**7英文数字语音识别**

1. 要求
2. 使用LibROSA包对语音信号进行特征提取。
3. 基于TensorFlow构建卷积神经网络模型，实现英文数字语音zero-nine的识别。
4. 数据集及训练结果下载地址https://pan.baidu.com/s/1A\_aBLuYlGn6fOnom-1XIww。
5. 步骤
6. **数据预处理**
7. Recording中的数据，0-9，每个语音390条数据；将每个类别的数据按照273：78：39（70%：20%：10%）随机划分，最后合为训练集、验证集、测试集
8. 对音频文件特征MFCC进行提取
9. 对特征向量进行归一化处理
10. **构建模型：ASRCNN**

输入是shape为[20,100]数据矩阵，经过四种size的卷积核：[2,100]、[3,100]、[4,100]、[5,100]。四种卷积核的个数都是64个，分别得到64个长度为18的向量、64个长度为19的向量、64个长度为17的向量、64个长度为16的向量。然后取每个向量的最大值。把这些最大特征拼凑在一块作为x经过卷积后的特征向量(size=256)，后接全连接层以及输出层。

1. **模型训练**

输出模型训练过程中训练集、验证集的损失和准确率值的变化情况，最后输出测试集的指标值，同时保留最终训练结果。

1. **模型使用**

加载模型，使用测试数据，输出音频识别结果。

1. 结果
2. **训练细节**

迭代次数，20.5K；499epoch；

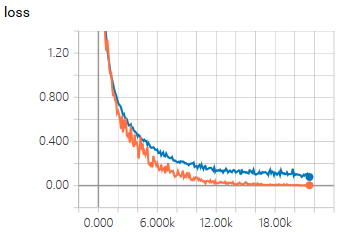
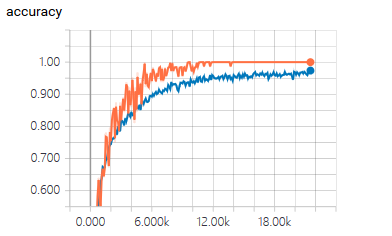
学习率，0.0001

Batch size，64

1. **训练及验证结果**

|  |  |  |
| --- | --- | --- |
|  | Accuracy | Loss |
| 训练 | 1.0 | 0.0033 |
| 验证 | 0.9679 | 0.1053 |

（蓝色为验证集）



1. **测试结果**

精度0.9657895，loss为0.1046659

