

Research of Aircraft Engine Sound Recognition Method Based on GMM-UBM

Haoge Yang, Chengli Sun

School of Information Engineering, Nanchang Hangkong University, Nanchang Jiangxi
Email: 1432683346@qq.com, sun_chengli@163.com

Received: Aug. 6th, 2017; accepted: Aug. 21st, 2017; published: Aug. 28th, 2017

Abstract

Gaussian mixture model-universal background model (GMM-UBM) is the most commonly used model in speaker recognition technology; the model has achieved very good results in many experiments. In this design, the GMM-UBM model is used in the abnormal sound detection. First, we process the aircraft engine sound signal, second extract the MFCC characteristic parameters, then train UBM model and last obtain the GMM-UBM model by MAP adaptive algorithm. The ultimate goal of the test indicates that the method could optimize the recognition rate decline due to interference change.

Keywords

Speaker Recognition, GMM-UBM, MFCC, Abnormal Sound Detection, MAP

基于GMM-UBM的飞机发动机声音识别方法研究

杨毫鹤, 孙成立

南昌航空大学信息工程学院, 江西 南昌
Email: 1432683346@qq.com, sun_chengli@163.com

收稿日期: 2017年8月6日; 录用日期: 2017年8月21日; 发布日期: 2017年8月28日

摘要

高斯混合模型-通用背景模型(Gaussian mixture model-universal background model, GMM-UBM)是说话人识别技术中最为常用的模型, 该模型在诸多试验中都取得了很好的效果。本设计探索把GMM-UBM

模型用在异常声音检测中, 通过对飞机发动机声音信号的处理, 提取梅尔频率倒谱(MFCC)特征参数, 训练UBM模型, 用MAP自适应的算法得到GMM-UBM模型, 用GMM-UBM模型检测识别发动机声音。实验证明, 该方法优化了由于外界干扰变化导致的识别率下降的问题。

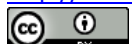
关键词

说话人识别, GMM-UBM, MFCC, 异常声音检测, MAP

Copyright © 2017 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着科学技术的发展, 用科学技术方法来保障人们所处环境的安全越来越得到重视。异常声音识别技术[1]是近年来语音识别的一个新兴的研究方向, 通过对声音信号的监测跟踪, 能及时判断出是否有异常事件发生, 进而采取相应的措施。一般来言, 当发动机出现异常时, 往往都伴随着异常的响动。因此我们可以通过对发动机的异常声音的监控识别, 来判断是否出现了异常, 以便提前得到处理, 可以预防危急事件的发生。

本文研究的飞机发动机声音识别属于非语音的识别问题。一直以来, 对语音识别的研究很多, 但是对于非语音方面的研究却很匮乏, 大多数是把基于单一的 GMM, SVM (Support Vector Machine), HMM (Hidden Markov Model)等作为分类器, 效果并不理想, 因此本文考虑用 GMM-UBM 的方法作为识别模型, 实验证明, 识别率比单一的分类器模型高很多。

2. 发动机异常声音识别系统结构

GMM-UBM [2]是一个高阶的 GMM, 它是根据训练声音音自适应得到的, 可以减小测试声音与训练声音不同所带来的影响, 提高系统识别率。在训练阶段, 首先对发动机声音进行预处理之后, 提取出发动机声音的 MFCC (Mel Frequency Cepstral Coefficients)特征[3], 然后用部分声音训练统一背景模型即 UBM 模型, 再通过最大后验准则[4]训练来得到在通用背景模型的每个高斯分量上进行自适应得到 GMM-UBM 模型。最后在识别阶段, 把待测试语音的特征, 与训练好的 GMM-UBM 结构模型匹配, 最终的输出评分为 GMM 和 UBM 的输出评分之差。图 1 为基于 GMM-UBM 的异常声音识别结构图。

3. 基本原理

3.1. 预处理和特征提取

一个完整的声音识别系统, 第一步是要对声音做预处理, 并提取出准确代表声音特性的特征, 预处理一般有预加重、分帧、加窗和端点检测, 把声音信号预处理之后, 就要提取声音特征, 常用的声音特征有基音周期[5], 线性预测系数 LPC (Line Prediction Coefficient) [6], 线性预测倒谱系数 LPCC (Line prediction Cepstral Coefficient) [7], 梅尔倒谱系数 MFCC[8]等等。由于 MFCC 特征表征人耳频率特性, 鲁棒性更好, 所以本设计用 MFCC 系数作为发动机声音特征。

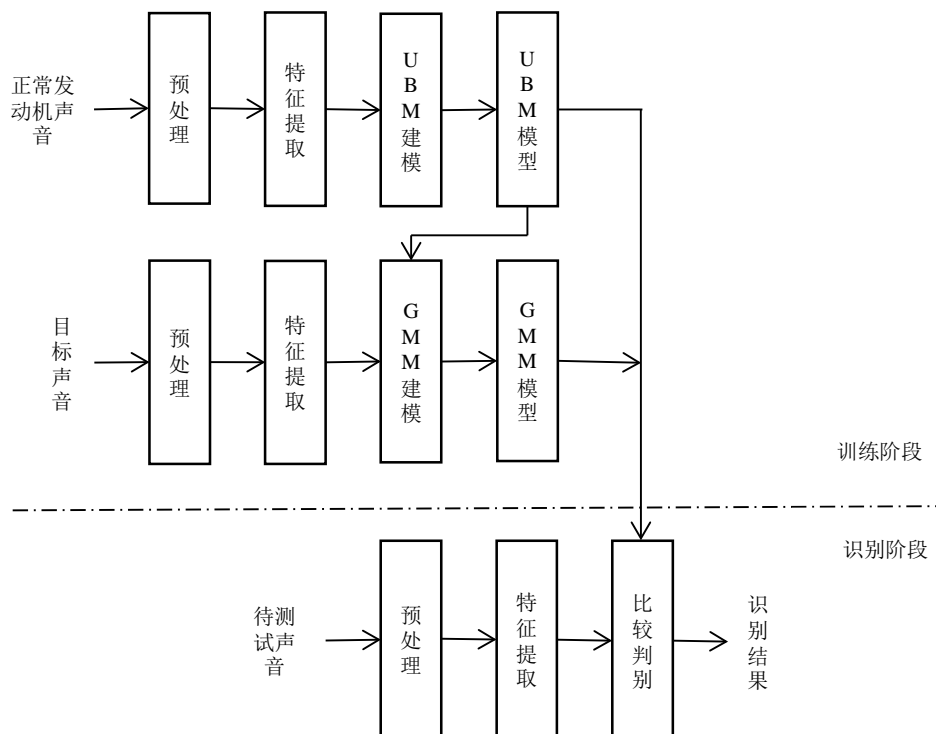


Figure 1. Recognition structure graph of GMM-UBM
图 1. GMM-UBM 识别结构图

3.2. GMM-UBM 模型

高斯混合模型就是对一定数量的高斯概率密度函数进行线性加权组合, M 阶高斯混合模型 GMM 用 M 个单高斯分布的线性组合来描述帧特征在特征空间的分布[9], 一个完整高斯混合模型是由参数均值矢量 μ_i 、协方差矩阵 Σ_i 和权重 ω_i 组成, 可以表示为:

$$\lambda = \{\omega_i, \mu_i, \Sigma_i | i = 1, 2, \dots, M\} \quad (1)$$

统一背景模型 UBM 也是高斯混合模型, 它用很多各种环境下的声音数据训练获得, 故 UBM 是所有声音特征及环境通道的共性反映。因此 UBM 的训练数据集越多、覆盖面越广, 最终的识别效果越好。每一类声音的 GMM 模型是用该类训练声音由 UBM 自适应得到, 本实验中用到的自适应方法是 MAP 自适应算法[10]。自适应算法可以很好的解决训练数据少的问题, 它可以利用少量数据通过修改输入声音的特征矢量, 或者调整模型的参数, 以此达到训练模型和测试数据相匹配的目的, 从而保证了系统良好的识别性能。由 UBM 模型通过 MAP 自适应得到的 GMM 就是 GMM-UBM 模型。

3.3. EM 算法

高斯混合模型的学习方法和 MAP 自适应估计中, 最大期望算法即 EM 算法(Expectation Maximization algorithm) [11]起着重要的作用。EM 算法是一种迭代算法, 通过 E 步和 M 步两大迭代步骤, 每次迭代都使极大似然函数增加。EM 算法流程图下图 2 所示:

设一组长度为 T 的音频特征序列为 $\mathbf{X} = \{X_1, X_2, \dots, X_T\}$, 按照最大似然估计准则, 找到 GMM 的参数, 使得 GMM 模型训练特征的似然概率为最大。用 EM 算法对 GMM 参数进行估计的迭代公式如下:

第 m 个 GMM 模型的权重 ω_m ,

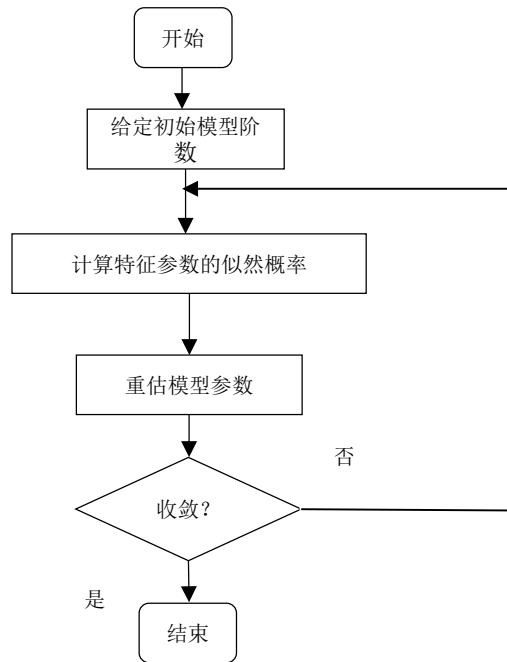


Figure 2. Flow chart of EM
图 2. EM 算法流程图

$$\omega_m = \frac{1}{T} \sum_{t=1}^T P(m|X_t, \lambda) \quad (2)$$

第 m 个 GMM 模型的均值 μ_m ,

$$\mu_m = \frac{\sum_{t=1}^T P(m|X_t, \lambda) X_t}{\sum_{t=1}^T P(m|X_t, \lambda)} \quad (3)$$

第 m 个 GMM 模型的方差 σ_m^2 ,

$$\sigma_m^2 = \frac{\sum_{t=1}^T P(m|X_t, \lambda) X_t^2}{\sum_{t=1}^T P(m|X_t, \lambda)} - \mu_m^2 \quad (4)$$

其中第 m 个 GMM 模型的后验概率为,

$$\sum_{t=1}^T P(m|X_t, \lambda) = \frac{\omega_m N(X_t | \mu_m, \Sigma_m)}{\sum_{i=1}^M \omega_i N(X_t | \mu_i, \Sigma_i)} \quad (5)$$

用 EM 算法修正 UBM 中各模型组件的参数结果如下:

修正之后的权重 ω'_m ,

$$\omega'_m = \left[\frac{\alpha_m^\omega n_m}{T} + (1 - \alpha_m^\omega) \omega_m \right] \gamma \quad (6)$$

修正后的均值 μ'_m ,

$$\mu'_m = \alpha_m^m E_m(X) + (1 - \alpha_m^m) \mu_m \quad (7)$$

修正之后的方差 $\sigma_m'^2$,

$$\sigma_m'^2 = \alpha_m^v E_m(X^2) + (1 - \alpha_m^v)(\sigma_m^2 + \mu_m^2) - \mu_m'^2 \quad (8)$$

其中, α_m^o , α_m^m , α_m^v 分别为权重、均值、方差的修正因子, γ 是权重规整因子, 使 ω_m' 的和为 1,

$$\alpha_m^\rho = \frac{n_m}{n_m + r^\rho} \rho = o, m, v \quad (9)$$

式中 r^ρ 为常数, 用来约束修正因子的变化尺度。

经过以上几个步骤的, 最终得到修正过的模型即为 GMM-UBM 模型。在识别阶段, 将待识别的声音特征向量分别计算与 UBM 模型、目标声音模型的似然度输出比, 采用对数评分时, 待识别声音的似然度即为两个模型似然度对数之差。

$$S = \frac{1}{T} \sum_{i=1}^T [\log P(X_i | \lambda_{TAR}) - \log P(X_i | \lambda_{UBM})] \quad (10)$$

式中, X_i 是待识别声音的某一帧音频的特征向量, λ_{TAR} 和 λ_{UBM} 分别表示目标声音模型和 UBM 模型, S 为最终识别的似然度评分。

4. 实验结果与分析

通过对网上数据的搜集和整理, 建立了一个飞机发动机声音数据库, 数据库包含了民航波音 727、军用战斗机 F-15 等十几种型号的飞机发动机声音, 在本课题中, 把发动机声音分为四种类型, 包括发动机启动声音、停止声音、运行声音和一些发动机坠毁声音。在实验中用到的发动机声音的总时长为 1474 s, 其中发动机启动声音有 347 s, 停止声音有 205 s, 正常运行声音 629 s, 和 293 s 坠毁声音, 声音信号的采样频率为 16 KHz, 单声道录音, 采用 16 Bit 量化。声音特征采用 MFCC 特征。根据这些声音信息, 进行了下面三个实验。

实验一比较相同的训练数据时, 混合度对识别率的影响。选取 61 个发动机音频作为 UBM 训练, 其中包括 12 个启动音频, 8 个熄火音频, 30 个正常运行音频以及 11 个坠毁音频, 总时长为 6 分 22 秒, 用于 MAP 自适应音频有 39 个包括 8 个启动音频, 8 个熄火音频, 13 个正常运行音频和 10 个坠毁音频总时长为 5 分 32 s, 待识别声音分别是 6 个启动音频, 6 个熄火音频, 29 个正常运行音频和 11 个坠毁音频。特征为 9 维 MFCC 参数时, 各类声音的识别率结果如下表 1:

由实验数据看出, 在其他条件相同的情况下, 混合度为 256 的时候总体识别效果较好, 但是在该实验条件下启动声音识别率最高仅为 50%, 相比其他类型的声音识别率较低。针对特定的一类声音来说, 随着混合度的增加, 识别率也随之增加, 但当混合度增加到一定值的时候, 识别率不再增加, 甚至开始下降。

Table 1. The recognition rate of different mixing degrees with 9-D feature
表 1. 9 维特征, 不同混合度的识别率

混合度	识 别 率					
	8	16	32	64	128	256
启动	33.3%	33.3%	50%	50%	50%	50%
熄火	66.6%	50%	50%	50%	50%	66.6%
运行	89.6%	96.5%	96.5%	96.5%	96.5%	93.1%
坠毁	63.6%	72.7%	72.7%	81.8%	81.8%	81.8%

实验二实验数据与实验一相同, 不同的是选择 12 维的 MFCC 特征参数, 识别结果如下表 2:

与实验一相比, 识别率整体上升了很多, 启动声音识别率增加到 66.6%, 正常声音和坠毁声音识别率达到 100%。由此可见, 特征维数的增加可以提高声音的识别率。

根据实验一和实验二对比可知, 选取 12 维 MFCC 特征参数, 混合度为 128 时得到较高的识别率。所以下面的实验验证混合度为 128 的时候, 选取 MFCC 静态系数、动态一阶差分系数和二阶差分系数以及它们的混合对识别率的影响, 这几种特征分别用 MFCC、 Δ MFCC、 $\Delta\Delta$ MFCC 和 HMFCC 来表示。实验结果如下表 3 所示。

由于 MFCC 系数只是梅尔倒谱特征的静态系数, 只表征了音频谱的即时信息, 所以识别率相对于一阶或二阶动态差分系数较低, 音频谱的动态信息是表征音频特征参量随时间变化的规律, 因而识别率比倒谱特征静态系数所得的识别率稍高。由实验结果可知, 二阶系数的识别率和一阶系数的识别率只是在启动声音上有较大的提高, 这是由于启动声音本身的特性决定的。实验中我们还可得知, 倒谱特征静态系数和一阶差分系数结合得到的识别率比较单一的特征系数识别率明显有提高, 而把静态系数和一阶、二阶差分系数组合得到的 MFCC 特征的识别率是最好的。

由实验可知, 当选择 9 维 MFCC 特征参数的时候, 随着混合度的增加, 识别效果也有所提高, 但当加到一定值后, 识别效果反而开始下降。这是由于 GMM 模型本质上是概率模型, 混合度越高模型越复杂, 用于训练所必须的数据量也大大增加, 而本实验的数据量相对较少, 若混合度增加过多, 会产生过度拟合, 从而导致识别性能下降。当特征参数的维数增加到 12 维时, 识别率有了明显的提高, 而且在 GMM 混合度为 128 时, 得到最好的识别效果。通过实验三对特征系数的选择来看, 差分系数 Δ MFCC 和 $\Delta\Delta$ MFCC 特征比静态系数 MFCC 特征识别率高, 结果证明, 用三个系数组合成的 12 维 MFCC 特征, 混合度为 128 的时候, 对发动机四种声音的识别率最高。

Table 2. The recognition rate of different mixing degrees with 12-D feature
表 2. 12 维特征, 不同混合度的识别率

	识 别 率					
混合度	8	16	32	64	128	256
启动	33.3%	50%	50%	66.6%	66.6%	66.6%
熄火	66.6%	66.6%	66.6%	83.3%	83.3%	83.3%
运行	89.6%	96.5%	96.5%	100%	100%	86.2%
坠毁	63.6%	72.7%	100%	100%	100%	100%

Table 3. The recognition rate of different feature
表 3. 不同特征的识别率

	识 别 率				
特征	MFCC	Δ MFCC	$\Delta\Delta$ MFCC	MFCC- Δ MFCC	HMFCC
启动	33.3%	33.3%	50%	50%	66.6%
熄火	50%	50%	50%	66.6%	83.3%
运行	31%	17.2%	27.6%	96.5%	100%
坠毁	16.6%	27.3%	27.3%	81.8%	100%

5. 结束语

GMM-UBM 在说话识别中运用的非常多, 而且也取得了很好的识别效果, 本文把说话人识别方法用在发动机声音识别中, 提出了用 GMM-UBM 模型实现对发动机声音的识别, 基于 GMM 的通用背景模型是一种自适应的过程, 可以弥补声音数量有限所导致的 GMM 混合度低, 辨识力差的缺点, 而且通过仿真实验可知, 把 GMM-UBM 模型用在飞机发动机声音识别是可行的, 实验也得出了较好的识别效果。下一步将继续探究飞机发动机声音的识别方法, 提高发动机启动声音和熄火声音的识别率。

致谢

本论文的工作是在我的导师孙成立副教授的悉心教导和亲切关怀下完成的, 老师严肃的科学态度, 严谨的治学精神, 精益求精的工作作风深深地感染和激励着我, 在此衷心的感谢老师对我的关心和指导。同时还要感谢国家自然科学基金项目(61362031)及航空科学基金(20145556011)对本文的资助。

参考文献 (References)

- [1] Xie, C., Cao, X.L. and He, L.L. (2012) Algorithm of Abnormal Audio Recognition Based on Improved MFCC. *Procedia Engineering*, **29**, 731-737. <https://doi.org/10.1016/j.proeng.2012.01.032>
- [2] 张正平. 基于 GMM-UBM 说话人模型的连续自适应算法研究[J]. 通信电源技术, 2016, 33(2).
- [3] Wu, J.Q. and Yu, J.J. (2011) A Improved Arithmetic of MFCC in Speech Recognition System. *IEEE Electronics, Communication and Control (ICECC)*, 719-722.
- [4] Harma, A., McKinney, M.F. and Skowronek, J. (2005) Automatic Surveillance of the Acoustic Activity in Our Living Environment. *Proc of IEEE ICME*, 634-637.
- [5] 赵力. 语音信号处理[M]. 北京: 机械工业出版社, 2012: 75-134.
- [6] 李燕萍. 说话人辨认中的特征参数提取和鲁棒性技术研究[D]: [硕士学位论文]. 南京: 南京理工大学, 2009.
- [7] 王炳锡, 屈丹, 彭焯. 实用语音识别基础[M]. 北京: 国防工业出版社, 2005.
- [8] 王伟, 邓辉文. 基于 MFCC 参数和 VQ 的说话人识别系统[J]. 仪器仪表学报, 2006, 27(6): 2253-2155.
- [9] Ito, A. and Aiba, A. (2009) Detection of Abnormal Sound Using Multi-Stage GMM for Surveillance Microphone. *International Conference on Information Assurance and Security*.
- [10] Mark, M.-W. and Rao, W. (2011) Utterance Partitioning with Caoustic Vector Resampling for GMM-SVM Speaker Verification. *Speech Communication*, **53**, 119-130. <https://doi.org/10.1016/j.specom.2010.06.011>
- [11] Ntalampiras, S., Potamitis, I. and Fakotakis, N. (2009) An Adaptive Framework for Acoustic Monitoring of Potential Hazards. *EURASIP Journal on Audio, Speech, and Music Processing*, **2**, 1-15.

期刊投稿者将享受如下服务:

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24 小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: csa@hanspub.org