



福州大学
FUZHOU UNIVERSITY

系统综合实践

语音识别实践

任课老师：刘莞玲

邮箱：liuwanling@fzu.edu.cn

平台：<http://met2.fzu.edu.cn/meol/index.do>

03

PART THREE



实验3-语音分类小实验

教学平台: <http://met2.fzu.edu.cn/meol/index.do>



1.1

课程简介----授课计划



序号	时间	授课内容	备注
1	第 2 周 0305	语音识别概念与行业发展介绍、实验环境准备	实验报告
2	第 3 周 0312	语音信号分类实验	实验报告
3	第 4 周 0319	传统语音识别模型搭建-1	实验报告
4	第 5 周 0326	传统语音识别模型搭建-2	实验报告
5	第 6 周 0402	基于深度学习的语音识别-1	实验报告
6	第 7 周 0409	基于深度学习的语音识别-2	实验报告
7	第 8 周 0416	基于Mind X SDK的语音识别	实验报告
8	第 9 周 0430	语音识别综合实践项目	实验报告
9	第10周	语音识别综合设计	课外完成
10	第11周	语音识别综合设计	课外完成
11	第12周	语音识别综合设计	课外完成
12	第13周	项目答辩考核	PPT++答辩+报告

实验课程介绍 - 语音分类

知识点	实验	实验简介	实验平台	课时	难度
语音分类	ModelArts语音分类实验	基于ModelArts自动学习实现语音分类	ModelArts	1	初级

ModelArts

总览

活动

自动学习

数据管理

开发环境

算法管理

训练管理

AI应用管理

部署上线

AI Gallery

专属资源池

总览

流程引导

评价

使用指南

月月享好礼

ModelArts 邀请好友 好礼1+1

AI算力、华为手表、手环、周边纪念品，随你拿！

购买套餐包

我的套餐包

产品定价

购买促销包 Hot!

新手入门

找云宝 (使用自动学习实现物体检测应用) 这里提供了一个“找云宝”样例 (“云宝” 是华为云的吉祥物)，帮助您快速熟悉ModelArts自动学习的使用过程。

1

准备工作

登录OBS服务，创建OBS桶和文件夹。(请自定义桶名称，不能与其他桶重名)

创建OBS桶和文件夹的操作指导请参见 [创建桶](#) 和 [新建文件夹](#)。为保证数据能正常访问，请务必保证创建的OBS桶与ModelArts在同一区域。

2

获取数据

单击 [数据集下载链接](#)，将“Yunbao-Data-Custom”数据集下载至本地并解压。


参考 [上传文件](#)，使用批量上传方式将“Yunbao-Data-Custom”文件夹下的所有文件上传至“test-modelarts/dataset-yunbao” OBS 路径下。获取的数据集分为2个子目录，分别为“eval”和“train”。“train”存储的数据用于

3

执行引导

完成前两步后，点击下面“开始引导”按钮，根据引导完成找云宝任务。

实验课程介绍 - 语音分类

 华为云 | 控制台 | 北京四

搜索

更多 | 中文 (简体) | hw34024505

ModelArts

总览

自动学习

数据管理

开发环境

算法管理

训练管理

AI应用管理


部署上线

AI Gallery

专属资源池

全局配置


自动学习



图像分类

识别一张图片中是否包含某种物体。


创建项目



物体检测

识别出图片中每个物体的位置及类别。


创建项目



预测分析

对结构化数据做出分类或数值预测。


创建项目



声音分类 **New!**

识别一段音频中是否包含某种声音。

创建项目



文本分类 **New!**

评价 | 使用指南 | 建议





创建声音分类项目

[< 返回自动学习](#)[😊 评价](#)

* 计费模式

按需计费

* 名称

exeML-4fa7

描述

0/256

数据集来源

新建数据集

已有数据集

目前支持的数据格式为：16bit的wav文件。

* 数据集名称

dataset-4fa7



* 数据集输入位置

请选择对象存储服务（OBS）路径



配置费用 **按需计费**

创建免费，使用阶段按训练及部署的时长收费。优先使用免费时长。[了解计费详情](#)



福州大学
FUZHOU UNIVERSITY

实验课程介绍 - 语音分类

本实验通过 ModelArts 自动学习实现动物叫声分类，主要面向业务开发者，无需专业的开发基础和编码能力，只需上传数据，通过自动学习界面引导和简单操作即可完成模型训练和部署。具体流程如图 2-1 所示。

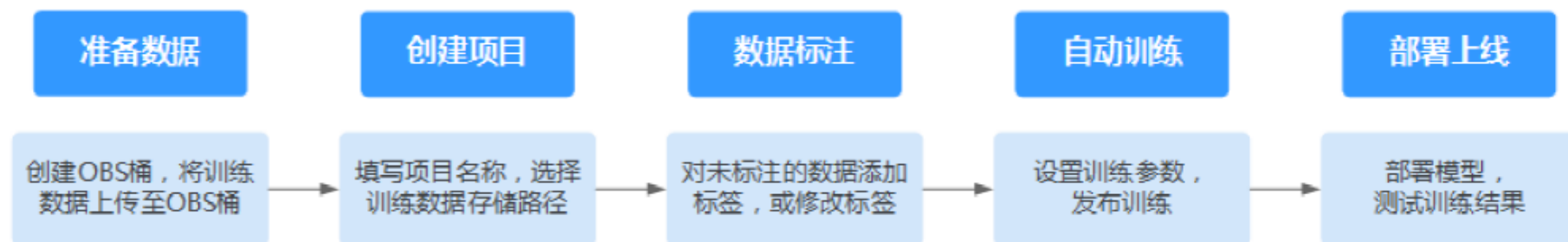


图2-1



您尚未开启敏感操作保护，存在安全风险，请您前往 [安全设置>敏感操作>操作保护](#) 开启敏感操作保护。 [前往开启](#)



华为云

控制台

北京四

OBS



更多

中文（简体）

hw34024505



产品管理

资源管理

自定义控制台



诚邀您参加控制台首页使用体验调研，您宝贵的意见和建议是我们持续提升产品体验的源动力。反馈即有机会获得2000码豆，码豆奖励规则：1. 码豆奖励将于问卷提交后的7个工作日内充值到您的账户中，可在[个人中心](#)查看。

收藏服务

弹性云服务器 ECS

云耀云服务器 HECS

裸金属服务器 BMS

弹性伸缩 AS

云硬盘 EVS

云备份 CBR

对象存储服务 OBS

虚拟私有云 VPC

弹性负载均衡 ELB

[查看更多服务](#)

最近访问的服务

ModelArts

对象存储服务 OBS

obs

OBS Browser+简介

OBS请求条件

obsutil使用

OBS目录导入操作



您好!

hw34024505

已实名认证

[费用中心](#)

使用过程中有任何问题，智能客服帮您解答
好的，知道了

[7天内待续费](#)

[未付订单](#)

[工单消息](#)

注册领8000元上云礼包

开年采购季



福州大学
FUZHOU UNIVERSITY

您尚未开启敏感操作保护，存在安全风险，请您前往 [安全设置>敏感操作>操作保护](#) 开启敏感操作保护。 [前往开启](#)



华为云

控制台

搜索



更多

中文（简体）

hw34024505



对象存储服务



总览 NEW



桶列表



并行文件系统



总用量 NEW



资源包管理 NEW



我的套餐



数据快递服务



云存储网关



媒体转码服务



CDN

桶列表 ? [开源声明](#)

评价

流程引导

使用指南

任务中心

创建桶

购买资源包

诚邀您参加对象存储服务使用体验调研，您宝贵的意见和建议是我们持续提升产品体验的源动力，感谢您的参与！

您还可以创建96个桶。

桶名称

请输入查询的关键词



桶名称	存储类别	区域	数据冗余存...	存储用量...	Data+...	对象数量...	创建时间	操作
422data	标准存储	华北-北京四	多AZ存储	0 byte		0	2021/04/22 ...	修改存储类别 删除
0422data	标准存储	华北-北京四	多AZ存储	0 byte		0	2021/04/22 ...	修改存储类别 删除
asr-data	标准存储	华北-北京四	多AZ存储	62.54 KB		2	2021/04/22 ...	修改存储类别 删除
dataasr-data1	标准存储	华北-北京四	多AZ存储	48.99 KB		2	2021/02/25 ...	修改存储类别 删除



福州大学
FUZHOU UNIVERSITY



复制桶配置

选择源桶

该项可选。选择后可复制源桶的以下配置信息：区域 / 数据冗余策略 / 存储类别 / 桶策略 / 默认加密 / 归档数据直读 / 企业项目 / 标签。

区域

华北-北京四

不同区域的云服务产品之间内网互不相通；请就近选择靠近您业务的区域，可减少网络时延，提高访问速度。[如何选择区域](#) ?

数据冗余存储策略

多AZ存储

单AZ存储



启用后不支持修改。多AZ存储采用相对较高计费标准。[价格详情](#)

桶名称

数据在同区域的多个AZ中存储，可用性更高。

默认存储类别

标准存储

适合高性能，高可靠，高可用，频繁访问场景

多AZ存储

单AZ存储

图片处理

低频访问存储

适合高可靠，低成本，较少访问场景

多AZ存储

单AZ存储

图片处理

归档存储

适合长期存储，基本不访问场景

单AZ存储

数据冗余有

创建桶时选择的存储类别会作为上传对象的默认存储类别。[了解存储类别差异](#) ?

桶策略

私有

公共读

公共读写

复制桶策略



任何用户都可以对桶内对象进行读操作，仅桶拥有者可以进行写操作。

默认加密

☐ 开启默认加密 ?

免费

建议开启默认加密，密钥管理全免费，核心数据更安全。

归档数据直读

开启

关闭



桶名称 下拉

存储类别 下拉

区域 下拉

数据冗余存...

存储用量...

Data+...

speech

--

华北-北京四

--

--



数据集

获取的数据集包含“train”和“test”两个目录，其中训练集位于 sound_classification\train 目录下，共 4 类动物叫声，分别是鸟（bird）、猫（cat）、狗（dog）和虎（tiger），每种动物 25 个叫声音频数据，共一百条音频数据。

测试集位于 sound_classification\test 目录下，每种动物提供 5 个叫声音频数据。音频数据格式均为 wav 格式，均可在各音频播放器直接播放试听。

指导书 > 实验3-语音分类 > data2 > 语音分类 >

名称

- test_data
- training_data

指导书 > 实验3-语音分类 > data2 > 语音分类 > training_data

名称

#

标题

- 1_1.wav
- 1_2.wav
- 1_3.wav
- 1_4.wav
- 1_5.wav
- 1_6.wav
- 1_7.wav
- 1_8.wav
- 1_9.wav
- 1_10.wav
- 2_1.wav

100个数据

指导书 > 实验3-语音分类 > data2 > 语音分类 > test_data

名称

#

标题

- 1_1.wav
- 1_2.wav
- 1_3.wav
- 1_4.wav
- 1_5.wav
- 1_6.wav
- 1_7.wav
- 1_8.wav
- 1_9.wav
- 1_10.wav

1-创建声音自动学习

☰

☁

🔗

📶

☁

📱

📁

☁

📶

🌐

📶

创建声音分类项目

返回自动学习

* 计费模式

按需计费

* 名称

syfl--2137

▲

描述

语音分类

4/256

数据集来源

新建数据集

已有数据集

目前支持的数据格式为：16bit的wav文件。

* 数据集名称

dataset-2137

✓

* 数据集输入位置

🔍

/speech/test/

📁



1.1-设置数据集输入输出位置

数据集输出位置

请选择文件夹。

speech

请输入名称查询

新建桶 >>

桶名称

0422data

422data

asr-data

dataasr-data1

speech

新建文件夹

输入对象名前缀搜索

名称	最后修改时间	类型	大小
outdata	--	文件夹	--
test	--	文件夹	--

4/256

新建数据集

已有数据集

目前支持的数据格式为：16bit的wav文件。

* 数据集名称

dataset-2137

* 数据集输入位置

/speech/inputdata/

* 数据集输出位置

/speech/outdata/

添加标签集

动物声音

添加标签

自行选择OBS桶

1.2上传数据



The screenshot displays the Huawei Cloud console interface. On the left is a sidebar with various service icons. The main area shows a message: "未查询到音频" (No audio files found) and "请添加并标注音频文件，以准备训练数据。" (Please add and label audio files to prepare training data). A blue link "添加音频" (Add Audio) is visible, with a red arrow pointing to it. A modal dialog box titled "添加音频" (Add Audio) is open on the right. It contains a button "添加音频" (Add Audio) with a red arrow pointing to it, and a text box stating: "仅支持WAV格式音频文件，单个音频文件不能超过4MB，且单次上传的音频文件总大小不能超过8MB。" (Only WAV format audio files are supported. Individual audio files cannot exceed 4MB, and the total size of audio files uploaded at one time cannot exceed 8MB). At the bottom of the dialog are "确定" (Confirm) and "取消" (Cancel) buttons.

华为云 控制台

添加音频

未查询到音频

请添加并标注音频文件，以准备训练数据。

添加音频

添加音频

仅支持WAV格式音频文件，单个音频文件不能超过4MB，且单次上传的音频文件总大小不能超过8MB。

确定 取消

1.2上传数据

台

除音频

同步数据源

添加音频

已添加 100 个音频文件 (27.47 MB) [清空](#)

仅支持WAV格式音频文件，单个音频文件不能超过4MB，且单次上传的音频文件总大小不能超过8MB。

bird_0-2_1595274333614.wav

×

bird_1-2_1595274333763.wav

×

bird_2-2_1595274333794.wav

×

bird_3-2_1595274333812.wav

×

bird_4-2_1595274333851.wav

×

bird_5-2_1595274333899.wav

×

bird_6-2_1595274333923.wav

×

确定

取消

更多

中文 (简体)

hw34024505

27

全部标签 1

标签	数量	快捷键	操作
四种动物叫声	0	--	

...

😊

?

📄

数据上传和标注



添加音频

添加音频

仅支持WAV格式音频文件，单个音频文件不能超过4MB，且单次上传的音频文件总大小不能超过8MB。

确定

取消



福州大学
FUZHOU UNIVERSITY

数据删除

已标注 35

未标注 10

+ 添加音频

删除音频

同步数据源

选择当前页

cat_18-2_1595274...

cat_16-2_1595274...

cat_19-2_1595274...

cat_10-2_1595274...

cat_12-2_1595274...

标签 已选择 0 个音频文件

--

快捷键 ? 标签

确定

取消

cat_18-2_1595274...

cat_16-2_1595274...

cat_19-2_1595274...

cat_10-2_1595274...

cat_12-2_1595274...

+ 添加音频

删除音频

同步数据源

选择当前页

bird_16-2_159527...

bird_17-2_159527...

bird_12-2_159527...

bird_18-2_159527...

bird_15-2_159527...

bird_13-2_159527...

bird_19-2_159527...

bird_11-2_159527...

bird_20-2_159527...

bird_14-2_159527...

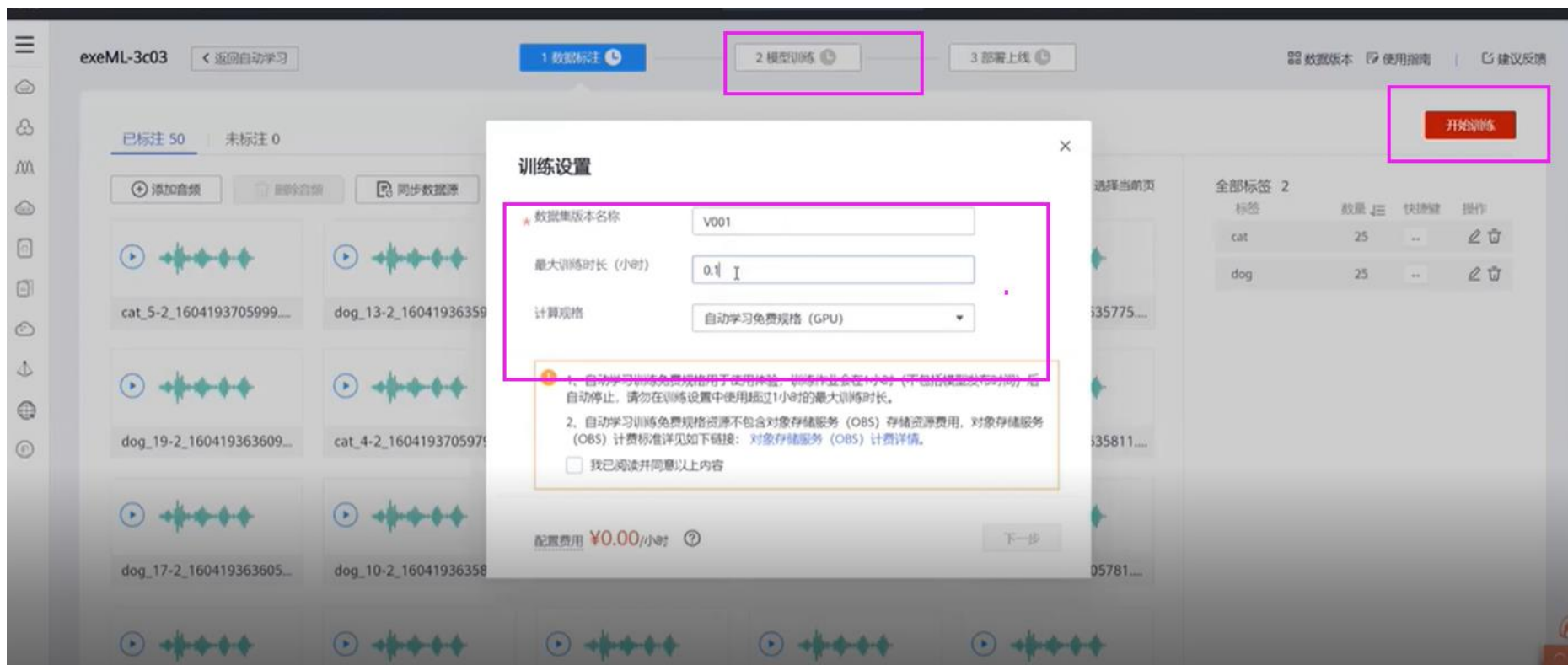
示签 3

签	数量	快捷键	操作
d	25	1	<div>✎</div> <div>🗑</div>
	10	2	<div>✎</div> <div>🗑</div>
中动物叫声	0	--	<div>✎</div> <div>🗑</div>

福州大学

FUZHOU UNIVERSITY

模型训练



开始模型训练



华为云

控制台

搜索



更多

中文 (简体)

hw34024505



exeML-b747

< 返回自动学习

1 数据标注

2 模型训练

3 部署管理

评价

使用指南

建议反馈

版本管理



训练中

V001 (543fb898-c3f...

2022/03/17 10:55:46 G...

准确率: --

停止

删除

训练详情



训练任务提交成功!

当前版本的训练状态为 **训练中**, 请耐心等待。

返回自动学习



福州大学
FUZHOU UNIVERSITY

模型训练结果

华为云

控制台

搜索

更多

中文 (简体)

hw34024505

27

exeML-b747

返回自动学习

1 数据标注

2 模型训练

3 部署版本管理

评价

使用指南

建议反馈

版本管理

✓

已完成

V001 (543fb898-c3f...

2022/03/17 10:55:46 G...

准确率: 95%

部署

删除

训练详情

已完成

开始时间

2022/03/17 10:56:24 GMT+08:00

训练时长

00:01:09

评估结果

召回率

0.950

精确率

0.958

准确率

0.950

F1值

0.949

训练参数

训练时长不大于 (min)

6

计算规格

自动学习免费规格 (GPU)

分类统计表

请输入关键字查询


标签名	F1值	精确率	召回率
tiger	1.000	1.000	1.000
cat	0.889	1.000	0.800
bird	0.909	0.833	1.000
dog	1.000	1.000	1.000


AI应用: exeML-b747_ExeML_66802c02 0.0.1

福州大学

FUZHOU UNIVERSITY

模型部署


 华为云 | 控制台

搜索 

更多


中文 (简体)


hw34024505


 27

exeML-b747

[< 返回自动学习](#)

1 数据标注 

2 模型训练 


3 部署版本管理 


[评价](#)

[使用指南](#)

[建议反馈](#)

版本管理



	V001 (543fb898-c3f... 2022/03/17 10:55:46 G...	<div>部署</div> <div>删除</div>
已完成	准确率: 95%	

训练详情

已完成

开始时间 2022/03/17 10:56:24
GMT+08:00

训练时长 00:01:09

评估结果

训练参数





 福州大学
FUZHOU UNIVERSITY

模型



exeML-b747

< 返回自动学习

1 数据标注

2 模型训练

3 部署版本管理

评价

使用指南

版本管理



已完成

V001 (543fb898-c3f...

2022/03/17 10:55:46 G...

准确率: 95%

部署

删除

训练详情

已完成

开始时间 2022/03/17 10:56:24

GMT+08:00



华为云

控制台

搜索



更多

中文 (简体)

hw340243



exeML-b747

< 返回自动学习

1 数据标注

2 模型训练

3 部署上线

评价

使用指南

版本管理



部署中

2022/03/17 11:09:34 ...

启动

删除

服务测试



部署任务提交成功!

当前版本的部署状态为 **部署中**, 请耐心等待。

返回自动学习

模型效果测试



华为云

控制台

搜索



更多

中文 (简体)

hw34024505



exeML-b747

< 返回自动学习

1 数据标注 ✓

2 模型训练 ✓

3 部署上线

评价

使用指南

建议反馈

版本管理



运行中

2022/03/17 11:09:34 ...

56 分钟 后停止 | 设置

停止

删除

服务测试

当服务状态为“ 运行中”时，才能进行预测。

服务名称: exeML-b747_ExeML_1647486572452218981

选择需要预测的文件

上传

重新预测

文件预览



cat_10-2_1595274027664.wav

预测结果

预测结果: cat

```
1  
2 predicted_label: "cat",  
3 : 0.999999  
4
```

exeML-b747

< 返回自动学习

1

版本管理



运行中

2022/03/17 11:09:34 ...

54 分钟 后停止 | 设置

停止

删除



福州大学
FUZHOU UNIVERSITY



华为云

控制台

北京四

搜索



更多

中文 (简体)

hw34024505



27



ModelArts



总览



活动



自动学习



数据管理



开发环境



算法管理



训练管理



AI应用管理



部署上线



AI Gallery



专属资源池

△ 目录



文本分类 **New!**

识别一段文本的类别。

创建项目

所有分类

请输入项目名称查询



项目名称

项目类型

训练状态 (个...

部署状态...

数据源

创建时间

描述

操作

exeML-b747

声音分类

已完成 1

停止 1

dataset-b748

2022/03/17 10:30:06...

语音分类--动物

删除

syfl--2137

声音分类

--

--

dataset-2137

2022/03/17 09:15:52...

语音分类

删除



福州大学
FUZHOU UNIVERSITY

利用Python语言和Librosa库来对动物声音进行语音分类

```
import os
import librosa
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.svm import SVC

# 设置数据集路径
dataset_path = "path/to/dataset"

# 定义一个函数来从数据集中加载所有音频文件
def load_data(dataset_path):
    # 存储特征值和标签
    features = []
    labels = []

    # 遍历数据集的所有类别
    for label in os.listdir(dataset_path):
        # 拼接类别文件夹路径
        class_path = os.path.join(dataset_path, label)
        # 遍历类别文件夹下的所有音频文件
        for audio_file in os.listdir(class_path):
            # 加载音频文件
            file_path = os.path.join(class_path, audio_file)
            x, sr = librosa.load(file_path)

            # 抽取MFCC特征
            mfccs = librosa.feature.mfcc(y=x, sr=sr, n_mfcc=40)
            mfccs_scaled = np.mean(mfccs.T, axis=0)

            # 将特征值和标签添加到列表中
            features.append(mfccs_scaled)
            labels.append(label)

    return np.array(features), np.array(labels)

# 加载数据集并将其划分为训练集和测试集
X, y = load_data(dataset_path)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)

# 训练SVM模型
model = SVC(kernel='linear')
model.fit(X_train, y_train)

# 在测试集上评估模型性能
accuracy = model.score(X_test, y_test)
print("Accuracy:", accuracy)
```

选做题：语音分类综合设计实验

1.复现基于MindX SDK的语音多分类

2.复现基于交叉熵的语音分类

3.开放创新基于开源数据集做分类实验 <https://www.openslr.org/resources.php>



OpenSLR

Open Speech and Language Resources

[Home](#) [Resources](#) [Contribute](#)

About OpenSLR

OpenSLR is a site devoted to hosting speech and language resources, such as training corpora for speech recognition, and software related to speech recognition. We intend to be a convenient place for anyone to put resources that they have created, so that they can be downloaded publicly.

Part of our goal is to mirror software available elsewhere, in order to provide a failover location. We are starting by mirroring some software which is used in the [Kaldi](#) scripts.

We aim to provide a central, hassle-free place for others to put their speech resources. For more information, see [here](#).

If you wanna keep updated, we have created a mailing list openslr-news@googlegroups.com in which we intend to send announcements about newest additions and other news.

For a list of resources, please click on [resources](#) above.

If you want to download things from this site, please download them one at a time, and please don't use any fancy software-- just download things from your browser or use 'wget'. We have a firewall rule to drop connections from hosts with more than 5 simultaneous connections, and certain types of download software may activate this rule.

Now, aside from the main site openslr.org, we also have a mirror in China that is available at openslr.magicdatatech.com. The mirror server is made available by [Magic Data Technology](#). Thank you Magic Data Technology!
Another mirror of the content is provided by [ELDA](#) in EU. The server is available as openslr.elda.org. Thank you ELDA!

Contact

jtrmal@gmail.com
(Jan "Yenda" Trmal)



福州大学
FUZHOU UNIVERSITY

注意事项

★ 镜像

公共镜像自定义镜像?

★ 镜像

公共镜像自定义镜像?

前往AI Gallery获取更多镜像

请输入镜像名称

名称

描述

☒

mindspore1.7.0-cann5.1.0-py3.7-euler2.8.3

Ascend+ARM算法开发和训练基础镜像, AI引擎预置MindSpore

☐

tensorflow1.15-cann5.1.0-py3.7-euler2.8.3

Ascend+ARM算法开发和训练基础镜像, AI引擎预置TensorFlow

☐

mindspore_1.9.0-cann_6.0.0-py_3.7-euler_2.8.3

Ascend+ARM algorithm development and training. MindSpore is preset in...

☐

mindspore_1.7.0-cann_6.0.1-py_3.7-euler_2.8.3

Ascend+ARM algorithm development and training. MindSpore is preset in...

☐

cylp0.91.4-cbcipy2.10-ortools9.0-cplex20.1.0-ubuntu18.04

CPU运筹优化求解器开发基础镜像, 预置cylp, cbcipy, ortools及cplex.

10 总条数: 25 < 1 2 3 >

☐

pytorch1.10.0-cann5.1.0-py3.7-euler2.8.3-ubuntu18.04

CPU、GPU通用算法开发和训练基础镜像, 预置AI引擎PyTorch1.4

☐

tensorflow1.15-mindspore1.7.0-cann5.1.0-euler2.8-aarch64

Ascend+ARM算法开发和训练基础镜像, AI引擎预置TensorFlow和MindSp...

10 总条数: 25 < 1 2 3 >

《系统综合实践》-----实验3

语音分类实验

序号	任务
1	查阅相关资料，理解语音分类相关知识
2	熟悉华为实验云平台、配置自动学习环境
3	完成语音语音分类的实验任务
4	语音分类综合设计实验（加分选做）
5	提交实验报告



Thank you!

Any questions?

