**2017.9.7 融合音频情感和视频表情分析问题解答**

**1、具体的低阶音频格式**

答：具体的低阶音频格式并没有展示出，利用Opensmile提取出的特征已是高阶特征。Opensmile直接对原始的音频数据进行了统计操作，比如max、min等，从原始的低阶特征中得到了高阶特征。下方表1可以双击查看具体内容。

表1. Opensmile提取出的高阶特征示例

****

**2、提取动态回归系数，比如a1是什么，要得到一个什么样的结果。需要得到动态后上面的格式会变成什么样子，以什么样的数据格式表示。动态音频特征从一个形式变换到另外一个形式。（关于这部分疑问后期将更新）**

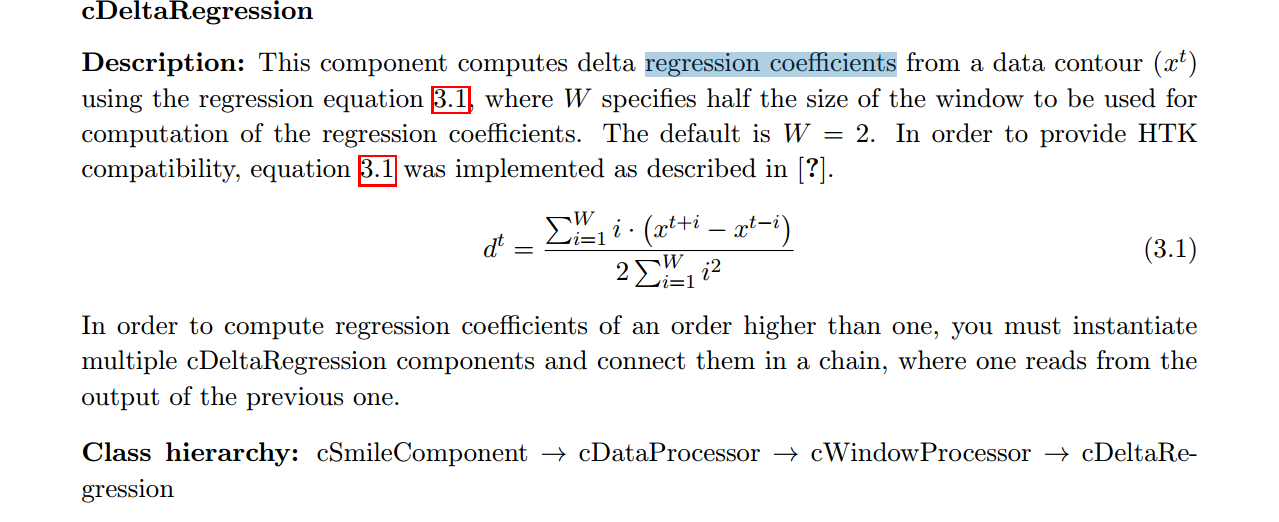
答：在语音模型中，在静态模型上加入时间参数会极大提升性能。

回归系数指的是所选择出来的变量的重要程度。比如此处代表某一个wav文件，就可以用提取出的特征，以多元线性回归的方式表示。

Yi=β0+β1X1+β2X2+…+βkXk

其中 k为解释变量的数目，βj（j=1,2,…,k)称为回归系数。回归系数的大小反应了该变量在总体中的重要程度。动态回归系数在原始的基础上加入了时间参数。

[**http://www1.icsi.berkeley.edu/Speech/docs/HTKBook3.2/node67\_mn.html**](http://www1.icsi.berkeley.edu/Speech/docs/HTKBook3.2/node67_mn.html)



**3、利用统计函数，有最大最小值统计低阶音频特征。针对上面的用不同统计函数还是统一的统计函数。最大函数是对mfcc提出来一个，还是以统计函数作用于每一个特征。高阶音频的特征是什么？**

答：此处的高阶特征计算为opensmile自动处理，将统计函数应用于每一个特征，高阶特征的数值见问题1的表1。在Opensmile里面处理时主要用到的统计函数如下：

max The maximum value of the contour  
min The minimum value of the contour  
range = max-min  
maxPos The absolute position of the maximum value (in frames)  
minPos The absolute position of the minimum value (in frames)  
amean The arithmetic mean of the contour  
linregc1 The slope (m) of a linear approximation of the contour  
linregc2 The offset (t) of a linear approximation of the contour

**4、最终加权融合后的结果是什么**

答：融合的结果是欺诈概率。视频欺诈模块与语音欺诈模块，分别通过SVM训练出两个模型，每个模型具有不同的准确率（此处的模型准确率为固定值，比如语音欺诈模块的准确率为60%，视频欺诈模块准确率80%）。加权融合模块的权重根据两个模块的准确率计算而得。

P(Video) = a（=0.8）

P(Audio) = b=（0.6）

W(Video) = =（0.8/1.4）

W(Audio) = =（0.6/1.4）

以上为训练出的模型准确率。当新的语音及视频进入后，首先提取出它们的特征，之后通过各自的SVM得出是否欺诈以及欺诈的概率。比如视频欺诈概率为0.8，语音欺诈概率为0.7。则根据W(Video)与W(Audio)得出的权重，计算出加权的权重。此处融合起来的欺诈概率为：

P=（0.8/1.4）\*0.8+（0.6/1.4）\*0.7

**5、实时欺诈识别，主要是情绪。两个模型识别出来的情绪，后面还有无相关处理。最后加权的结果也是情绪，是不是缺乏步骤。**

答：视频欺诈系统与音频欺诈系统，均以抽取的特征利用交叉验证方式训练svm分类器。此处SVM标注的的label是“欺诈”与“非欺诈”两个类别。因而输出的SVM直接输出是否欺诈，以及欺诈的概率。libsvm中svmpredict函数指定-b参数后，可输出概率值。标准的SVM的无阈值输出。Platt利用sigmoid-fitting方法，将标准SVM的输出结果进行后处理，转换成后验概率。SVM输出概率的原理说明参见：http://blog.csdn.net/funny75/article/details/50154391

之前我的理解有误，从此处输出的不是某种情绪结果，而是欺诈的概率。由于之前刚接手专利，许多背景不是特别清楚，给郭老师带来不便尽请谅解！非常抱歉！

**CQT 专利问题部分**

**1、如果d\_ri = d\_rc的情况该作何解释？**

答：如果两者距离相等则无法判断被试是否隐藏了真实情绪，但是该情况出现的可能性非常小。计算距离的时候，两个距离恰巧相等的情况几乎小概率，一般多为d\_rc更大或者d\_ri更大。若d\_rc更大则说真话，d\_ri更大说假话存在欺诈可能性。