

**情感计算课程报告**

**报告题目 基于开源数据的语音情感识别**

**主讲教师 陶建华**

**学生姓名 陈航艇 学号 201718001027009**

**学位类别 直博**

**学科专业 信号与信息处理专业**

**研究所（院系） 中国科学院声学所（电子与通信工程学院）**

**报告日期 2018年5月26日**

**中国科学院大学人工智能技术学院制**

一、选题

基于开源数据的语音情感识别

二、问题分析

语音是人类最有效、最自然的交流沟通手段。除了包含语义信息外,还携带有丰富的情感信息。语音情感识别就是让计算机能够通过语音信号识别说话者的情感状态，是情感计算的重要组成部分 。

语音情感识别本质上属于模式识别。目前基于情感语音的模型方兴未艾。我们打算基于**IEMOCAP**英语情感语料库，提取韵律，音质，频谱等声学特征以及局部，全局等统计特征，进行多类特征组合，利用CNN+LSTM 的深度学习方法进行训练，目标是识别中立、高兴、生气、悲伤这四种基本情绪。

三、数据集描述

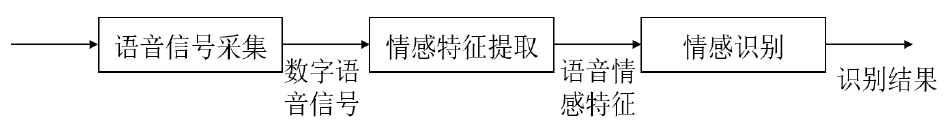
我们选定了**IEMOCAP**英语情感语料库进行实验。

该数据集分为5个部分，每个部分均有一男声和一女声的语料，每个语料都根据剧本或设定以表达出特定的情感。总共有10039个utterance，每个utterance的长度约4.5秒，然而为了便于与其他论文做比较，本次作业只挑选了包含中立、高兴、生气、悲伤四种情感的utterance，一共4490个。音频格：采样率：16kHZ，单声道，16比特采样率，WAV格式。

四、系统概要

1. 框架

本系统拟采用统计模式识别的特征提取-分类框架，包括语音信号采集、情感特征提取和情感识别三部分，框架示意图如下：



1. 设计

将语音作为输入的数字语音信号，按照25ms帧长、10ms帧移，进行预加重、分帧、加窗，按帧提取低级描述特征（LLD），基线系统拟采用MFCC(Mel-frequency cepstral coefficients )特征，特征值在每个utterance上进行z-normalization。

情感分类器部分采用循环神经网络（RNN），其中，使用长短期记忆单元（LSTM）的RNN对情感有更好的建模能力。本系统拟使用[1]中提出的一种新的LSTM结构（A-LSTM），这个结构是每隔几帧添加一个融合前面几帧(文中是3帧)LSTM cell状态信息的new cell，作为下一帧的前状态，这样就避免了LSTM中t时刻cell状态只依赖t-1时刻cell直接输出的假设，能够更好建模情感识别这种需要长时间信息的模型。

为了克服一个utterance只有一个情感标注，却对应很多帧的问题，之前的办法将每一帧都按照utterance的标注结果标注，[3]中提出的Attention Based Weighted Pooling是给每一帧RNN的输出赋予权重，而后融合作为预测结果，该权值和RNN的参数一起训练，这样就给静音、噪声、不包含情感的帧赋予了比较低的权值，给有情感的帧赋予了比较高的权值，该方法的实验结果明显好于不使用Weighted Pooling的方法，本系统也会对RNN的每帧结果使用Attention Based Weighted Pooling，而后送入softmax层进行分类。

五、预期目标

**基础目标：**

搭建分类框架，其中情感分类器使用长短期记忆单元（LSTM）的卷积神经网络（RNN），从而实现语音情感识别的功能。

**提高目标：**

改进分类器，以实现更好的情感识别效果，并于最初的方法进行比较。

评价标准：MAF（unweighted average F-score）, MAP（unweighted average precision）, accuracy.

对比对象：传统LSTM，平均LSTM

改进目标：MAF比另外两种方法提高至少5%，MAP比另外两种方法提高至少2%，准确率比另外两种方法提高4%。

六、论文查阅

1. Tao, F., & Liu, G. (2017). Advanced lstm: a study about better time dependency modeling in emotion recognition.
2. Tao, F., Liu, G., & Zhao, Q. (2018). An ensemble framework of voice-based emotion recognition system for films and tv programs.
3. Mirsamadi, S., Barsoum, E., & Zhang, C. (2017). Automatic speech emotion recognition using recurrent neural networks with local attention. IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE.
4. About LSTM: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

七、人员安排

|  |  |  |
| --- | --- | --- |
| **学号** | **姓名** | **专业** |
| 201718001027009 | 陈航艇 | 信号与信息处理 |
| 2017E8001061045 | 柳宗铭 | 电子与通信工程 |
| 201728013229040 | 姚永珍 | 计算机系统结构 |
| 201718013229001 | 陈羽丰 | 计算机系统结构 |

八、进度安排

|  |  |
| --- | --- |
| **时间** | **进度** |
| 05.05 | 搭建好语音情感识别分类框架，实现基础目标。 |
| 05.10 | 进一步阅读文献，确定优化方案 |
| 05.28 | 完成提高目标，撰写第二次书面汇报，以及进行汇报。 |