DOI:10.13656/j.cnki.gxkx.2004.04.012
广西科学 Guangxi Sciences 2004, 11 (4): 320322

# 基于人工神经网络的汉语数字语音识别系统*

# The Mandarin Digit Speech Recognition Based on Artificial Nerual Networks

卢小春　　胡维平　　王修信
Lu Xiaochun　Hu Weiping　Wang Xiuxin

(广西师范大学物理与信息工程学院　桂林市育才路3号　541004)
(Coll. of Phy. &Info. Tech., Guangxi Normal Univ., 3 Yucailu, Guilin, Guangxi, 541004, China)

摘要　利用改进的有序聚类算法得到解决时间规整问题的新算法,在此基础上建立了基于人工神经网络的普通话数字语音识别系统。对基于人工神经网络的算法和基于动态时间伸缩的算法作比较识别实验,结果表明,基于人工神经网络的语音识别算法的识别性能优于传统的动态时间伸缩算法.

关键词　语音识别　人工神经网络　有序聚类算法　时间规整　孤立词
中图法分类号　TN912.34; TP391.42

**Abstract**　There is a time alignment problem in artificial neural network based speech recognition. In this paper, we adopt an improved sequential cluster method to solve the problem. A mandarin digit speech recognition system is established based on this method. The experiment results demonstrate that the improved time alignment method can improve the recognition rate effectively.

**Key words**　speech recogrition, artificial nerual network, sequential cluster, time alignment, isolated word

语音识别应用的模式匹配和模型训练技术主要有动态时间伸缩技术(Dynamic Time Warping,简称DTW)、隐马尔可夫模型(Hidden Markov Model,简称HMM)和人工神经网络(Artificial Neural Network,简称ANN)[1]. DTW算法是较早的一种模式匹配和模型训练技术,它应用动态规划方法成功解决了语音信号特征参数序列比较时时长不等的难题,在孤立词语音识别中获得了良好性能;HMM模型是语音识别技术中的主流,它的优点在于对动态时间序列有极强的建模能力,一般用于非特定人、大词汇量、连续语音的识别系统,但HMM方法存在分类决策能力弱,需要语音信号的先验统计知识等缺点;人工神经网络具有比较好的分类能力,模拟了人类神经元活动的原理,具有自学习、联想、对比、推理、分类和概括能力.人工神经网络为很好地解决语音识别模式分类问题提供了新的途径.数字(0-9)孤立词的识别是语音识别的一个基本问题,很多语音识别的新理论新方法往往都以此作为最初的尝试。本文构建了一个特定人孤立数字汉语语音识别系统,并将人工神经网络和DTW算法2种方案作为比较.

2003-12-12 收稿。
*广西师范大学2002年科研基金资助项目。

## 1　基于人工神经网络的语音识别

由于不同的汉语孤立词,或不同人说相同的汉语词语时,发音长短、清浊音比例等都是变化的(即输入汉语语音词组信号的帧数不同),而大多数神经网络分类器的输入结构是固定的,利用神经网络进行汉语孤立词语音识别时,存在着时间规整这一难题,这就意味着必须设法从可变长度的输入语音信号中提取相同帧数的特征矢量序列,才能满足分类器的使用要求.解决时间规整问题的方法主要有2类:一种是对语音信号进行线性时间规整;另外一种是采用非线性时间规整方法.但前者可导致相同词组中的音素或类音素无法对准,因此本文采用的是非线性时间规整算法来解决时间对准问题,有序聚类算法[2]就是效果较好的一种非线性时间规整算法,其思路是从语音信号特征矢量序列的局部出发,将相邻的且具有相似或相近声学特征的语音端进行合并,从而除去输入语音信号中的冗余信息,提取出对分类识别有用的语音特征.

时间规整的作用,一方面从输入不等长的语音信号特征矢量序列中提取固定长度的特征矢量序列;另

一方面是可大幅度降低特征矢量帧数，从而减小后端神经网络输入的节点数. 这种网络输入矢量的降维处理依靠合并若干语音帧实现，并非依靠减小语音的特征参数，因此这种降维方法对语音特征描述所造成的损失较小.

时间规整网络的结构如图 1 所示，假设网络的输入层有 $n$ 个节点，$n$ 是输入语音信号的帧数. 每个节点均有一个与之联系的输入矢量 $A_k^0(k = 1, 2, \cdots, n)$. $A_k^0$ 是第 $k$ 帧语音信号的特征矢量，进入第一层后，将距离最近的 2 个顺序矢量以加权平均合并，其余矢量不变，这样第一层就具有 $n-1$ 个节点以及与之联系的 $n-1$ 个矢量 $A_k^1(k = 1, 2, \cdots, n-1)$. 依此类推，在经过 $n-N$ 步合并后，最终网络的输出层具有 $N$ 个节点以及与之联系的 $N$ 个矢量 $A_k^{n-N}(k = 1, 2, \cdots, N)$. 时间规整网络对特征矢量序列的聚类合并过程，从整体上看又是对输入语音信号的一个分段过程.

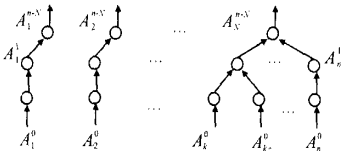

图 1　时间规整网络的结构

Fig. 1　Frame work of time wrapping network

其具体算法如下：

设 $A_1^0, A_2^0, \cdots, A_n^0$ 是输入语音信号的特征矢量，$m_k^i$ 表示矢量 $A_k^i$ 所代表的语音帧数，其中 $i = 0, 1, \cdots, n-N$, $k = 1, 2, \cdots, n-i$.

当 $i = 0$ 时有 $m_k^0 = 1(k = 1, 2, \cdots, n)$，并以 $d_k^i$ 表示矢量 $A_k^i$ 与 $A_{k+1}^i$ 之间的距离，即 $d_k^i = \| A_k^i - A_{k+1}^i \|$，从 $i = 0$ 开始重复下述过程直至 $i = n-N-1$.

步骤 **1**　计算 $d_k^i(k = 1, 2, \cdots, n-i-1)$，并找出 $j$，使得 $d_j^i < d_k^i$ 对所有的 $k \neq j$ 成立.

步骤 **2**　计算 $A_k^{i+1}(k = 1, 2, \cdots, n-i-1) =$

$$\begin{cases} A_k^{i+1} = A_k^i, & k < j, \\ A_k^{i+1} = \dfrac{m_k^i A_k^i + m_{k+1}^i A_{k+1}^i}{m_k^i + m_k^i + 1}, & k = j, \\ A_k^{i+1} = A_{k+1}^i, & k < j; \end{cases}$$

步骤 **3**　计算 $m_k^{i+1}(k = 1, 2, \cdots, n-1-1) =$

$$\begin{cases} m_k^{i+1} = m_k^i, & k < j, \\ m_k^{i+1} = m_k^i + m_{k+1}^i, & k = j, \\ m_k^{i+1} = m_{k+1}^i, & k > j. \end{cases}$$

## 2　系统模型的建立

系统模型的建立包括：语音数据的采样、特征向量的选择、语音数据的预处理、识别方案的确定. 本文采用 LPC 预测系数组成的特征向量. 语音信号的预处理包括分帧、预加重、加窗、求短时自相关系数，对语音信号预处理的一些参数指标为：采样速率：8kHz；量化精度：16 位；语音帧长：20ms；帧移：7.5ms；预加重系数：0.97；加窗函数：hamming 窗.

短时自相关函数阶数：10；LPC 参数是采用 Durbin 递推算法提取的，阶数为 10 阶.

### 2.1　基于人工神经网络的识别模型

神经网络算法系统流程图如图 2 所示，其中，神经网络的训练算法采用 BP 算法. 在本系统中，对语音信号的特征矢量序列进行时间规整作 2 种比较方案：一是从第一帧到最末帧都进行规整；二是保留开始前两帧和最末两帧不参与规整. 并进行实验比较，如表 1 所示，实验表明后一种方案明显地提高识别率.

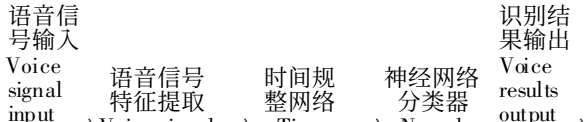


图 2　神经网络算法的识别流程图

Fig. 2　Flow chart of the recognition system of neural network algoritum

BP 网络由输入层、隐含层和输出层构成。(1)输入层：12×11 个节点，其中 12 为经过时间规整后的语音帧数，11 为每一帧的维数(即特征向量的阶数)，对应 12×11 输入阵列；(2)隐含层：30 个节点；(3)输出层：输出层作为分类器输出，取 10 个节点，对应 09 十个数字输出类别.

### 2.2　基于 **DTW** 算法识别模型

基于 DTW 算法的孤立词语音识别见文献[3, 4].

动态规划(DP)算法中有 4 种确定最佳路径的选择：固定起点、固定终点；固定起点、松弛终点；松弛起点、固定终点；松弛起点、松弛终点[3]. 在传统的算法中，松弛终点也是被限制在模板的最大帧数内. 由于考虑到某一测试模式大于参考模式但又属于该模板的情况，本文对松弛终点作了一定改进，将模板的长度放宽 45 帧.

动态时间伸缩算法的识别流程图如图 3 所示.

## 3　实验分析与讨论

本文进行人工神经网络技术的语音识别实验和动态时间伸缩(DTW)算法的识别实验，实验结果如表 1 所示. 实验中选取“0”“9”十个汉语语音为识别语音，每个数字重复说 30 次，总共 300 个样本，将 150

个样本作为训练集, 另外的 150 个样本作为测试集. 表2 为在采用神经网络的语音识别实验中 2 种时间规整方法的比较.

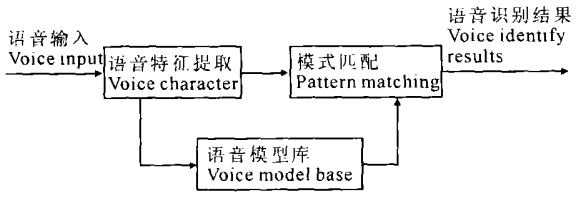


图 3 DTW 系统识别流程图

Fig. 3 The flow chart of the recognition system of DTW algorithm

表 **1** 人工神经网络和 **DTW** 算法

**Table 1 Neural network and DTW algorithm**

| 识别方案 Recognition project | 训练集的识别率 Recognition rate of trainning set(%) | 测试集的识别率 Recognition rate of testing set(%) |
|---|---|---|
| DTW 算法 DTW algorithm | 87.3 | 82 |
| 神经网络 Neural network | 100 | 96 |

表 **2** 神经网络的语音识别

**Table 2 The recognition system of neural network algorithm**

| 时间规整方法 Time wrapping algorithm | 训练集的识别率 Recognition rate of trainning set (%) | 测试集的识别率 Recognition rate of testing set (%) |
|---|---|---|
| 未改进的 Before improved | 100 | 84.7 |
| 改进后的 After improved | 100 | 96 |

表 1 的实验结果表明, 未改进的时间规整方法与改进后的时间规整方法在进行训练集的识别率都是 100%, 而对没有参加训练的测试集的识别率则有很大差别, 后者的效果非常明显. 分析其原因, 是由于在采样语音数据时所采用的端点检测算法并不是非常严格, 语音起点和终点的检测有一定的偏差, 即前面几帧和最后几帧中有可能包含非语音的信息, 它们之间的相关性不强, 如果进行合并就有可能丢失起始音和终止音的一些信息. 因此, 改进的有序聚类算法可以有效的保留语音发音变化区, 而压缩了语音的稳定区, 更加突出语音信号自身的特性, 提高识别率.

## 4 结束语

实验结果表明, 该系统利用有序聚类算法有效解决了神经网络语音识别中的时间规整难题, 系统识别性能明显得到改善. 同时证实了神经网络具有很强的分类、辨识能力, 其识别性能优于采用传统的 DTW 算法语音识别方法. 然而, 由于神经网络分类器的结构往往是固定的, 即必须首先将语音特征矢量序列进行时间规整统一长度之后才能馈入分类器进行识别, 因此, 目前人工神经网络在实时系统中的应用还有一定的难度.

## 参考文献

1 陈 方, 高 升. 语音识别技术及发展. 电信科学, 1996, 10: 5457.

2 史笑兴, 顾明亮, 王太君, 等. 有序聚类方法及其在神经网络语音识别中的应用. 电路与系统学报, 2000, 6: 99103.

3 杨行峻, 迟惠生. 语音信号数字处理. 北京: 电子工业出版社, 1995.

4 Rabine L R, et al. Fundamentals of Speech Recognition. Prentice-Hall Internation Inc, 1993.

(责任编辑: 黎贞崇)

## 中国科研人员主导性增强

中国科学技术信息研究所公布的“2003 年度中国科技论文统计结果”显示, 在 2003 年《科学引文索引》(SCI)收录的中国大陆的科技论文中, 国际合作产生的论文为 11739 篇, 占我国发表论文总数的 23.6%, 比 2002 年增加了 2.2 个百分点。我国作者为第一作者的国际合作论文 5942 篇, 合作伙伴涉及 67 个国家或地区; 其他国家作者为第一作者、我国作者参与工作的国际合作论文为 5797 篇, 合作伙伴涉及 65 个国家或地区。中国科学家为第一作者的国际合作论文首次超过其他国家作者为第一作者的论文, 表明在国际科技合作研究中, 中国科研人员的主导性正在增强。

“2003 年度中国科技论文统计结果”还显示, 19982002 年我国作者发表的国际论文中, 有 61.2%的论文在 5 年内间接被引用了 1 次, 其中被累计引用次数超过 100 次的有 18 篇。2003 年我国被 SCI 收录的国际论文被引用数, 由 2002 年的 24154 篇增加到 31168 篇; 被引用次数由 51766 次增加到 72131 次, 分别增长 29.0%和 39.3%, 表明我国科学家科技论文的影响力迅速增长。

据《科学时报》