

基于语音识别系统中 DTW 算法改进技术研究

Discussion of DTW Programming Improved Way On Speech Recognition

(重庆三峡学院)陈立万

Chen, Liwan

摘要: 动态时间规整 (Dynamic Time Warping) 是语音识别中的一种经典算法, 由于算法简单又有效, 因此在实现孤立词识别系统中获得了广泛的应用, 但仍然存在一些不足, 本文从提高识别率和识别速度入手提出对 DTW 算法进行改进研究, 以达到更满意的语音识别效果。

关键词: 语音识别; 端点检测; DTW 算法

中图分类号: TP391.42 **文献标识码:** A

Abstract: Dynamic Time Warping is a kind of classical programming in speech recognition system, because the programming is simple and valid, therefore, it is achieved extensive application in speech recognition of single word, but there are some questions, the paper raises improved ways from high distinguishable rate and speed, so that it will attain satisfied effect in speech recognition.

Keyword: Speech Recognition Beginning Inspect DTW Programming

技术创新

引言

通过语音传递信息是人类最重要, 最有效和最方便的交换信息的形式, 语音识别主要指让机器转达人说的话, 即在各种情况下, 准确地识别出语音的内容, 从而根据其信息, 执行人的各种意图。孤立词识别系统就是识别孤立发音的词, DTW 是把时间规整和距离测度计算结合起来的一种非线性规整技术, 它尽管对大词汇量、连续语音、非特定人语音识别效果较差, 但对孤立词语音识别, DTW 算法较为简洁, 正确识别率也较高, 因此, DTW 算法在孤立词语音识别系统有较广泛的应用, 因而对其进一步的研究以提高其正确识别速度具有很强的实用价值。

1 孤立词语音识别系统

孤立词识别系统, 一般是以孤立词为识别单位, 即直接取孤立词为识别基元, 孤立词语音识别系统流程可用下列框图 (1) 表示:

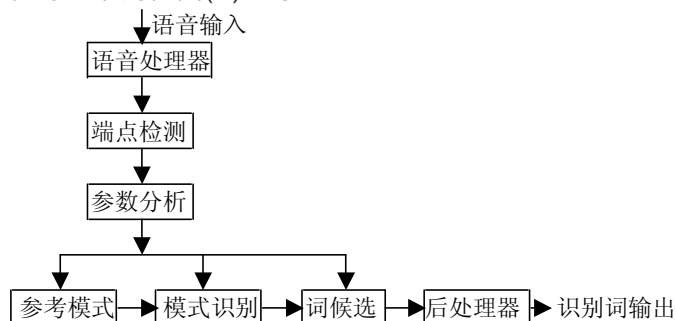


图 1 语音识别系统框图

其中模式识别部分是将输入语音特征参数信息与训练时预存的参数模型 (或模板) 进行比较匹配, 可用 DTW 技术法 (或 HMM 法), 这里着重讨论 DTW 技术方法, 它是把时间规整和距离测度计算结合起来的一种非线性规整技术, DTW 算法的质量就是运用动态规划的思想, 利用局部最佳化的处理来自动寻找一条路径, 两个特征矢量之间的积累失真量最小, 从而避免由于时长不同而可能引入的误差。即寻找一个规整函数 $i_m = \phi(i_n)$, 将测试矢量的时间轴 n 非线性地映射到参考模板的时间轴 m 上, 并使该函数满足:

$$D = \min_{i_n=1}^N d(T(i_n), R(\phi(i_m)))$$
。其中 D 是处于最优时间规整情况下两矢量的距离, 参数模板有 M 帧矢量 $\{R(1) \dots R(M)\}$, 测试模板有 N 帧矢量 $\{T(1) \dots T(N)\}$, $d(T(i_n), R(i_m))$ 是在 (i_n, i_m) 平面上矩形参考模板与测试模板之间的畸变度量, 即表示 T 中的第 i_n 帧特征与 R 中的第 i_m 帧特征之间的畸变度量, 其中 $i_n = \frac{N}{M} i_m$ 。

端点检测的目的是从包含语音的一般信号中确定出语音的起点以及终点, 有效的端点检测不仅能使处理时间最小, 而且能排除无声段的噪声干扰, 从而使识别系统具有良好的识别性能, 端点检测的成功与否甚至在某种程度上直接决定了整个语音识别系统的成败。

在孤立词语音识别系统中, 识别精度和计算复杂度的降低 (提高识别速度) 是主要的指标, 要解决这两个问题关键是特征的选择和提取, 失真测度的选择以及匹配算法的有效性。下面就如何提高识别精度和识别速度进行讨论。

陈立万: 副教授

2 DTW 算法改进技术

1、端点检测

(1) 双门限端点检测

用两个门限进行端点检测,高门限被超过基本值可以确定语音的起点端点,低门限用于确定语音真正的起点端点,但只是低门限被成功检测未必就是语音的起始端点,也有可能是短时的噪音;当高门限已经确定语音起始端点后,再返回去利用低门限才能确定语音的真正起点;语音结束点与起始点判别类似,有时噪声的能量相当大,可能超过高门限,但是噪声一般持续时间比较短,可以用持续时间来决定噪声还是语音。

(2) 动态窗长语音端点检测

语音端点检测时,语音为静音段,采用较长的窗;语音和静音的过渡段时采用较小的窗,可以确切判断语音的起始点;一旦确定语音的起始点就改用常规窗长,因汉语音节末尾都是浊音,只用短时能量就可以较好地判断一个语音的末点,有流程图(2)为:

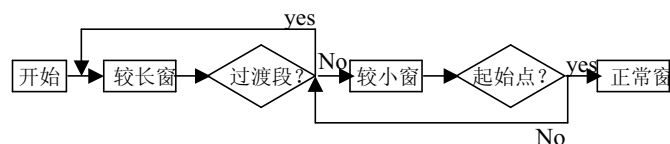


图2 端点检测动态窗流程图

(3) 噪音下的端点检测

实际中的语音识别系统有一定的噪声,当噪声增加时,单纯用过零率和能量值积作为检测端点,端点的检测准确率将下降,若采用当前帧与前一帧的过零率与能量值积的差作为检验端点,便可提高端点检测的正确率。

2、DTW 算法中的整体路径约束

动态 DTW 搜查方法需要存储较大的矩阵,计算量大,在实际应用中,可使路径约束斜率在 1/2-2 的范围,如图(3)所示:

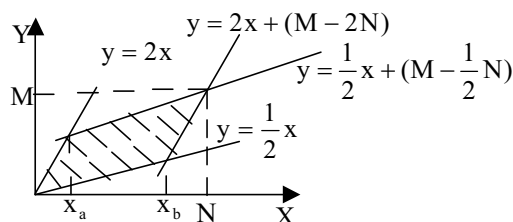


图3 整体、路径约束图

这时待测模板的每一帧不再需要与参考模板的每一帧进行比较,只需与阴影菱形范围之内的帧比较。

当 $X_a = X_b$ 时:

$$\begin{cases} \frac{1}{2}x \rightarrow 2x & x \leq x_a \\ 2x + (M-2N) \rightarrow \frac{1}{2}x + (M-\frac{1}{2}N) & x \neq x_a \end{cases}$$

$x_a \neq x_b$ 时比较分三段:

$$\begin{cases} \frac{1}{2}x \rightarrow 2x & x \leq x_a \\ \frac{1}{2}x \rightarrow \frac{1}{2}x + (M-\frac{1}{2}N) & x_a \neq x_b \\ 2x + (M-2N) \rightarrow \frac{1}{2}x + (M-\frac{1}{2}N) & x \neq x_b \end{cases}$$

当 $x_a \neq x_b$ 时:

$$\begin{cases} \frac{1}{2}x \rightarrow 2x & x \leq x_b \\ \frac{1}{2}x \rightarrow \frac{1}{2}x + (M-\frac{1}{2}N) & x_b \neq x_a \\ 2x + (M-2N) \rightarrow \frac{1}{2}x + (M-\frac{1}{2}N) & x \neq x_a \end{cases}$$

这样,路径的累积距离为:

$D(i_n, i_m) = d(T(i_n), R(i_m)) + \min\{D(i_{n-1}, i_m), D(i_{n-1}, i_{m-1}), D(i_{n-1}, i_{m+1})\}$, 由于 X 轴上每前进一帧,只需要用到前一列的累积距离,所以只需两个矢量 D 和 d 分别保存前一列的累积距离和当前列的累积距离,而不用保存整个矩阵。通过不断更新数据来实现,一直进行到待测模板的最后一帧,矢量 D 的最后一个元素即为两个模板经过动态规划后的匹配距离,这样可大大减小存储空间,减小计算量,从而提高识别速度。

3、搜索宽度受限的 DTW 算法

在实际应用中 DTW 算法加入了一些搜索限制条件,实际的搜索范围是在一定的宽度之内,如图(4)所示的对角线附近的带状区域。

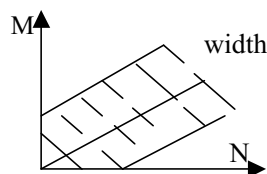


图4 对角线附近的带状区域图

在这个范围之内按动态规划路径计算累积匹配距离,可以进一步减少存储空间,减少计算量,提高识别速度。

4、放宽端点限制的 DTW 算法

对于普通 DTW 对端点检测比较敏感,端点信息是作为一组独立的参数提供给识别算法,它要求两个比较模式起点对起点,终点对终点,对端点检测的精度要求比较高,当环境噪声比较大或语音由摩擦音构成时,端点检测不易进行,放宽端点限制方法不严格要求端点对齐,克服了由于端点算法不精确造成的测试模式和参考模式起点终点不能对齐的问题,一般情况下,起点和终点在纵横两个方向只要放宽 2-3 帧就可以,也就是起点可以在 (1,1)、(1,2)、(1,3)、(2,1)、(3,1),终点也可类似放松,放宽端点后的区域限制图(5):

在放松端点限制的动态时间规整算法中,累积距离矩阵中的元素 (1,1)、(1,2)、(1,3)、(2,1)、(3,1),不

是根据局部判决函数计算得到的,而是直接将帧匹配距离矩阵的元素填入,自动从其中选择最小的一个作为起点,对于终点也是从松弛终点的允许范围内选择一个最小值作为参考模式和未知模式的匹配距离。

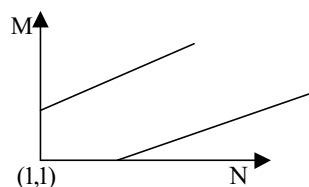


图5 放宽端点后的区域限制

3 实验研究

上述算法改进均可通过实验进行证明,这里不一一说明,现就整体路径约束的 DTW 算法的两个实验进行比较可以看出其性能的优劣。实验一是对语音控制媒体播放系统指令词的识别,其中包括:播放、前进、停止、前进、倒退、打开、关闭。实验二是对单个字放、停、前、后、开、关的识别。实验所采用的数据都是通过计算机的声卡录音,采样频率是 8000HZ,语音数据在不同时间录音。实验一中的数据是来自一个女声发音的 30 组样本,也就是每个指令词有 30 个不同的发音,总共是 180 个样本。实验二中的数据是来自一个女声发音的 20 组数据,总共 120 个语音作为实验数据。每个实验都以第一组的发音作为训练模板,其余样本作为待测数据。语音特征参数采用基于听觉系统的 MFCC 及其相应的差分倒谱参数,每帧语音提取 24 个特征参数。实验一和实验二采用不同样本集时对应传统 DTW 算法和整体路径约束 DTW 算法的识别率如表

识别率比较表

正确率	高效 DTW 算法	传统 DTW 算法
实验一	100%	100%
实验二	96%	98. 90%

虽然区域限制的 DTW 算法识别率稍低于传统的 DTW 算法,但是在识别率接近的情况下其识别速度明显优于传统 DTW 算法。在实验一中用区域限制的 DTW 算法识别出一个待测模式大概需要 2 秒(matlab 环境, p 533 的 CPU, 128M 的 RAM),而普通的 DTW 算法大概需要 3 倍的时间。实验验证整体路径约束的 DTW 算法能够减少匹配算法的计算量,提高了识别速度,提高了系统的实用性。

结束语

另外对 DTW 算法可精心加权改进,若从整体上加权,它能实现每一个短时畸变对整体相似度量度的贡献,若从局部约束路径进行加权,它能提高识别效果;若采用非线性分段得到两个模板之间的对应点考虑,时间规整函数某些点的值可事先得到,利用这些

先验知识大大缩小了 DTW 算法的搜索空间,加快匹配速度;采用二级识别,以减少识别速度与识别率的矛盾,区分易混淆的词,先用精确度低速度快的算法,从所有参考词中选出少数词,随后用精确度高速度慢的方法在筛选出的少量词中确定最后结果;某一种方法对某一部分词识别效果不佳,再利用第二种不同方法对这些词的其他特点进行检测区分,从而增大识别正确率。

由于对语音识别的需要来自不同的领域和部门,因此,对语音识别系统应具有的性能和指标提出了差异极大的要求,语音识别系统要根据一定的指标要求设计,每一种方法都有各自的特点,需要针对不同的识别任务,不同参考构成的模式及其特点选择最为合适的方法。

参考文献:

- [1]吕孟军,郭琪,张家峰等.航空发动机状态智能识别[J].微计算机信息,2005,2: 60-62.
- [2]胡光锐,语音处理与识别,上海科学技术文献出版社,1994.
- [3]赵力,语音信号处理,机械工业出版社,2003.
- [4](美)Parson T W.语音处理.文成义等译,国防工业出版社,1996.
- [5](美)Rabiner L R.等.Fundamentals of Speech Recognition.清华大学出版社,1999.

作者简介:陈立万,男,(1964-),汉族,本科,副教授,主要从事语音信号处理研究. E-mail:chenliwan@163.net (404000 重庆三峡学院应用技术学院)陈立万

(Chong qing three gorges college 404000) chen,Liwan
(投稿日期:2005.8.6) (修稿日期:2005.8.16)

(接 91 页)

- [2]亨特(Hunt,C.),《TCP/IP 网络管理》;翟炯等译,电子工业出版社,1997.8
- [3]柳宁,孟高勇,胡捍英等.基于不可分离多径瑞利信道的最佳二元检测[J].微计算机信息,2005,4:180-181
- [4]杨锐,《移动设备.NET 应用程序设计》,机械工业出版社,2003.5
- [5]李永隆,《PDA 程序设计》,清华大学出版社
- [6]汪晓平,钟军,《Visual Basic 网络通信协议分析与应用实现》,人民邮电出版社,2003.1

作者简介:臧佳锋,男,1977-,汉族,在职硕士,从事嵌入式系统设计与开发工作. email: zangjf_2008@yahoo.com.cn.

Author brief introduction: Zang, Jiafeng, Male,1977-, the Han nationality,The graduate in active service,majoring in the design and development of embedded system. (214122 江苏无锡江南大学信息工程学院) 臧佳锋 戴月明

(The School of Information Engineering, Southern Yangtze University, Wuxi, 214122) Zang,Jiafeng Dai,Yueming

联系方式: (214122 无锡市蠡湖大道 1800 号信息工程学院) 臧佳锋

(投稿日期:2005.6.26) (修稿日期:2005.7.6)