**8 声纹识别**

1. 目的
   1. 掌握声纹识别的基本原理。
   2. 掌握声模型纹识别的实现方法。
2. 要求

基于TensorFlow框架搭建声纹识别模型并训练，实现对未知人的声音识别。训练好的模型（https://pan.baidu.com/s/1dO3L8WNBVFuS4dVUsxuDiw）。

1. 工具及数据集
2. 主要工具

TensorFlow-gpu-1.12.0、librosa等

1. 数据集

CSTR VCTK语料库包括109名母语为英语且具有各种口音的语音数据。每位发言者读出大约400个句子。语料库下载地址：<https://datashare.is.ed.ac.uk/handle/10283/2651>。

使用其中90%作为模型训练数据，10%作为测试数据。

1. 步骤
2. 数据预处理（data\_preprocess.py）

按9：1划分训练集与数据集并采用滑动窗口方法利用librosa语音处理库对语音的[MFCC](https://zh.wikipedia.org/wiki/%E6%A2%85%E5%B0%94%E9%A2%91%E7%8E%87%E5%80%92%E8%B0%B1%E7%B3%BB%E6%95%B0)（梅尔频率倒谱系数）特征进行提取，并保存为.npy格式的文件至train\_tisv和test\_tisv两个文件夹，文件表示每一位说话者的声纹特征。

1. 构建训练模型

1)设置输入批次（utiles.py）

在步骤1中生成的.npy文件中，随机选择N位说话者并对应选择M条口语声纹特征，组成的batch大小N×M。

2）构建网络（model.py)

使用3层LSTM网络，输出作为Embedding d-Vector，再进行L2正则化，得到的向量就是说话人的声纹表征。

1. 计算相似度(utiles.py)

2019-04-23 10-52-25屏幕截图

1. 模型训练(model.py)

指定相关参数，如模型保存位置、GPU编号等(utiles.py)开始训练。

1. 模型测试(model.py, utiles.py)

利用已处理好的测试数据进行模型的测试使用，并根据评价指标EER分析实验结果。

1. 结果
2. 训练细节

每批次说话者的数量，4；每个说话者的话语数量，5；

优化器，SGD；

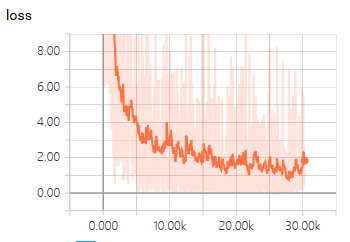
初始学习率，0.02；

迭代次数，30k

训练时间，30k

1. 训练结果

Loss最低达到0.005



1. 测试结果

错误率，EER：0.05

错误接受率（FAR） ，0.05

错误拒绝率（FRR），0.05

