# 第三次实验报告

廖宇轩 无77 2017011223

分工：完成任务三

## 原理

首先要明确的是，训练任务是一个分类问题，其赋予整个视频关键帧和音频信息一个分类结果，由此看来，数据集实在不大，由此我们必须基于人的经验先指导对象进行初级的特征化，然后进一步压缩为更低维度的特征，最后进行最终的分类。我们借助于传统的数据处理手段和深度学习共同完成。

视频是一个时序数据，要紧密联系其特点选用合适的模型。序列结构可以抽象为特殊的树形图，为此我们可以应用图神经网络来生成对象的嵌入（Embedding），并同时进行分类（Classification），特别地，本次运用的是GGNN（Gated Graph Sequence Neural Networks）

注意到数据集是定长序列（30帧图像），为了使音频也具有同样性质，我们预处理音频，预先提取每一秒的音频信息，这里采用<https://github.com/pchao6/Speech_Feature_Extraction>的成果：

由于我们感兴趣的面部和姿态信息的预处理，已经由数据集给出，这里稍作处理就可以运用：

身体姿态信息中，我们比较关心的是头部和上肢相对于身体的动作，需要做一个差分，为了明确我们人为设置的判断标准（动作的幅度），序列间也做差分，而使模型不太去关注肢体的具体位置。

面部信息中，我们其一为了降低数据处理复杂程度，其二为了使模型更关注面部表情而非“面部”，也采用<https://github.com/CMU-Perceptual-Computing-Lab/openpose/blob/master/doc/output.md#reading-saved-results>的成果，给出面部信息。

将以上三者，合成为一个特征矢量，通过分配权重来消除其维度差异的影响，然后代入模型进行训练的到分类器。

实现方法

结果展示

结果分析

问题与

不足