K$), 即聚类中心。聚类中心的个数即为 RBF 网络的隐层节点数。

每个隐层节点的高斯宽度 σ_i 值,可通过计算该节点到所在的聚类中心的距离获得,具体计算为:

$$\sigma_i = |x - z_i| / \sqrt{-2 \ln R_i(x)}, \quad i=1,2,\dots,K \quad (7)$$

式中, x 为输入样本, z_i 为聚类中心。

3 系统实现

3.1 说话人识别过程

在基于人工免疫聚类算法的 RBF 网络的说话人识别实验中,所用语音数据来源于 Timit 数据库^[10],对不同的人采用不同内容的连续语音对网络进行训

练,保证所用语音与文本无关,训练语音长度为 20s。参加试验者为 32 人,每人 15 段不同的语音,将说话人的语音分为两部分,其中 6 段用来训练网络,9 段用来识别身份。对两组语音分别求取 12 维的语音特征 MFCC 并保存。

首先对输入语音依次进行端点检测、预加重、分帧、加汉明窗,然后逐帧计算 12 维 MFCC 系数。语音数据帧长为 $n=256$ 点,帧移 128 点,采用 $1-0.95Z^{-1}$ 高通滤波器进行预加重,去除噪音,形成语音特征矢量序列;其次对人工免疫算法参数进行设置,取种群规模 $N=32 \times 6=192$ 个,192 个训练样本被分为 48 个组,产生 48 个聚类中心,即隐层节点数可取 48 个,其中,交叉概率 $P_c=0.8$,变异概率 $P_m=0.001$, α 设定为 5.0×10^{-8} ,因实验结果要求分辨 32 个说话人的身份,所以输出节点设为 32 个,所以构建基于免疫算法的 RBF 网络的结构为: 12-48-32 结构。

本系统采用 Visual C++ 作为开发平台,程序对于网络的输出采用的准则为:对 32 维输出节点选取其最大值,然后置 1,其他输出置 0。误差采用公式: $E = \frac{1}{2} \sum_{k=1}^N e_k^2$, N 为样本数。程序设置当误差 $E < 10^{-5}$ 或训练次数大于 1000 次时,训练结束,即程序结束。

识别时,在网络中输入待识别者的 12 维语音特征矢量,分别调用每个人已保存好的网络权值 W_{ij} ,计算输出结果得到一个 32 个维的矩阵,输出为 1 的节点即为对应的所要识别的说话人。

3.2 实验结果分析

在 Matlab6.5 实验环境中,利用提取的语音数据进行说话人识别实验,实验结果如下:

(1) 聚类码本适应度的比较

码本是指所提取语音信号的特征样本,适应度是免疫算法中描述抗体性能的重要指标,抗体的适应度值越大,表明抗体和抗原的匹配性越好。在本文算法中,将码本对应于免疫算法中的抗体,则码本的适应度是求解过程中对待求的语音信号的特征参量进行优胜劣汰的判断指标,根据这个指标,可在候选的语音信号的特征参量群体中获得最优解。

图 4 给出了 15 个说话人在聚类中心为 48 时聚类中心产生的适应度分布曲线,从图中可以看出,使用免疫算法能得到较高的适应度,其次是遗传算法, K 均值聚类算法产生的适应度最低,由此可见,免疫聚类算法明显优于一般聚类算法。

(2) 识别率比较

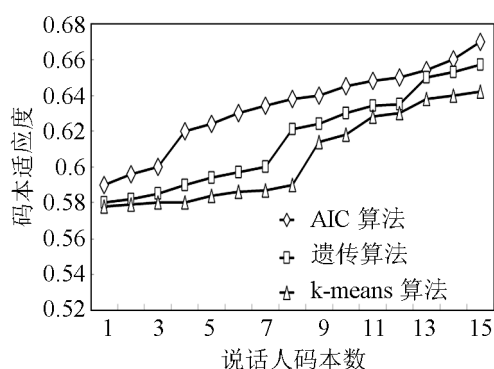


图4 聚类码本适应度比较

Fig.4 Comparison of clustering code book adaptability

为了进一步比较各种聚类算法的识别效果,对说话人识别系统进行识别率实验,定义说话人识别系统识别率如下:

识别率=识别正确的帧数/识别的总帧数

在聚类中心数为 48 时,其识别结果如图 5 所示:

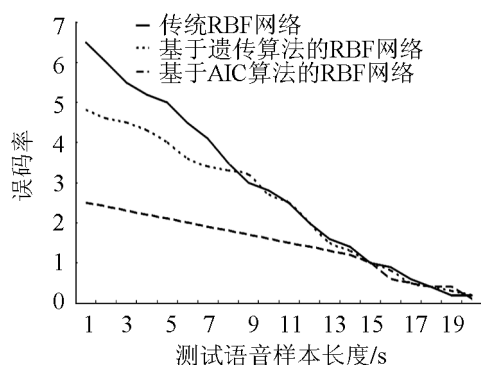


图5 不同测试语音下的误码率比较

Fig.5 Comparison of error detection rate between different voice samples

图 5 显示了不同测试语音长度条件下,不同训练算法的误码率曲线,本文介绍的利用免疫算法对语音特征参数聚类的说话人识别系统的误码率明显低于遗传聚类算法及一般 RBF 网络。

4 结束语

根据人工免疫机制设计了一种基于 RBF 网络的说话人识别系统。该系统利用人工免疫聚类算法

确定 RBF 网络初始中心的数量和位置,使得网络的局部搜索能力得到提高,收敛速度增快,避免了退化现象的出现。实验证明,基于免疫算法的 RBF 网络可以得到较高的说话人识别率。

参 考 文 献

- [1] Benzeghiba M, De Mori R. Automatic speech recognition and speech variability: A review[J]. speech communication, 2007, 49: 763-786.
- [2] 关海欣, 曾庆宁. 多通道子空间算法在说话人识别中的应用[J]. 声学技术, 2008, 27(3): 396-402.
GUAN Haixin, ZENG Qinning. Application of Multi-channel sub-space algorithm in Speaker Recognition[J]. Technical Acoustics, 2008, 27(3): 396-402.
- [3] 邱作春, 曾庆宁. 独立分量分析在说话人识别技术中的应用[J]. 声学技术, 2008, 27(6): 863-866.
QIU Zuochun, ZENG Qinning. Application of independent component analysis in speaker recognition[J]. Technical Acoustics, 2008, 27(6): 863-866.
- [4] 丁永生, 任立红. 人工免疫系统:理论与应用[J]. 模式识别与人工智能, 2000, 13(1): 28-67.
DING Yongshen, REN Lihong. Artificial immune system: theory and application[J]. Pattern Recognition and Artificial Intelligence, 2000, 13(1): 28-67.
- [5] KIM S, ERIKSSON T. A pitch synchronous feature extraction method for speaker recognition[C]. IEEE, Acoustics Speech and Signal Proceedings, 2004, (1): 405-408.
- [6] 王华朋, 杨洪臣. 声纹识别特征的提取方法研究[J]. 中国人民公安大学学报(自然科学版), 2008, (1): 28-30.
WANG Huapeng, YANG Hongcheng. A Research of Feature Extraction Method for Sound Print Recognition[J]. Journal of Chinese People's Public Security University(Nature Science Edition), 2008, (1): 28-30.
- [7] Ahmed Mezghani, Douglas. Speaker verification using a new representation based on a CMFCC and fomants[J]. IEEE Electrical and Computer Engineering, 2005, 22: 1469-1472.
- [8] 关海欣, 曾庆宁. 基于子带分析稳健的说话人识别[J]. 声学技术, 2009, 28(2): 142-147.
GUAN Haixin, ZENG Qinning. Speaker recognition based on robust sub-band analysis[J]. Technical Acoustics, 2009, 28(2): 142-147.
- [9] 刘韬, 王耀才, 王致杰. 一种基于人工免疫系统的聚类算法[J]. 计算机工程与设计, 2004, 25(11): 2051-2053.
LIU Tao, WANG Yaocai, WANG Zhijie. A cluster algorithm based on artificial immune system[J]. Computer Project and Design, 2004, 25(11): 2051-2053.
- [10] Ortega Garcla J O, Rodríguez J G, et al. AHUMADA: A large speech corpus in Spanish for speaker characterization and identification[J]. Speech Communication, 2000, 31: 255-264.