评估抑郁量表的真实性研究

魏楚扬[1]，李嘉康[1]，王子祎[1]，华有伟[1]
1. 兰州大学 信息科学与工程学院，兰州 730000
摘要：抑郁量表是大规模筛查评估抑郁症的主流方案之一。实际中，患者在填写量表时出于隐私、利益、病耻感等因素的考虑，可能会出现欺骗、隐瞒、敷衍等行为，导致量表填写结果失真，进而影响量表的评估和治疗。本项目基于心理活动和外显行为间的关联关系，侧重于真实性与注意力机制的相关性研究。首先关注被试作答的认真程度这一有关量表填写真实性的子问题。在开发一款智能量表作答app的基础上，借助多模态融合的思想，分别利用android端收集的手势操作数据进行行为建模、人脸图像数据进行情绪评估、环境音频数据进行噪音监测三种方法，量化被试作答量表过程当中的认真程度，为量表的真实性评估提供了一个重要参考。
关键词： 抑郁量表；多模态融合；注意力机制；情绪检测；信号处理

A Study to Assess the Veracity of Depression Scales

Wei C Y[1], Li J K[1], Wang Z Y[1], Hua Y W[1], Xu C Y[2]
1. School of Information Science & Engineering, Lanzhou University, Lanzhou 730000, China
Abstract: The Depression Scale is one of the mainstream options for large-scale screening assessment of depression. In practice, patients may deceive, conceal, or fool when filling out the scales due to privacy, interest, and stigma of illness. This leads to distortion of the results in scale completion, which in turn affects the assessment and treatment of the scale. Our project focuses on the correlation study of authenticity and attention mechanism based on the correlation between mental activities and external behaviors. The first focus is on the sub question of conscientiousness of the participant's responses, which is related to the authenticity of the completed scale. Based on the development of the smart scale, an app we developed, we use the multimodal fusion idea to quantify the attentiveness of subjects in the process of responding to the scale by using three methods: behavioral modeling with gesture data collected from android, emotional assessment with face image data, and ambient noise monitoring with environmental audio data collection. Our work provides an important reference for the authenticity assessment of the scale.
Key words: Depression Scale; Multimodal Learning Analysis; Attention Mechanism; Emotion Detection; Signal Processing

抑郁症是一种常见的精神疾病，抑郁自评量表是对门诊病人的粗筛、情绪状态评定的有效方法之一，但在缺少医师正确的指导下填写量表的结果存在真实性偏差，我们项目的目的为研究抑郁量表受试者在填写量表过程中外显的特征与其填写的量表的真实性之间的相关性，并开发一款能收集被试在填写量表过程中的有效信息，并在进行相关特征分析的基础上给予被试真实性评价的app。项目研究的基本思路为：首先通过研究相关文献，探究被试填写量表过程中有关真实性的相关特征，并以此为理论基础收集相关特征数据，再根据机器学习、深度学习模型完成模型构建，最终通过融合模型的部署与计算，开发出一款评估量表填写真实性的app。

1 背景介绍

据世界卫生组织（WHO）发布的《抑郁症及其他常见精神障碍》，全球有超过3.5亿人罹患抑郁症，近十年来患者增速约18%。目前中国泛抑郁人数逾9500万。抑郁会极大地影响患者的身心健康，严重者甚至会威胁生命。2020年最新统计数据显示，我国约有4万名精神科医生，每10万人中仅有3.5名。由于患者众多且缺乏医师，外加国人对抑郁的偏见观念，抑郁症从产生到发现再到治疗康复的周期长、效率低等问题迟迟得不到解决。

抑郁自评量表（Self-Rating Depression Scale，SDS），如PHQ-9，是对门诊病人的粗筛、情绪状态评定的有效方法之一，量表含有20个反映抑郁主观感受的项目，如：忧郁、易哭、睡眠障碍、食欲减退、性欲减退、体重减轻等，被试（患者）通过如实填写量表能够有效地诊断出抑郁评级，根据评定结果也能获得相应的治疗康复方案[1]。但是，在缺少医师正确的指导下填写量表存在的结果真实性问题值得我们思考：

1) 量表以问题为导向考察形式单一，不能与被试（患者）建立情感联结。被试（患者）在填写过程中容易陷入烦闷、悲伤、焦虑等情绪中，可能因失去耐心胡乱填写、因逻辑混乱产生误解与遗忘，影响统计结果的真实性。
2) 被试出于隐私、利益、病耻感等因素的考虑，产生态度不正、有意不合作以及敷衍的表现。这样的作答会影响结果的正确性与准确度，进而会影响数据采集者做出的决策。

基于量表填写真实性不易量化，我们设计了一款记录用户填写过程、科学评估量表填写真实性的app。从用户填写量表的手势操作、面部表情、环境声音三个角度建模，进而进行模型融合，以提高模型的信度与效度。由于刻意隐瞒、不合作等深层态度不易考量，我们的项目着重探究被试填写量表时的认真程度与注意力机制，基于前人对相关领域的研究，展开我们的实践。

2 模型准备

2.1 基本假设

以下假设均使用于本文中的所有模型

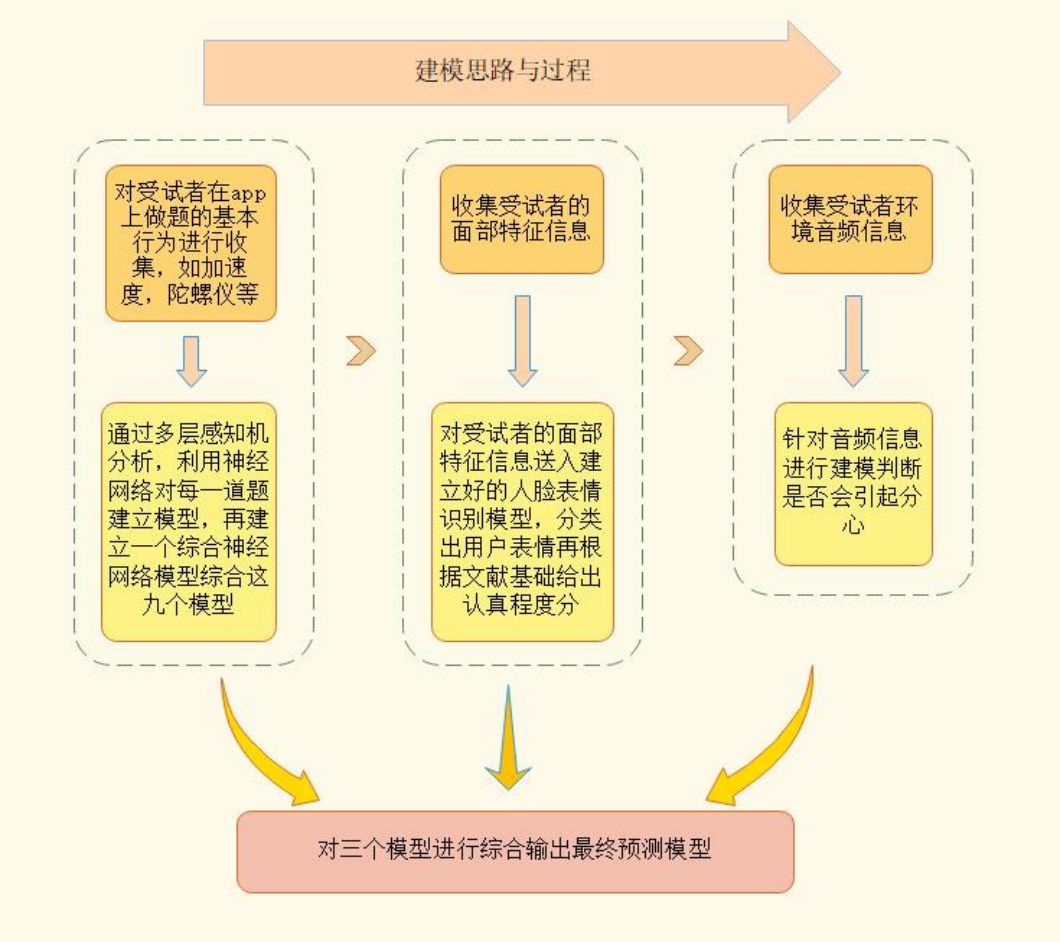
- 所给的数据均是真实准确可靠的，不考虑在数据统计时存在偏差。
- 收集的数据符合一定的客观真实分布。
- 所有受试者均按照要求完成PHQ9量表，并根据自身答题的真实情况进行自评。

2.2 符号说明

2.3 模型简介

整体模型的大致思路与过程如下图所示，主要是分别利用不同的信息建立行为分析模型，

表情识别模型，噪声监测模型，最后利用多模态融合思路构建出一个最终模型。



3 行为分析模型

3.1 建模策略

行为分析模型集中分析受试者在填写量表中的所体现的可量化行为，即找出与受试者不认真填写表现相关的特征，并搜集这些特征构建模型。基于传感器技术的发展[2]，我们能够收集到的相关特征有许多，每道题做题时间[3]，改动次数，加速度，陀螺仪等特征信息[4]。考虑到PHQ-9量表共有九道题，每道题的内容、考察方向都不相同，且受试者读题时间、思考时间也不相同。因此我们针对每一道题构建了一个神经网络模型，以此更精确地研究采集到的相关数据。最后把9个模型的预测结果综合，训练成一个模型算出最终的模型预测得分。

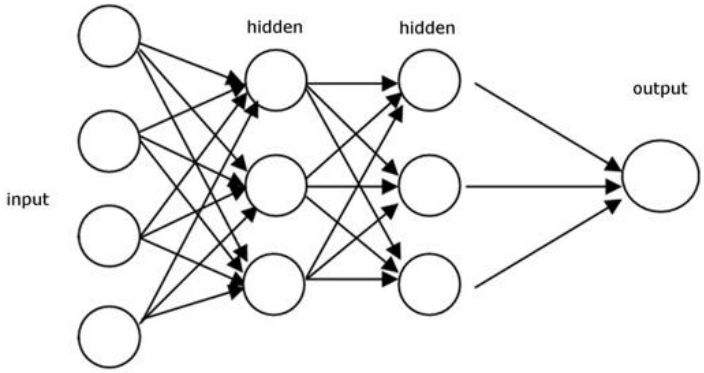
3.2 模型分析

3.2.1 神经网络方法

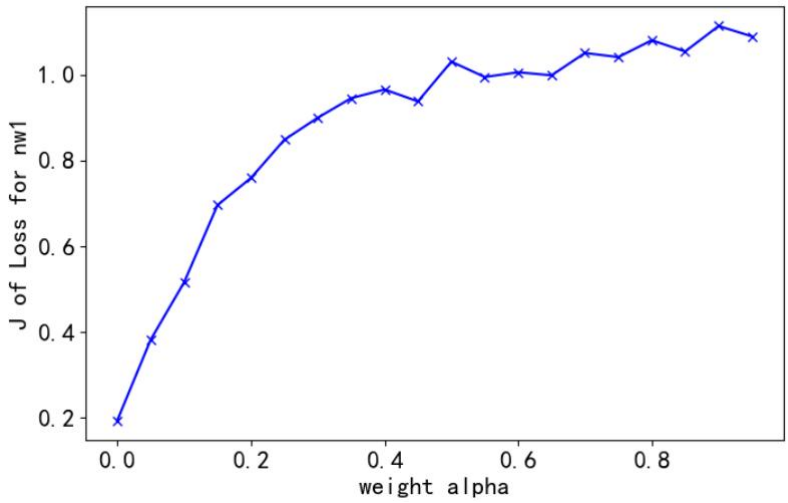
神经网络模型是一个十分强大的机器学习算法。神经网络的模型类似脑细胞传递神经信号的方式。神经网络模型是以神经元的数学模型为基础来描述的。简单地讲，它是一个数学模型，神经网络模型由网络拓扑。

其中大体分为三层，第一层是输入层，第二层是隐藏层，第三层为输入层，其中，hidden layer的层数不固定，在某些简单问题中，hidden layer的层数可能为0，仅有input layer

和output layer；在复杂问题中，hidden layer的层数也可能成百上千。模型中每一层的节点称为“神经元”。位于input layer的神经元对应着训练数据的特征。hidden layer和output layer中的神经元由activation function（激活函数）表达。



由于数据量较小，为了防止过拟合和更好的训练在这个场景中我们为每道题的模型设置的隐藏层数为20层，，激活函数选为ReLU，Solver选为Adma，learning rate初始值为0.001，指数系数power_t设置为0.5，权重系数由于对结果影响较大我们进行了实验来获取J_of_loss最低时alpha又相对合理的权重系数，下图为第一题的神经网络模型的J_of_loss随着alpha的变化而变化的曲线。



可以看到当weight_alpha在0.43时取最优值。其余参数调整如上类似。

3.2.2 数据预处理

由于量表的填写是基于app的，受试者填写的数据会传到服务器再传到本地进行训练，因此数据格式并不能直接满足训练要求，需要进行一定的处理。

下图为从服务器直接得到的数据，ID为题号，USERNAME为用户名，ANSWER为选择的选项，ANSWERTIME为答题所花时间，UPDATE_TIMES为更新改动次数，LINEAR_ACCELEROMETER为线性加速度数据，AVERAGE_GX,GY,GZ分别为加速度的X轴，Y轴，Z轴由陀螺仪返回的各轴加速度，EVAL是用户是否认真做的标签，AGE为年龄，SEX为性别，EDU为学历。

ID	USERNAME	ANSWER	ANSWERTIM	UPDATE_TIM	LINEAR_ACCELEROMETER	AVERAGE_GX	AVERAGE_GY	AVERAGE_GZ	AGE	SEX	EDU	EVAL
1	ID1634816050416	有几天	5.887	1	0.0103662	0.0108721	0.00721674	-0.00516656	35~50	女	专科	非常认真
2	ID1634816050416	没有	3.257	1	0.0188188	0.0210039	0.0305158	-0.00433675	35~50	女	专科	非常认真
3	ID1634816050416	没有	3.542	1	0.013444	0.0172415	0.023329	-0.00389487	35~50	女	专科	非常认真
4	ID1634816050416	没有	3.628	1	0.013681	-0.0256836	-0.0561472	-0.00559891	35~50	女	专科	非常认真
5	ID1634816050416	没有	5.14	1	0.012849	-0.00028384	-0.00291656	-0.00261268	35~50	女	专科	非常认真
6	ID1634816050416	没有	3.308	1	0.00616205	-0.0125937	-0.0306592	-0.00262601	35~50	女	专科	非常认真
7	ID1634816050416	没有	2.493	1	0.0172857	-0.00516794	-0.0120542	-0.00179507	35~50	女	专科	非常认真
8	ID1634816050416	没有	3.67	1	0.0041978	-0.00385433	-0.00916732	-0.00198794	35~50	女	专科	非常认真
9	ID1634816050416	没有	3.183	1	0.0079759	-0.00189694	-0.00320719	-0.00106278	35~50	女	专科	非常认真
1	ID1634816367555	没有	1.894	1	0.0866907	-0.0165678	-0.282443	-0.0659637	18~24	男	本科	非常认真
2	ID1634816367555	没有	0.896	1	0.0909003	0.0640134	0.0547536	0.0064416	18~24	男	本科	非常认真
3	ID1634816367555	没有	1.315	1	0.0809068	-0.0641792	0.0854933	-0.0125516	18~24	男	本科	非常认真
4	ID1634816367555	没有	1.066	1	0.0911454	-0.0576028	0.0151749	0.0167234	18~24	男	本科	非常认真
5	ID1634816367555	没有	0.985	1	0.0713983	-0.0187916	0.0479744	-0.00637055	18~24	男	本科	非常认真
6	ID1634816367555	没有	1.083	1	0.0851398	-0.0668935	0.0715389	0.0312789	18~24	男	本科	非常认真
7	ID1634816367555	没有	0.967	1	0.0973675	0.00168109	-0.0305329	-0.00503238	18~24	男	本科	非常认真
8	ID1634816367555	没有	1.001	1	0.0909809	-0.00409365	0.0122809	-0.00818728	18~24	男	本科	非常认真
9	ID1634816367555	没有	1.017	1	0.0890347	-0.0204666	0.0714618	-0.0181173	18~24	男	本科	非常认真
1	ID1634817708559	没有	1.881	4	0.0801673	-0.164343	-0.0643295	-0.198563	18~24	男	本科	比较认真
2	ID1634817708559	一半以上时间	1.218	2	0.0415029	0.111291	0.0141458	0.0435818	18~24	男	本科	比较认真

由于需要每一道题的数据所以需要把所有受试者相同题目的数据整理在一起如下图所示:

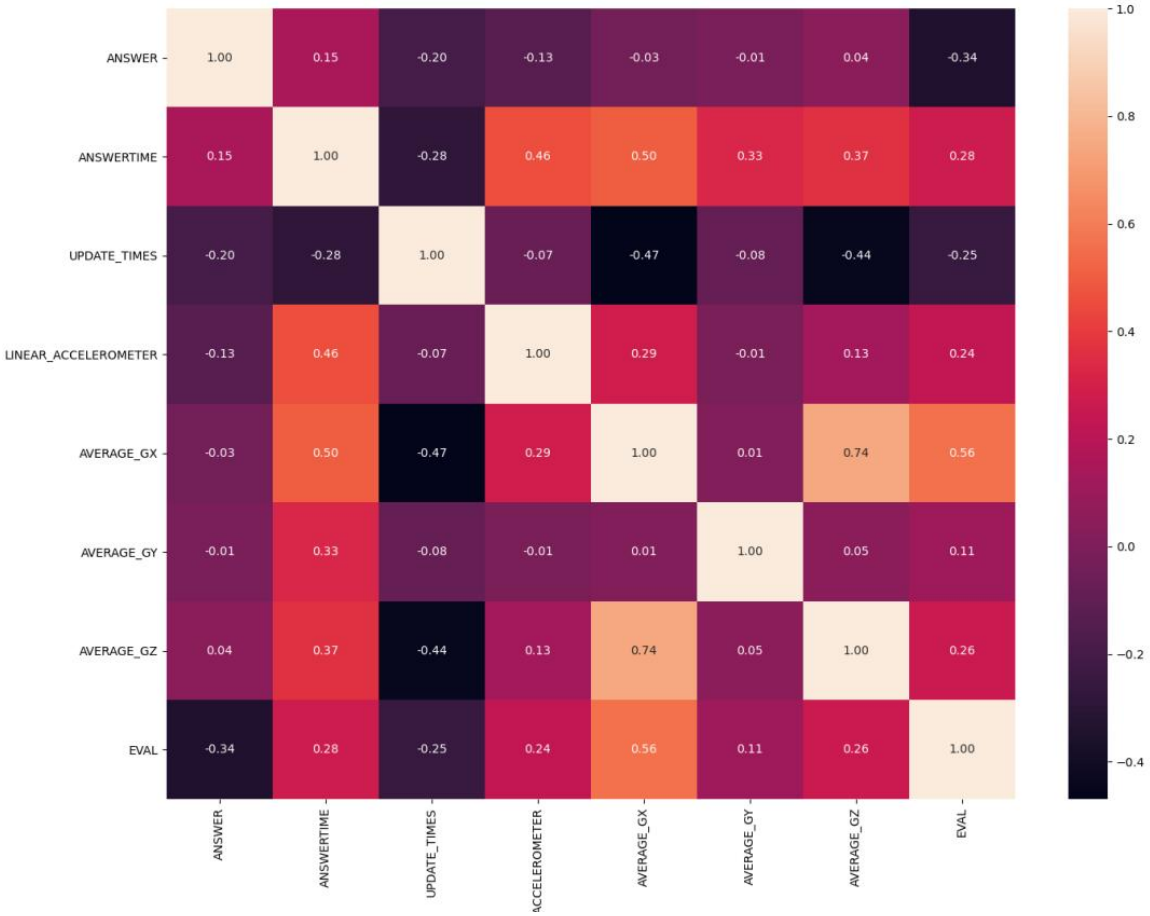
ID	ANSWER	NSWERTIM	DATE_TIM	_ACCELER	VERAGE_G	VERAGE_G	VERAGE_G	EVAL
1	2	5.887	1	0.010366	0.010872	0.007217	-0.00517	5
1	1	1.894	1	0.086691	-0.01657	-0.28244	-0.06596	5
1	1	1.881	4	0.080167	-0.16434	-0.06433	-0.19856	4
1	2	11.513	1	0.170277	-0.02541	-0.02711	-0.01344	3
1	1	1.417	1	0.110445	-0.00072	-0.00016	-0.00046	5
1	1	8.037	1	0.039784	0.000476	0.003845	0.000385	4
1	2	1.52	2	0.060996	-0.30586	-0.05117	-0.23394	4
1	2	5.787	4	0.034699	0.011365	-0.14172	0.001671	4
1	2	19.556	1	0.13664	-0.01351	-0.00841	0.01271	4
1	2	2.45	1	0.036891	-0.13661	-0.01478	-0.1841	4
1	1	2.551	2	0.03046	-0.14487	-0.00983	0.028044	1
1	3	2.556	2	0.04125	-0.17794	-0.06175	-0.17012	1
1	4	1.356	1	0.013286	-0.19725	-0.10896	0.041168	1
1	1	3.281	1	0.051444	0.035868	-0.15976	0.102265	4
1	2	1.672	1	0.051893	-0.09187	-0.44733	-0.10431	2
1	1	1.67	4	0.016032	-0.39489	-0.04698	-0.27125	1
1	2	3.478	2	0.024021	-0.09902	-0.1574	-0.03774	4
1	1	2.251	2	0.031111	-0.20281	-0.24632	-0.09757	4
1	2	11.854	1	0.038644	0.023156	-0.02022	0.020145	5
1	2	4.141	1	0.007167	-0.02918	-0.02117	0.000973	4
1	2	9.424	1	0.078969	-0.00782	-0.04298	-0.0275	5
1	1	2.201	3	0.109406	-0.10932	-0.14318	0.016632	4
1	1	3.62	1	0.036184	-0.16007	0.000788	-0.01632	5

同时需要人工剔除一些不需要研究的特征譬如年龄教育水平等，在完成这些工作之后紧接着对一些范围变化较大的数据进行归一化，譬如加速度，考虑到有些受试者在高速运行的情况下填写量表加速度会非常大，严重影响训练模型的效果，因此需要对特定特征信息进行归一化，同时有一些数据譬如性别，EVAL，ANSWER均为汉字编码无法送入神经网络学习，需要进行one hot处理，对其进行1,2,3.....等编码，譬如从非常不认真到很认真的5个状态分别编码1-5处理完之后送入模型训练即可。

3.3 数据分析

根据数据，在不真实/不认真完成量表的情况下，被试者完成每道题的速度会很快，且修改次数会普遍较多，手机重力加速度和陀螺仪的值会相对（而不是绝对）较大。在认真专注的情况下，男生女生完成每道题的速度会在各自性别的合理区间内波动（男生审题，思考速度会普遍大于女生），修改次数会较少，手机重力加速度和陀螺仪的值会比较小。

其中针对第一题的各项特征与最后结果EVAL关联的热度图如下所示：



可以看到，ANSWERTIME对EVAL有明显的正向关系，UPDATE_TIME呈明显的负向关系，加速度传感器与陀螺仪传感器的稳定性弱，有待更大规模数据的统计研究。

3.4 实验结论

我们把收集到的数据进行3-7 Cross-Validation验证方法，虽然EVAL只有五种，非常认真，认真，一般，不认真，非常不认真，但是我们将其映射到0-20,20-40，40-60,60-80，80-100，分数越高代表越认真，如果数据label为一般，但是它的得分在40-60之间我们则认为预测accurate，最后模型准确率在71.35%下。

4 表情识别模型

4.1 模型构建

4.1.1 模型分析

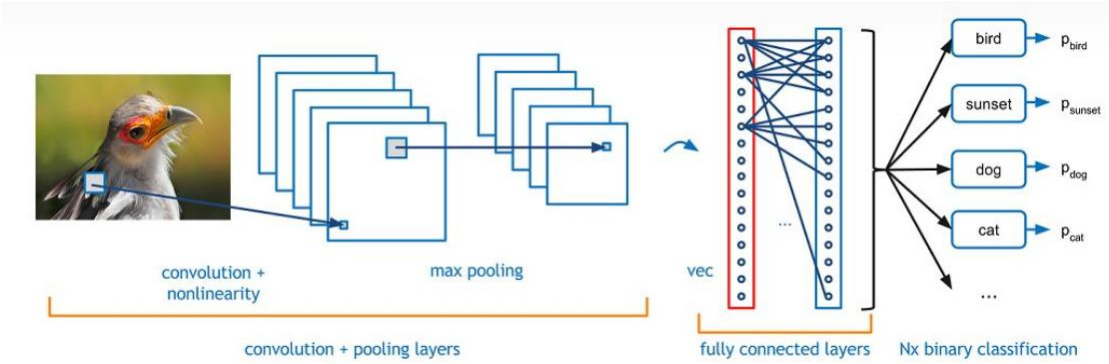
表情识别模型分析主要集中在受试者在做题过程中的表情（微表情）信息上，基于表情分类的思想与研究[6]，我们通过采集被试答题时的面部表情，来探究被试在填写量表过程中的专注度。

近年来，基于卷积神经网络CNN的面部表情识别算法在FER数据集中取得了显著效果[7]。我们的思路是先用CNN深度卷积神经网络模型接收FER2013人脸表情数据完成模型训练，对输入的表情数据进行表情分类，接着借助[6]的评分系统，将用户在不同专注度情况下收集到

的人脸信息进行图像分割，再输出分类后接着评分，最后将每道题的得分加权算出一个综合评分，该评分即为认真程度评分。

4.1.2 CNN模型构建

CNN模型是一种人工神经网络，类似于行为模型中所用到的多层感知机，不过CNN的网络更深，它的主要结构可以分为3层：



- 卷积层(Convolutional Layer)-主要作用是提取特征。
- 池化层(Max Pooling Layer)-主要作用是下采样(downsampling)，却不会损坏识别结果。
- 全连接层(Fully Connected Layer)-主要作用是分类。

在这里我们基于FER2013人脸表情数据集构建CNN神经网络分类出人脸表情，在模型构建过程中，我们将batch_size设置为64，num_epoch设置为50，由于FER2013的图片皆为灰度48x48格式的jpg照片我们需要相应的设置target_size和color_mode。我们首先将数据进行两次卷积，in_channels为32和64核大小皆为(3,3)，激活函数为relu，接着进行池化，池化核大小为(2,2),然后做dropout（0.25）。随后再次放入卷积层，in_channels为128，核大小不变，激活函数为ReLU，接着进行池化、卷积、池化，完毕后dropout（0.25）。然后我们扁平化张量，flatten结束后加入全连接输入channels为1024，激活函数为ReLU，接着dropout(0.5)，最后放入一个维度为7的全连接层,激活函数为softmax，即构建模型完毕。最后输出的标签有fear，disgust，sad，angry，happy，neutral，surprise。构建模型的过程中我们借鉴了Kaggle Fer2013人脸表情数据挑战的一位挑战者[8]的思路，通过CV 0.7-0.3的交叉验证方法，我们的人脸表情分类准确率高达91.3%。

4.1.3 人脸裁剪分割

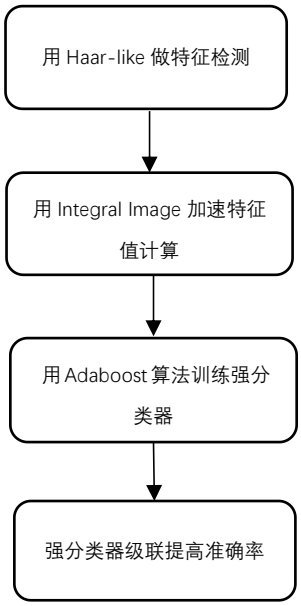
由于在PHQ9量表填写过程中的收集到的受试者表情信息并不能直接输入到CNN模型中进行分类，因为FER2013的训练集是灰度48x48如右所示：

手机摄像头能够收集到的照片如下图所示：

该照片中包含了其它大量的冗余信息，如窗帘、教室支架、座位、衣服、门等，因此我们必须用到一个人脸裁剪模型，来过滤冗余信息，提高模型精确度。这里我们可以用到OpenCV中CascadeClassifier库，它是OpenCV中objdetect模块中用来做目标检测的级联分类器的一个类。它的功能是利用滑动窗口机制与级联分类器的方式，利用Harr-Like特征确定人脸的位置。

在所有缩放尺度下，这些特征组成了boosting分类器使用的全部“原材料”。他们从原始灰度图像的积分图中快速计算得出。

在这里我们构建人脸提取模型时先利用load()函数加载cv库中已经内置好的haarcascade_frontalface_alt2.xml文件，接着我们将图片灰度化，resize成48x48，然后调用detectMultiScale()函数实现多尺度检测其中参数考虑到实际情况，一开始我们将size从最大逐步缩小，因为一开始我们的剪裁效果并不是非常理想，在size减小的同时我们调整了ScaleFactor，让函数可以继续缩小检测框而达不到下限，最后我们将参数设置为如下scaleFactor=1.3, minNeighbors=4, minSize=(30, 30)。



经过该模型处理最终我们输出的剪裁后非灰度图像为如下图所示:

4.1.4 得分模型

由于CNN卷积模型只能输出一个表情的分类,而不能直观的告诉我们最后的专注度得分,因此我们借鉴文献的评分机制，为每一个表情的类附上一个权重系数加入最后得分，其中result[0]到result[6]分别代表模型对Angry,disgust,fear,happy,neutral,sad,surprise的预测概率，构建的得分模型如下所示：

$$FinalScore = 50 + 25 * (0.6 * result[0] + (-0.3) * result[1] + retult[2] * 0.5 + result[3] * (-0.4) + result[4] * 1 + result[5] * 0.4 + result[6] * 1.65$$

其中对于相关系数的构建是我们根据受试者在真实情况下自行调整的,与文献中不太符合的是Angry系数在文献里是负值，但是真实情况下许多受试者在认真填写认真思考的情况下表情偏Angry，sad，甚至是fear，相反的happy情况在文献里是正权重，但是许多不认真的受试者是呈现嬉皮笑脸的情况，因此我们将相关权重调成负值，其余权重诸如此类调整，最后得到上述模型。

4.2 结果总结

我们将受试者人脸表情数据集按照CV交叉验证0.7-0.3的思路，对数据进行验证，其中由于数据量比较局限,无法真正意义上的泛化,因此我们将其评价指标划分为50以下为不认真做，50以上为认真做，其中分数越高越认真。经过测试发现按照这样的评价标准，准确率在85.7上下波动。

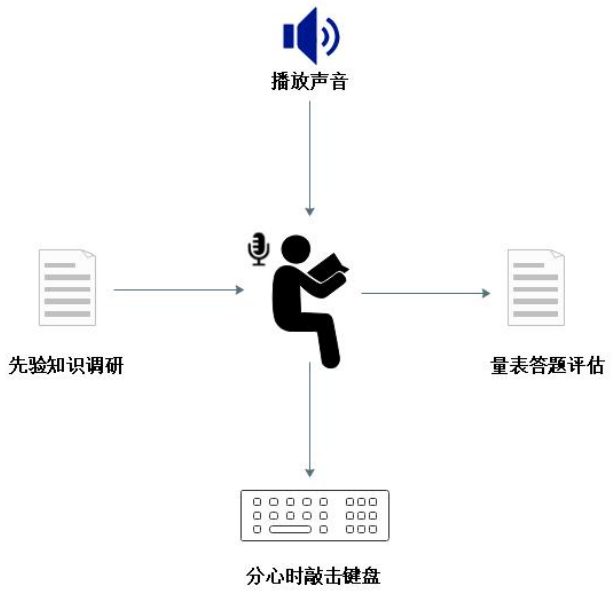
如下为最终模型输出的得分和处理后的人脸照片数据：

5 噪声监测模型

环境噪音会引发人的注意力偏差，噪音作为一种客观存在可量化的指标，经研究证实，人们会被一定波动频率的声音分散注意力，音频信号已经逐渐被学者用于注意力机制的研究当中[9][10]。 由于填写量表的真实性可从一定程度上利用专注力与认真程度表达，我们将Jeffrey Pronk等人有关通过测量环境噪声对持续注意力的多模态学习分析的模型[4]应用到实际场景中。我们的目标是构建一个模型来预测环境声音是否让被试分心，并将这个模型与前面两个模型结合起来，在多模态模型中提高模型的准确性。

5.1 建模策略

我们的模型是基于Jeffrey Pronk团队有关远程学习中噪声对注意力影响的实验与建模获取的。模型构建的数据通过实验所得，实验过程如下图所示：



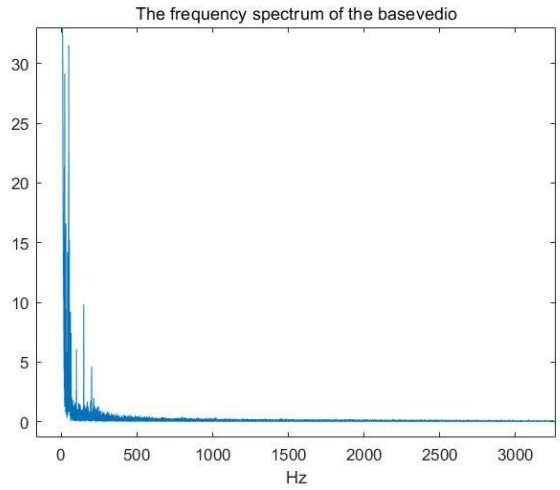
首先要求参与者填写有关一些先验知识的量表，基于参与者的先验知识我们要求其阅读文章，持续45-90分钟，在阅读学习的过程中，参与人员需根据自身情况进行客观分析，当判断自己分心时通过敲击键盘反馈自身状态，设备将记录下分心时间。实验过程中一共会播放9段分贝、频率不同的声音，音频文件可以在[11]中找到。音频片段以大约5分钟的随机时间间隔手动播放。每个参与者都不会两次听到相同的音频片段。在阅读文章后，参与者会做与阅读前相同的问卷，我们通过比较两次测试答案的差异，以验证参与者在实验过程中是否专注。

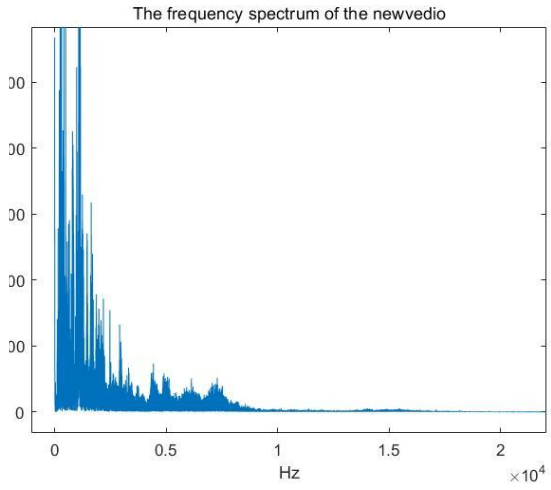
由于每段音频都被打上了标签，Jeffrey Pronk团队根据收集到的录音数据进行训练，最终得到了一个对音频文件的分心判别模型。

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
resizing (Resizing)          (None, 32, 32, 1)         0
_________________________________________________________________
normalization (Normalization (None, 32, 32, 1)         3
_________________________________________________________________
conv2d (Conv2D)              (None, 30, 30, 32)        320
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 28, 28, 64)        18496
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 14, 14, 64)        0
_________________________________________________________________
dropout (Dropout)            (None, 14, 14, 64)        0
_________________________________________________________________
flatten (Flatten)            (None, 12544)             0
_________________________________________________________________
dense (Dense)                (None, 128)               1605760
_________________________________________________________________
dropout_1 (Dropout)          (None, 128)               0
_________________________________________________________________
dense_1 (Dense)              (None, 2)                 258
=================================================================
Total params: 1,624,837
Trainable params: 1,624,834
Non-trainable params: 3
```

5.2 数据分析

我们的项目采用上述模型，通过Android端调用用户录音机，采集被试填写量表过程当中的音频数据，由于Android端无法后台调用录音机，我们让被试在作答前录制10秒的环境声音。先将原始音频文件进行滤波，以去除低频段上的工频干扰以及设备工作期间的工作噪音，接着经过文件格式转换与重采样等操作后，我们将音频数据取进行切割，每一秒作为一个片段带入模型进行运算，运算结果根据阈值进行叠加最后取平均，进而得到音频模型注意力判别得分。





上左图为手机录音机录制的无声环境的频谱图，右图为我们采集到一段嘈杂环境下的录音频谱，嘈杂环境中有人声、歌曲声、风扇转动声与键盘敲击声。我们将收集到的语音数据减去左图中的基线频谱，放入模型中进行训练。

5.3 结果总结

我们先将噪音数据放入模芯，可以看到，模型将10秒的环境录音分成了10段，每一段有一个专注度预测结果，数组中第一个元素返回的是注意力分散度，第二个元素返回的是专注程度。

$$DistractionRate + \text{AttentionRate} = 1$$

模型设置了专注度评估阈值为0.5，当预测注意力分散度低于0.5，即判断环境声音不会对注意力产生影响。最终我们对分心程度和专注度的综合计算伪代码为：

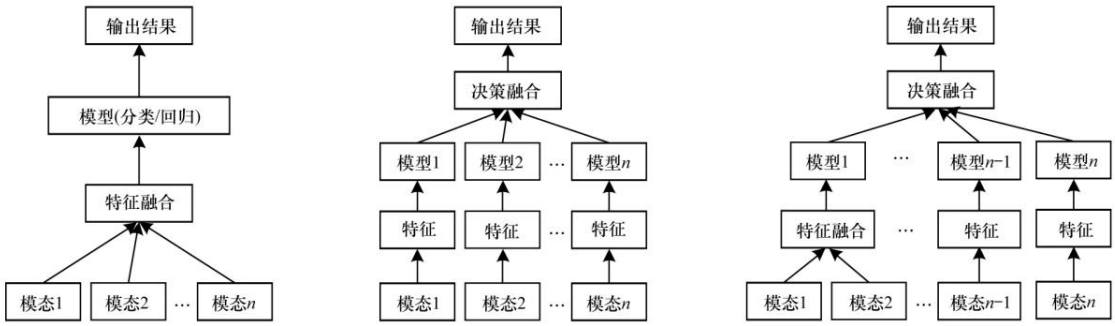
```
Begin
   For i: = 0 To Fragmentsize - 1{
      If(DistractionRate[i] <= 0.5)    Then{
         avg: += 0;
      }
      else{
         avg: += (DistractionRate[i] - 0.5) / 0.5;
      }
   }
   avg_Distra: = avg / FragmentSize * 100;
   avg_Atten: = (1 - avg / FragmentSize) * 100;
End
```

模型对嘈杂环境音频数据的平均分心程度判准高达99.8%，故判定在此段声音下被使的注意力难以集中，答题认真程度与真实性受到严重影响。

```
Fragment: 0 | start: 0 | end 44100 | fragment: 44100 | Prediction: [9.99985695e-01 1.43156985e-05] | Derived: 0.9999857
**
Fragment: 1 | start: 44100 | end 88200 | fragment: 44100 | Prediction: [1.0000000e+00 5.0430764e-13] | Derived: 1.0
**
Fragment: 2 | start: 88200 | end 132300 | fragment: 44100 | Prediction: [9.999223e-01 7.767953e-05] | Derived: 0.9999223
**
Fragment: 3 | start: 132300 | end 176400 | fragment: 44100 | Prediction: [0.99898046 0.00101947] | Derived: 0.99898046
**
Fragment: 4 | start: 176400 | end 220500 | fragment: 44100 | Prediction: [0.9976799  0.00232016] | Derived: 0.9976799
**
Fragment: 5 | start: 220500 | end 264600 | fragment: 44100 | Prediction: [9.9971002e-01 2.8999135e-04] | Derived: 0.99971
**
Fragment: 6 | start: 264600 | end 308700 | fragment: 44100 | Prediction: [0.998949   0.00105107] | Derived: 0.998949
**
Fragment: 7 | start: 308700 | end 352800 | fragment: 44100 | Prediction: [9.999924e-01 7.648143e-06] | Derived: 0.9999924
**
Fragment: 8 | start: 352800 | end 396900 | fragment: 44100 | Prediction: [9.9999416e-01 5.7836455e-06] | Derived: 0.99999416
**
Fragment: 9 | start: 396900 | end 441000 | fragment: 44100 | Prediction: [9.993974e-01 6.025660e-04] | Derived: 0.9993974
分心程度: 0.9986528158187866
专注度： 0.001347184181213379
```

6 多模态融合

多模态融合是一种模型融合的方法，模态是指事物发生或存在的方式，多模态是指两个或者两个以上的模态的各种形式的组合。对每一种信息的来源或者形式，都可以称为一种模态（Modality）。



(a) 早期融合方法　(b)晚期融合方法　(c)混合融合方法

由于我们生活在一个多领域相互交融的复杂环境汇总，听到的声音、看到的实物、闻到的味道等等都是一种模态。因此研究人员开始关注如何将多领域数据进行融合实现异质互补，例如语音识别的研究表明，视觉模态提供了嘴的唇部运动和发音信息，包括张开和关闭，有助于提高语音识别性能。可见，利用多种模式的综合语义对深度学习研究具有重要意义。深度学习中的多模态融合技术（Multimodality Fusion Technology，MFT）[12] 是模型在分析和识别任务时处理不同形式数据的过程。多模态数据的融合可为模型决策提供更多信息，从而提高决策总体结果的准确率，其目标是建立能够处理和关联来自多种模态信息的模型。[13]

在我们的模型中，单单只有行为模型远远不够，它更多的表现了用户在填写量表中受试者所处的综合环境因素以及客观因素，但是缺少了面部的特征信息，以及可量化的最重要的环境因素。因此综合三个模型可以让他们进行互补，丰富特征信息，于是我们的想法是对三个模型进行融合，其中融合过程我们借鉴了Giuseppe团队[5] 在他们的实验过程中进行多模态融合的方法:

Case	Multi-modal score
P >= 0.5	P
P < 0.5 and A >= 0.5	((87.78 * P) + (62.07 * A)) / (87.78 + 62.07)
P < 0.5 and A < 0.5	((87.78 * P) + (59.26 * A)) / (87.78 + 59.26)

Table 3: Multi-modal model (P: score of the mobile movement tracking model output, A: ambient audio model output)

我们根据自己的实验的结果比对最后综合模型的结果调整了参数，如下公式所示：

$$FinalScore = 0.63 * Behavior + 0.21 * FacialExpression + 0.16 * AmbientSound$$

其中Behavior是行为模型分析输出的综合得分，FacialExpression是人脸模型输出的结果，AmbientSound是环境噪声模型输出的结果。

经验证，如果按照如下评价标准评价：划分50以下为不认真做，50以上为认真做，其中分数越高越认真，则模型准确率可以到达75.4%。

7 Android 端开发

Android 端开发的详细内容见附件《软件操作手册》、《工程代码》。

8 项目总结

本项目以抑郁量表的填写为切入点，基于心理活动和外显行为间的关联关系，侧重于量表填写真实性与注意力机制的相关性研究。采用多模态融合的思路，在开发一款量表填写app的基础上，通过收集用户填写过程中的手势操作、面部表情以及环境声音综合建模，量化被试作答量表过程当中的认真程度，为量表的真实性评估提供了一个重要参考。

项目科学性：

项目具有认知科学与认知心理学的发展为心理活动与外显行为的关联关系提供了理论依据；基于前后端技术及人工智能的发展，运用机器学习的方法分析app端采集到的数据存在可行性。

项目创新性：

本项目以注意力机制为判定用户填写量表认真程度的指标，将有关真实性的研究剖析开来，从用户自身与环境的角度出发着重探究评估用户专注度的方法。项目开发集成了app、前后端、机器学习等多种技术，从平台搭建到数据收集，从模型构建到应用反馈，搭建出一套完整的闭环控制系统。

项目意义：

- 本项目为抑郁自评量表填写的信度提供了一份量化指标，有利于帮助医师在量表填写质量可监测的前提下进行规模化的抑郁症初筛。
- 有关量表真实性的评估方法并不局限于PHQ-9，本项目的研究思路与方法具有较强的移植性。
- 本项目利用传感器数据、人脸图像、音频数据进行综合建模、组建多模态的思想，可广泛应用于复杂的、具有不确定性的场景中。

参考文献

[1] Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001.

[2] Schneider J , D Börner, Rosmalen P V , et al. Augmenting the Senses: A Review on Sensor-Based Learning Support[J]. Sensors, 2015.

[3] 静远 梁 . Detecting Deception through Reaction Time[J]. Advances in Psychology, 2019, 09(10):1735-1747.

[4] Pronk J. Multimodal learning analytics on sustained attention by measuring ambient noise[J]. 2021.

[5] Giuseppe D. Assessing learner’s distraction in a multimodal platform for sustained attention in the remote learning context using mobile devices sensors[J]. 2021.

[6] 郭晓旭. 基于微表情识别的学生课堂专注度分析系统研究[D]. 云南师范大学, 2019.

[7] Trigueros D S , Meng L , Hartnett M . Face Recognition: From Traditional to Deep Learning Methods[J]. 2018.

[8] https://www.kaggle.com/deadskull7/fer2013

[9] Jka B , Ming Z , Xla B , et al. Monitoring distraction of construction workers caused by noise using a wearable Electroencephalography (EEG) device[J]. Automation in Construction, 125.

[10] Martin J . Sound Distraction Effect on Memory Processing Knowledge Familiarity[J]. 2016.

[11] https://github.com/MultimodalLearningAnalytics/ambient-audio-tracking

[12] Ramachandram D , Taylor G W . Deep Multimodal Learning: A Survey on Recent Advances and Trends[J]. IEEE Signal Processing Magazine, 2017, 34(6):96-108.

[13] 何俊, 张彩庆, 李小珍,等. 面向深度学习的多模态融合技术研究综述[J]. 计算机工程, 2020, 46(5):11.