

项目功能

语音分类，从给定的几类语音中使用模型学习，并且预测某个给定语音属于的类型，默认数据为四句诗的分类："望门投止思张俭","忍死须臾待杜根","我自横刀向天笑","去留肝胆两昆仑".

数据类型

data文件夹放的是训练数据，test文件夹放的是测试数据，data和test文件夹下的文件结构相同
data文件夹下放的是n个文件夹，每个文件夹代表一个类，文件夹的名字为类名，文件夹下面放属于该类的语音，文件格式为wav

可以按照自己的需求调整文件夹内的东西，但是文件结构不能改变

语音长度不宜过长，有可能会引发一些异常导致程序无法正常运行，但是一句诗的时间一般而言是没有问题的

程序功能

model.py

定义了rnn模型，内部由两部分构成，第一部分是LSTM，负责学习分析到输入数据沿时间方向的信息，第二部分是一个深度神经分类网络，相比最初的版本（两层），我使用了更深的三层神经网络加上激活函数和批量归一化，经过这样的改进，训练集上的收敛速度显著快于最初版本，在数据量增加时，同时解决了最初版本在数量级较大时难以收敛的问题。通过巧妙地加深加宽神经网络，这个模型可以学习识别语音的更多信息从而更准确地实现语音分类

train.py

训练rnn模型，调用dataloader读取data中的数据，训练epoch次，epoch默认为三十，可以增加epoch使得训练更为充分，在训练集上准确性更高，但有可能导致过拟合，需要参考测试集的准确性做取舍

每训练10次，会跑两次预测，输出模型分别在训练集和测试集的准确性，输出频率可调整

dataloader.py

基于pytorch自带的读取模型拓展的数据读取模块，可以根据给出的data_path不同，读取不同的文件夹的数据，文件结构需符合数据类型部分所提到的结构

predict.py

根据给定的rnn模型和测试数据类型(data或者test)，进行预测，输出准确率

如果detail为true，会输出预测细节

arguments.py

记录程序中用到的参数

其他未提到的py文件属于第二个project的展示功能，需要进一步配置更多环境，所以在这里就不再提及，可以忽略

使用方法

安装环境

1. 安装cuda，gpu驱动
2. 安装python，理论上3.6, 3.7都可以

3. 安装pytorch, 去官网根据cuda和python版本安装即可 (下面是测试方法)

```
dararara@dararara-G3-3579:~/playground/audio-classification$ python
Python 3.6.8 (default, Oct  7 2019, 12:59:55)
[GCC 8.3.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> import torch
>>> torch.cuda.is_available()
True
>>> 
```

4. 安装相关库, numpy, scipy,, python_speech_features, 可以使用pip install numpy, scipy, python_speech_features
5. 到此为止, 理论上环境就配好了

测试方法

python train.py 训练模型, 存储为rnn.pth

python predict.py 测验模型, 读取rnn.pth, 输出预测结果

```
dararara@dararara-G3-3579:~/playground/audio-classification$ python3 predict.py
8
predict.py:55: UserWarning: Implicit dimension choice for softmax has been deprecated. Change the call to include dim=X as an argument.
  prob = torch.nn.functional.softmax(test_pred)
predict class is  sen4  true class:  sen4
predict class is  sen4  true class:  sen4
predict class is  cat  true class:  sen3
predict class is  sen3  true class:  sen3
predict class is  dog  true class:  dog
predict class is  dog  true class:  dog
predict class is  cat  true class:  cat
predict class is  dog  true class:  cat
acc:  6 / 8
dararara@dararara-G3-3579:~/playground/audio-classification$ 
```

如果cuda安装失败, 可以删去代码中的cuda(), 本项目基于第二次project进行了模型性能的改进, 功能与第二次project相同