

基于改进 DTW 的机器人语音指令识别系统研究

胡亚洲¹, 王新民¹, 曹江涛²

(1. 西北工业大学 自动化学院, 陕西 西安 710129;

2. 辽宁石油化工大学 信息与控制工程学院, 辽宁 抚顺 113001)

摘 要:动态时间规整(Dynamic Time Warping)是语音识别中的一种经典算法,该算法简单有效,在实现孤立词识别系统中得到了广泛的应用。为了提高机器人语音识别系统的识别率和识别速度,文中采用了一种改进的 DTW 语音识别算法。在 MATLAB 7.0 环境下,对改进的语音端点检测和改进的 DTW 算法进行仿真实验,实验证明改进的算法提高了识别率,并且减少了识别所用的时间;将该算法移植到机器人上,在安静的环境下进行试验,结果表明机器人能准确而又快速地识别语音内容。最后,得到了改进的语音识别算法能够有效提高识别率和识别速度的结论。

关键词:孤立词语音识别;动态时间规整;语音端点检测

中图分类号:TP31

文献标识码:A

文章编号:1673-629X(2013)07-0070-03

doi:10.3969/j.issn.1673-629X.2013.07.017

Study of Voice Command Recognition System for Robot Based on Improved DTW

HU Ya-zhou¹, WANG Xin-min¹, CAO Jiang-tao²

(1. College of Automation, Northwestern Polytechnical University, Xi'an 710129, China;

2. College of Information and Control Engineering, Liaoning Shihua University, Fushun 113001, China)

Abstract:Dynamic Time Warping (DTW) is a kind of classical algorithm in speech recognition system. Because of the simple and valid, DTW method has been widely used in speech recognition, especially for the recognition of isolated word. In order to achieve the purpose of increasing the recognition rate and recognition speed of robot voice recognition system, an improved DTW speech recognition algorithm is performed here. Improved voice endpoint detection and improved DTW algorithm simulation experiment has been presented in this proposed identification algorithm in MATLAB 7.0, the experimental results show that the proposed method can improve the recognition rate and reduce the recognition time. Finally, the algorithm is transplanted to the robot, in a quiet environment the practical application turns out that the robot can quickly identify the content of the speech. In conclusion, the improved speech recognition algorithm can effectively improve the recognition rate and recognition speed.

Key words:isolated word speech recognition; DTW; voice endpoint detection

0 引言

近年来,语音识别理论研究飞速发展,其相应的现实应用大大地提高了人们的工作效率和生活质量。常用的语音识别算法有隐马尔科夫(HMM)算法、人工神经网络(ANN)算法、动态时间规划(DTW)算法、向量量化(VQ)算法等。HMM算法适用于大词汇量、连续语音识别,其不足在于结构复杂,需要大量的训练样本^[1,2]。ANN算法人们还在研究,其缺点是当训练数

据较多的时候,学习速度会非常慢。VQ算法优点是识别计算量较小,缺点是没有利用语音特征序列的时间顺序从而对识别率有较大的影响。而DTW是一种小词汇量,孤立词语音识别的相对较成熟的技术,它是把时间规整和间距离计算联系在一起的非线性动态规划算法,相对比较简单而且有效^[3]。该算法的最大优点在于解决了语音信号具有较大随机性的关键问题^[4,5]。

语音识别先要进行语音端点检测,准确地判断出

收稿日期:2012-10-08

修回日期:2013-01-12

网络出版时间:2013-04-08

基金项目:辽宁省高校创新团队支持计划项目(LT2010058)

作者简介:胡亚洲(1987-),男,硕士研究生,研究方向为控制系统建模及仿真;王新民,教授,硕士,研究方向为飞行控制与仿真技术、先进控制理论及应用。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20130408.1709.047.html>

实时输入的声音信号的起始点和终止点;再进行特征参数提取,建立声学模型;最后进行匹配和识别^[6]。文中针对基于改进的 DTW 孤立词语音识别提出了一些改进的算法:首先,在语音端点检测中采用了将动态改变窗长、改变短时过零率门限的方法相结合的端点检测算法;其次,在语音特征参数提取中采用梅尔频率倒谱系数(MFCC)算法;最后,在语音识别模板匹配中,采用一种改进的 DTW 算法。通过仿真实验对改进的端点检测算法和改进的 DTW 算法进行了验证,实验表明改进算法提高了识别率和识别速度。

1 语音指令识别系统组成

1.1 语音端点检测

端点检测就是准确地判断出实时输入的语音信号的起始点和终止点。语音端点检测是语音识别中非常重要的一个环节。常用的端点检测算法有将语音能量和过零率相结合的双门限算法和将短时能量和过零率的乘积构成能频值算法^[7]。而文中采用了一种将动态改变窗长、改变短时过零率门限的方法结合起来的算法。

1.2 语音特征参数提取

在语音信号的频域特征参数中,目前使用最广泛的是线性预测倒谱系数(LPCC)和梅尔频率倒谱系数(MFCC)。文中采用了 MFCC,因为 MFCC 充分考虑了人的听觉特性,而且没有任何前提假设,MFCC 参数具有良好的识别性能和抗噪能力,但其计算量和计算精度要求高^[8]。

1.3 语音模板匹配

模板匹配是语音识别系统中的一个很重要的模块,也得到了特别的重视。常用的技术有三种:隐马尔科夫(HMM)算法、向量量化(VQ)算法、动态时间规整(DTW)算法。文中采用了一种在一般 DTW 算法上改进的 DTW 算法。

2 改进方案的设计

2.1 改进的端点检测

在语音识别中,端点检测是很关键的一步^[9]。它的准确性很大程度上直接影响着后续的语音识别工作能否有效地进行,而且 DTW 算法对端点检测的准确度有很高的要求。

为了达到提高端点检测精度和处理速度的目的,提出了一种改进的语音端点检测算法:将动态改变窗长和改变短时过零率门限相结合的方法,即在语音信号的静音段采用大窗长、高门限;在语音信号的过渡段采用小窗长、低门限;确定语音信号的起点和终点后采用常规窗、中等门限(如图 1 所示)。

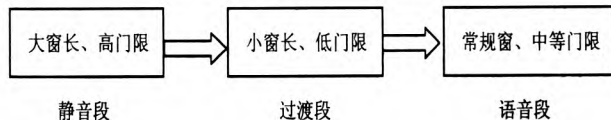


图 1 改进的端点检测算法示意图

2.2 改进的 DTW 算法

由于一般的 DTW 算法在匹配过程中限定了弯折的斜率,许多格点实际上是到达不了的^[10]。因此平行四边形之外的格点对应的帧匹配距离是不需要计算的,也没有必要保存所有的帧匹配距离矩阵和累计距离矩阵,并且每一列格点上的匹配计算只用到了前一列的三个网格,充分利用这两个特点可以减少计算量和存储空间的需求^[11]。因此,基于减少计算量和节省存储空间的考虑,特提出了一种改进的 DTW 算法。

如图 2 所示,如果把实际的动态弯曲分为三段(1, X_a), ($X_a + 1$, X_b), ($X_b + 1$, N), 其中:

$$\begin{cases} X_a = \frac{1}{3}(2M - N) \\ X_b = \frac{2}{3}(2N - M) \end{cases} \quad (1)$$

X_a 和 X_b 都取最相近的整数,可以推出 M 和 N 的长度限制条件:

$$\begin{cases} 2M - N \geq 3 \\ 2N - M \geq 2 \end{cases} \quad (2)$$

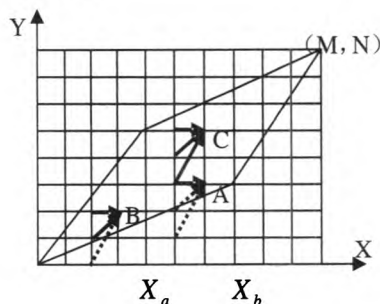


图 2 改进的 DTW 搜索途径

当不满足以上条件时,认为两者差别实在太大,无法进行动态时间规整。上述条件使得搜索路径由原来的矩形区域变成了一个面积小于矩形的平行四边形区域。由于匹配过程遵循路径最短的原则,可以认为,最短路径一定不太过于倾斜,因此,为了减少计算量和提高匹配速度,可以约束搜索路径的斜率在一定的范围内^[12,13]。文中采用搜索路径的斜率在 0.5~2 之间,所以在矩形区域内,路径如果已经通过了格点(m_i, n_i),那它之前的格点只能是($m_i - 1, n_i$), ($m_i - 1, n_i - 1$), ($m_i - 1, n_i - 2$)之间的一个;如果在平行四边形区域内,它之前的格点可能是($m_i - 1, n_i$) (如图 2 中的 A),或是($m_i - 1, n_i$), ($m_i - 1, n_i - 1$)之间的一个(如图 2 中的 B),或是($m_i - 1, n_i$), ($m_i - 1, n_i - 1$), ($m_i - 1, n_i - 2$)之间的一个(如图 2 中的 C),避免了

整个过程中每次搜索路径按三点匹配的繁琐,可以大大缩短匹配中的计算量,提高匹配速度。因此,在平行四边形搜索区域内和搜索路径斜率在 0.5 ~ 2 之间这两个条件下,如果 $d[T(m_i), R(n_i)]$ 表示两帧之间的距离,则有:

$$D[T(m_i), R(n_i)] = d[T(m_i), R(n_i)] + D[(m_{i-1}, n_{i-1})]$$
 (3)

$$D[(m_{i-1}, n_{i-1})] = \min \left\{ \begin{matrix} D[(m_{i-1}, n_i)] \\ D[(m_{i-1}, n_{i-1})] \\ D[(m_{i-1}, n_{i-2})] \end{matrix} \right\}$$
 (4)

按照这种方法计算出的最小累计距离的路径即为最佳路径。

3 实验结果及分析

3.1 实验平台介绍

机器人是一款用于家庭教育、娱乐、休闲、研究的机器人,能够实现语音识别,人脸识别,流程图编程,唱歌跳舞运动演示,触摸屏游戏和网络教学等众多资源,集娱乐和教育于一体;可实现识别非特定人的语音,陪人聊天;可实现识别各种不同人脸,主动与人打招呼;可实现通过流程图编程向机器人中下载程序,控制机器人运行等。图 3 为机器人试验平台。

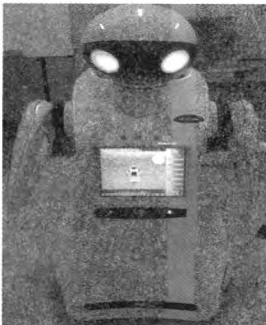


图 3 机器人试验平台

3.2 MATLAB 仿真实验分析

在 MATLAB 7.0 的环境下进行了各项改进算法的仿真实验,试验中采用了 10 组语音数据信号,改进后的端点检测仿真结果如图 4(左侧为改进前端点检测图,右侧为改进后端点检测图)所示:

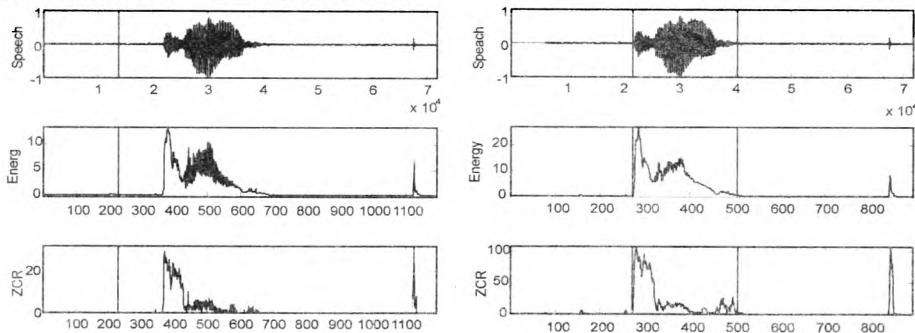


图 4 改进前后端点检测图

改进后的端点检测算法能有效地识别出语音信号的起始点和终止点,为语音信号的特征提取提供更高的精确度。

一般 DTW 与改进的 DTW 的比较:

比较结果如表 1 和表 2 所示。

表 1 DTW 和 DTW 改进算法测试时间(s)

测试语音	1	2	3	4	5	6	7	8	9	10
一般 DTW	13.7	12.9	12.4	12.3	13.4	13.2	11.1	11.4	14.4	12.5
改进的 DTW	10.2	10.1	10.5	9.7	9.9	9.8	9.2	10.1	9.2	9.8

表 2 DTW 和 DTW 改进算法识别率

测试语音	1	2	3	4	5	6	7	8	9	10
一般 DTW	90%	92%	85%	91%	89%	87%	91%	93%	91%	88%
改进的 DTW	92%	96%	92%	96%	95%	93%	95%	97%	94%	93%

对其中的匹配试验结果进行对比,从表中可以得到,这两种方法都能实现语音识别,但是改进的 DTW 算法能更快地更准确地识别语音信息。

4 结束语

图 4 表明,改进后的端点检测算法能有效地快速地检测到语音信号的端点,为特征参数的提取提供更高的精确度。另外,从表 1 和表 2 可以看出改进的 DTW 算法减少了测试的时间,提高了识别率。在迎宾机器人语音识别测试中,机器人能够准确地快速地识别语音信号,从而验证了改进后的算法能有效地提高语音信号的识别率和识别速度。实验证明,改进的 DTW 算法比一般的 DTW 算法在识别率和识别速度上都有一定的提高。高识别率、识别速度的算法的研究,有利于提高语音识别的速度和精度,使语音识别技术更成熟。以上的结论,都是在比较安静的实验室环境中得到的,在嘈杂环境中该算法对语音信号的识别率和识别速度问题,有待于以后的研究。

参考文献:

[1] 洪淑月,施晓钟,徐皓.改进的小波变换 HMM 语音识别算法[J].浙江师范大学学报(自然科学版),2011,34(4):398-403.

[2] 韩普,姜杰.HMM 在自然语言处理领域中的应用研究[J].计算机技术与发展,2010,20(2):245-248.

[3] 刘长明,任一峰.语音识别中 DTW 特征匹配的改进算法研究[J].中北大学学报,2006,27(1):37-40.

[4] 唐尧.基于 DSP 平台的语音识别算法的研究与实现[D].南京:南京航空航

(下转第 76 页)

功率;Cap:风电场的开机总容量; n :所有样本个数。

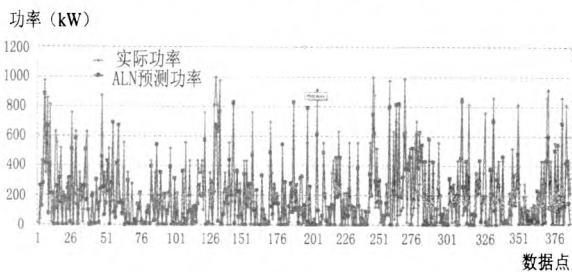


图 5 1MW 风机功率预测结果

由上式得到风机容量和预测精度的对应关系,如表 1 所示。

表 1 四种不同型号风机预测精度统计

风电机组容量	预测精度
600KW	85.14%
750KW	86.97%
1 MW	85.52%
1.3 MW	85.77%

以上仿真结果表明,所提出的基于 ALN 的风功率预测算法完全满足电网对风功率预测系统的要求。值得强调的是,文中的仿真只是基于 3 个月的历史风速,随着风场运行测风数据的不断积累,模型将不断提高其预测精度。

4 结束语

(1)风速和风力发电功率预测对电网经济稳定运行具有重要意义;同时提高了风电场参与市场竞争的能力。

(2)设计并实现了基于神经网络和经典统计学的“自适应线性逻辑网络(ALN)”风电场功率预测模型,具有易于分析训练、结果稳定可控、执行速度快、消耗硬件资源少、预测精度高等特点。

(3)虽然风速和发电功率预测难度较大,预测精度还可进一步提高。比如获取更多历史数据、增加测风手段、加入 NWP 数据对突发天气进行有效预测等。

(4)下一步工作将对中长期的风功率预测进行研究,同时还将研究风场集群的集中预报技术。

参考文献:

[1] 韩民晓,崔军立,姚蜀军. 大量风电引入电网时的频率控制特性[J]. 电力系统自动化,2008,32(1):29-33.

[2] 孙元章,吴俊,李国杰. 基于风速预测和随机规划的含风电场电力系统动态经济调度[J]. 中国电机工程学报,2009,29(4):41-47.

[3] 白永祥,房大中,侯佑华,等. 内蒙古电网区域风电功率预测系统[J]. 电网技术,2010,34(10):157-162.

[4] 杨秀媛,肖洋,陈树勇. 风电场风速和发电功率预测研究[J]. 中国电机工程学报,2005,25(11):1-5.

[5] Soman S S,Zareipour H,Malik O,et al. A review of wind power and wind speed forecasting methods with different time horizons[C]//Proceeding of the 2010 IEEE North American Power Symposium (NAPS). TX, USA:IEEE,2010:1-8.

[6] Foley A M,Leahy P G,McKeogh E J. Wind power forecasting and prediction methods[C]//Proceedings of the 2010 9th IEEE International Conference on Environment and Electrical Engineering (EEEIC). Prague, Czech:IEEE,2010:61-64.

[7] 冯利民,范国英,郑太一,等. 吉林电网风电调度自动化系统设计[J]. 电力系统自动化,2011,35(11):39-43.

[8] 侯昭武,周小珠. 智能化无线风能监测仪的设计[J]. 计算机技术与发展,2011,21(3):232-235.

[9] Zhong J,Hou Y,Wu F F. Wind power forecasting and integration to power grids[C]//Proceedings of the 2010 IEEE International Conference on Green Circuits and Systems (ICGCS). Shanghai, China:IEEE,2010:555-560.

[10] Sideratos G,Hatziaargyriou N D. An Advanced Statistical Method for Wind Power Forecasting[J]. IEEE Transactions on Power Systems,2007,22(1):258-265.

[11] 董安正,赵国藩. 人工神经网络在短期资料风速估计方面的应用[J]. 工程力学,2003,20(5):10-13.

[12] Kariniotakis G,Stavarakakis G,Nogaret E. Wind power forecasting using advanced neural network models[J]. IEEE Trans. on Energy Conversion,1996,11(4):762-767.

(上接第 72 页)

天大学,2007.

[5] 李绍梅,刘力雄,陈鸿昶. 事实说话人辨识系统改进的 DTW 算法[J]. 计算机工程,2008,34(4):218-219.

[6] 韩纪庆,张磊,郑铁然. 语音信号处理[M]. 北京:清华大学出版社,2003.

[7] 江官星. 非特定人孤立词语音识别系统的研究[D]. 成都:西南交通大学,2006.

[8] 马莉,党幼云. 特定人孤立词语音识别系统的仿真与分析[J]. 西安工程科技学院学报,2007,21(3):10-12.

[9] Lee L S,Tseng C Y,Liu F H. Special speech recognition approaches for the highly confusing Mandarin syllables based

on Hidden Markovs[J]. Computer Speech and Language, 1991,5(2):181-201.

[10] 文翰,黄国顺. 语音识别中的 DTW 算法改进研究[J]. 微型计算机信息(测控自动化),2010,26(7):195-197.

[11] 时晓. 孤立词语音识别系统设计研究[D]. 杭州:浙江大学,2006.


[12] 彭辉,魏玮,陆建华. 特定人孤立词的语音识别系统研究[J]. 控制工程,2011,18(30):397-404.

[13] 苏昊,王民,李宝. 一种改进的 DTW 语音识别系统[J]. 中国西部科技,2011,10(1):93-94.

基于改进DTW的机器人语音指令识别系统研究

作者：胡亚洲, 王新民, 曹江涛, HU Ya-zhou, WANG Xin-min, CAO Jiang-tao

作者单位：胡亚洲,王新民,HU Ya-zhou,WANG Xin-min(西北工业大学自动化学院,陕西西安,710129), 曹江涛,CAO Jiang-tao(辽宁石油化工大学信息与控制工程学院,辽宁抚顺,113001)

刊名：计算机技术与发展

英文刊名：Computer Technology and Development

年, 卷(期):2013, 23(7)

本文链接: http://d.wanfangdata.com.cn/Periodical_wjfz201307017.aspx