PAJRD20170275---一种融合音频情感和视频表情分析的实时欺诈识别系统

**暂定名称**：一种融合音频分析和视频分析的欺诈识别方法及系统

**背景技术**：

在反欺诈识别中，对语音中隐含的情绪进行分析难度较大，目前还没有成熟稳定的技术应用于语音分析以识别欺诈,。

微表情识别比较完善，但是实际生产样本数据很难收集，提取欺诈识别相关的微表情特征比较困难，实际效果堪忧。

目前市场上还没有同时进行语音和表情分析的欺诈识别系统。

**技术方案**

一种融合音频分析和视频分析的欺诈识别方法，该方法包括：

接收当前用户的语音视频数据，从该语音视频数据中分离出音频及视频；

从音频中提取声音特征，从视频中提取表情特征；

将提取的声音特征及表情特征分别输入预先训练的语音识别模型和表情识别模型分别得到当前用户的欺诈概率；

根据每个模型的权重对两个模型输出的欺诈概率进行加权融合，得到当前用户的欺诈识别结果。

**从音频中提取声音特征，包括以下步骤**：

第一特征提取步骤：从音频中提取大量的低阶音频特征，包括MFCC（Mel倒谱系数）语音特征、Pitch（音高）、Zero Crossing Rate（过零率）；

第二特征提取步骤：从上述音频低阶特征中提取动态回归系数（例如分别提取MFCC、Pitch、过零率的动态回归系数），得到动态音频特征；

第三特征提取步骤：利用统计函数从所述低阶音频特征及动态音频特征中提取高阶音频特征；及

筛选步骤：利用特征筛选算法从所述高阶音频特征中筛选出最优特征子集。

**从视频中提取表情特征，包括以下步骤**：

基础特征提取步骤：从视频中提取面部表情特征，包括面部动作（AU）、眼球朝向、头部朝向；

高阶特征构造步骤：统计各面部表情特征在视频中出现的次数及持续的时长，根据统计结果构造表情高阶特征；

筛选步骤：利用特征筛选算法对所述情高阶特征进行筛选出最优特征子集。

**所述语音识别模型及表情识别模型的训练，包括以下步骤**：

**视频切割步骤**：以情绪为单位对样本视频数据进行切割，得到一系列包含音频及视频的视频片段样本，给每个视频片段样本分配一个欺诈标签（“有欺诈” OR “无欺诈”）；

**音频视频分离步骤**：从所述视频片段样本中分离出音频样本及视频样本；

**语音识别模型训练步骤**，包括：

第一特征提取步骤：从每个音频样本中提取大量的低阶音频特征，包括MFCC（Mel倒谱系数）语音特征、Pitch（音高）、Zero Crossing Rate（过零率）；

第二特征提取步骤：从上述音频低阶特征中提取动态回归系数；

第三特征提取步骤：利用统计函数（如max-最大，min-最小, skewness-峰度， kurtosis-偏度）对所述低阶音频特征及动态回归系数进行处理，从而构建高阶音频特征；

筛选步骤：利用特征筛选算法从所述高阶音频特征中筛选出最优特征子集；

训练步骤：利用带情绪标注的各音频样本的最优特征子集对SVM进行训练得到语音识别模型，交叉验证确定语音识别模型的最优模型参数。

**表情识别模型训练步骤**，包括：

基础特征提取步骤：从每个视频片段中提取面部表情特征，包括面部动作（AU）、眼球朝向、头部朝向；

高阶特征构造步骤：统计各面部表情特征在每个视频片段中出现的次数及持续的时长，根据统计结果构造表情高阶特征；

筛选步骤：利用特征筛选算法对所述情高阶特征进行筛选出最优特征子集；

训练步骤：利用每个视频样本的最优特征子集对SVM进行训练得到表情识别模型，交叉验证确定表情识别模型的最优模型参数。

第一特征提取步骤中，从音频中提取低阶音频特征可以使用opensmile软件。

第二特征提取步骤中，动态回归系数表示所选择出来的变量（例如上述低阶音频特征）的重要程度。比如，某个低阶音频特征以一个波形（wav）文件表示，则该波形文件就可以以多元线性回归的方式表示：

Yi=β0+β1X1+β2X2+…+βkXk

其中 k为变量（X1~Xk）的数目，βj（j=1,2,…,k)称为动态回归系数。动态回归系数的大小反应了该变量在所有变量中的重要程度。

第三特征提取步骤是构建高阶音频特征。如通过将max、min、skewness、kurtosis函数应用于每一个低阶音频特征及动态回归系数，得到每个低阶音频特征及动态回归系数的最大值、最小值、峰度和偏度，再对提取得到的这些值进行组合、变换，得到高阶音频特征集合。

偏度又称偏斜系数、偏态系数，是用来帮助判断数据序列的分布规律性的指标。峰度（系数）是一个用于衡量离群数据离群度的指标。峰度（系数）越大，说明该数据系列中的极端值越多。

当所述高阶音频特征集合包含的特征数非常大时，可能仅仅有少部分特征会对结果产生影响，因此可以使用特征选择算法降低特征的数量。假设有n个高阶音频特征，那么，其有2n-1种可能的特征子集，如果特征选择需要去穷举所有2n种可能的特征子集，对于n比较大的情况，计算的代价太大，无法真正实现。因此可以通过一些算法实现特征的选择。上述特征筛选步骤中的特征筛选算法可以为mlxtend库中的SequentialFeatureSelector(SFS)贪婪搜索算法，正向搜索/反向搜索（forward/backwardsearch）、过滤特征选择（filter feature selection），或是其他可用的特征筛选算法。

参上述说明，语音识别模型与表情识别模型，是分别通过SVM训练得到的，训练得到的每个模型有一个预测准确率。例如，假设语音识别模型的准确率为60%，表情识别模型的准确率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）（表情识别模型的权重）

W(Audio) = =（0.6/1.4）（语音识别模型的权重）

假设一段某用户的实时语音及视频，经过语音识别模型、表情识别模型预测得到的欺诈概率分别为0.8、0.7。则根据W(Video)与W(Audio)进行加权融合计算，最终得到该用户的欺诈概率为：

P=（0.8/1.4）\*0.8+（0.6/1.4）\*0.7

本发明的系统包括能够实现上述方法、步骤的应用软件，以及安装该应用软件的存储介质或能够运行该应用软件的电子装置。