task report

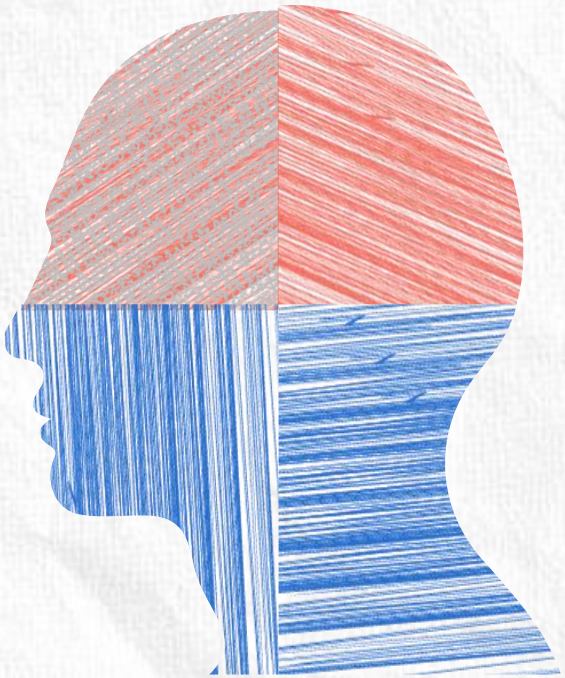
BUPT Zhang Yang

个人介绍

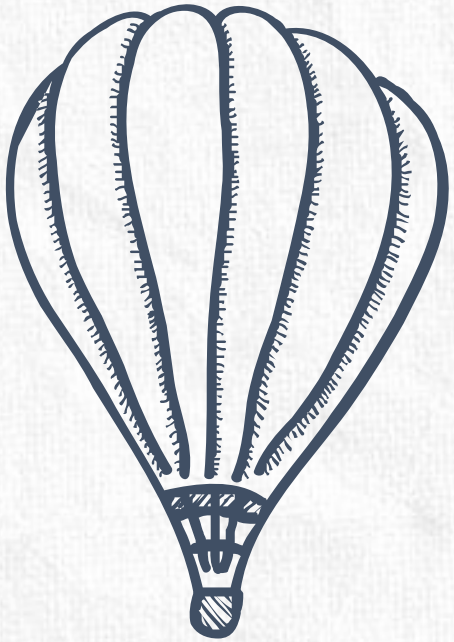
BYR

北邮信通院通信专业大三学生

张阳



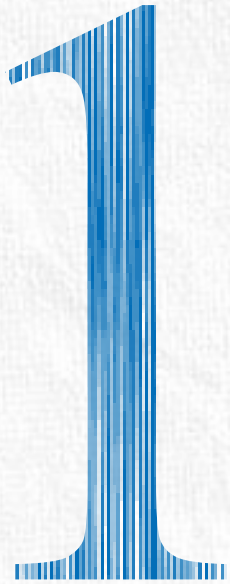
CONTENT



1 ML and DL

2 Speaker Recognition and kaldi

3 test



Machine Learning and Deep Learning

了解了机器学习和深度学习的基本概念，学习了深度学习的基本模型和训练方法

Relationship between ML and DL

深度学习与机器学习的关系

人工智能是一个大概念，机器学习则是人工智能领域的一个小分支，如果说AI是一个合集，那么ML就是AI的子集。

任何通过数据训练的学习算法的相关研究都属于机器学习，包括很多已经发展多年的技术，比如线性回归（Linear Regression）、K均值（K-means，基于原型的目标函数聚类方法）、决策树（Decision Trees，运用概率分析的一种图解法）、随机森林（Random Forest，运用概率分析的一种图解法）、PCA（Principal Component Analysis，主成分分析）、SVM（Support Vector Machine，支持向量机）以及ANN（Artificial Neural Networks，人工神经网络）。

而人工神经网络则是深度学习的起源。

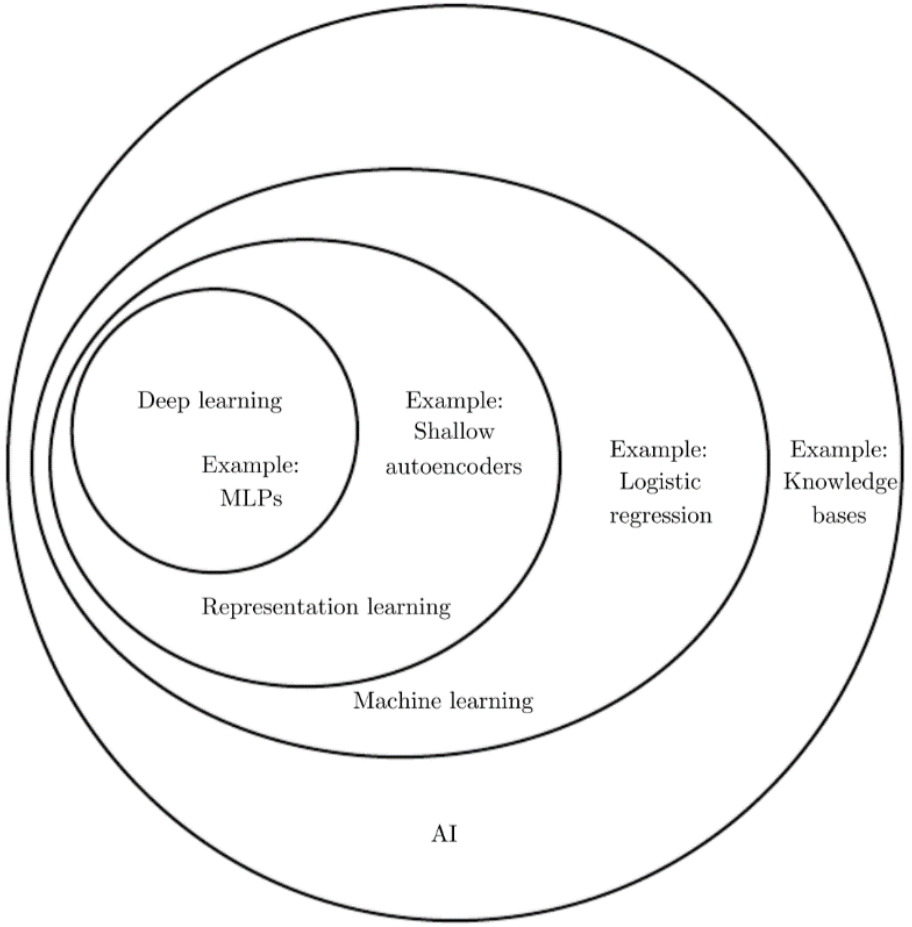
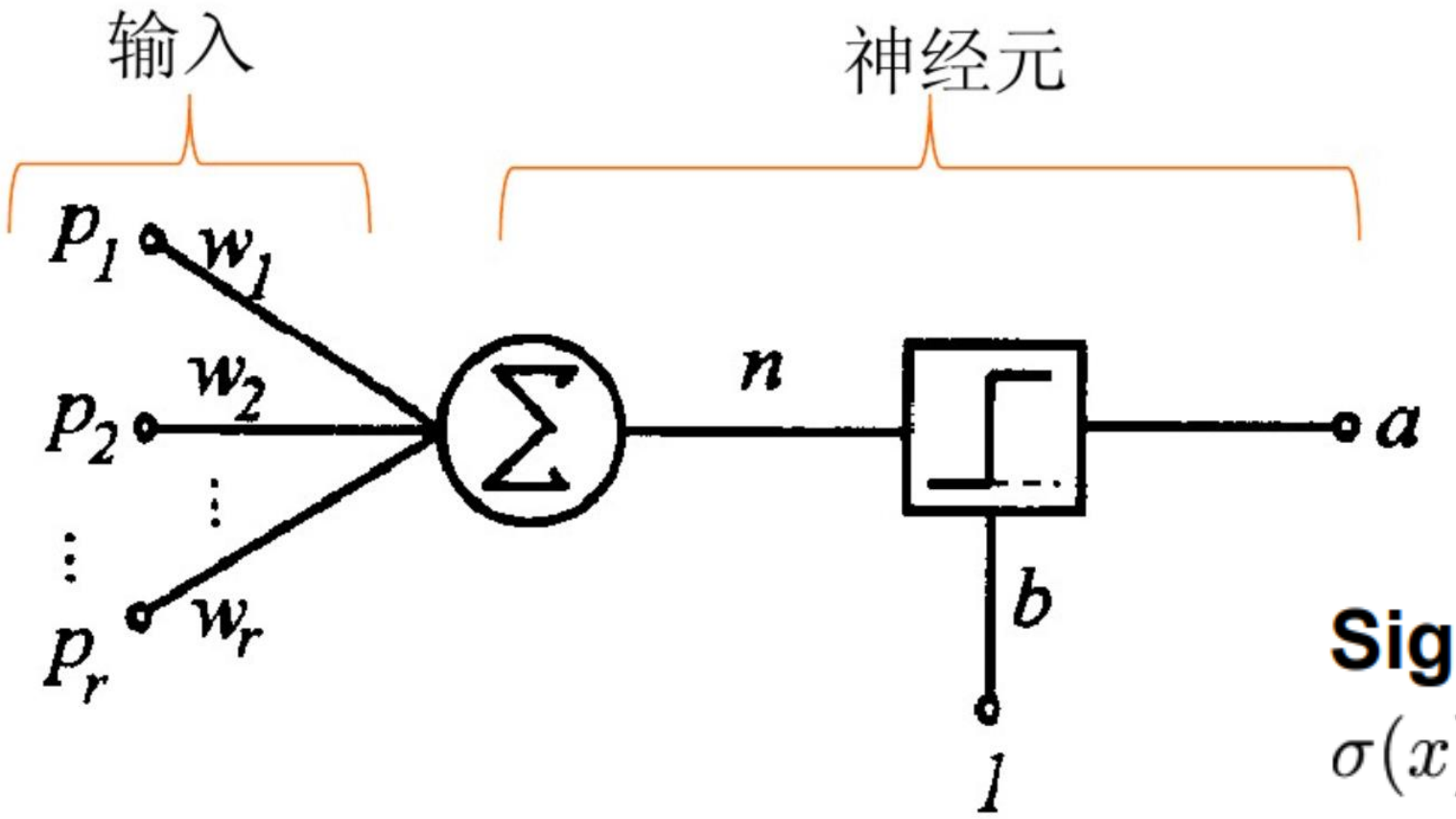


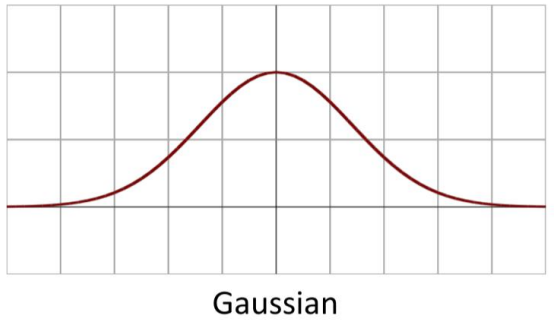
Figure 1.4: A Venn diagram showing how deep learning is a kind of representation learning, which is in turn a kind of machine learning, which is used for many but not all approaches to AI. Each section of the Venn diagram includes an example of an AI technology.

神经网络模型与激活函数

训练神经网络就是计算神经元的权值和阈值

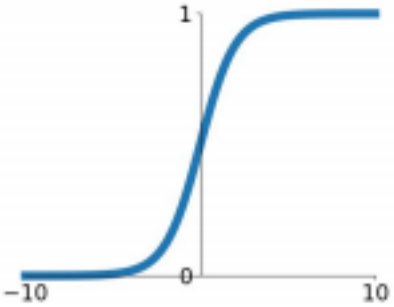


常用的激活函数



Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

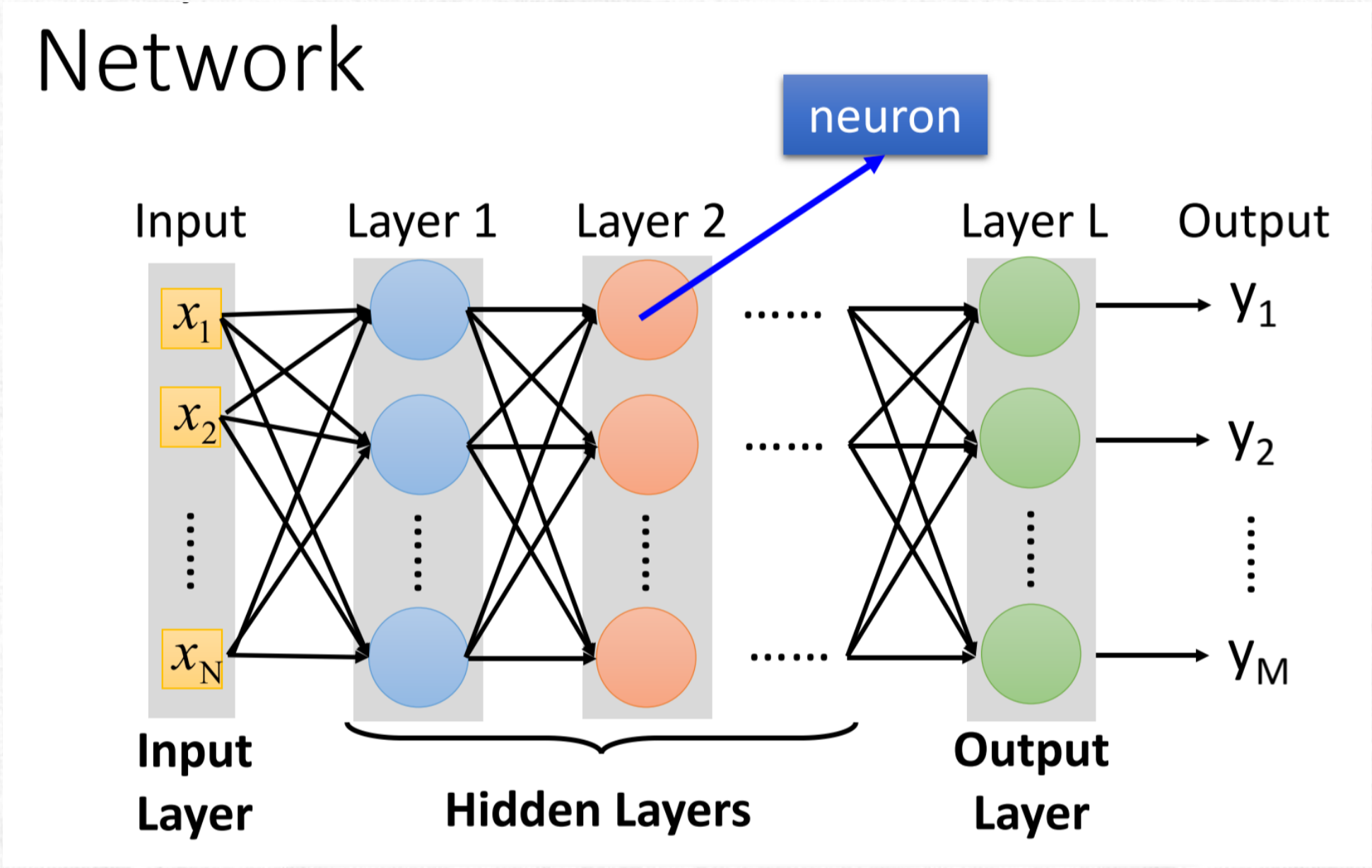


Multi-Layer Perceptron

多层神经网络

神经网络的学习过程，就是根据训练数据来调整神经元之间的连接权以及每个功能神经元的阈值

换而言之，神经网络学到的东西，蕴含在连接权和阈值之间



CNN 卷积神经网络

输入二维矩阵or图片（input)

卷积（Convolution)

使用滤波器（卷积核）

池化（pooling)

降低图像的大小

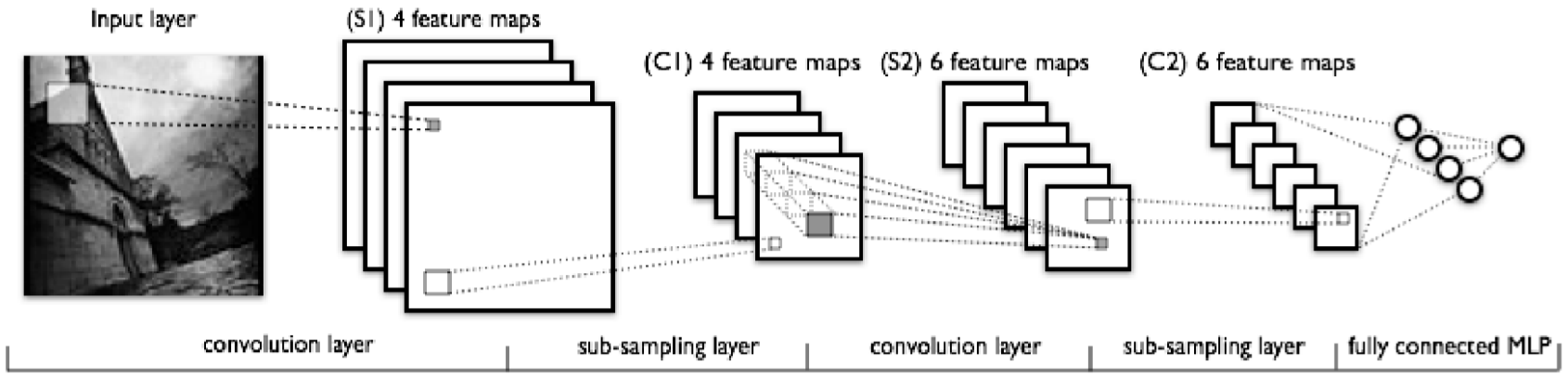
Flatten

把二维特征映射图连接成一个列向量，
作为训练分类器的输入特征

卷积神经网络
是一个适用于图像识别的神经网络。

它模仿人类识别图像的多层过程：瞳孔摄入像素；大脑皮层某些细胞初步处理，发现形状边缘、方向；抽象判定形状（如圆形、方形）；进一步抽象判定（如判断物体是气球）。

卷积和池化可以
重复多次



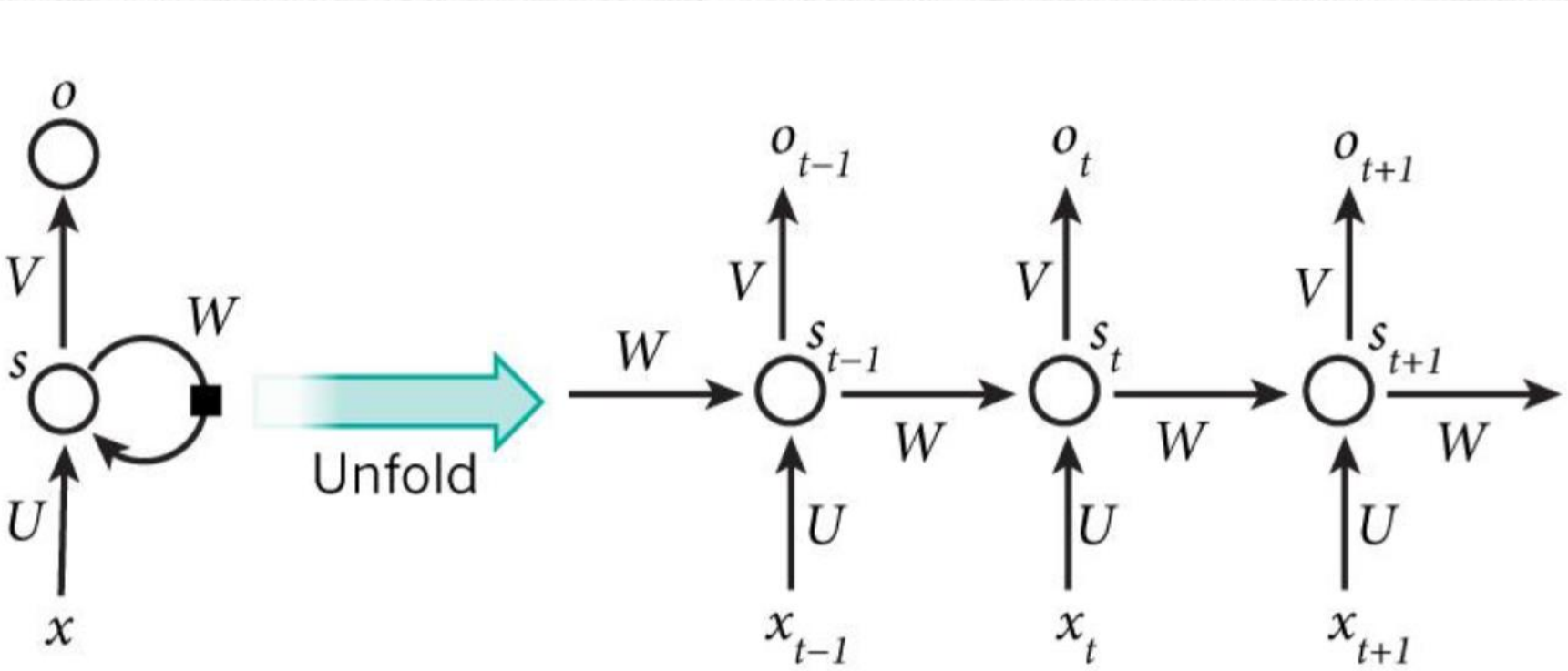
RNN 循环神经网络

序列的当前输出与之前的输出有关，

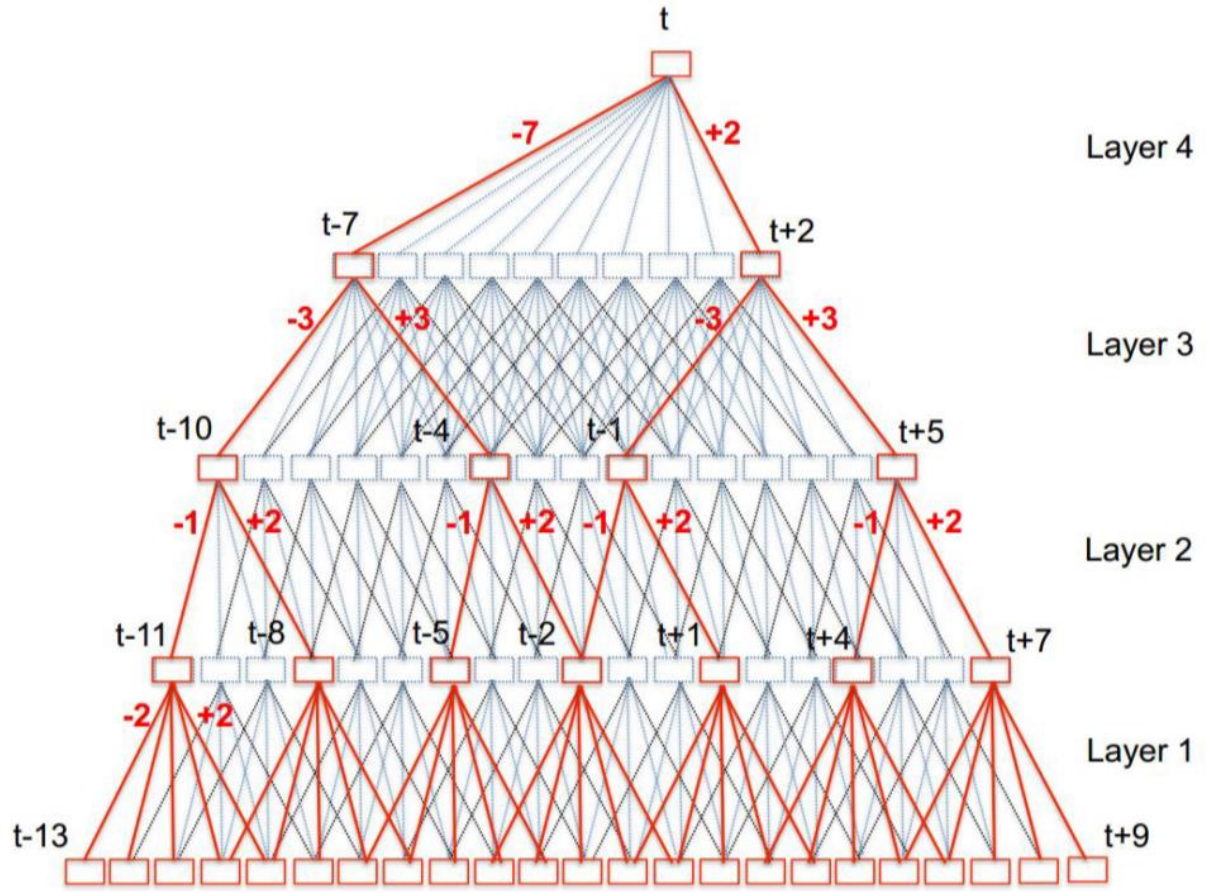
是一个有**反馈**的神经网络结构

适用于自然语言处理（NLP）

RNN在时域上展开



TDNN 延时神经网络



输入层输入的数据包含多帧

例如图中延时为4，同时输入5帧数据

该模型适用于语音识别

BP 算法（误差逆向传播）

BP算法是训练神经网络（计算神经网络中的权值和阈值）的一个关键算法，其基本思想分为两步：

1.正向传播时，输入样本从输入层进入网络，经隐层逐层传递至输出层，得到输出层。计算输出误差

2.反向传播时，将输出误差按原通路反传计算，通过隐层反向，直至输入层，在反传过程中将误差分摊给各层的各个单元，获得各层各单元的误差信号，并将其作为修正各单元权值的根据修正权值

Speaker Recognition and kaldi

了解与学习了语音识别的原理和训练模型的方法

说话人识别

分类

1. 有文本相关的(Text-Dependent)

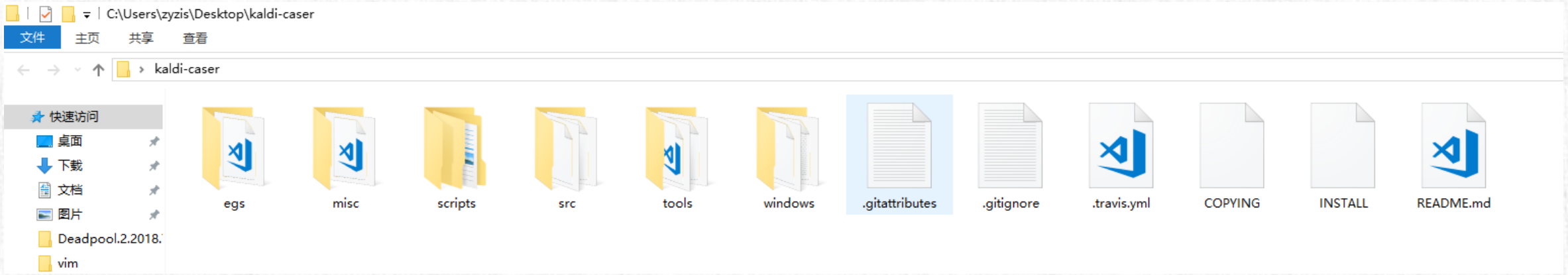
2.文本无关的(Text-Independent)

识别方法：

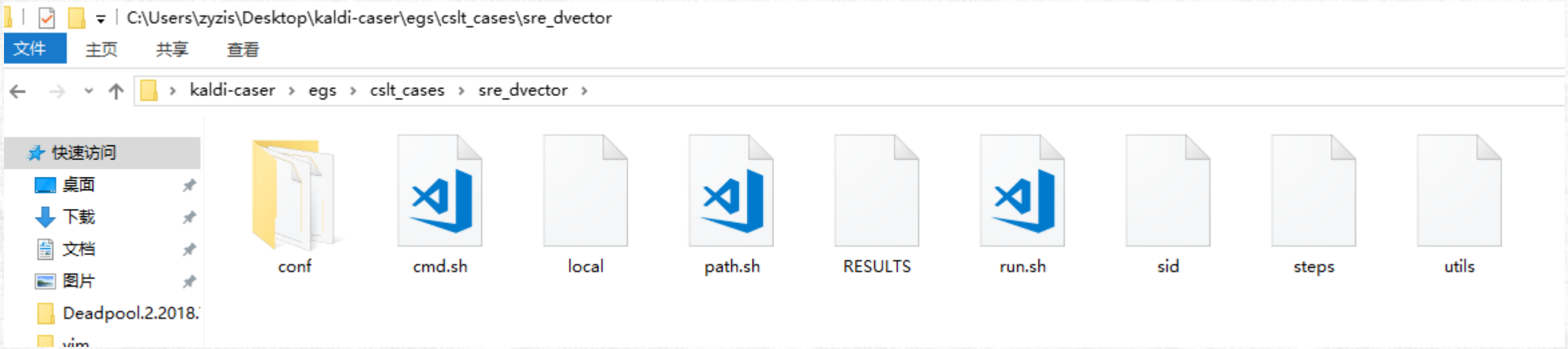
1. i-vector
2. D-vector
3. end-to-end

kaldi

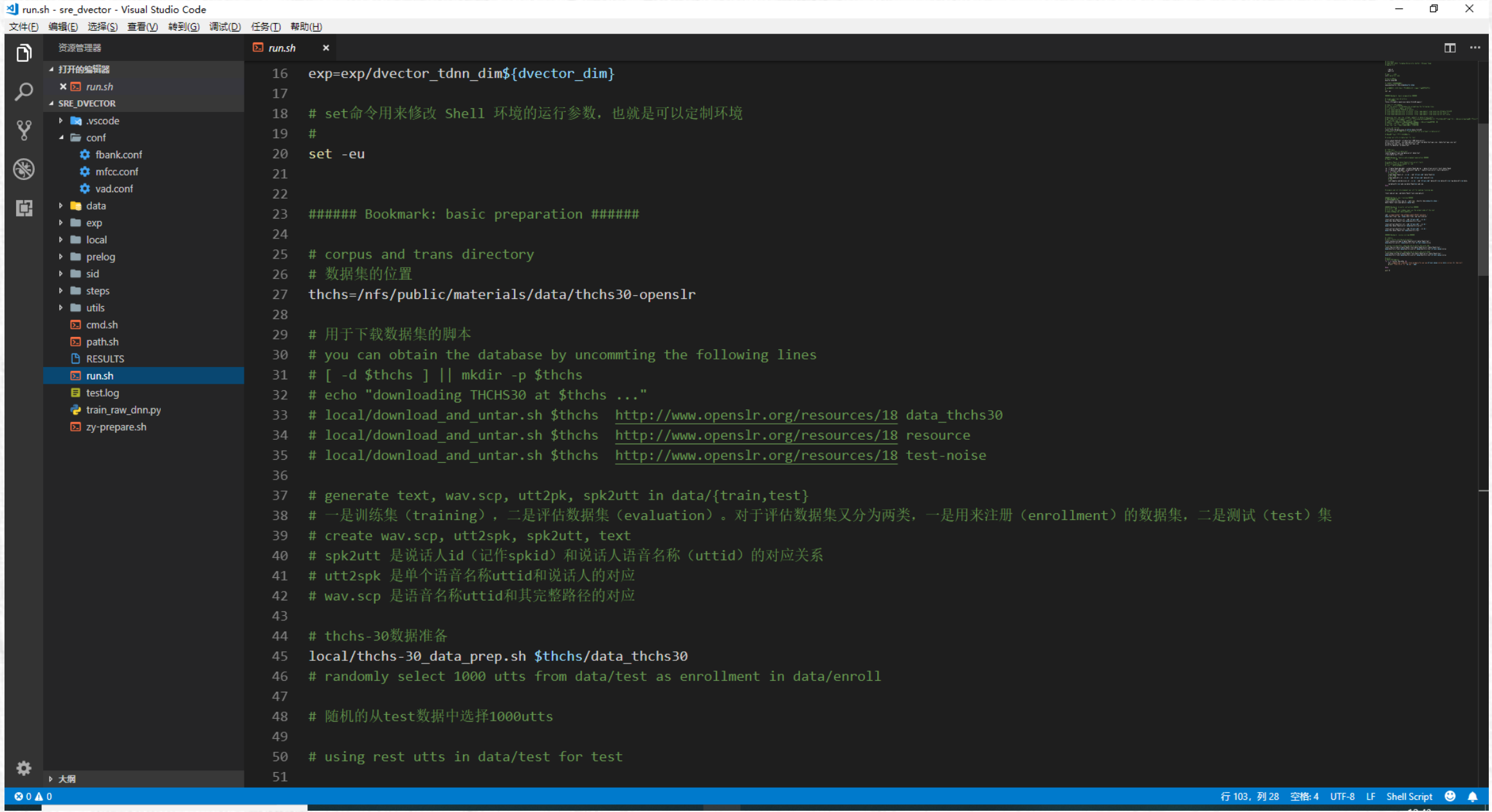
kaildi



ege文件夹下的代码



Data preparation数据准备

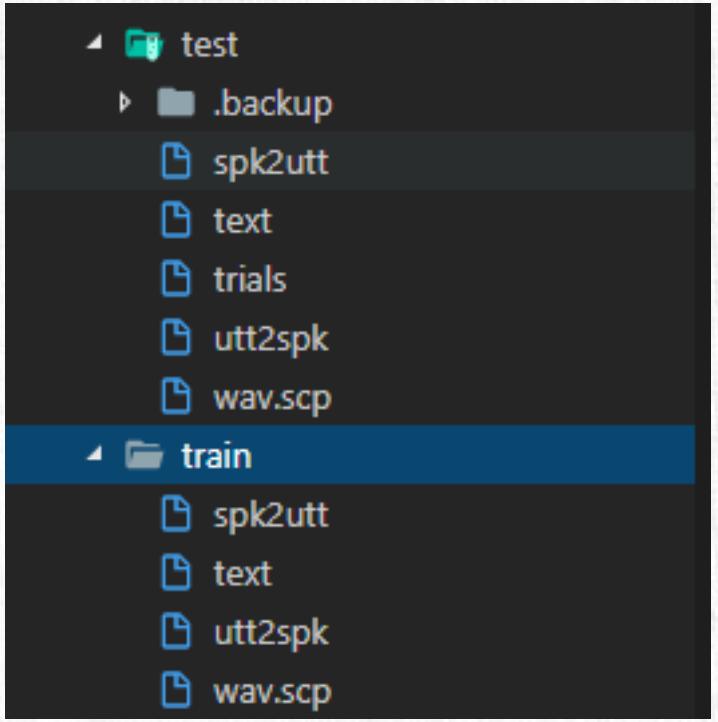


Data preparation数据准备

对于这些数据集中，一是训练集（training），二是评估数据集（evaluation）。对于评估数据集又分为两类，一是用来注册（enrollment）的数据集，二是测试（test）集

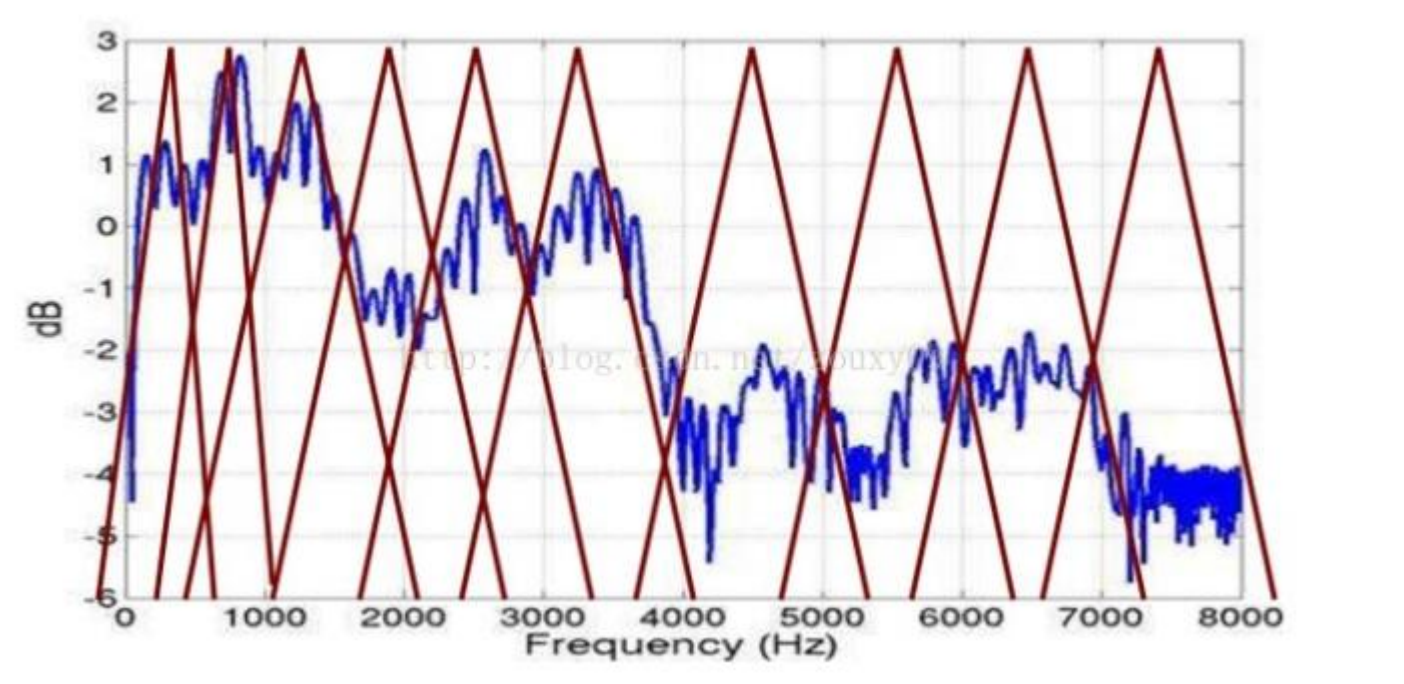
准备的过程中生成了以下类型的文件：
根据kaldi官方文档，我们可以得知，图中所示的四种文件格式和其作用。

1. Text 包含了每个语音数据的文字信息
2. spk2utt 是说话人id（记作spkid）和说话人语音名称（uttid）的对应关系
3. utt2spk 是单个语音名称uttid和说话人的对应
4. wav.scp 是语音名称uttid和其完整路径的对应



feature and alignment generation特征提取与对齐

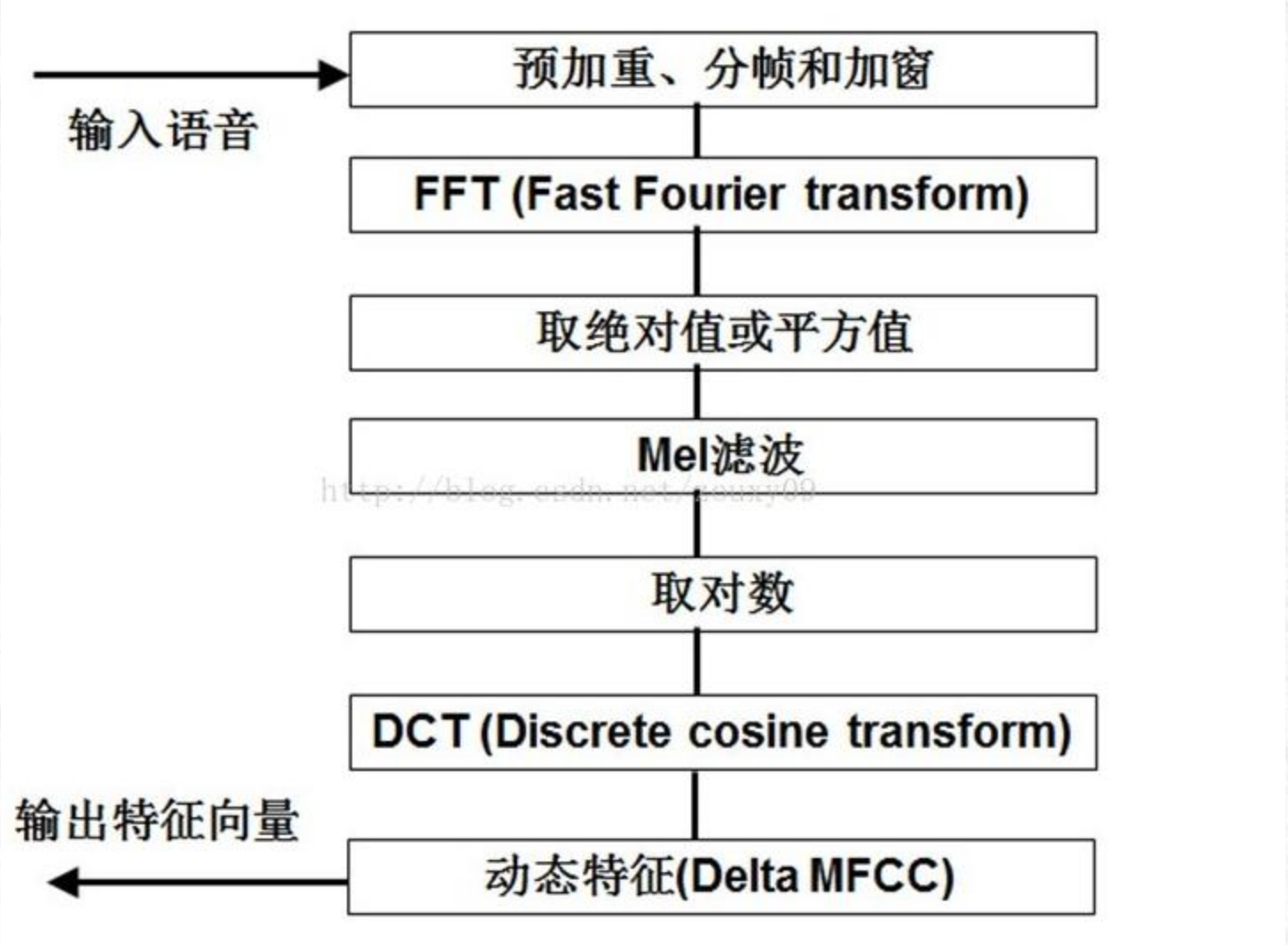
特征提取的好坏直接影响系统的效能。根据人耳的听觉特性提取具有辨识度的音频信息，使用shell脚本对已经准备好的train enroll test数据进行遍历，运行make_mfcc.sh脚本对语音进行mfcc特征提取，运行make_fbank.sh脚本对数据进行fblank滤波器带特征提取，运行compute_vad_decision.sh进行vad（语音活动检测）



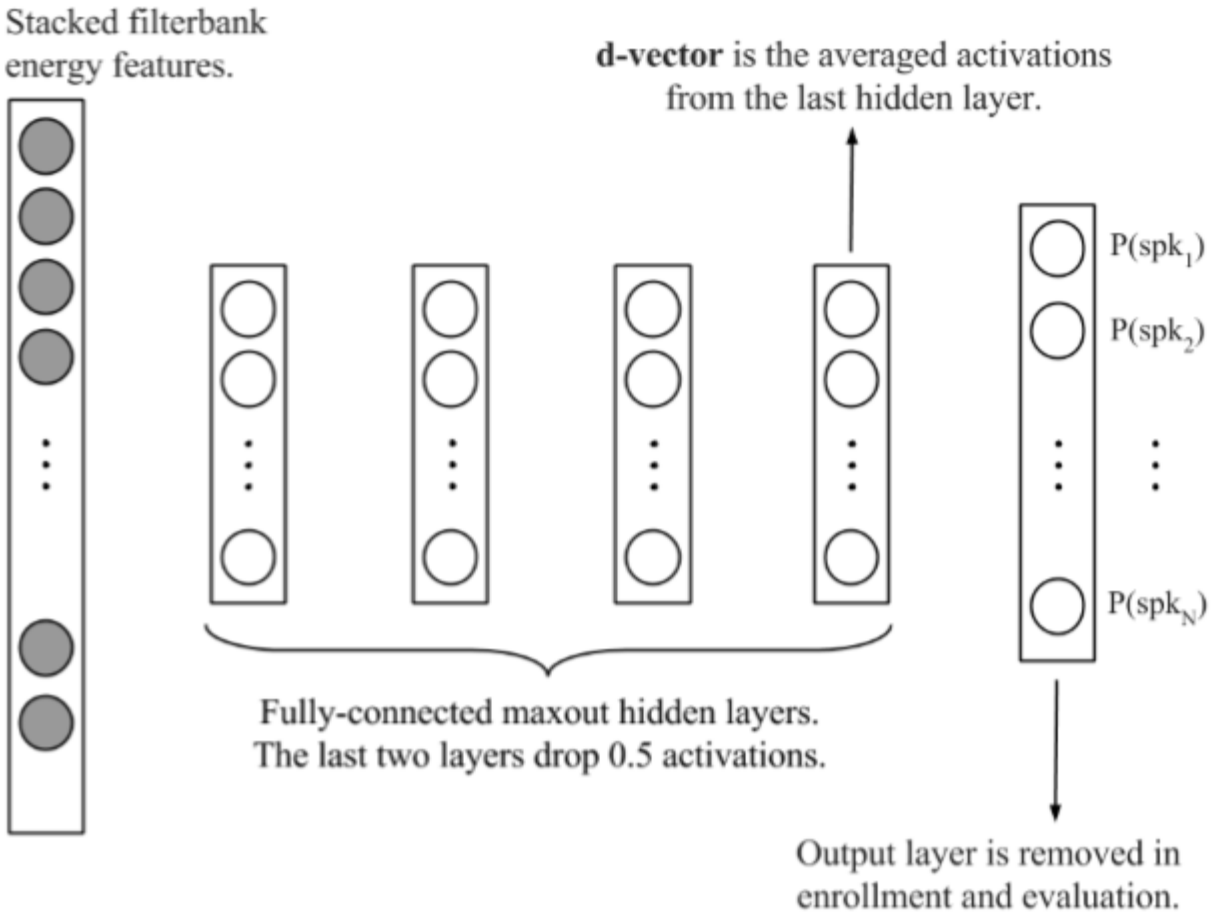
feature and alignment generation特征提取与对齐

对信号进行

1. 预加重——提高语音的高频部分
2. 分帧——由于傅里叶变换要求的是平稳的语音信号，因此要将将语音信号分成小段小段的
3. 加窗处理——加上一个窗函数，例如汉明窗，消除吉布斯效应



DNN-training训练深度神经网络



DNN-training训练深度神经网络

train_raw_dnn.py ×

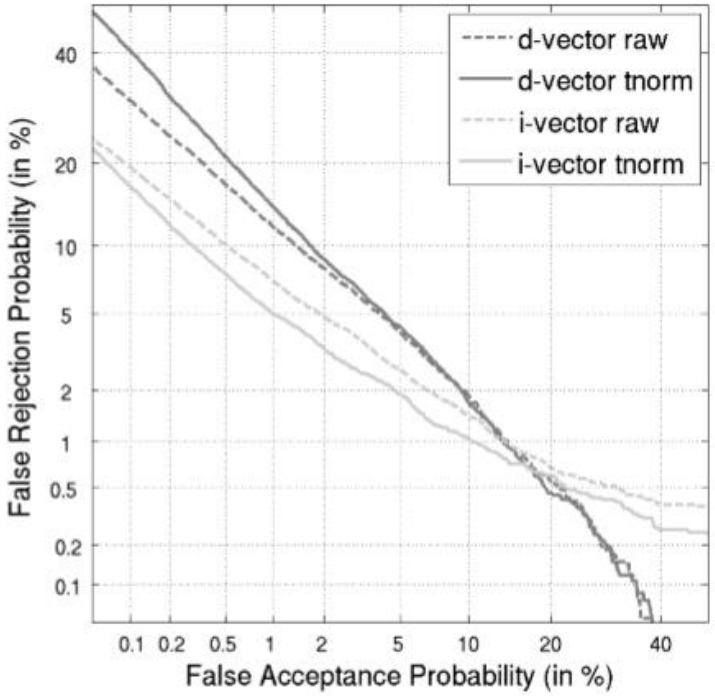
```
22  import libs.nnet3.train.frame_level_objf as train_lib
23  import libs.nnet3.report.log_parse as nnet3_log_parse
24
25
26  logger = logging.getLogger('libs')
27  logger.setLevel(logging.INFO)
28  handler = logging.StreamHandler()
29  handler.setLevel(logging.INFO)
30  formatter = logging.Formatter("%(asctime)s [%(pathname)s:%(lineno)s - "···
32  handler.setFormatter(formatter)
33  logger.addHandler(handler)
34  logger.info('Starting raw DNN trainer (train_raw_dnn.py)')
35
36
37  def get_args():···
121
122
123 def process_args(args):···
179
180
181 def train(args, run_opts):···
503
504
505 def main():···
520
521
522 if __name__ == "__main__":
523     main()
524
```

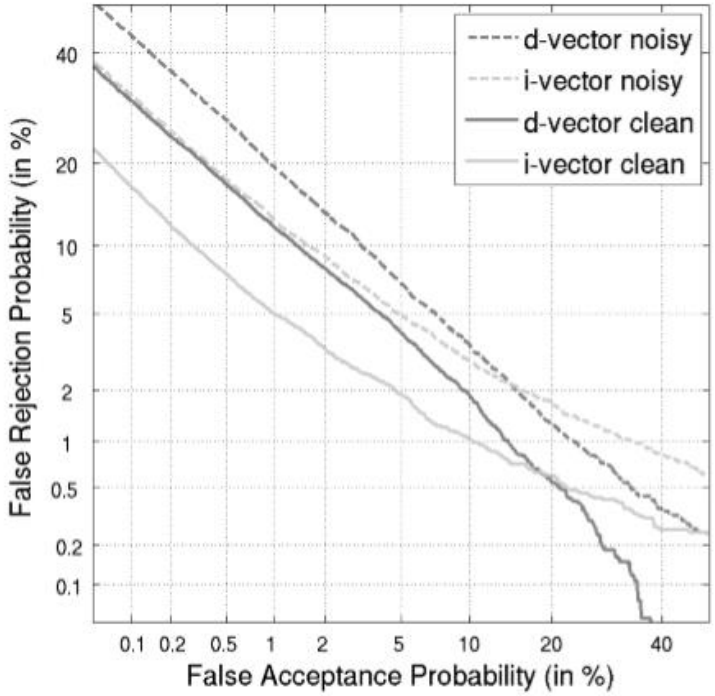
train_raw_dnn.py

1. 参数获取函数，获取参数，控制训练的运行方式、是否使用GPU、训练的次数等信息，
2. 训练函数，控制具体的训练流程和算法，
3. Main函数，使用try和expect语句进行整体函数的调用和流程的控制。

02

scoring 评分





02 scoring 评分

```
cosine eer:  2.74
   lda eer:  1.34
  plda eer:  2.34
```

End-to-End

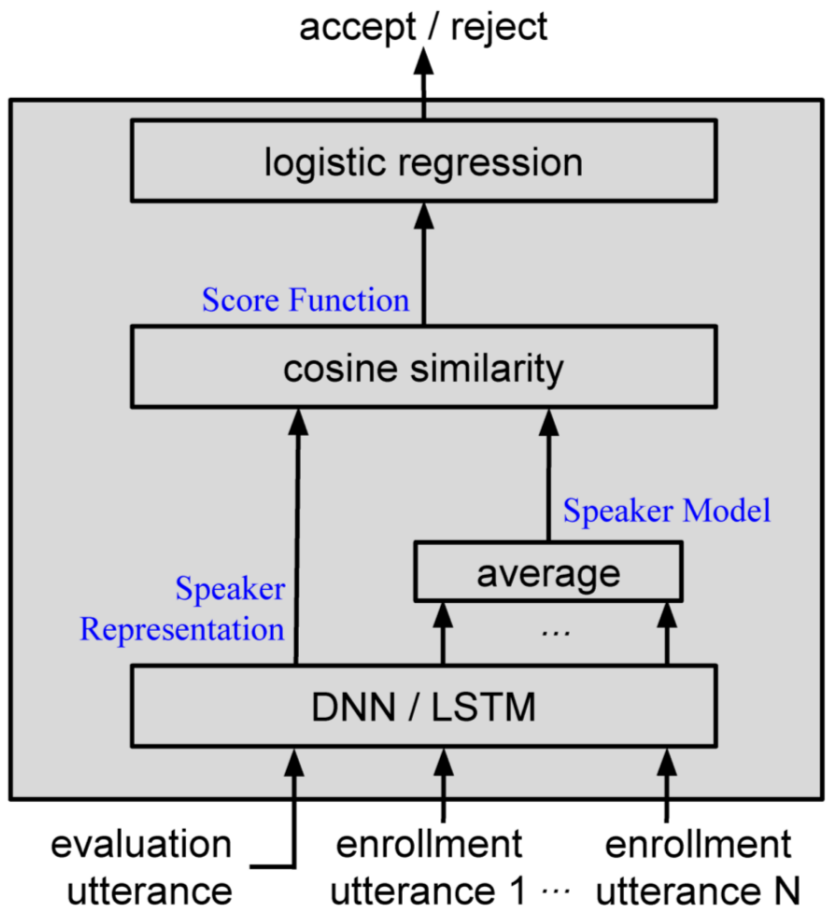


Figure 2: *End-to-end architecture, combining training of speaker representations (DNN/LSTM box), estimation of a speaker model based on up to N "enrollment" utterances (in blue), and verification (cosine similarity/logistic regression boxes).*

传统机器学习的流程往往由多个独立的模块组成

由于如今计算机硬件和深度神经网络的成熟，我们可以使用“端到端”只有一个模块的实现识别
训练时使用BP算法对模型进行修正

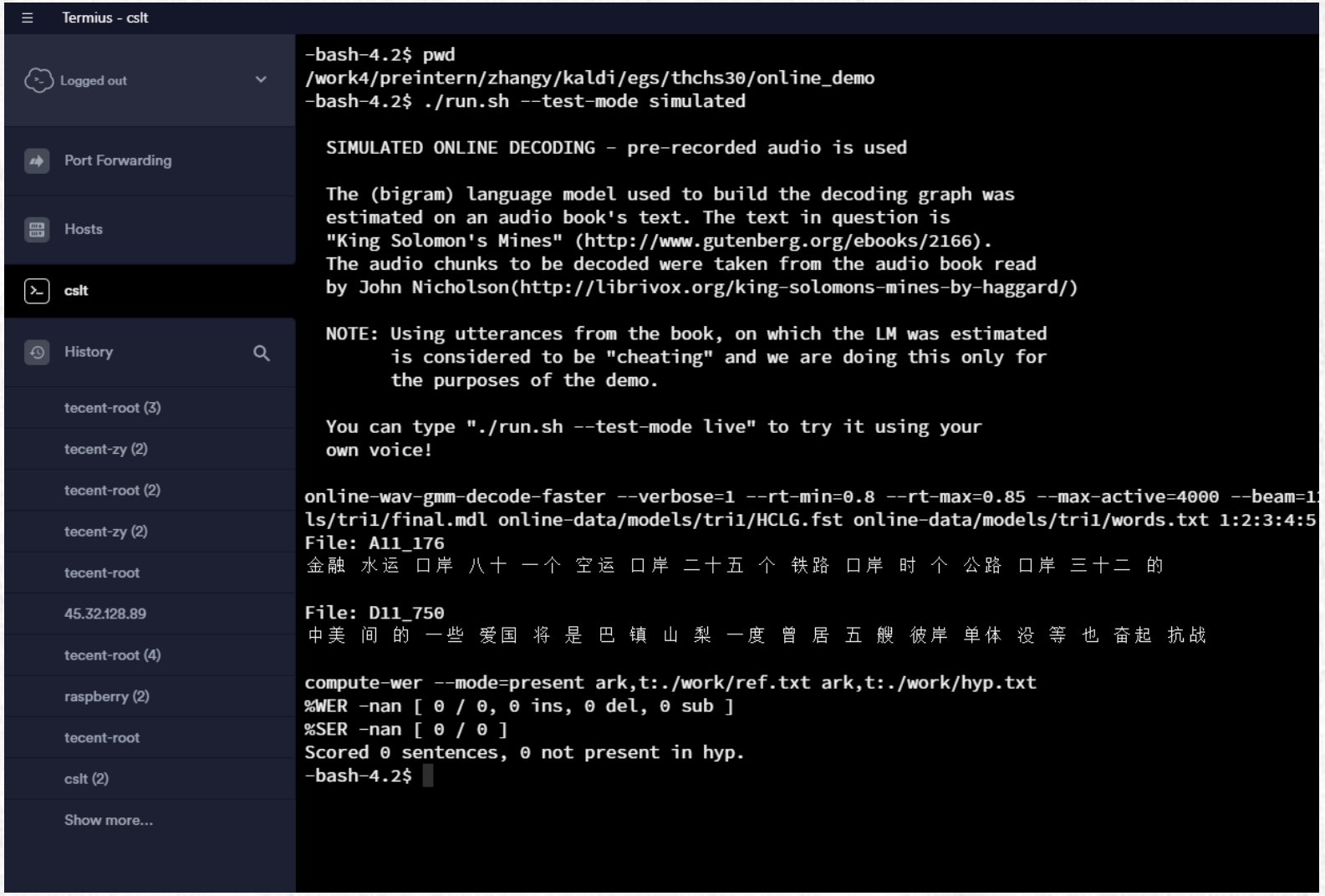
如图所示，神经网络采用深度神经网络和长短期记忆网络。
训练时通过一个评分函数，对于结果只有reject和accept

test

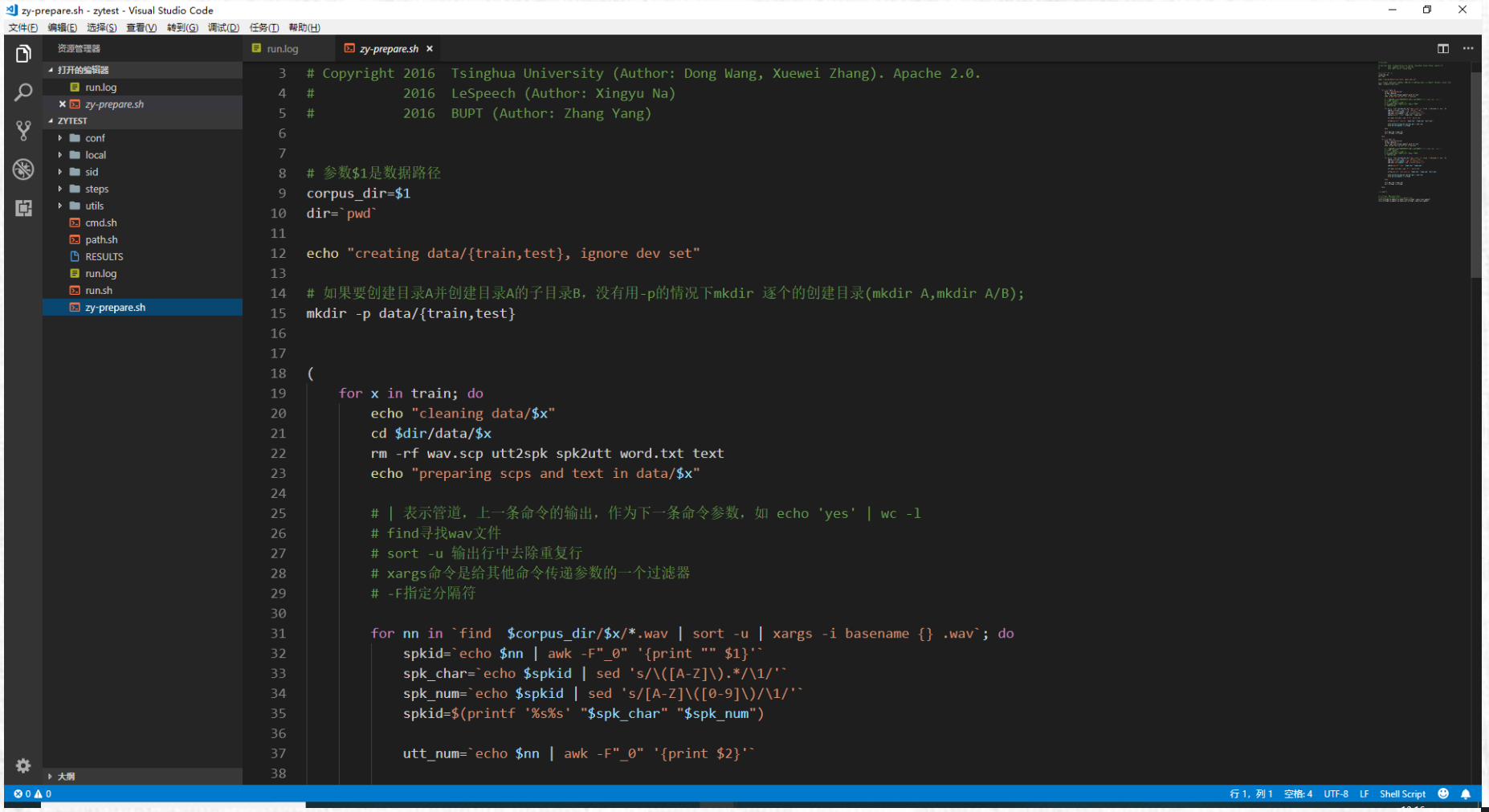
在服务器上运行和修改了部分代码，重现了老师的研究结果和尝试训练自己的模型

中文语音识别

输入 bash run.sh --test-mode simulated

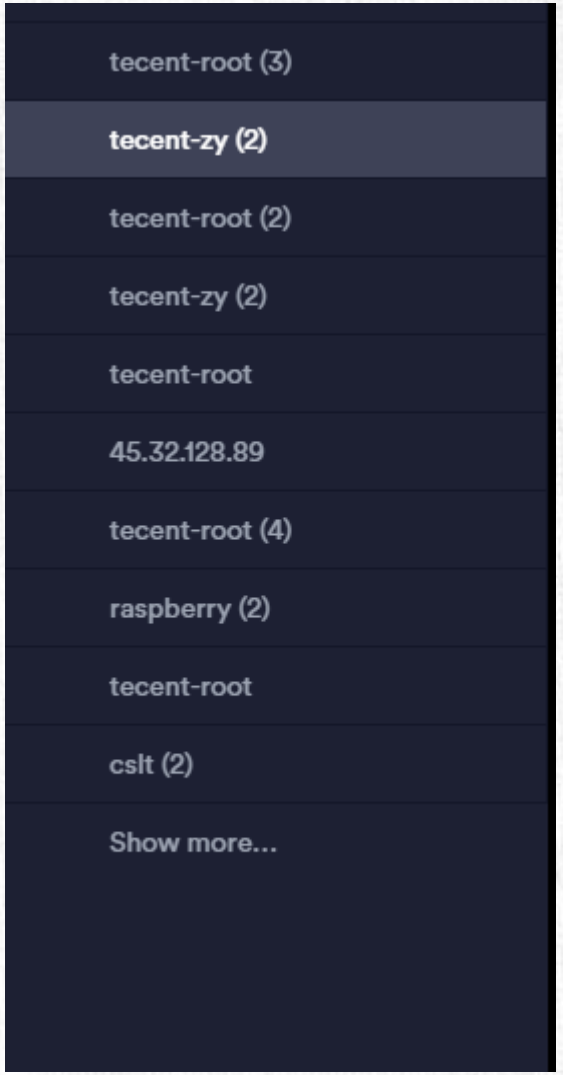


维吾尔族语言声纹识别




```
cosine eer:  3.61
   lda eer:  0.94
  plda eer:  2.21
```

我的收获



最后一个部分是我的收获和体会。

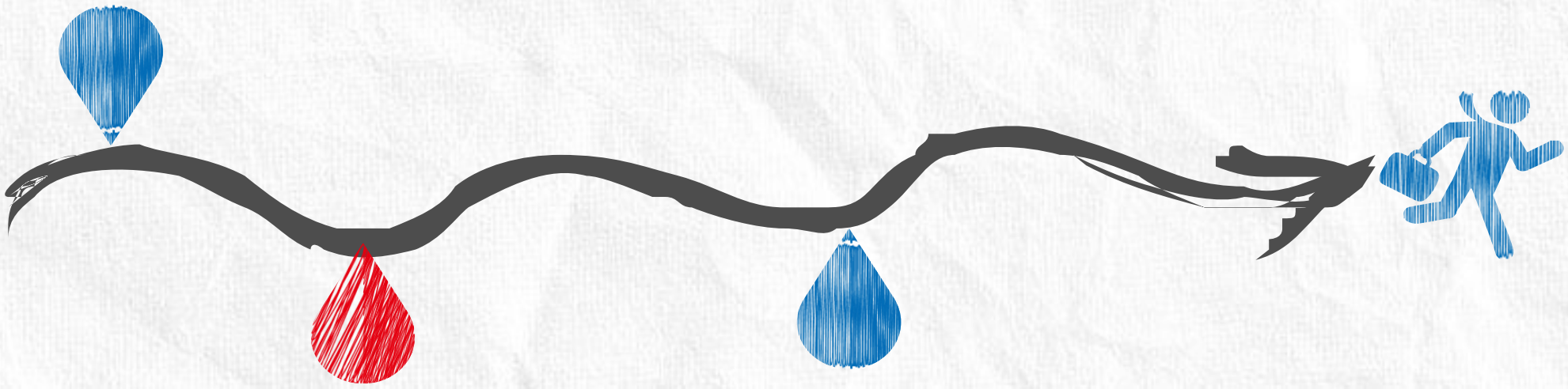
在这“不算很短的”预实习阶段中，这是我第一次接触到Linux远程服务器，并且还是多个高性能的服务器集群。在好奇心的的驱动下，我也在腾讯云上购买了属于自己的服务器和树莓派等尝试进行云计算。在暑假闲暇的时候尝试搭建了自己的个人博客和git服务器。尝试做在线语音识别，但由于自己的服务器计算性能不足，失败了。

同时，因为自己是通信专业的学生，很多知识比如声学模型和傅里叶变换、马尔科夫链等，在我先前的专业课程中有所了解和学习，在运行kaldi中进行说话人识别的过程中，我更加深入的了解了这些通信专业课所学知识在语音识别和说话人识别领域的应用。

最后

感谢老师给我不止有2周的实习机会和服务器的使用机会

我的能力与水平有限，还有很多知识不会，但未来我将会认真学习



THANKS

BUPT 张阳