

文献引用格式: 吕丹璇, 吕家威. Python 在声音特征提取与分类中的实现方法研究 [J]. 电声技术, 2024, 48(4):20–22, 27.

LYU D X, LYU J W. Research on implementation methods of Python in sound feature extraction and classification[J]. Audio Engineering, 2024, 48(4):20–22, 27.

中图分类号: TN912.34

文献标识码: A

DOI: 10.16311/j. audioe. 2024. 04. 006

Python 在声音特征提取与分类中的实现方法研究

吕丹璇, 吕家威

(河南测绘职业学院, 河南 郑州 450000)

摘要: 设计基于 Python 的声音特征提取与分类方法, 以 ESC-50 数据集为基础, 结合 Librosa 和 TensorFlow 库实现了声音信号的分析和处理。首先, 介绍声音分析的总体架构, 包括数据预处理、特征提取、模型构建和评估等关键步骤。其次, 详细探讨基于 Librosa 的声音特征提取方法, 特别是梅尔频率倒谱系数 (Mel Frequency Cepstral Coefficients, MFCC) 的实现原理和基于 TensorFlow 的分类方法, 如支持向量机的使用。最后, 通过实验验证了所提方法的有效性, 评估了分类器的准确率、精确率及召回率等性能指标。实验表明, 基于 Python 的声音特征提取与分类方法表现出良好的性能, 为声音信号处理领域的研究和应用提供了重要的参考和支持。

关键词: 特征提取; 声音分类; Python; Librosa 库

Research on Implementation Methods of Python in Sound Feature Extraction and Classification

LYU Danxuan, LYU Jiawei

(Henan College of Surveying and Mapping, Zhengzhou 450000, China)

Abstract: Design a method of sound feature extraction and classification based on Python, based on ESC-50 data set, the analysis and processing of sound signals are realized by combining Librosa and TensorFlow library. Firstly, the overall framework of sound analysis is introduced, including key steps such as data preprocessing, feature extraction, model construction and evaluation. Secondly, the voice feature extraction method based on Librosa is discussed in detail, especially the implementation principle of Mel Frequency Cepstral Coefficients (MFCC) and the classification method based on TensorFlow, such as the use of support vector machine. Finally, the effectiveness of the proposed method is verified by experiments, and the performance indexes such as accuracy, precision and recall of the classifier are evaluated. Experiments show that the method of sound feature extraction and classification based on Python shows good performance, which provides important reference and support for the research and application of sound signal processing.

Keywords: feature extraction; sound classification; Python; Librosa library

0 引言

在数字化时代, 声音信号的分析和处理在许多领域扮演着重要角色, 如语音识别^[1]、音乐信息检索^[2]以及环境声音分类等^[3]。随着 Python 语言在科学计算和机器学习领域的日益流行, 其灵活性和强大的工具库使其成为声音特征提取^[4]、声音分类^[5]、文本分类^[6]等领域研究和实践的理想选择之

一。Librosa 作为 Python 中常用的音频处理工具库, 具有丰富的功能和易用性。TensorFlow 则提供了强大的机器学习和深度学习功能, 为声音分类任务提供了可靠的工具支持。文章聚焦于基于 Python 的声音分析方法, 首先介绍基于 Python 的声音分析的总体架构; 其次探讨基于 Librosa 和 TensorFlow 等工具库的声音特征提取和分类方法; 最后通过实验验证所提方法的有效性。文章的研究为声音信

作者简介: 吕丹璇(1992—), 女, 本科, 助教, 研究方向为计算机应用技术。

号处理领域的相关研究提供新的实施方法,促进Python在声音分析技术中的推广和发展。

1 基于 Python 声音分析的总体架构

在 Python 中,对声音信号分析的技术路线如图 1 所示,包括数据采集、数据预处理、特征提取、数据标记、模型构建以及训练与测试等。

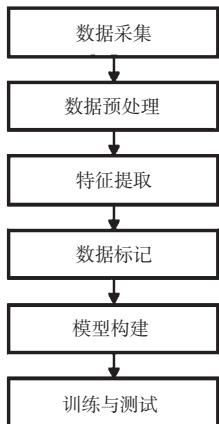


图 1 总体框架

数据采集是从声音信号数据库或实际采集设备中获取声音信号数据集。首先使用 Python 编程语言中的 NumPy 和 Pandas 等库预处理原始声音数据,包括数据清洗、去噪及格式转换等操作。其次利用 Librosa 库提取声音信号的特征,如梅尔频率倒谱系数(Mel Frequency Cepstral Coefficient, MFCC)、色度图谱(Chromagram)等。Python 的函数式编程风格和灵活性可用于快速开发和测试各种特征提取算法。再次标记提取的特征数据,以便后续的分类任务。Python 的 Scikit-learn 库提供了方便的工具来进行数据标记和预处理。最后使用 TensorFlow 等深度学习框架构建声音分类模型,并训练和测试构建的声音分类模型。

2 基于 MFCC 和支持向量机的声音信号分析方法

声音的特征提取与分类是声信号分析的关键,文章研究了 Python 中基于 Librosa 库的特征提取和基于 TensorFlow 的声音分类方法。

2.1 函数库介绍

2.1.1 Librosa

Librosa 是一个用于音频和音乐分析的 Python

库,提供了丰富的功能来进行声音特征提取。Librosa 库中常用的函数如表 1 所示,通过这些函数可以提取出声音信号的多种特征,为后续的声音分类任务提供丰富的特征信息。

表 1 Librosa 库中常用的声音特征提取函数及其功能

| 函数 | 功能 |
|----------------|----------|
| stft | 短时傅里叶变换 |
| melspectrogram | 计算梅尔频谱图 |
| mfcc | 梅尔频率倒谱系数 |
| chroma_stft | 计算色度图谱 |
| tempogram | 计算临时图谱 |

2.1.2 TensorFlow

TensorFlow 是由 Google 开发的一个开源机器学习框架,广泛应用于深度学习领域。TensorFlow 提供了丰富的应用程序接口(Application Programming Interface, API)和工具,使得构建、训练及部署各种深度学习模型更加简单和高效。TensorFlow 库中常用的分类方法有逻辑回归、支持向量机、随机森林、卷积神经网络、循环神经网络以及长短期记忆网络等。逻辑回归是一种基本的线性分类算法,适用于二分类问题。支持向量机是一种经典的监督学习算法,可以用于二分类和多分类问题。随机森林是一种集成学习算法,通过多个决策树的组合提升分类性能。卷积神经网络是一种深度学习模型,广泛用于图像分类和语音识别等任务。循环神经网络适用于序列数据的分类,如文本分类和时间序列分析等。长短期记忆网络是一种特殊的循环神经网络结构,适用于处理序列数据的长期依赖关系。这些方法可以应用于各种声音特征分类任务,具有不同的优缺点和适用场景。

2.2 实施方法与原理

文章以 mfcc 函数和支持向量机为例,研究 Python 在声音特征提取和分类中的实现方法。一方面,使用 Librosa 库中的 mfcc 函数,将音频信号转换为 MFCC 特征;另一方面,使用 TensorFlow 实现支持向量机,通过寻找最优的超平面来划分不同类别的数据,使得两类数据之间的间隔最大化。

在 MFCC 方法中^[7-8],设原始音频信号为 $x(t)$,其短时傅里叶变换为 $X(t,\omega)$ 。利用三角滤波器组将频谱图 $|X(t,\omega)|$ 映射到梅尔频率空间,得到梅尔频谱图 $S(t,\omega)$ 为

$$S(t, m) = \log \left[\sum_{\omega=0}^{f_s/2} |X(t, \omega)| H_m(\omega) \right] \quad (1)$$

式中: $H_m(\omega)$ 为第 m 个三角滤波器, f_s 为采样率。

对梅尔频谱图进行离散余弦变换, 得到 MFCC 特征向量为

$$MFCC(t, n) = \sum_{m=1}^M S(t, m) \cos \left[\frac{\pi n(m-0.5)}{M} \right] \quad (2)$$

式中: M 为滤波器组的数量。

在支持向量机中^[9-10], 给定训练集 $\{(x_i, y_i)\}_{i=1}^N$, 其中 x_i 是特征向量, y_i 是类别标签, $y_i \in \{-1, 1\}$ 。支持向量机的目标是找到一个超平面 $w \cdot x + b = 0$, 使得所有训练样本到该超平面的距离(间隔)最大化, 并且能够正确地将不同类别的样本分开。间隔的计算方法为

$$\text{margin} = \frac{2}{\|w\|} \quad (3)$$

支持向量机的优化目标可以转化为凸二次规划问题, 即

$$\begin{aligned} & \min \frac{1}{2} \|w\|^2 \\ & \text{s.t. } y_i(w \cdot x_i + b) \geq 1, \forall i \end{aligned} \quad (4)$$

3 实验与分析

为明确 Python 在声音特征提取和分类中的应用效果, 文章使用 ESC-50 数据集进行分析验证。该数据集是一个广泛用于环境声音分类研究的公开数据集, 包含来自 50 种不同环境的音频样本, 涵盖日常生活中各种常见的环境声音, 如动物叫声、交通声、自然声等。

基于 ESC-50 数据集, 文章采用 Librosa、TensorFlow 以及支持向量机提取与分类声音特征, 实施方法如下。第一, 数据准备。下载 ESC-50 数据集, 使用 Python 中的 Pandas 库加载数据集的标注信息。第二,

数据预处理。使用 Librosa 库读取音频文件, 将音频信号转换为时域表示形式, 并进行预处理。第三, 特征提取。使用 Librosa 库中的 mfcc 函数提取音频信号的 MFCC 特征, 得到每个音频样本的 MFCC 特征向量。第四, 数据标签提取。根据数据集的标注信息, 提取每个音频样本的类别标签。将类别标签转换为数值编码, 以便支持向量机进行处理。第五, 数据划分。将 80% 的数据作为训练集, 20% 的数据作为测试集。第六, 模型构建。利用 TensorFlow 构建支持向量机模型, 将训练集的特征矩阵和标签向量作为输入训练支持向量机模型。第七, 模型训练。调用支持向量机分类器的 fit 函数, 使用训练集训练模型, 调整模型参数以提升分类性能。

为了评估效果, 本实验使用 Scikit-learn 库提供的 accuracy_score 计算分类器的准确率, 用 precision_score 计算分类器的精确率, 使用 recall_score 计算召回率, 结果如表 2 所示。

根据表 2, 计算得到平均准确率为 87%, 平均精确率为 85%, 平均召回率为 89%。实验结果表明, 基于 Python 实现的声音特征提取和分类方法在准确率、精确率及召回率方面都表现出良好的性能。本实验为 Python 在声音分析领域的应用提供了重要的参考和支持。

4 结语

文章主要研究基于 Python 的声音特征提取与分类方法, 利用 Librosa 和 TensorFlow 库实现声音信号的分析和处理, 重点分析 MFCC 特征提取和支持向量机分类方法的应用, 展示 Python 在声音分析领域的灵活性和强大功能。通过在 ESC-50 数据集上的实验分析, 验证了所提方法在声音特征提取与分类任务中具有良好的性能表现, 也进一步表明 Python 在声音信号处理领域具有巨大的应用潜力。

表 2 实验结果 单位: %

| 编号 | 准确率 | 精确率 | 召回率 |
|----|-----|-----|-----|
| 1 | 85 | 82 | 88 |
| 2 | 89 | 88 | 90 |
| 3 | 83 | 80 | 86 |
| 4 | 91 | 90 | 92 |
| 5 | 87 | 85 | 89 |

(下转第 27 页)

以实现；粒径比率越大，分离效果越好。

参考文献：

- [1] PENG J, LONG M, FEI Y C, et al. Progress in micro-scale acoustic manipulation based on surface acoustic wave[J]. Journal of Integration Technology, 2013(5):44–49.
- [2] SHI J, YAZDI S, LIN S C S, et al. Three-dimensional continuous particle focusing in a microfluidic channel via Standing Surface Acoustic Waves (SSAW) [J]. Lab on A Chip, 2011, 11(14):2319–2324.
- [3] KING L V. On the acoustic radiation on spheres[J]. Proceedings of the Royal Society A Mathematical Physical & Engineering Sciences, 1934, 147(861):212–240.
- [4] ZHU X, KIM E S. Microfluidic motion generation with acoustic waves[J]. Sensors and Actuators A Physical, 1998, 66 (1–3):355–360.
- [5] SHI J, MAO X, AHMED D, et al. Focusing microparticles in a microfluidic channel with Standing Surface Acoustic Waves (SSAW) [J]. Lab on A Chip, 2008, 8(2):221–223.
- [6] SHI J, HUA H, STRATTON Z, et al. Continuous particle separation in a microfluidic channel via Standing Surface

Acoustic Waves (SSAW) [J]. Lab on A Chip, 2009(23):3354–3359.

- [7] GOR'KOV L P. On the forces acting on a small particle in an acoustical field in an ideal fluid[J]. Soviet Physics Doklady, 1962, 6(1):773.
- [8] YANG W, LIU P, LI Q. Discrete simulation of particle manipulation in micro-fluid with acoustic[J]. Powder Technology, 2019, 356:618–627.
- [9] WU Y, YANG W J, ZHU F H, et al. Numerical study of particle separation with Standing Surface Acoustic Waves (SSAW) [J]. Powder Technology, 2022, 395:103–110.
- [10] CUNDALL P A. A computer model for simulating progressive large-scale movements in blocky rock systems[J]. Proceedings of the International Symposium on Rock Fracture, 1971, 2:129–136.
- [11] CHU K W, WANG B, XU D L, et al. CFD-DEM simulation of the gas-solid flow in a cyclone separator[J]. Chemical Engineering Science, 2011, 66(5):834–847.
- [12] GAN J Q, ZHOU Z Y, YU A B. CFD-DEM modeling of gas fluidization of fine ellipsoidal particles[J]. Aiche Journal, 2016, 62(1):62–77.

编辑：葛存

(上接第 22 页)

参考文献：

- [1] 禹琳琳.语音识别技术及应用综述[J].现代电子技术, 2013, 36(13):43–45.
- [2] 袁婷婷, 曹萌萌.以语音识别为基础的音乐检索系统[J].科技通报, 2015, 31(7):156–159.
- [3] 张科, 苏雨, 王靖宇, 等.基于融合特征以及卷积神经网络的环境声音分类系统研究[J].西北工业大学学报, 2020, 38(1):162–169.
- [4] 何月顺, 周良, 丁秋林.基于 Python 的音频数据处理研究与应用[J].微机发展, 2003, 13(11):51–53.
- [5] 董绍江, 刘伟.DenseNet 结合空间通道注意力机制的环境声音分类[J].重庆理工大学学报(自然科学), 2023, 37

(11):179–187.

- [6] 孙强, 李建华, 李生红.基于 Python 的文本分类系统开发研究[J].计算机应用与软件, 2011, 28(3):13–14.
- [7] 傅云, 景新幸, 叶懋.MFCC 特征改进算法在语音识别中的应用[J].计算机工程与科学, 2009, 31(12):146–148.
- [8] 周萍, 李晓盼, 李杰, 等.混合 MFCC 特征参数应用于语音情感能识别[J].计算机测量与控制, 2013, 21(7):1966–1968.
- [9] 丁世飞, 齐丙娟, 谭红艳.支持向量机理论与算法研究综述[J].电子科技大学学报, 2011, 40(1):2–10.
- [10] 顾亚祥, 丁世飞.支持向量机研究进展[J].计算机科学, 2011, 38(2):14–17.

编辑：郭芳园