



哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

2022 年秋季学期

《视听觉信息理解》课程

听觉部分实验

姓名	姚舜宇
学号	1190202107
班号	1903602

1. 实验目的

掌握一种说话人识别/语音识别工具包的使用方法，如 Kaldi

能够读懂，并按照要求修改说话人确认算法的源代码

能够利用该工具包，完成指定的实验方案，并优化实验结果

能够在实验报告中条理清晰地、详实专业地汇报实验过程，总结实验结果

2. 实验过程

1. 下载实验指导书中给出的 kaldi 源码以及数据集，并按照给定的连接进行环境的配置与安装

2. 进入实验指导书中给出的 egs/voxceleb/v2 文件夹，阅读 run.sh 文件。该文件是脚本文件，用于批处理，描述了实验的训练与测试过程

3. 该 run.sh 文件的流程如下：

首先准备数据集，然后对数据集提取 MFCC 特征

对数据进行增强，利用回响和三种噪声增强了训练数据，并测试增强效果

通过归一化的方法除去静音帧

生成的 egs 用于训练 x-vector 模型

用 embedding 层提取特征

用 LDA 降维

用 PLDA 打分

计算 EER

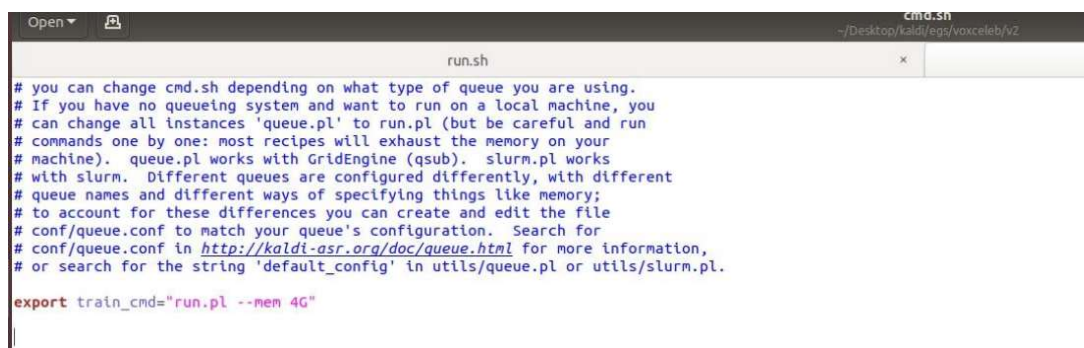
4. 由于实验指导书中已经给定模型和数据集，即要求修改此文件使得能在所给的数据集上进行复现测试，因此对其改动如下：

只保留 stage1 对测试数据集的处理，保留 stage9 中的“Extract x-vectors used

in the evaluation"部分，保留 stage11 和 stage12 部分用于进行评分测试。其他部分可以删除。详见 sh 文件

5. 测试过程如下：

首先修改 cmd.sh 文件如下：

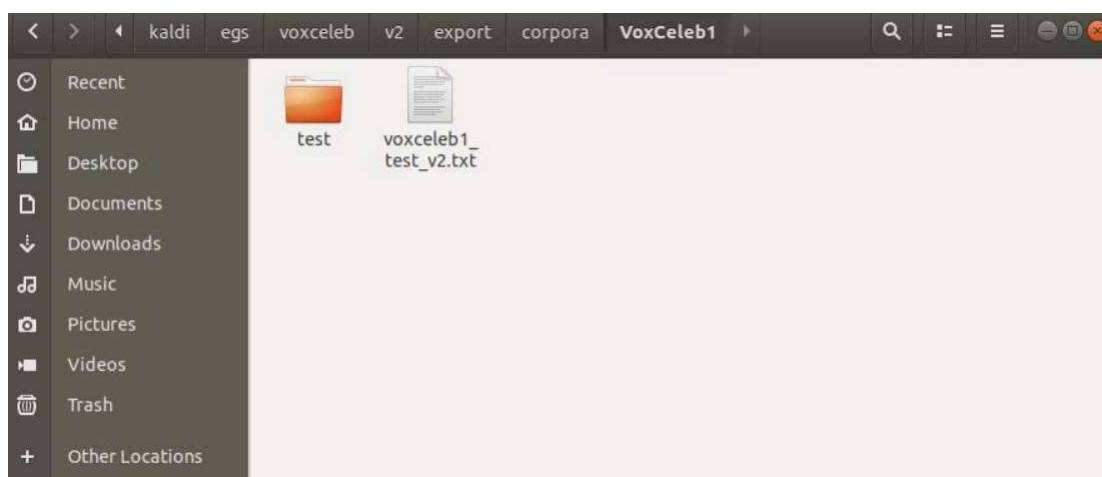


```
Open ▾  cmd.sh
~/Desktop/kaldi/egs/voxceleb/v2

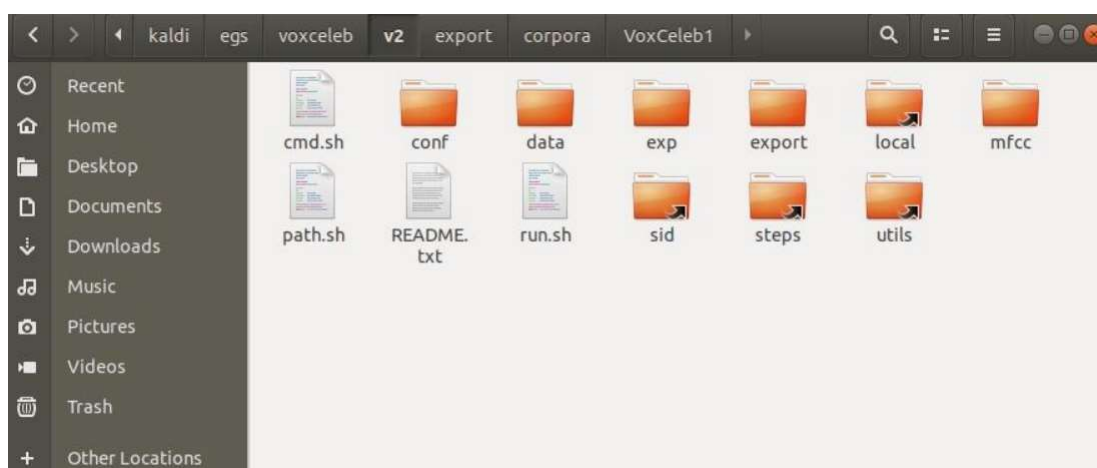
run.sh
# you can change cmd.sh depending on what type of queue you are using.
# If you have no queueing system and want to run on a local machine, you
# can change all instances 'queue.pl' to run.pl (but be careful and run
# commands one by one: most recipes will exhaust the memory on your
# machine). queue.pl works with GridEngine (qsub). slurm.pl works
# with slurm. Different queues are configured differently, with different
# queue names and different ways of specifying things like memory;
# to account for these differences you can create and edit the file
# conf/queue.conf to match your queue's configuration. Search for
# conf/queue.conf in http://kaldi-asr.org/doc/queue.html for more information,
# or search for the string 'default_config' in utils/queue.pl or utils/slurm.pl.

export train_cmd="run.pl --mem 4G"
```

分别下载模型与数据，测试数据放在如下文件夹中：



模型放在如下文件：



然后手动调用 `make_voxceleb1_v2.pl` 对测试集数据进行处理，并生成后续所需的 `scp` 文件。

在根目录下执行 `bash run.sh` 命令，即可以运行。

3. 实验结果

最终得到的实验结果如下所示：

```
EER: 3.128%  
minDCF(p-target=0.01): 0.3258  
minDCF(p-target=0.001): 0.5003
```