**实验四、声纹识别系统实现及验证**

1. **实验目的**

本次实验利用公开数据库实现声纹识别系统并进行验证。该实验旨在贯通课程学习中声纹识别的相关内容。通过实现前端特征提取、深度神经网络、池化层、损失函数等必要步骤，掌握声纹识别的整体流程，并了解部分优化方法。

1. **实验平台及编程环境**

python3，pytorch，librosa，numpy，pandas

1. **实验内容（标注“\*”内容为选做实验）：**
2. **数据准备：**

实践内容基于中文语料库AiShell（https://www.openslr.org/33/）[实现。该数据集包含340](https://www.openslr.org/33/)实现。该数据集包含340)位说话人的训练数据，40位说话人的开发数据，以及20位说话人的测试数据。本次实验共选取300位说话人，每位说话人10条语音作为训练数据；20位说话人，每位说话人10条语音作为测试数据。

1.1 数据配置

将**config.yaml**中的**root**选项修改为具体的数据集位置。

文本

描述已自动生成

文本

描述已自动生成

1. **实现基于x-vector的声纹识别系统（2学时）：**

补全xvector.py、pooling.py、loss.py中的缺失代码，使其可以正常运行；以课程中所学习的声纹识别训练过程为依据，调用必要函数，训练基本的声纹识别系统。（第一部分实验采用前端特征：mfcc，深度神经网络：TDNN，池化层：global average pooling，损失：softmax）

2.1 全局平均池化层（GlobalAveragePooling）的实现

根据**pooling.py**中**GlobalAveragePooling**类的定义，以及提示的forward函数的输入输出维度，补全**forward**函数。

文本

描述已自动生成

2.2 x-vector中深度神经网络的实现

在**xvector.py**文件中，已经基本完成了Xvector各层的定义，但仍然缺少池化层以及最终的线性层。请参考各层的定义代码，为模型添加池化层和最终的线性层，补全**\_\_init\_\_**函数。

电脑萤幕截图

描述已自动生成

调用Xvectorc定义的各层，补全Xvector类的**forward**函数。

电脑萤幕的截图

描述已自动生成

2.3 Softmax损失函数的实现

损失函数的计算主要分为两步：1) 线性层将嵌入映射到说话人标签空间上；2) 采用交叉熵计算损失。请以此为依据补全**loss.py**文件中Softmax的**forward**函数。

文本

描述已自动生成

2.4 基于x-vector的声纹识别系统的训练

在**train.py**中完成模型、损失函数、学习器的定义。

文本

描述已自动生成

在**train.py**中完成嵌入的提取、损失的计算、正确率的计算以及反向传递。

文本

描述已自动生成

2.5 请给出模型在训练过程中的输出截图或视频。

1. **声纹识别系统的验证（1学时）：**

3.1 基于x-vector的声纹识别系统的验证

在**EER.py**中，定义并导入模型(xvector\_gap\_softmax\_checkpoint.pth)；为测试数据生成深度嵌入。

文本

描述已自动生成

在**verification**函数中计算注册数据嵌入和测试数据嵌入之间的余弦相似度。

文本

描述已自动生成

通过完成**roc**函数来绘制ROC曲线。

文本

描述已自动生成

等错误率（Equal Error Rate EER）是声纹识别中常用的评价指标。请根据EER.py的相关代码**get\_eer**函数，简述等错误率（EER）的计算流程。并在ROC曲线上画出EER的对应点。

文本

描述已自动生成

3.2 请给出模型验证结果的截图以及绘制的ROC曲线。

1. **声纹识别优化方法的研究\*（1学时）：**

目前，在x-vector的基础上，涌现了许多有效的优化方法。其中，Am-softmax以及ECAPA-TDNN分别在损失函数及神经网络架构上进行了改进。项目文件中已经进行了实现，请尝试选择一至两种方法进行调用，替换训练代码中的对应部分，并简要说明方法的优势。（注：<https://arxiv.org/pdf/1801.05599.pdf>； <https://arxiv.org/pdf/2005.07143.pdf>）

4.1 Am-softmax的实现

Am-softmax的相关代码保存在**loss.py**文件中，请在**train.py**文件中对其进行调用并运行，给出训练过程中的输出截图或视频。

文本

描述已自动生成

4.2 ECAPA-TDNN的实现

ECAPA-TDNN的相关代码保存在**ecapa.py**文件中，请在**train.py**文件中对其进行调用并运行，给出训练过程中的输出截图或视频。（提示：在使用ECAPA-TDNN计算语音的深度嵌入时，需要转置特征的第1维和第2维）

1. **参考资料**

[1] SpeechBrain官方文档及代码

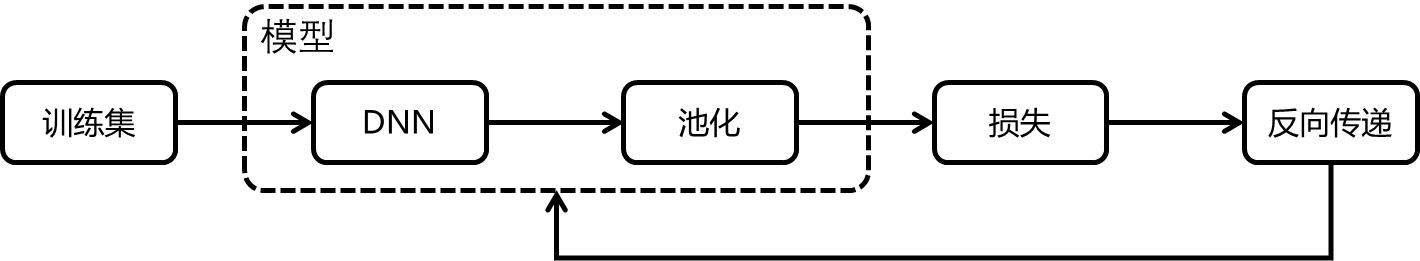
<https://speechbrain.readthedocs.io/en/latest/index.html>

<https://github.com/speechbrain/speechbrain>；

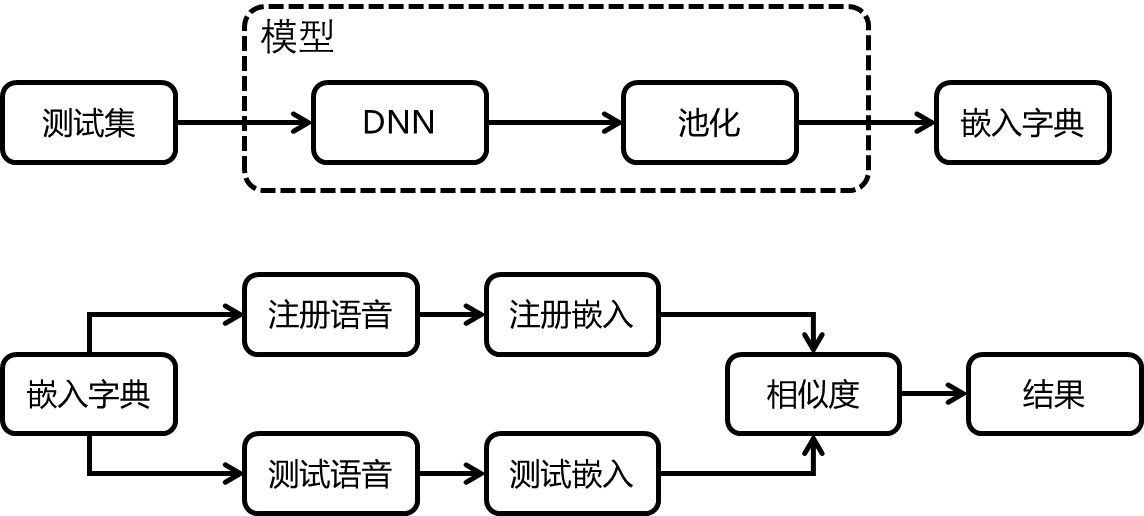
[2] 声纹识别系统搭建

<https://github.com/zengchang233/asv_beginner/tree/master>

1. **声纹识别及模型搭建参考**
2. 声纹识别系统训练流程



1. 声纹识别系统应用流程



1. Pytorch模型搭建参考

<https://pytorch.org/tutorials/beginner/blitz/neural_networks_tutorial.html#sphx-glr-beginner-blitz-neural-networks-tutorial-py>

图形用户界面, 文本, 应用程序

描述已自动生成

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成