

**2023年秋季学期  
《视听觉信息理解》课程**

**听觉部分实验：**

**说话人确认实验**

|  |  |
| --- | --- |
| **姓名** | **刘天瑞** |
| **学号** | **7203610121** |
| **班号** | **20W0362** |
| **专业** | **人工智能（视听觉信息处理）** |
| **指导老师** | **郑铁然** |
| **时间** | **2023.10.11** |

1. **实验目的**
2. 掌握一种说话人识别/语音识别工具包的使用方法（如 Kaldi）；
3. 能够读懂，并按要求修改说话人确认算法的源代码；
4. 能够利用该识别工具包，完成指定的实验方案，并优化实验结果；
5. 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；
6. 提出有创新性的改进方案，并进行验证。
7. **选择的实验平台**

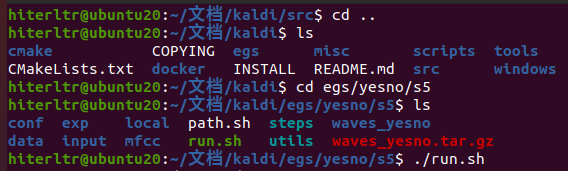
语音识别工具包框架Kaldi（使用WFST来实现解码算法）；VMware Workstation 17；Ubuntu 20.04。

1. **实验方案步骤过程**
2. 首先下载实验指导书中给出的kaldi源码以及测试数据集，并按照给定的github链接进行相关环境的配置与安装；

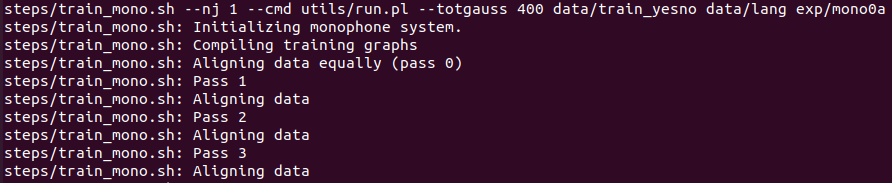
安装步骤：



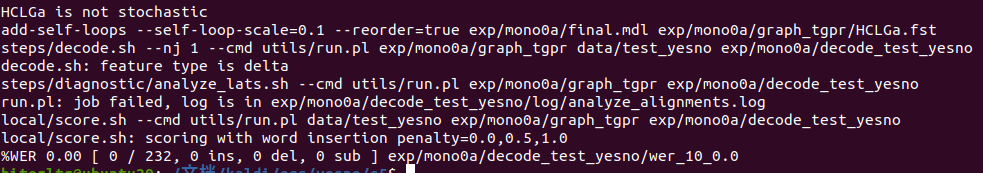
下载完成后进入tools目录，执行 extras/check\_dependencies.sh 检查有哪些必要的库没有安装，然后再执行其所推荐的安装方法安装，这里值得注意的是，安装MKL数学库时发生GPG密钥缺失无法解决，手动下载安装后续也会出现软件源无法定位的问题，所以先忽略该库直接执行编译：make -j 4，之后再进入src目录，执行./configure --mathlib=ATLAS，相当于更换了数学核心库，除了使得Kaldi相对慢一些无区别。然后再执行make depend -j 4（电脑虚拟机处理器数目），执行编译make -j 4，安装完成。使用yesno测试成功无误：



中间运算过程：



识别结果：



1. 然后进入实验指导书中给出的kaldi/egs/voxceleb/v2目录，分析run.sh脚本文件。说话人识别算法结构与流程描述：基于VoxCeleb1数据集的说话人识别系统，即名为VGGVox的嵌套系统。该系统以直接从原始音频中提取的短时语谱图为训练对象，无需其他预处理。采用深度神经网络主干结构来提取帧级特征，然后池化成句子级别的说话人嵌入向量。整个模型使用对比损失进行训练。为了提高模型性能，该文首先使用softmax层和交叉熵进行了识别预训练。并且对于评估方法，该系统使用的测试集是VoxCeleb1数据集，其本身就含有配对信息。使用的性能度量标准为：等错误率（EER）：即错分正负样本概率相等的值以及损失函数：

在这里插入图片描述这两个指标经常被用于评估身份验证系统。

主干网络：

-VGG-M：该网络是在已有VGG-M CNN网络的基础上稍作修改后形成的。这里的VGG-M网络将原始的VGG-M网络的最后一层全连接层fc6替换成了两层——一层全连接层与一个平均池化层。这样修改后的网络与时间位置无关，与频率有关，这正是语音信号与图像的区别之处。

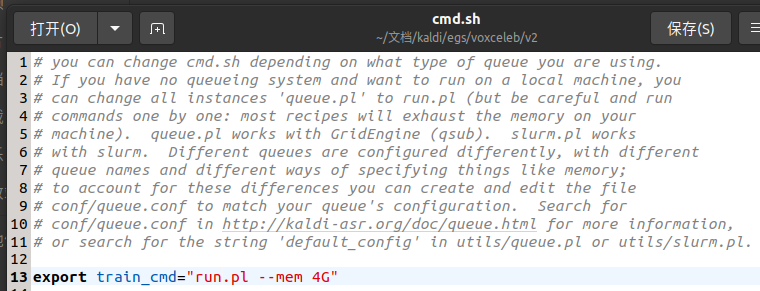
-ResNets：采用了ResNet-34和ResNet-50两种网络架构，为了适应输入层的频谱图对各层稍微做了修改。

该脚本文件用于批处理，描述了实验训练与测试的过程，该run.sh文件的具体流程内容如下所示：首先准备好数据集，然后对数据集提取MFCC特征；对训练数据进行增强，通过利用回响以及三种噪声增强了训练数据，并且测试了增强效果；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。

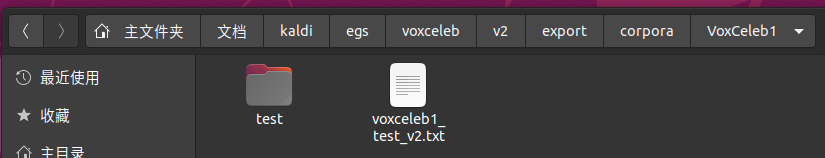
1. 由于实验指导书中已经全部给定了所需要的具体模型：X-vector pretrained model放在了0007\_voxceleb\_v2\_1a文件中以及测试数据集test放在了voxceleb1\_test\_v2文件中，即要求修改run.sh脚本文件使得能够在所给的测试数据集上进行算法复现测试得出实验结果，因此我对其进行了一些改动如下所示：保留下stage1对测试数据集的处理部分，保留下stage9中的“Extract x-vectors used in the evaluation”部分，保留下stage11和stage12部分用于进行输出评分测试。其他stage部分均可以删除，修改完脚本文件并且适当加上了中文标注后如下图所示：



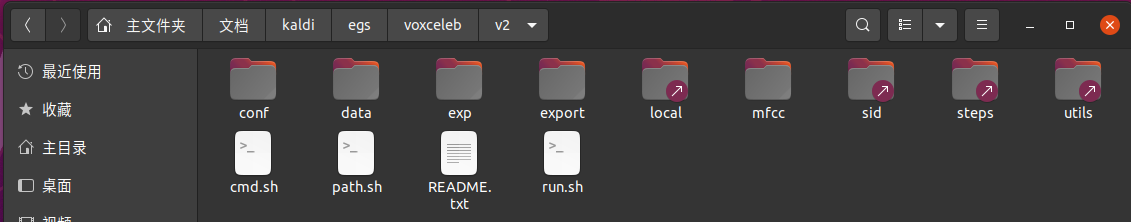
1. 最后是测试的过程，首先需要替换修改cmd.sh脚本文件如下图所示：



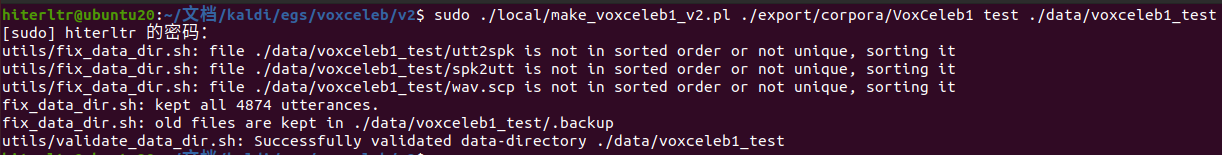
然后分别下载实验指导书链接里的现成模型与测试数据，其中测试数据集需要放置在如下图所示目录中：

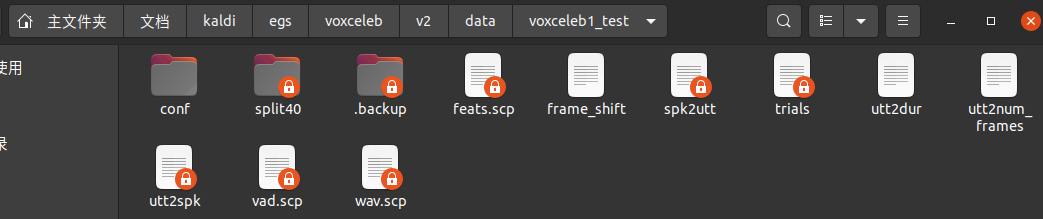


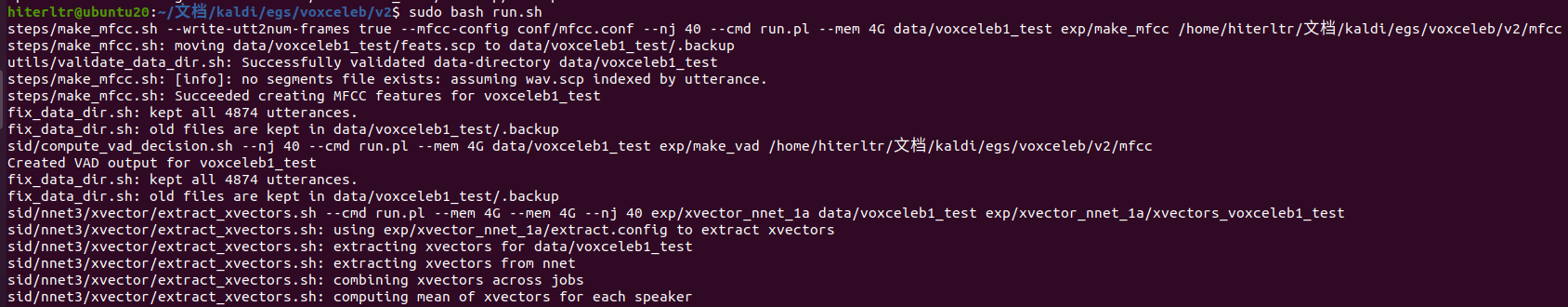
而将具体模型的文件夹0007\_voxceleb\_v2\_1a中的exp文件夹放入如下图所示目录中：



其次需要我们手动调用make\_voxceleb1\_v2.pl对测试集数据进行处理，同时生成后续测试所需的wav.scp文件。最后在v2目录下执行sudo bash run.sh命令，即可以成功运行：

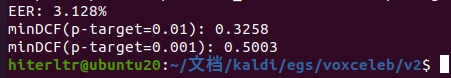






1. **实验结果**

最终得到的实验结果如下图所示：



1. **实验心得体会**

本次实验我觉得总体来说较为简单，我大多数时间都花费在编译配置安装Kaldi环境，因为经过很多次从头开始，编译过程比较漫长，要有耐心，在配置前置所需库时遇到阻遏，需要自己想方设法绕过去，同时避免依赖冲突。而如果绕不过去，就应该找到可替代的解决方案。

1. **参考资料**

<http://kaldi-asr.org/doc/index.html>

<https://github.com/kaldi-asr/kaldi>

[最大规模开源说话人识别语料集——VoxCeleb-CSDN博客](https://blog.csdn.net/weixin_44532659/article/details/104433770?ops_request_misc=%257B%2522request%255Fid%2522%253A%2522169693513716800226555993%2522%252C%2522scm%2522%253A%252220140713.130102334.pc%255Fall.%2522%257D&request_id=169693513716800226555993&biz_id=0&utm_medium=distribute.pc_search_result.none-task-blog-2~all~first_rank_ecpm_v1~rank_v31_ecpm-2-104433770-null-null.142%5ev95%5einsert_down1&utm_term=voxceleb%E8%AF%B4%E8%AF%9D%E4%BA%BA%E7%A1%AE%E8%AE%A4%E5%AE%9E%E9%AA%8C%20&spm=1018.2226.3001.4187)

[语音识别--kaldi环境搭建（基于Ubuntu系统）-CSDN博客](https://blog.csdn.net/weixin_38468077/article/details/103842314?ops_request_misc=%257B%2522request%255Fid%2522%253A%2522169698947716800184180793%2522%252C%2522scm%2522%253A%252220140713.130102334..%2522%257D&request_id=169698947716800184180793&biz_id=0&utm_medium=distribute.pc_search_result.none-task-blog-2~all~top_positive~default-1-103842314-null-null.142%5ev95%5einsert_down1&utm_term=kaldi&spm=1018.2226.3001.4187)

[Linux环境下kaldi的安装及测试\_could not find mkl-CSDN博客](https://blog.csdn.net/junbaba_/article/details/109235288?ops_request_misc=%257B%2522request%255Fid%2522%253A%2522169692396416800186585997%2522%252C%2522scm%2522%253A%252220140713.130102334.pc%255Fall.%2522%257D&request_id=169692396416800186585997&biz_id=0&utm_medium=distribute.pc_search_result.none-task-blog-2~all~first_rank_ecpm_v1~rank_v31_ecpm-4-109235288-null-null.142%5ev95%5einsert_down1&utm_term=Configuring%20MKL%20library%20directory%3A%20***configure%20failed%3A%20Could%20not%20find%20the%20MKL%20library%20directory.&spm=1018.2226.3001.4187)