

语音信号 MFCC 特征提取案例教学实践

沈希忠

(上海应用技术大学电气与电子工程学院, 上海 201418)

摘要: 数字信号处理 (DSP) 课程的核心知识点是傅立叶变换 (FT), 尤其是快速傅里叶变换 (FFT), 其应用广泛, 至今不衰。MFCC 是语音信号处理的关键内容, 也是 FT、FFT 的一个典型应用, 在很多公开算法中具有举足轻重的作用。本文基于语音信号的 MFCC 案例教学, 讲述其教学过程和实践活动, 描述 MFCC 的算法仿真各个模块, 算法用 python 软件实现, 并总结 MFCC 的实际应用, 探讨信号算法的教学思路。由于 MFCC 是一个综合性的案例, 在教学中引入该案例教学法, 可以辅助学生进一步掌握语音信号表示、信号分段、FT、FFT、DCT 等知识点, 同时利用 FFT 进行算法构建和 Python 编程教学活动, 提高学生对知识的理解和应用能力。

关键词: 数字信号处理; MFCC; 案例法; 教学方法

一、引言

“数字信号处理 (DSP)” 课程在上海应用技术大学电气与电子工程学院开设已经二十多年, 积累了不少经验, 这是电子信息专业的一门专业基础课程, 其内容包括信号的时域表示、Z 变换、傅里叶变换 (FT)、快速傅立叶变换 (FFT)、滤波器设计、DCT、相关分析等。该课程的目标是, 通过上述知识的讲解让学生能够在时域表达信号, 能够通过 FT 变换进行频谱分析, 实现数字序列的变换与处理, 提取有用的特征信息, 以满足工程需求。其中, 傅里叶变换 (FT) 是当前应用最为广泛的信号变换分析方法, 快速傅里叶变换 (FFT) 是实现 DFT 的快速高效算法。

传统 DSP 教学模式以课堂理论讲授为主, 课堂讲解以 PPT、黑板等形式呈现, 同时附加实验, 实验中以 MatLAB、Python 等计算机语言为主进行编程训练, 最后进行考试。如知识点 FFT 的课堂教学, 先从 FT 的计算公式开始公式推导, 推导基 2 DIT FFT 的相关公式, 并画出 8 点 FFT 流程图, 然后让学生编程实验, 观察信号的频率。这样的课程教学往往缺乏实际应用背景, 教学细节脱离实际, 使得学生难以在实践中理解并应用课程知识。而且, DSP 课程的知识点较多, 传统教学缺乏知识点之间的关联, 学生难以全面掌握知识体系, 难以真正应用到工程中。

近年来, 案例教学法深得教师和学生的喜爱, 也受到广泛的研讨, 同时在各高校的课堂教学和实践活动中广泛应用案例教学, 如文献基于 MATLAB 设计了遥感 DSP 综合课程实验。

本文在多年教学活动中设计一个声学信号处理的综合应用案例——梅尔频率倒谱系数 (MFCC, Mel Frequency Cepstral Coefficient) 的计算, 并用于语音、鸟叫声的特征提取。该案例深度剖析数字信号处理课程的各个知识点, 主要有语音信号的表示、语音分段、FFT、DCT、滤波器等, 讲解各个知识点在 MFCC 案例中如何应用。MFCC 案例给出 MFCC 的算法流程和 Python 实现方法, 给出课内课外编程实践内容和时间分配, 提高学生对相关知识的全面认知, 也为特征提取实际应用提供经验借鉴。

二、MFCC 算法流程

MFCC 的 Python 程序实现见图 1 所示, 其 MFCC 计算步骤如图 1 所示, 主要分为 5 步。相关的知识点如下:

知识点 1. 信号的表示。 声学信号经过麦克风 (声学传感器) 采集, 可以表达为一个数字信号, 这是一个不平稳的信号, 是不断变化的, 如 $x(n) = x(t)|_{t=nT}$, $T = \frac{1}{f_s}$ 为采样周期, f_s 为采样频率。

知识点 2. 信号的分割。 将信号分割为时间尺度较短的帧, 如 20–40 ms, 在帧内可以认为信号变化不大, 认为是平稳的。注意: 帧太短, 没有足够的样本获取频谱估计, 反之, 帧太长, 信号在整个帧中变化太大。

知识点 3. FFT 谱。 对每一个信号帧利用 FFT 计算其功率谱, 这个知识点是整个算法的关键, 但如果用 MatLAB 或者 python 语言就是一条命令, 很简单。

知识点 4. 滤波器 (Mel 谱的计算)。 将 Mel 滤波器组用于功率谱, 并对将每个滤波器中的能量求和。设置一簇周期图箱并汇总起来, 计算不同频率区域存在多少能量。这个周期图箱就是 Mel 滤波器组。随着频率越来越高, 滤波器会变得更宽, 在每个滤波器作用频率范围内, 只计算大约有多少频域能量。梅尔频率准确地表达了滤波器组的各个频率点及其宽度。

知识点 5. 序列运算, 即对数运算。 离散信号的序列运算也是很重要的。

知识点 6. DCT。 DCT 是一种离散余弦变换, 一般用于图像压缩。此处, 因为滤波器组是重叠的, 滤波器组的能量彼此之间相关, 故用 DCT 对能量进行去相关。一般计算中, DCT 共 26 个系数, 保留编号从 2 到 13, 丢弃其余系数。26 个 DCT 系数中仅保留了 12 个, 是因为较高的 DCT 系数代表了滤波器组能量的快速变化。



图 1 MFCC 计算图

三、MFCC 计算案例

设时域信号为 $s(n)$, 分帧后为 $s_i(n)$, 其中 $n = 1 - 400$ (如果帧有 400 个样本), i 为帧数。当计算复数 DFT 时, 得到帧 i 的功率谱 $S_i(k)$ 。语音信号的采样频率为 f_s 。计算案例时的几点要点和学生分组情况如下:

1. 信号分帧。计算中可以让学生分组尝试, 如设置帧长为 20、25、30、35、40ms, 采样频率 f_s 也可以不同, 如 8、16kHz, 帧步长也可以不同, 如 5、10、15ms 等, 由此造成最后的 MFCC 长度不一。如帧长为 25ms, 16kHz 采样时, 信号帧长为 $0.025 \times 16000 = 400$ 个样本。帧步长通常约为 10ms, 即 160 个样本, 每个帧可以提取一组 12 个 MFCC 系数。

2. 帧的离散傅里叶变换

$$S_i(k) = \sum_{n=1}^N s_i(n) \hat{h}(n) e^{-j2\pi kn/N}, 1 \leq k \leq K$$

其中 $\hat{h}(n)$ 是 N 个样本长窗口, 可以尝试不同的窗口函数, 如 Hamming、Hanning、矩形窗、Bessel 窗、Kaiser 窗等, K 是 DFT 的长度。基于周期图的语音帧 $s_i(n)$ 的功率谱估计为

$$P_i(k) = \frac{1}{N} |S_i(k)|^2$$

称为功率谱的周期图估计, 即取复傅里叶变换的绝对值, 对

结果进行平方。通常执行 512 点 FFT, 只保留前 257 个系数。

计算 Mel 间隔的滤波器组, 滤波器组可以设置为 20、24、28、32、36、40 个三角形滤波器, 其取值方法见图 1。将每个滤波器组与功率谱 $P_i(k)$ 相乘, 将系数相加, 得到 26 个数字。

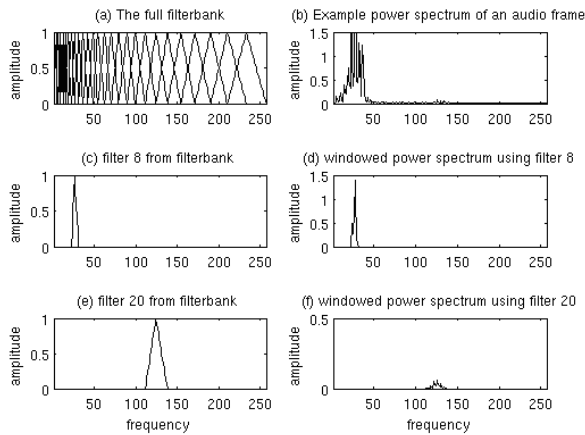


图 2、Mel 滤波组和窗口功率谱图

DCT 计算。DCT 计算时, 如果对 M 个对数滤波器组能量进行 DCT, 得到 M 个倒谱系数, 仅保留 M 个系数中的 $2 - M/2$ 个系数, 得到每帧 $\frac{M}{2} - 1$ 个特征数字, 这个特征就是梅尔频率倒谱系数, 即 MFCC 值。

四、教学反思

MFCC 案例的教学方法能使学生在案例中复习和巩固的主要知识点如图 4 所示, 其中 Python 知识点课堂上不讲, 也不辅导, 完全靠学生自学, 数字信号中的知识点见表 1, 占用 2 个课时, 学生上机实践时间 2 个课时, 课外要求学生自己 Python 编程, 估计 4 个学时, 如果还要进行分析, 另外要 4 个学时, 如果考虑写报告还要 4 学时, 另外由于部分学生 Python 不熟悉, 还要额外 8 学时, 因此做多需要 24 学时。学生在 MFCC 案例教学中能够进一步掌握数字信号处理相关知识, 积累 FFT 应用经验, 提高编程能力。

实践中 MFCC 案例教学灵活讲解, 进一步深化理论, 综合应用各个知识点, 强调实用。

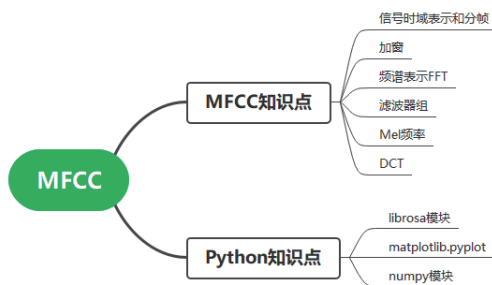


图 4 MFCC 的知识点

表 1 MFCC 案例教学时间分配和程序设计

序号	知识点	实践方法	时间分配
1	信号的时域表示和分段	复习时域信号的表示, 复习 Python 编程知识	20 分钟
2	加窗	复习窗函数的定义, 了解窗函数的影响	5 分钟
3	频谱变换	复习 FFT, 复习功率谱的计算方法	20 分钟

4	滤波器组	复习滤波器组的含义, 计算发方法	20 分钟
5	Mel 频率	复习 Mel 频率的计算方法	5 分钟
6	DCT	复习 DCT 的定义, 计算方法	20 分钟

由于数字信号处理课程存在“理论复杂难懂”“教学中其理论与实践经常脱节”等问题, 通过 MFCC 案例的讲解, 结合语音信号、鸟叫声等各种声学信号, 同时在实践中配合不同的分组方法, 做到紧密结合工程实践, 引导学生分组设计, 主动开展实践活动, 巩固课堂知识, 拓宽知识面, 提高发现问题、分析与解决问题的能力, 从而提高学生的创新能力。

MFCC 案例设计的考核评价表如表格 2 所示。案例设计中, 对班级进行分组, 每组可能获得不同的声学信号、不同的分段方法、不同的窗函数等, 其中信号如语音信号 MFCC、鸟叫声 MFCC、轴承振动 MFCC 等, 每组 3-5 人, 设组长 1 人。

表格 2 MFCC 案例设计评价表

项目	案例设计内容	考核指标	设计和分析要求	分值比例 %
1	案例设计过程	程序设计 仿真内容 数据可视化	设计解决方案 设计案例情况 Python 工具	10 10 10
2	案例报告	报告格式和规范 仿真和数据处理 数据分析		10 10 10
3	案例答辩	个人与团队 自述 理论与实践情况		10 10 20

五、结语

本文基于 MFCC 案例教学法, 对“数字信号处理”课程的核心内容 FFT 引入一个应用算法, 即语音信号的 MFCC 特征提取, 通过 MFCC 的算法流程讲解, 用 Python 实现方法, 同时进行演示说明。

针对信号处理课程中的问题, 如理论复杂难懂, 并与实践脱钩等问题, 探讨了案例教学思路。本文以语音信号处理的 MFCC 算法为例, 具有一定的典型性。

该案例在数字信号处理课程中涉及知识面较广, 是一个综合性的案例, 几乎涉及信号处理本科阶段的所有内容, 因此该案例能够深度剖析教学过程中理论讲解和实践内容讲解, 同时该案例贯穿多个知识点, 能够提高学生的综合理解, 通过引导学生可以参与一些具体的工程实践活动, 提高学生对于数字信号处理的综合理解能力, 通过 Python 编程提高学生的计算机编程能力。

参考文献:

- [1] 沈希忠. 数字信号处理 [M]. 北京: 机械工业出版社, 2014.
- [2] 闭吕庆, 陈思谕, and 等. 数字信号处理"课程案例教学实践——FFT 原理与应用的深度剖析 [J]. 电子元器件与信息技术, 2023 (7): 183-186.
- [3] 刘洪, 林洁馨, and 等. “数字信号处理”课程中的案例式教学法研究 [J]. 教育教学论坛, 2022 (3): 133-136.
- [4] 杨桃丽, and 于瀚雯. 基于 MATLAB 的数字信号处理综合课程实验 [J]. 实验科学与技术, 2023, 22 (1): 57-61, 67.

作者简介: 沈希忠, 1968 年生, 男, 汉族, 上海, 博士, 上海应用技术大学电子信息, 教授, 主要研究方向为信号处理和人工智能。