

基于EBF网络的特征映射器的研究 及其在鲁棒话者识别中的应用

作 者：李 昕

专 业：控制理论与控制工程

导 师：费敏锐



上海大学出版社

· 上海 ·

2003 年上海大学博士学位论文

基于 EBF 网络的特征映射器的研究 及其在鲁棒话者识别中的应用

作 者： 李 昕

专 业： 控制理论与控制工程

导 师： 费敏锐

上海大学出版社

• 上海 •

Shanghai University Doctorial Dissertation (2003)

要 容 内

Feature Mappers based on Elliptical Basis Function Networks for Robust Speaker Recognition

Candidate: Li Xin

Major: Control Theory and Control Engineering

Supervisor: Prof. Fei Min-rui

Shanghai University Press

• Shanghai •

图书在版编目(CIP)数据

2003 年上海大学博士学位论文集/上海大学研究生部
编. —上海: 上海大学出版社, 2005.8

ISBN 7-81058-898-2

I. 2... II. 上... III. 上海大学-博士-学位论文-汇编
—2003 IV. G643.8

中国版本图书馆 CIP 数据核字 (2005) 第 091219 号

2003 年上海大学博士学位论文

上海大学出版社出版

上海市上大路 99 号 邮政编码 200444

出版人: 姚铁军

*

上海锦佳装订厂印刷

开本 850×1168 1/32 印张 160 字数 3732 千字

2006 年 5 月第 1 版 2006 年 5 月第 1 次印刷

ISBN 7-81058-898-2 / G · 383

全套定价: 260.00 元

内容提要

《2003 年上海大学博士学位论文》内容包括：“强磁场对合金中析出相凝固行为的影响”、“炼钢电弧炉全程动态模型与仿真研究”、“气膜软接触连铸技术的基础研究”、“蓝色有机薄膜电致发光器件及界面特性的研究”、“多元氧化锆(ZrO_2)基陶瓷的相图计算和材料制备”、“非线性材料中的空穴生成与增长问题”、“现代通信系统中微波滤波器小型化的研究”、“数字图像和音频中隐藏信息的检测和主动攻击”、“孤子与辐射场的相互作用及对数型非线性介质中空间孤子族的理论研究”、“输电当量电价方法论及其应用”、“新型改质植物油和环境友好润滑剂的摩擦学性能和应用性研究”、“移动通信中若干随机现象的建模及分析”、“广义逆函数值 Padé 逼近的理论与方法及在 Fredholm 积分方程中的应用”、“时滞微分方程 Hopf 分歧分析、周期解的计算及其数值动力系统”、“Thiele-Werner 型连分式复向量有理插值若干问题及应用”、“矩阵广义逆理论、计算及其应用的若干问题”、“全局优化的几种确定性方法”、“Stokes 方程和 Navier-Stokes 方程谱方法、抛物型和双曲型方程时空谱方法”、“新型的微波电路及 EBG 结构的理论分析”、“无穷域问题的谱方法研究”、“细小工业管道群管束间狭窄空间检测机器人系统的研究”、“蓝色有机薄膜电致发光材料及器件的研究”、“氧化铝衬底上金刚石的厚膜生长及其微电子学应用研究”、“掺氮类金刚石薄膜的性能和应用研究”、“立体视频帧估计与内插技术及立体视频编码系统研究”、“基于 EBF 网络的特征映射器的研究及其在鲁棒话者识别中的应用”。

上海大学

本论文经答辩委员会全体委员审查, 确认符合上海大学博士学位论文质量要求.

答辩委员会名单:

主任: 吴启迪	教授, 同济大学	200092
委员: 柴佩琪	教授, 同济大学计算机系	200092
杨 杰	教授, 上海交通大学电子信息学院	200030
顾幸生	教授, 华东理工大学自动化系	200237
曹奇新	教授, 上海交通大学机器人研究所	200030
宋志坚	教授, 上海复旦大学	200032
曹家麟	教授, 上海大学	200072
导师: 费敏锐	教授, 上海大学	200072

评阅人名单:

柴佩琪	教授, 同济大学计算机系	200092
杨 杰	教授, 上海交通大学电子信息学院	200030
王朔中	教授, 上海大学电子信息工程系	200072

评议人名单:

顾幸生	教授, 华东理工大学自动化系	200237
汤天浩	教授, 上海海运学院	200135
丁永生	教授, 上海东华大学	200051
李 斌	教授, 上海大学	200072

答辩委员会对论文的评语

李昕同学的博士学位论文《基于 EBF 网络的特征映射器的研究及其在鲁棒话者识别中的应用》，面向身份辨别、司法鉴定、语音检索、声控假肢等众多应用领域中话者识别的重要理论和方法，开展了深入和有效的研究，具有重要的理论意义和实用价值。

该论文的创新性工作和成果主要体现在以下几个方面：

1. 提出用 RBF 网络的扩展形式椭圆基函数 (EBF) 网络来进行话者识别的思想；

2. 针对传统的聚类算法（如 K 均值法）需预先确定聚类数目的问题，提出一种扩展的次胜者受罚的竞争学习 (RPCL) 算法，使之能通过竞争学习自动确定聚类的数目，改进了聚类的代表性，提高了话者识别率；

3. 针对 EBF 网络存在的次优化问题，提出结合期望值最大 (EM) 算法优化 EBF 网络结构和参数的策略。实验表明，该方法可以提高话者识别的精度；

4. 针对倒谱均值互减法 (CMS) 存在的问题，提出一种基于 EBF 神经网络的非线性特征映射器，实现特征映射与恢复。实验表明，使用该方法的识别率较未经补偿时提高近一倍，较之 CMS 补偿后的识别率也有显著的提高。

5. 针对目前神经网络鲁棒话者识别存在的问题, 提出了切实可行的改进方法和结构, 开发出了基于神经网络的电话话者身份识别系统软件.

李昕同学论文写作规范, 答辩中表达清楚, 且能正确地回答答辩委员会所提出的问题. 综上所述, 李昕同学已具备了坚实宽广的基础理论和系统深入的专业知识, 具有较强的独立科研能力.

答辩委员会表决结果

经答辩委员会表决, 全票同意通过李昕同学的博士学位论文答辩, 建议授予工学博士学位.

答辩委员会主席: **吴启迪**

2003 年 3 月 12 日

摘 要

话者识别是用语音信号和预先提取的话者特征来确定或鉴别说话人的身份。鲁棒话者识别技术能够在输入语音质量退化,或语音的音素特性、分割特性或声学特性在训练和测试环境中不同时,仍然保持较高的识别率。提高话者识别系统的鲁棒性是目前国际上在此领域内的研究重点和热点之一。

神经网络是 20 世纪 80 年代以来的另一大研究热点。它以其高度并行性、任意非线性逼近能力、分布存储、容错性以及自学习、自适应能力在诸多领域得到广泛的应用。人们已经认识到,语音信号中存在着非线性性质,因而很自然地,神经网络技术成为话者识别研究中的一个重要方法。从识别角度看,目前国际上绝大多数的研究都集中在各种神经网络方法上。

文章在对神经网络、话者识别,尤其是对鲁棒话者识别的研究现状和发展进行认真详细分析的基础上,利用神经网络构建特定话者的特征映射器来改善话者识别系统的鲁棒性。这包括两个方面的研究工作:一是选择合适的神经网络类型,对其参数进行优化,使其对于语音信号特征在样本空间中发生混叠时,仍然保持较好的样本表征能力;二是通过补偿的方式,对在不同环境或信道影响下发生畸变的失真信号予以恢复,使其更接近于未失真信号,从而提高可区分程度。文章从上述这两个方面出发,对基于 EBF 神经网络的鲁棒话者识别作了以下 4 方面的深入研究。

首先,比较并说明了 RBF 神经网络结构以及相对于 BP 网络在识别方面的优势,并对径向基核函数的选择作了进一步的

讨论,由此引入了椭圆基函数(EBF)网络,它在RBF网络结构中以全协方差矩阵取代原来的对角协方差矩阵,使之在不同的输入方向上的函数宽度也不同,能够在不增加基函数数目的情况下表征复杂的分布。

其次,影响EBF网络性能的一个非常重要的因素是聚类。文章在分析了目前聚类的基本方法的基础上,指出以K平均法等为代表的分割聚类虽然在理论上找到类的所有划分并使得目标函数最小就可以获得最优解,但由于对于相当尺度的数据集列举所有可能的划分并不现实,因而多数分割方法采用迭代优化技术作为变通。该方法有可能导致收敛至局部最小;另外分割聚类方法一般必须事先确定类的数目,这在实际应用中也是难以事先获知的。因而本文提出一种结合全协方差矩阵的扩展的次胜者受罚竞争学习(RPCL)算法。其优点在于,当类存在互相重叠以及输入矢量含有非独立分量时,仍能够确定适当的聚类数目。同时,由于其在最初的训练中保留了大部分初设单元,使这些单元能够参与以后的竞争,从而降低了原算法对初始条件的敏感度。由于将扩展的RPCL与EBF网络结合了起来,充分考虑到现实中数据形态的复杂性,抛弃了以往RBF网络的等方差的约束,对复杂数据的描述更为准确,表征能力更强。

第三,文章在对期望值最大(EM)算法的思想、收敛性以及收敛速度问题作了认真细致的研究的基础上,指出EM算法在参数优化方面具有无须在每次迭代时都计算逆赫斯阵、接近于超线性收敛、无须设置步长参数就可以达到似然度的单调收敛等优点。基于这些优点,本文采用EM算法来优化EBF网络的均值及协方差矩阵参数,并给出了应用EM算法估计EBF网络参数的具体公式。并且通过话者识别实验,表明利用EM算法

优化的 EBF 网络, 其性能要优于普通的 RBF 网络和 EBF 网络。

第四, 文章通过对鲁棒语音技术, 尤其是通过电话线路的鲁棒话者识别技术的特点的分析, 对鲁棒话者识别的补偿技术提出了一种新的设想和方法。文章指出了在鲁棒语音技术中最常用的倒谱均值互减法 (CMS) 及其改进方法所存在的问题, 强调该方法及其改进一般都基于未失真语音的倒谱均值为零和信道作用为线性关系这两个假设, 而这两个假设在大多数情况下是不能成立的。由此, 本文提出了一种基于 EBF 网络的特定话者的特征映射器, 将失真语音映射回未失真语音来进行话者识别。文章详细阐述了该映射器设计的思想以及构建的步骤, 并通过文本独立的话者识别实验表明使用该映射器及其恢复步骤, 能够有效地复原未失真语音特征。将经过复原后的数据输入话者模型中, 可以极大地降低识别错误率, 显著地改善识别性能。另外, 文章也就构建通用特征映射器的设想进行了简单的讨论和实验验证, 指出在目前研究水平下, 设计特定话者或特定环境的映射器更为现实。

最后, 文章还将上述扩展的 RPCL 和 EM 算法结合起来, 对 EBF 网络参数加以优化, 并以此为基础开发了话者识别系统软件。该系统软件可以对语音数据库进行训练、测试和修改, 也可以加入新的话者语音, 自动地进行采样、特征提取和训练, 并能够对未知话者的语音进行测试和判决, 显示接受或拒绝未知话者。在现场试验中, 该系统表现出识别性能良好, 具有广阔的应用前景。

关键词: 鲁棒话者识别, EBF 网络, 扩展的 RPCL 算法, EM 算法, 特征映射器

Abstract

Speaker recognition is to verify or identify individual's identity by his/her voice and features extracted in advance. Robust speaker recognition techniques attempt to maintain the performance of a recognition system under the adverse conditions such as degradation of input voice quality or characteristic mismatch of phoneme, segmentation or acoustics between training and testing environments. Many researchers of the world in this field focus on how to improve the robustness of a speaker recognition system now.

Another research hotspot is neural networks. The neural networks are widely used by its capabilities of high parallel, non-linearity, distributed storage, fault tolerance, self-adaptation and self-learning. Since people were cognizant of the non-linearity within the voice signal, neural network techniques became an important approach of speaker recognition.

Based on the advancement of speaker recognition and neural networks, a specific speaker feature mapper based on elliptical basis function neural networks is presented in this paper. There are two ways to improve the robustness of the speaker recognition system. One is to select a proper type of neural networks, to optimize the parameters, so that it can maintain high performance of sample representation even if the features of voice signal was overlapped on sample space. Another way is the method of compensation. To recover the clean signal from the distorted signal which influenced

by different environment or channel. The main idea, innovative concepts, approaches and neural structures are described as follows.

Firstly, the article describe the advantages of radial basis function (RBF) neural networks by comparing with other kinds of neural nets such as the BP networks and the conventional probabilistic neural networks. At the meantime, we discuss the selection of radial basis function. Then the elliptical basis function (EBF) neural network is inducted. Instead of diagonal covariance matrices of RBF nets, full covariance matrices were used in EBF networks. It has different function widths in different directions so that complex distributions could be represented without the need for using a large number of basis functions.

Secondly, clustering is a very important factor which affects the performance of EBF networks. The best theoretical solution to a partitional clustering problem is to find all possible partitions and choose the partition that minimizes the objective function. However, the enumeration of all possible partitions is impractical even for data sets of moderate size. Therefore, as an alternative, most partitional approaches use an iterative optimization technique. And this could be cause local minimum. Classical partitional approach such as K-means also have the problem that have to predefine the number of clusters before training. And in most case, the number could not to be known in advance. An extended rival penalized competitive learning (RPCL) algorithm, which incorporates full covariance matrices, is proposed in this paper. By using this algorithm, an appropriate number of clusters are determined even if the clusters

are overlapped or input vectors contain dependent components. At the meantime, it alleviates the initial conditional sensitivity of the original RPCL algorithm, since the most initial units are kept during the first several iterations in the extended RPCL algorithm, they could be attended the later competition. The idea for combining the extended RPCL algorithm and the EBF networks, which considered the complexity of the data forms in practice and discarded the constraint of equal covariance of the conventional RBF networks, has better representation for complicated data.

Thirdly, this paper discuss the concept, convergence and convergence rate of expectation-maximization (EM) algorithm, and point out the EM algorithm has advantages in parameter optimization. The EM algorithm needs not to calculate the Hessian and it can approximate a super-linear method. Compare with those gradient ascent methods, the EM algorithm convergence monotonously without setting the parameter of step size. By reason of those virtues, we apply the EM algorithm to optimize the parameters of the EBF networks in this paper. The speaker recognition experiment results show that by using EM algorithm to optimize the EBF networks, system have superior performance than RBF networks and EBF networks without EM.

Fourthly, the article point out the disadvantages of cepstral mean subtraction (CMS), which is used in speech processing popularly. The CMS is based on two assumptions: the cepstral mean of clean speech is zero and the channel is linear. In this paper, we propose a novel compensation method that does not rely on any of

the above assumptions. We propose a novel specific speaker feature mapper based on EBF network can recuperate the clean data from the distorted speech data. In this paper, we describe the idea and the steps for constructing the mapper in detail. By using the mapper and the procedures of recuperation, we can recover the clean features effectively. Experiment results show that it can reduce the equal error rate enormously. At the end of the paper, the article point out that it is more practical to construct a speaker specific or an environment specific mapper than to construct a universal mapper through an experiment.

Finally, the article developed speaker recognition system software by applying a hybrid method with the extended RPCL and EM algorithm to optimize the parameters of the EBF networks. The software can train, test and modify the data from the voice database, It can add new speaker voice, and to sample and to extract the features automatically. It also can test and judge the unknown voice, show to accept or to reject the speaker. The software works well in practice and has a wide prospect.

Keywords: robust speaker recognition, EBF networks, extended RPCL algorithm, EM algorithm, feature mapper

目 录

第一章 绪论	1
1.1 引言	1
1.2 文本独立(自由文本)话者识别的基本方法	5
1.3 从话者识别到鲁棒话者识别	7
1.4 目前鲁棒话者识别的主要技术路线和研究现状	9
1.5 本文的研究内容与背景	16
1.6 本文的内容编排	18
第二章 话者识别的基本原理、方法与 EBF 网络的研究	20
2.1 话者识别的方法与步骤	20
2.2 径向基函数(RBF)网络用于模式识别的 优越性及其基函数的选择	23
2.3 径向基函数(RBF)网络的拓展 ——椭圆基函数(EBF)网络	34
2.4 EBF 网络参数的优化问题	38
2.5 本章小结	39
第三章 扩展的 RPCL 算法与聚类数目的确定	41
3.1 传统聚类算法及存在的问题	41
3.2 次胜者受罚的竞争学习算法(RPCL): 思路、 步骤及其缺陷	43
3.3 RPCL 算法存在的问题及其改进	47
3.4 聚类和话者识别实验	50