

文章编号: 2095-4980(2015)06-0942-05

改进的高效动态时间规整算法语音识别系统

王新胜¹, 巩捷甫¹, 喻明艳²

(1. 哈尔滨工业大学 信息与电气工程学院, 山东 威海 264209; 2. 浙江大学 宁波理工学院, 浙江 宁波 315100)

摘要: 动态时间规整算法是结合了动态时间规整(DTW)技术和距离测度计算技术的一种非线性规整算法, 在语音识别模板匹配中有重要的应用。为此提出一种改进的高效动态时间规整算法, 其能有效加快搜索路径的寻找。基于 Matlab 实现了隐马尔可夫算法、高效动态时间规整算法和改进的高效动态时间规整算法的语音识别系统, 同时进行了算法的仿真实验。实验结果表明, 基于改进高效动态时间规整算法的训练速度远大于基于隐马尔可夫算法和高效动态时间规整算法的训练速度, 而识别率下降很小, 对于小词汇量非连续语音识别中高效动态时间规整算法的识别率为 97.56%, 隐马尔可夫算法的识别率为 97.14%, 改进高效动态时间规整算法的识别率为 96.43%。

关键词: 语音识别; 动态时间规整; 隐马尔可夫

中图分类号: TN912.34; TP391.42 **文献标识码:** A **doi:** 10.11805/TKYDA201506.0942

Improved high efficiency Dynamic Time Warping for speech recognition

WANG Xinsheng¹, GONG Jiefu¹, YU Mingyan²

(1. School of Information and Electrical Engineering, Harbin Institute of Technology, Weihai Shandong 264209, China;
2. Ningbo Institute of Technology, Zhejiang University, Ningbo Zhejiang 315100, China)

Abstract: The Dynamic Time Warping(DTW) algorithm is a nonlinear warping algorithm combined with dynamic time warping technique and distance measurement computing method. It has an important application in speech recognition based on template matching. This paper proposes an improved efficient dynamic time warping algorithm, which can effectively speed up the path searching in speech recognition. Based on Matlab, the hidden Markov algorithm, efficient dynamic time warping algorithm and improved dynamic time warping algorithm speech recognition systems are implemented. Simulation experiments based on above algorithms are performed. The experimental results show that the training speed of improved efficient dynamic time warping algorithm is much faster than that of efficient dynamic time warping algorithm and hidden Markov algorithm, but the recognition rate of decline based on improved efficient dynamic time warping algorithm is very low. For small vocabulary continuous speech recognition, the recognition rate is 97.56% for efficient dynamic time warping algorithm, 97.14% for hidden Markov recognition algorithm, and 96.43% for improved efficient dynamic time warping algorithm respectively.

Key words: speech recognition; Dynamic Time Warping; hidden Markov

语音识别的研究开始于 20 世纪 50 年代初期^[1], 第 1 个用来识别英文数字的语音识别系统在贝尔实验室研制出来。20 世纪 60 年代, 随着计算机技术的发展, 语音识别研究也得到了发展, 这个阶段主要解决语音识别中语音信号产生模型问题, 1959 年英国的 Fry 等人研究成功了第 1 个计算机语音识别系统^[2]。20 世纪 70 年代以后, 语音识别技术得到了迅猛发展, 在理论方面, 线性预测分析技术得到进一步的发展, 动态时间规整技术^[3]趋于成熟, 与此同时提出了矢量量化技术^[4-5]和隐马尔可夫模型(Hidden Markov Model, HMM)技术^[6-7]。根据识别系统的识别词汇量大小, 可将语音识别系统分为 3 类: 小词汇量语音识别系统、中等词汇量语音识别系统以及大量词汇量的语音识别系统。本文主要研究小词汇量的语音识别系统, 这个系统通常包括几个至几十个识别词汇, 主要用于语音密码。语音识别过程包括语音识别的前期处理步骤, 如语音信号的采集与数字化、语音信号的预处理和语音信号的特征参数提取, 语音信号后期匹配, 如动态时间规整算法、矢量量化算法和隐马尔可夫算法等。

收稿日期: 2014-12-03; 修回日期: 2014-12-26

基金项目: 国家自然科学基金资助项目(61201307)

在小词汇量语音识别系统中,最为简单而有效的后期匹配算法为动态时间规整(DTW)算法,该算法通过动态规划的思想,很好解决了发音中长短不一致的模板匹配问题。本文在高效 DTW 算法的基础上,通过改进其算法的匹配路径范围,使识别速度在识别率降低很小的情况下大幅度提高。

1 高效 DTW 算法原理

语音识别过程本质上是一个语音信号的模式识别过程^[8],其主要包括训练和识别这 2 个过程,但不管是前期的训练还是后期的测试识别,在语音信号数字化处理后对语音信号的预处理是一项非常重要的工作^[9]。预处理的目的是消除原始语音信号中的噪声,从而获得比较纯净的训练和识别的样本数据。语音信号的处理技术包括预加重、分帧加窗以及端点检测等。无论在训练和建立模板阶段还是在识别阶段,都首先采用端点法确定语音的起点和终点。DTW 算法主要用于后端匹配处理,下面给出其简单介绍。

假设训练语音信号 R 被分成 m 帧数据, $R(m)$ 为训练者语音的第 m 帧数据,测试语音信号 T 被分成 n 帧数据, $T(n)$ 为测试者的第 n 帧数据。训练者语音每帧数据与测试者语音每帧数据可以进行距离测度计算,度量的结果表征其相似度,即距离越小,相似度越高。如果 $n=m$ 则可以直接计算,否则需要将 $T(n)$ 和 $R(m)$ 对齐,考虑到语音中各帧数据在不同情况下的持续时间会产生或长或短的变化情况,一般采用基于动态规划的方法。若把测试模板的各个帧在 1 个二维直角坐标系中的横轴上标出,把训练模板的各个帧在纵轴上标出,通过这些表示帧号的整数坐标画出一一些纵横线的网格,网格上的每 1 个交叉点 (n, m) 表示测试模板中某一帧与训练模板中某一帧的交汇点。为了进行测度计算,首先必须找到一条通过此网格的路径,路径通过的格点即为测试和训练模板中进行计算的帧号。路径的寻找不是随意的,因为虽然任何一种语音的发音快慢都有可能变化,但是其各部分的先后次序不可能改变,因此所选的路径必定是从左下角出发到右上角结束。假设一条可能的路径通过的格点为 $(n_1, m_1), (n_2, m_2), \dots, (n_i, m_i), (n_n, m_m)$, 而路径可用离散分段函数表示为式(1)。为了使路径不至于过分倾斜,约束斜率在 $0.5 \sim 2$ 的范围内。假定路径已经经过了网格点 (n_{i-1}, m_{j-1}) 则下一个网格点满足式(2)的约束。因此,求上述问题的最佳路径问题可归结为满足式(2)的约束下,求式(1)所示最佳函数,使得沿最佳路径的累积距离达到最小。DTW 算法不断地搜索计算测试帧矢量与训练模板帧矢量的距离以寻找最优的匹配路径,这保证了它们之间存在最大的相似性。

$$m_i = \Phi(n_j) \quad (1)$$

$$\begin{cases} (n_i, m_j) = (n_{i-1} + 1, m_{j-1} + 2) \\ (n_i, m_j) = (n_{i-1} + 1, m_{j-1} + 1), \text{约束斜率在 } 0.5 \sim 2 \\ (n_i, m_j) = (n_{i-1} + 1, m_{j-1}) \end{cases} \quad (2)$$

上面说明了 DTW 算法的基本原理,依据统计结果,一般情况下最佳路径都在一定的范围内,所以在实际的应用中,为了减少计算量,提高匹配速度,一般将搜索区间限制在一定的范围内。图 1 中将区间限制在边长斜率为 $1/2 \sim 2$ 的平行四边形范围内^[10]。

高效 DTW 算法的匹配路径,将区间缩小至图 1 中区域,分别用斜率为 $1/2$ 和 2 的 2 条直线以及起始点及终止点包裹着整个搜寻区域。经过计算得到 x_a 和 x_b , 见式(3),

$$\begin{cases} x_a = (2m - n) / 3 \\ x_b = (2n - m) / 3 \end{cases} \quad (3)$$

式中: x_a 为 A 点横坐标; x_b 为 B 点横坐标; n 和 m 分别为测试语音帧数和训练语音帧数。

当 $x_b > x_a$, 搜索区域被限定在如下 3 个区域内:

$$\begin{cases} y_{\min} = 0.5x \\ y_{\max} = 2x \end{cases}, \quad x = 1 : x_a \quad (4)$$

$$\begin{cases} y_{\max} = 0.5(x - n) + m \\ y_{\min} = 0.5x \end{cases}, \quad x = (x_a + 1) : x_b \quad (5)$$

$$\begin{cases} y_{\max} = 0.5(x - n) + m \\ y_{\min} = 2(x - n) + m \end{cases}, \quad x = (x_b + 1) : n \quad (6)$$

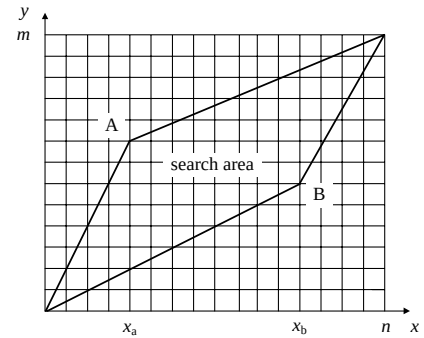


Fig.1 Schematic of constrained search path for DTW algorithm
图 1 DTW 算法约束搜索路径示意图

当 $x_a > x_b$ 时, 搜索区域被限定在如下 3 个区域内:

$$\begin{cases} y_{\max} = 2x \\ y_{\min} = 0.5x \end{cases}, \quad x = 1: x_b \quad (7)$$

$$\begin{cases} y_{\max} = 2x \\ y_{\min} = 2(x-n) + m \end{cases}, \quad x = (x_b + 1): x_a \quad (8)$$

$$\begin{cases} y_{\max} = 0.5(x-n) + m \\ y_{\min} = 2(x-n) + m \end{cases}, \quad x = (x_a + 1): n \quad (9)$$

当 $x_a = x_b$ 时, 搜索区域被限定在如下 2 个区域内:

$$\begin{cases} y_{\max} = 2x \\ y_{\min} = 0.5x \end{cases}, \quad x = 1: x_a \quad (10)$$

$$\begin{cases} y_{\max} = 0.5(x-n) + m \\ y_{\min} = 2(x-n) + m \end{cases}, \quad x = (x_a + 1): n \quad (11)$$

2 改进高效 DTW 算法搜索匹配路径

改进高效 DTW 算法搜索匹配路径进一步缩小到这样一个区域, 分别用斜率为 2/3 和 3/2 的 2 条直线以及起始点及终止点包裹出整个搜寻区域。经过计算得到以下 2 点, 见式(12)。

$$\begin{cases} x_a = (m - 2n/3)6/5 \\ x_b = (3n/2 - m)6/5 \end{cases} \quad (12)$$

当 $x_b > x_a$ 时, 整个的搜索区间被限制在如下区域中:

$$\begin{cases} y_{\min} = 2x/3 \\ y_{\max} = 3x/2 \end{cases}, \quad x = 1: x_a \quad (13)$$

$$\begin{cases} y_{\max} = 2x/3 + m - 2n/3 \\ y_{\min} = 2x/3 \end{cases}, \quad x = (x_a + 1): x_b \quad (14)$$

$$\begin{cases} y_{\max} = 2x/3 + m - 2n/3 \\ y_{\min} = 3x/2 + m - 3n/2 \end{cases}, \quad x = (x_b + 1): n \quad (15)$$

当 $x_a > x_b$ 时, 整个的搜索区间被限制在如下区域中:

$$\begin{cases} y_{\max} = 3x/2 \\ y_{\min} = 2x/3 \end{cases}, \quad x = 1: x_b \quad (16)$$

$$\begin{cases} y_{\max} = 3x/2 \\ y_{\min} = 3x/2 + m - 3n/2 \end{cases}, \quad x = (x_b + 1): x_a \quad (17)$$

$$\begin{cases} y_{\max} = 2x/3 + m - 2n/3 \\ y_{\min} = 3x/2 + m - 3n/2 \end{cases}, \quad x = (x_a + 1): n \quad (18)$$

当 $x_a = x_b$ 时, 整个的搜索区间被限制在如下区域中:

$$\begin{cases} y_{\max} = 3x/2 \\ y_{\min} = 2x/3 \end{cases}, \quad x = 1: x_a \quad (19)$$

$$\begin{cases} y_{\max} = 2x/3 + m - 2n/3 \\ y_{\min} = 3x/2 + m - 3n/2 \end{cases}, \quad x = (x_a + 1): n \quad (20)$$

3 实验结果与分析

3.1 语音识别系统实现及测试

在 Matlab 环境下, 实现了高效 DTW 算法和改进高效 DTW 算法, 具体流程见图 2。为了测试改进的高效 DTW 算法的有效性, 请一位女士录制 0~9 这 10 个数字, 每个数字分别录制 3 个 wav 格式的音频文件, 共计 30 个语音文件, 逐一进行测试, 全部识别正确, 其识别速度对比见表 1。

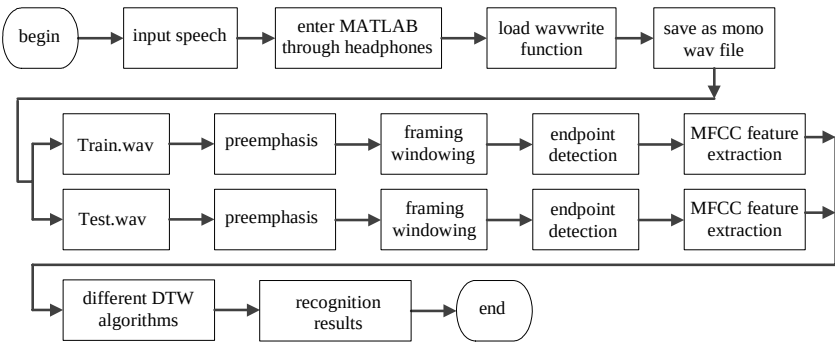


Fig.2 Flow chart of Matlab-based DTW speech recognition system
图 2 基于 Matlab 的 DTW 相关语音识别系统流程图

表 1 0~9 数字不同 DTW 算法识别速度对比

Table1 Comparison of different DTW algorithm recognition speeds for number 0-9

number	0	1	2	3	4	5	6	7	8	9
efficient DTW recognition time/s	5.41	9.05	7.18	8.90	9.23	11.21	5.91	10.96	10.83	11.10
improved efficiency DTW recognition time/s	2.96	4.56	3.68	4.66	4.98	6.20	3.25	6.06	6.15	6.20

从表 1 中能够看出改进的高效 DTW 算法识别速度提高了 43%。同时还设计了说话人识别实验,共 7 个人分别进行训练和测试,结果也表明改进的 DTW 算法在识别率变化很小情况下,识别速度有较明显的调高。

3.2 不同匹配算法性能比较

这里为了分析改进的高效 DTW 算法的优势,进行了高效 DTW 算法、改进高效 DTW 算法和 HMM 算法的对比分析,具体包括特定人数字识别比较(结果见表 2)、特定人汉字识别比较(结果见表 3)和非特定人汉字识别比较(结果见表 4)。

表 2 各种识别算法的数字识别比较

Table2 Comparison of the digital recognition of different recognition algorithms

algorithm	training time/s	recognition time/s	recognition rate/%
efficient DTW algorithm	1.663 68	30.827 8	97.149 2
improved efficiency DTW algorithm	1.389 20	20.787 3	90.000 0
HMM algorithm	273.000 00	28.895 5	88.571 4

从表 2 可以看出,在训练时间方面,由于 HMM 算法使用迭代收敛的方法进行训练,因此基于 HMM 算法的时间最长,远高于其他 2 种方法;在识别时间方面,HMM 算法的识别时间短于高效 DTW 算法而长于改进高效 DTW 算法,这说明改进的效果明显。在识别率方面,由于改进高效的 DTW 算法减小了搜索区域,所以有可能漏掉正确的匹配路径,所以识别率比高效 DTW 算法低。

表 3 各种识别算法的特定人汉字识别比较

Table3 Comparison of the specific characters of different recognition algorithms

algorithm	training time/s	recognition time/s	recognition rate/%
efficient DTW algorithm	1.44	19.11	97.14
improved efficiency DTW algorithm	1.01	12.82	96.43
HMM algorithm	275.50	18.31	97.86

表 3 中的数据是基于 10 个汉字的识别结果(2 个不同的人识别结果的平均值),这 10 个汉字均为 2 个字的短语。从表 3 中的数据能进一步看出,改进高效 DTW 算法的识别和训练速度较其他方法提高很大,但识别率降低很少(降低 1%左右)。

表 4 各种识别算法的非特定人汉字识别比较

Table4 Comparison of the non-specific characters of different recognition algorithms

algorithm	training time/s	recognition time/s	recognition rate/%
efficient DTW algorithm	1.69	30.68	70.00
improved efficiency DTW algorithm	1.41	20.31	65.71
HMM algorithm	286.00	28.76	51.43

表 4 给出基于几种不同算法的语音识别系统对非特定人汉字识别情况,采用完全相同的语音进行训练,识别也是利用完全相同的语音进行测试识别,其中训练语音是 1 个人,而测试语音是另外 1 个人。从表 4 也能够看出改进高效 DTW 算法的识别速度较其他方法有较大提高,而识别率降低很少。

4 结论

语音识别技术在未来信息通信中将发挥越来越重要的作用,人机通过语音实现信息交互越发具有发展潜力。本文对高效 DTW 算法的语音识别技术进行改进,其有效缩短 DTW 算法搜索路径的时间。基于 Matlab 实现了高效 DTW 算法、改进高效 DTW 算法和 HMM 算法,并且进行了简单数字识别、特定人简单汉字识别和非特定人简单汉字识别的实验。实验结果证明改进高效 DTW 算法在识别率降低很少情况下,有效提高训练和识别速度。

参考文献:

- [1] Davis K H, Balashek S A. Recognition of spoken digits[J]. Journal of the Acoustical Society of America, 1952, 24(6): 637-642.
- [2] Fry D B, Denes P. Theoretical aspects of mechanical speech recognition[J]. Journal of the British Institution of Radio Engineers, 1959, 19(4): 211-229.
- [3] Sakoe H, Chiba S. Dynamic programming algorithm optimization for spoken word recognition[J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1978, 26(1): 43-49.
- [4] Gray R. Vector Quantization[J]. IEEE ASSP Magazine, 1984, 1(1): 4-29.
- [5] Makhoul J, Roucos S, Gish H. Vector Quantization in Speech Coding[J]. Proceedings of the IEEE, 1985, 73(11): 1551-1588.
- [6] Richter A. Modeling of Continuous Speech Observations[C]// Advances in Speech Processing Conference. Zurich, Switzerland: IBM Europe Institute, 1986: 18-24.
- [7] Ney H, Noll A. Phoneme modelling using continuous mixture densities[C]// IEEE International Conference on Acoustics, Speech and Signal Processing. New York: IEEE, 1988: 437-440.
- [8] Chanwoo K, Seo K D. Robust DTW-based recognition algorithm for hand-held consumer devices[J]. IEEE Transactions on Consumer Electronics, 2005, 51(2): 699-709.
- [9] 高维深. 基于 HMM/ANN 混合模型的非特定人语音识别研究[D]. 成都: 电子科技大学, 2013. (GAO Weishen. Study of speaker-independent speech recognition base on HMM/ANN model[D]. Chengdu, Sichuan, China: University of Electronic Science and Technology of China, 2013.)
- [10] 姜天辰, 夏仁平. 放宽端点限制的 DTW 法用于孤立词识别[J]. 计算机工程与应用, 1998, 56(2): 57-66. (JIANG Tianchen, XIA Renping. Unlimited on the endpoint of the DTW method used for isolated word recognition[J]. Computer Engineering and Applications, 1998, 56(2): 57-66.)

作者简介:



王新胜(1978-), 男, 山东省威海市人, 博士, 主要从事太赫兹集成电路设计及嵌入式研究开发工作. email: xswang@hit.edu.cn.

巩捷甫(1991-), 男, 山东省威海市人, 硕士, 主要研究方向为计算机通信.

喻明艳(1964-), 男, 杭州市人, 教授, 主要研究方向为多核处理器架构和并行编程.