depression speech abstract&feature

A Novel Decision Tree for Depression Recognition in Speech

Depression is a common mental disorder worldwide which causes a range of serious outcomes. The diagnosis of depression relies on patient-reported scales and psychiatrist’s interview which may lead to subjective bias. In recent years, more and more researchers are devoted to depression recognition in speech , which may be an effective and objective indicator. This study proposes a new speech segment fusion method based on decision tree to improve the depression recognition accuracy and conducts a validation on a sample of 52 subjects (23 depressed patients and 29 healthy controls). The recognition accuracy are 75.8% and 68.5% for male and female respectively on gender-dependent models. It can be concluded from the data that the proposed decision tree model can improve the depression classification performance.

feature： Through Open SMILE extracte 1582 dimensional features and three energy related features : energy, low frequency ratio and temporal time. The features that are extracted from the silent segment are also called pause features, or each pause reflects the brain response of the speaker at that time.Pause features can distinguish normal and depression people。

Automated speech-based screening of depression using deep convolutional neural networks

Early detection and treatment of depression is essential in promoting remission, preventing relapse, and reducing the emotional burden of the disease. Current diagnoses are primarily subjective, inconsistent across professionals, and expensive for individuals who may be in urgent need of help. This paper proposes a novel approach to automated depression detection in speech using convolutional neural network (CNN) and multipart interactive training. The model was tested using 2568 voice samples obtained from 77 non-depressed and 30 depressed individuals. In experiment conducted, data were applied to residual CNNs in the form of spectrograms—images auto-generated from audio samples. The experimental results obtained using different ResNet architectures gave a promising baseline accuracy reaching 77%.

特征：频谱图

Automatic Assessment of Depression from Speech via a Hierarchical Attention Transfer Network and Attention Autoencoders

Early interventions in mental health conditions such as Major Depressive Disorder (MDD) are critical to improved health outcomes, as they can help reduce the burden of the disease. As the efficient diagnosis of depression severity is therefore highly desirable, the use of behavioural cues such as speech characteristics in diagnosis is attracting increasing interest in the field of quantitative mental health research. However, despite the widespread use of machine learning methods in the depression analysis community, the lack of adequate labelled data has become a bottleneck preventing the broader application of techniques such as deep learning. Accordingly, we herein describe a deep learning approach that combines unsupervised learning, knowledge transfer and hierarchical attention for the task of speech-based depression severity measurement. Our novel approach, a Hierarchical Attention Transfer Network (HATN), uses hierarchical attention autoencoders to learn attention from a source task, followed by speech recognition, and then transfers this knowledge into a depression analysis system. Experiments based on the depression sub-challenge dataset of the Audio/Visual Emotion Challenge (AVEC) 2017 demonstrate the effectiveness of our proposed model. On the test set, our technique outperformed other speech-based systems presented in the literature, achieving a Root Mean Square Error (RMSE) of 5.51 and a Mean Absolute Error (MAE) of 4.20 on a Patient Health Questionnaire (PHQ)-8 scale [0, 24]. To the best of our knowledge, these scores represent the best-known speech results on the AVEC 2017 depression corpus to date.

特征：语谱图

Automatic Detection of Depression in Speech Using Ensemble Convolutional Neural Networks

This paper proposes a speech-based method for automatic depression classification. The system is based on ensemble learning for Convolutional Neural Networks (CNNs) and is evaluated using the data and the experimental protocol provided in the Depression Classification Sub-Challenge (DCC) at the 2016 Audio–Visual Emotion Challenge (AVEC-2016). In the pre-processing phase, speech files are represented as a sequence of log-spectrograms and randomly sampled to balance positive and negative samples. For the classification task itself, first, a more suitable architecture for this task, based on One-Dimensional Convolutional Neural Networks, is built. Secondly , several of these CNN-based models are trained with different initializations and then the corresponding individual predictions are fused by using an Ensemble Averaging algorithm and combined per speaker to get an appropriate final decision. The proposed ensemble system achieves satisfactory results on the DCC at the A VEC-2016 in comparison with a reference system based on Support V ector Machines and hand-crafted features, with a

CNN+LSTM-based system called DepAudionet, and with the case of a single CNN-based classifier.

feature : log-spectrogram

基于语音的抑郁症识别

抑郁症是世界范围内常见的精神疾病之一,抑郁症患者往往长期伴随情绪低落,如悲伤内疚、低自尊、兴趣丧失、功能减退等,对个人、家庭及社会造成了巨大损失.抑郁症的发病原因复杂,临床诊断存在一定的困难,有必要寻找一种更加便捷、客观、高效的方式来辅助抑郁症的快速识别.语音作为一个相对客观且容易获得的变量,具有其潜在的价值.本研究旨在构建基于语音的抑郁症识别模型,探究语音与抑郁症之间的关系.收集了103名被试(45名抑郁症患者,58名健康人)的语音数据,实验组为临床确诊的抑郁症患者,年龄在23.844.6岁之间,控制组为健康人,年龄为20.1~41.7岁.我们采用了3(情绪状态:正性、中性、负性)×3(任务类型:语言问答、文本朗读、图片描述)的实验设计,运用机器学习的分类算法——逻辑回归(LR)来构建抑郁识别模型.实验结果表明,语音的抑郁识别精度可以达到82.9%.本文采用机器学习方法,基于语音变量建立有效的抑郁症自动识别模型,为抑郁症的辅助识别提供客观的指标和依据.

feature:选取了26个在抑郁症研究中应用较为广泛的语音特征作为研究对象, 包括强度、响度、过零率、清浊比率、基频、基频包络 、8个线性谱对以及12个梅尔倒谱系数。在静态特征基础上计算一阶导数和长时特征。采用专门的特征提取软件openSMILE,一共获得了988个语音特征.

Detecting Depression Using an Ensemble Logistic Regression Model Based on Multiple Speech Features

Early intervention for depression is very important to ease the disease burden, but current diagnostic methods are still limited. Tis study investigated automatic depressed speech classification in a sample of 170 native Chinese subjects (85 healthy controls and 85 depressed patients). Te classification performances of prosodic, spectral, and glottal speech features were analyzed in recognition of depression. We proposed an ensemble logistic regression model for detecting depression (ELRDD) in speech. Te logistic regression, which was superior in recognition of depression, was selected as the base classifier. Tis ensemble model extracted many speech features from different aspects and ensured diversity of the base classifier. ELRDD provided better classification results than the other compared classifiers. A technique for identifying depression based on ELRDD, ELRDD-E, was here suggested and tested. It offered encouraging outcomes, revealing a high accuracy level of 75.00% for females and 81.82% for males, as well as an advantageous sensitivity/specificity ratio of 79.25%/70.59% for females and 78.13%/85.29% for males.

feature： Te glottal features were calculated using the TTK Aparat toolbox, and the prosodic and spectral features were calculated using the openSMILE .

Giving Voice to Vulnerable Children: Machine Learning Analysis of Speech Detects Anxiety and Depression in Early Childhood

This paper presents a new approach for identifying young children with internalizing disorders using a 3-min speech task. We show that machine learning analysis of audio data from the task can be used to identify children with an internalizing disorder with 80% accuracy (54% sensitivity, 93% specificity). The speech features most discriminative of internalizing disorder are analyzed in detail, showing that affected children exhibit especially low-pitch voices, with repeatable speech inflections and content, and high-pitched response to surprising stimuli relative to controls. This new tool is shown to outperform clinical thresholds on parent-reported child symptoms, which identify children with an internalizing disorder with lower accuracy (67–77% versus 80%), and similar specificity (85–100% versus 93%), and sensitivity (0–58% versus 54%) in this sample. These results point toward the future use of this approach for screening children for internalizing disorders so that interventions can be deployed when they have the highest chance for long-term success.

HIERARCHICAL ATTENTION TRANSFER NETWORKS FOR DEPRESSION ASSESSMENT FROM SPEECH

A growing area of mental health research is the search for speech-based objective markers for conditions such as depression. However, when combined with machine learning, this search can be challenging due to a limited amount of annotated training data. In this paper, we propose a novel crosstask approach which transfers attention mechanisms from speech recognition to aid depression severity measurement. This transfer is applied in a two-level hierarchical network which mirrors the natural hierarchical structure of speech. Experiments based on the Distress Analysis Interview Corpus – Wizard of Oz (DAIC-WOZ) dataset, as used in the 2017 Audio/Visual Emotion Challenge, demonstrate the effectiveness of our Hierarchical Attention Transfer Network. On the development set, the proposed approach achieves a root mean square error (RMSE) of 3.85, and a mean absolute error (MAE) of 2.99, on a Patient Health Questionnaire (PHQ)-8 scale [0, 24], while on the test set, it achieves an RMSE of 5.66 and an MAE of 4.28. To the best of our knowledge, these scores represent the best-known speech-only results to date on this corpus.

Improvement on Speech Depression Recognition Based on Deep Networks

To reduce the burden of clinicians diagnosing a large number of depressive symptoms, the field of artificial intelligence researchers are increasingly interested in designing automatic recognition systems for depression. Depressed patient have different speech signal from normal people. Here, we present a deep model, Depression AudioNet, which encodes depression-related features in the vocal tract and provides a more comprehensive audio representation. Firstly, the Mel-frequency cepstral coefficients (**MFCCs**) were extracted from raw audio data. Secondly, the robust emotions features were acquired by Multiscale Audio Delta Normalization (**MADN**), which is a data processing algorithm we proposed. Finally, the MFCCs and the emotions features of two adjacent segments of local audio were fed into the Depression AudioNet in turn to train the network. This method solves the problem of less training data and low precision by increasing the length information of the sample without reducing the number of samples. Experiments are conducted on A VEC2014 dataset, and the results shows that the proposed method is more effective and accurate than the existing speech depression recognition algorithms.

Investigation of the Accuracy of Depression Prediction Based on Speech Processing

The present study investigates the accuracy of prediction of depression based on speech processing. Depression is one of the most widespread psychiatric disorders, but early detection of depression is difficult. The Beck Depression Inventory II (BDI) is a self-assessment questionnaire and can accurately predict the severity of depression. BDI is most often used for early detection. There is no known objective biomarker for depression, but the state alters the speech of the individual suffering from it, providing an opportunity for speech-based detection. In the current study, we investigated the accuracy of prediction of depression severity based on speech signal processing, and how accurately it can predict the severity of depression compared to the Beck Depression Inventory.

feature : LLD

MFCC-based Recurrent Neural Network for Automatic Clinical Depression Recognition and Assessment from Speech

Clinical depression or Major Depressive Disorder (MDD) is a common and serious medical illness. In this paper, a deep recurrent neural network-based framework is presented to detect depression and to predict its severity level from speech. Low-level and high-level audio features are extracted from audio recordings to predict the 24 scores of the Patient Health Questionnaire and the binary class of depression diagnosis. To overcome the problem of the small size of Speech Depression Recognition (SDR) datasets, expanding training labels and transferred features are considered. The proposed approach outperforms the state-of-art approaches on the DAIC-WOZ database with an overall accuracy of 76.27% and a root mean square error of 0.4 in assessing depression, while a root mean square error of 0.168 is achieved in predicting the depression severity levels. The proposed framework has several advantages (fastness, non-invasiveness, and non-intrusion), which makes it convenient for real-time applications. The performances of the proposed approach are evaluated under a multi-modal and a multi-features experiments. MFCC based high-level features hold relevant information related to depression. Y et, adding visual action units and different other acoustic features further boosts the classification results by 20% and 10% to reach an accuracy of 95.6% and 86%, respectively. Considering visual-facial modality needs to be carefully studied as it sparks patient privacy concerns while adding more acoustic features increases the computation time.

feature : MFCC

Natural Language Processing Methods for Acoustic and Landmark Event-based Features in Speech-based Depression Detection

This paper proposes a framework for analyzing speech as a sequence of acoustic events, and investigates its application to depression detection. In this framework, acoustic space regions are tokenized to ‘words’ representing speech events at fixed or irregular intervals. This tokenization allows the exploitation of acoustic word features using proven natural language processing methods. A key advantage of this framework is its ability to accommodate heterogeneous event types: herein we combine acoustic words and speech landmarks, which are articulation-related speech events. Another advantage is the option to fuse such heterogeneous events at various levels, including the embedding level. Evaluation of the proposed framework on both controlled laboratory-grade supervised audio recordings as well as unsupervised selfadministered smartphone recordings highlight the merits of the proposed framework across both datasets, with the proposed landmark-dependent acoustic words achieving improvements in F1(depressed) of up to 15% and 13% for SH2-FS and DAIC-WOZ respectively, relative to acoustic speech baseline approaches.

Tracking D9XXdepression severity from audio and video based on speech articulatory coordination

The ability to track depression severity over time using passive sensing of speech would enable frequent and inexpensive monitoring, allowing rapid assessment of treatment efficacy as well as improved long term care of individuals at high risk for depression. In this paper an algorithm is proposed that estimates the articulatory coordination of speech from audio and video signals, and uses these coordination features to learn a prediction model to track depression severity with treatment. In addition, the algorithm is able to adapt its prediction model to an individual’s baseline data in order to improve tracking accuracy. The algorithm is evaluated on two data sets. The first is the Wyss Institute Biomarkers for Depression (WIBD) multi-modal data set, which includes audio and video speech recordings. The second data set was collected by Mundt et al (2007) and contains audio speech recordings only. The data sets are comprised of patients undergoing treatment for depression as well as control subjects. In its within-subject tracking of clinical Hamilton depression (HAM-D) ratings, the algorithm achieves root mean squared error (RMSE) of 5.49 with Spearman correlation of r = 0.63 on the WIBD data set, and achieves RMSE = 5.99 with r = 0.48 on the Mundt data set.

feature :

- Low-level features:Audio formant frequencies and

Video facial action units.

- High-level features