

**数据挖掘实习三**

**中国地质大学（武汉）地理与信息工程学院**

**2021年 7 月 2 日**

# Lab\_3 基于监督学习的情感识别

## 实习目的

1、结合实际应用理解监督学习分类过程；

2、理论结合实践，采用一种分类方法实现表情类别预测；

3、通过进一步查阅文献，了解相关研究方向的最新研究进展。

## 二、实习题目

基于监督学习的情感识别。

#### 【实验数据】

实验数据为之前标注并处理的数据集。

训练数据可采用已裁剪出人脸的一系列视频帧数据或第二次实验已经降维过的一系列视频帧数据。

#### 【实验要求】

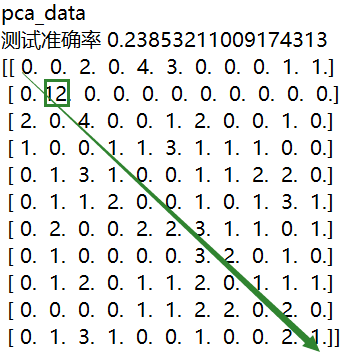
1. 要求采用一种分类器（决策树、贝叶斯、SVM、神经网络、深度学习等），自动识别面部表情。
2. 要求对实验结果进行详细分析，例如同组同学之间的准确率对比，实验不足等等。
3. 要求绘出混淆矩阵、计算准确率等结果图。

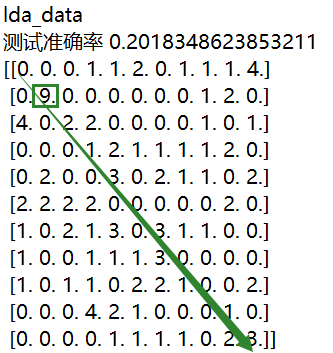
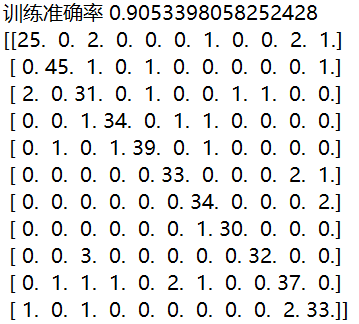
#### 【过程描述】

我选择了**逻辑回归、支持向量机、卷积神经网络**来对数据进行分类。而我采用的数据有四种来源，分别是**PCA降维数据、LDA降维数据、MY\_LDA降维数据，以及8人脸数据**。降维后的数据主要应用在逻辑回归与支持向量机模型，卷积神经网络输入的是8通道人脸图片。

**1、逻辑回归**

采用的回归模型是sklearn封装好的LogisticRegression，用降维后的三种数据进行测试的准确率是**PCA (23.85%)，LDA (20.18%)，MY\_LDA (8.25%)**。它们的混淆矩阵如下，**行代表所属的正确类别，列代表预测的类别**，0-10代表的类别为："anger", "disgust", "fear", "happiness", "neutral", "sadness","surprise", "contempt", "anxiety", "helplessness", "disappointment"。

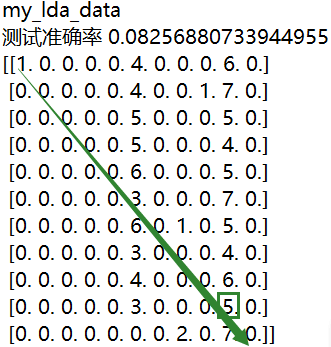
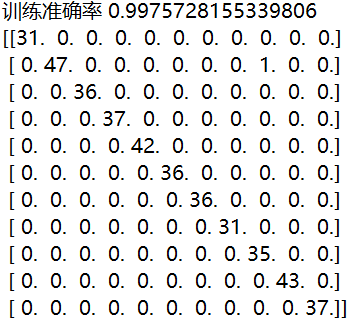
 

图1 三种数据在逻辑回归模型下的混淆矩阵

从混淆矩阵中，我们可以看出：

**① PCA与LDA的情感类别1（disgust）准确率最高，在训练集中情感类别1（disgust）也是数据量最多的（48个）；**

**② MY\_LDA降维的数据集上准确率不高，大多预测到情感类别5（sadness）与情感类别9（helplessness），原因不明。**

**2、支持向量机**

2.1 LinearSVC

LinearSVC模型中有一个系数**C**叫惩罚系数，表示对误差的宽容度；**C越高，说明越不能容忍出现误差**。它还有一个系数**penalty**叫正则化参数，有l1和l2两种可选，其中**l1正则化可以产生稀疏的效果**。因此，我在对线性支持向量机模型探索的时候，我观察了不同参数C和penalty在不同降维数据集上的准确率表现。

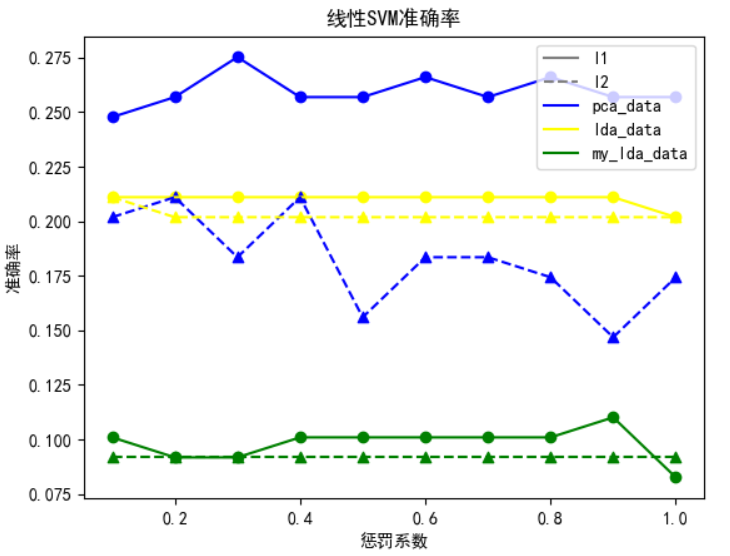


图2 不同参数不同降维方法的线性SVM模型的准确率

从准确率看：正则化方法l1 > l2，降维方法PCA > LDA > MY\_LDA；

从峰值看：三种降维方法的准确率均有提升，其中PCA（**27.5%** > 23.85%）、LDA（**21%** > 20.18%）、MY\_LDA（**12%** > 8.25%）。其中，PCA的峰值在**penalty=0.3**时取到。

2.2 NuSVC与SVC

在非线性支持向量机的实验中，有两种模型可选NuSVC与SVC（它们的唯一区别是损失函数度量方式稍有不同，NuSVC用nu、SVC用C来控制惩罚力度），而产生非线性映射的核函数有高斯核函数（rbf）、线性核函数（linear）、多项式核函数（poly）、sigmod核函数（sigmoid）四种。

以下的实验围绕上述不同类别展开，其中惩罚力度参数的选取，在一系列预实验的对比下决定取**0.39**。原因有三点，一是参数变化带来的准确率的改变不大；二是在线性模型的时候已经观察过了，这里主要看的是核函数的变化；三是在取0.39的时候PCA表现最佳。

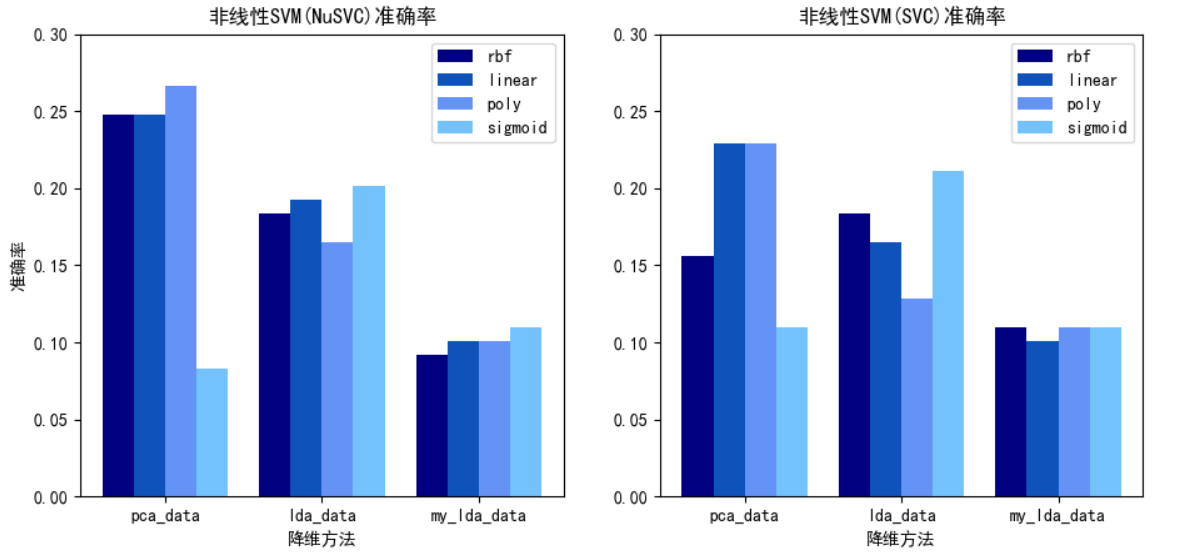


图3 不同核函数与模型在不同降维方法下的准确率对比

从准确率看：NuSVC在PCA降维的数据集上表现更好，准确率最高达26.5%；

SVC在LDA、MY\_LDA降维的数据集上表现更好，准确率最高达22%、12%；

总的来说：不同降维方法下的数据，在经过不同的模型、不同的核函数计算后，准确率多有不同，需综合考量以选择一个最佳的模型达到准确率的极限。

**3、卷积神经网络**

由于卷积神经网络输入的都是图片，本实习也将采用图片的方式输入，舍弃降维后的数据。一个视频用裁剪出来的**8张人脸**代替，**图片大小以模型的标准输入为准**。

3.1 LeNet

识别手写数字的卷积神经网络LeNet的输入是单通道的32\*32大小的数字图片，这里稍作修改，把输入的通道从1改为8，装载数据的时候把人脸图片大小压缩成32\*32，以及修改预测的类别总数为11，一个新的可以识别视频情感的卷积神经网络完成。

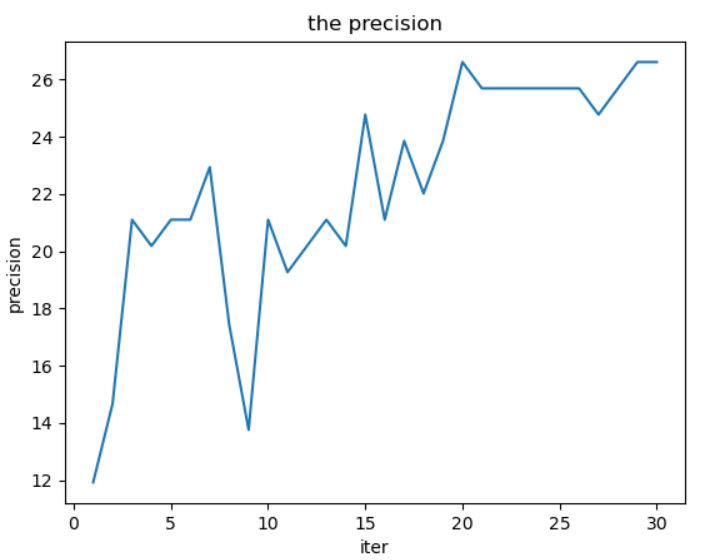
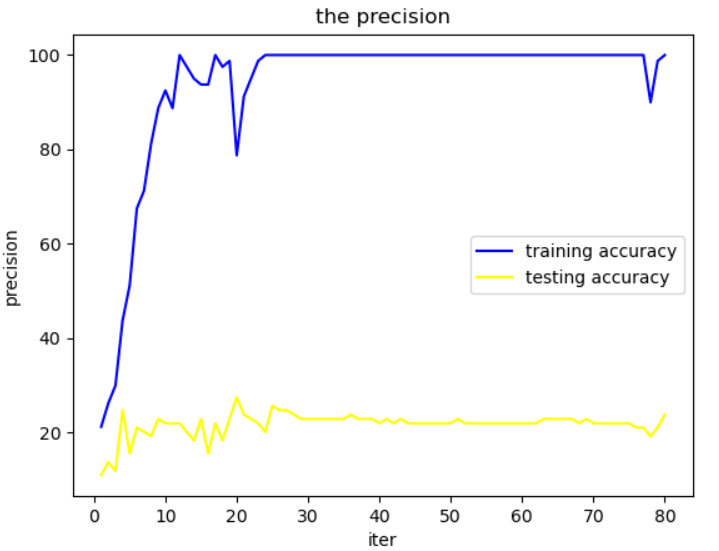
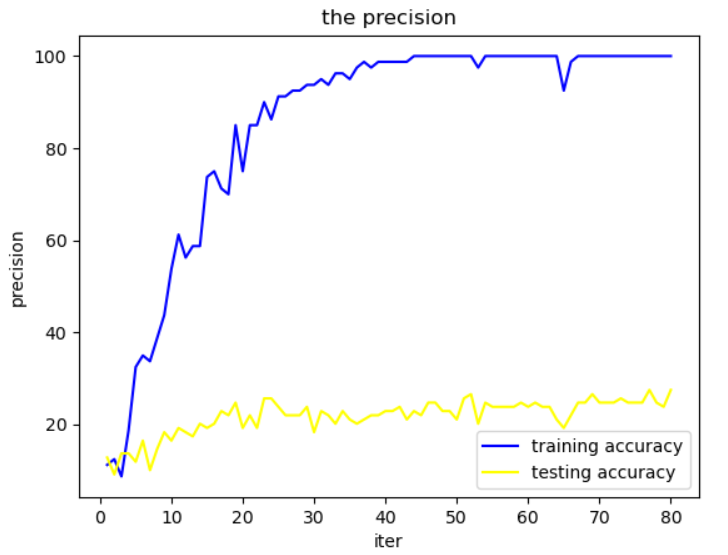


图4 随迭代次数增加的LeNet模型准确率（此后纵轴单位%）

从图4知，LeNet模型的最高准确率在26.5%左右，在**VGG踩坑**后，我又调整了batch\_size与learning\_rate大小，绘出了训练精度与测试精度随迭代次数变化的折线图。



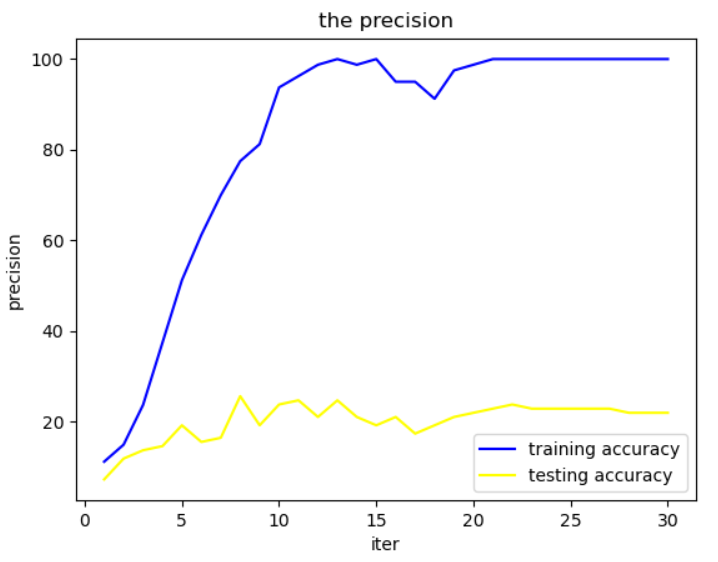
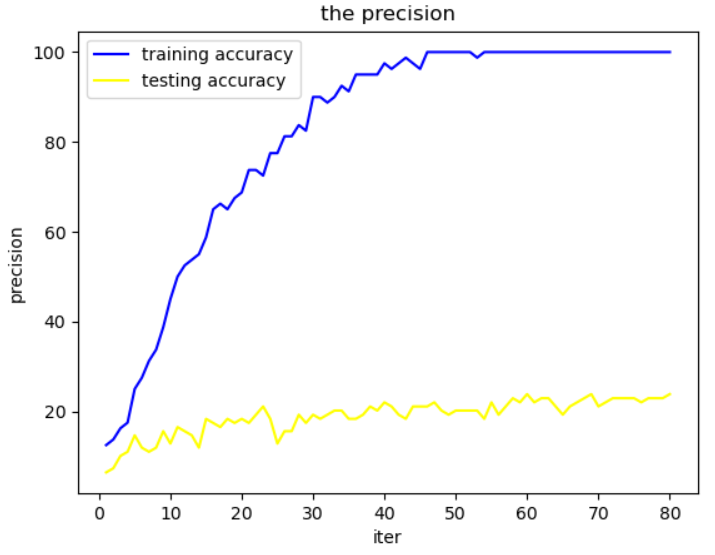


图5 第一行batch\_size=10，第二行batch\_size=20

第一列learning\_rate=0.001，第二列learning\_rate=0.0001

3.2 VGG

VGG模型的修改与上述相似，我用了VGG16、VGG19，却得到了与LeNet图三截然不同的折线图（如图6），**准确率随迭代次数增加而震荡且减少**。

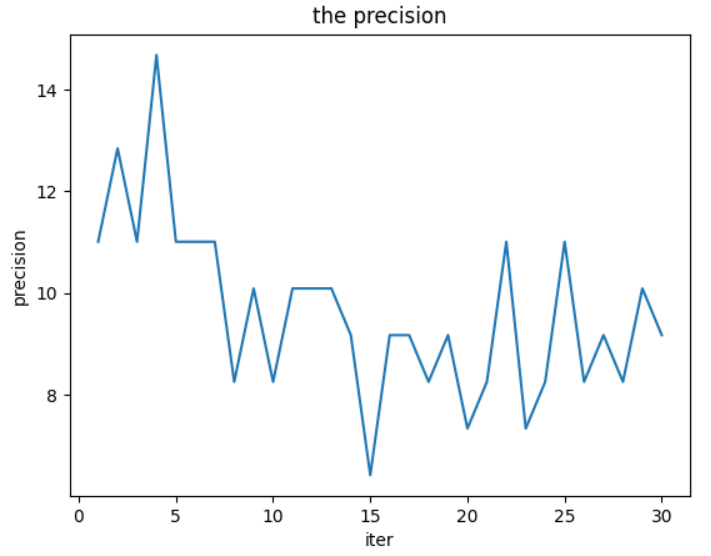
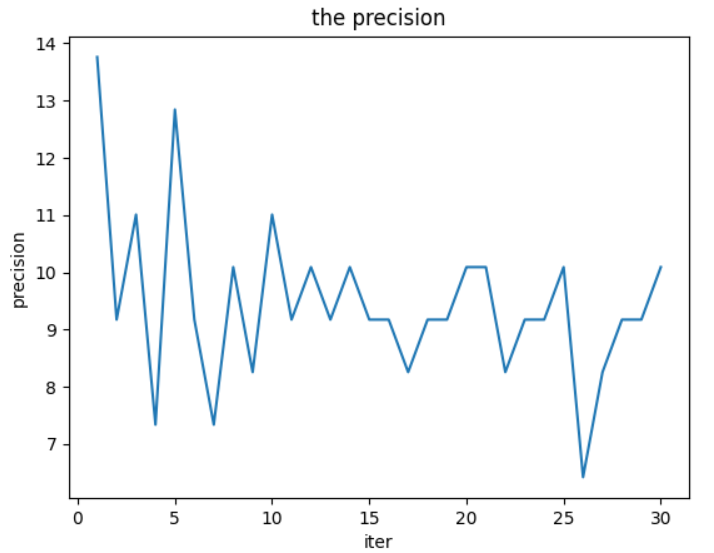


图6 左为VGG16，右为VGG19

为寻找原因，我又增大了迭代的次数，增加了对训练精度的监控，我得到的是训练、测试精度都震荡的折线图（如图7），助教张浩宇学长告诉我可能是**学习率的原因**，让我除以10再试试。

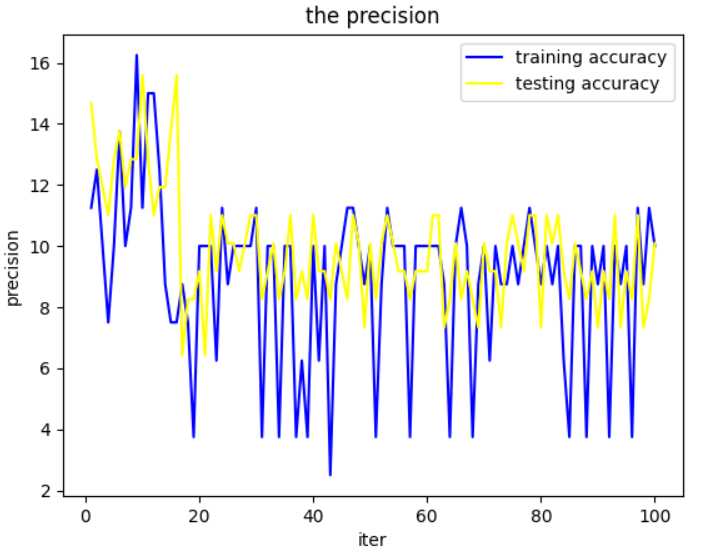


图7 VGG19训练、测试精度折线图

我试了试，确实是这个原因，我用的learning\_rate是0.001，在LeNet模型上还适用，但是VGG就不行了，我还尝试了改变batch\_size的大小，Dropout的概率，得到以下不成对照的图片。

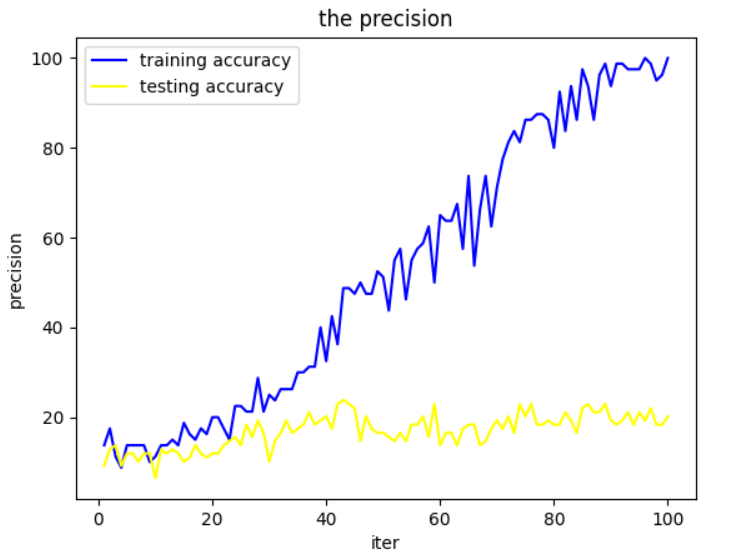
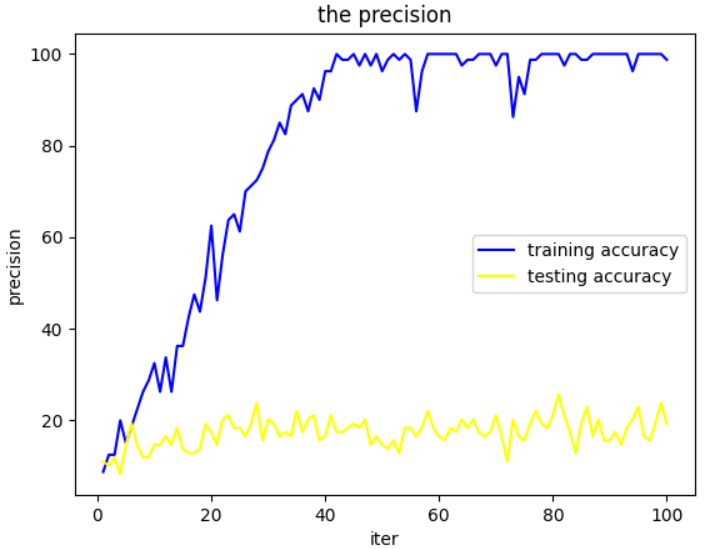


图8 参数不重要，准确率都不高的折线图

**4、总结分析**

试图弄清楚**准确率为什么都这么低**的我对逻辑回归+PCA组的结果进行了深入分析，我把正确的类别作为第一层文件夹，预测的类别作为第二层，来存储代表每个视频的8个人脸。

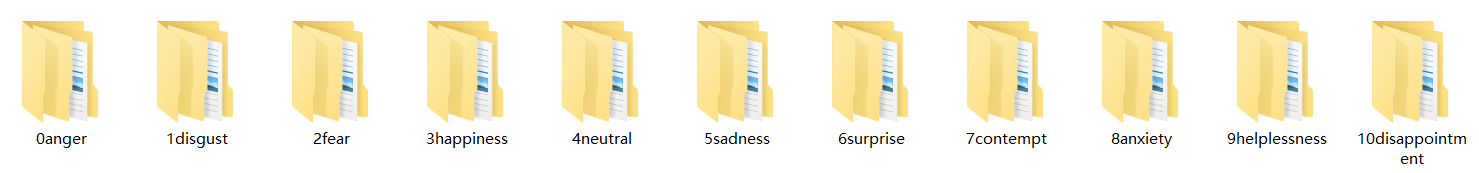


图9 代表正确类别的第一层文件夹

从下图知，情感类别1（disgust）全部预测正确，我们来看看‘1disgust\1disgust’文件夹下的数据是怎样的，怎么能识别全正确，以及在其他降维方法中这个类的准确率是最高的。



图10 逻辑回归+PCA的混淆矩阵

当我打开文件夹，我惊愕的发现，怎么disgust是这样的？！这，根本都看不出来，我当初是怎么标的disgust来着，应该有厌恶、讨厌的面部表情的，这，这，这。



图11 ‘1disgust\1disgust’文件夹里

于是我又查看了其他的类别，结果大抵如此。你能判别下图正确的标签是左边的还是右边的吗？



图12 左（正确）右（预测）

综上，我认为准确率普遍偏低的原因是：**一个视频的情感过于丰富、我们的标注不是很准确、剪裁出的人脸损失了情感**。

#### 【实习遗憾】

计划中，还想自己写一个卷积神经网络的模型，8层卷积2层全连接，用上BN与Dropout，再尝试一下残差的设计，卷积用3\*3与5\*5两种，设计成 [2, 2, 2, 2] 或 [2, 3, 3] 的结构。

## 三、本次实习小结

当人工智能的知识碰撞到《西部世界》的记忆，我意识到自己宛如大自然创造的智能生物，思考在自我与模型之间反复跳转、映射，我反思自己是否存在对已学习知识掌握很好但举一反三能力不足的现象，如同模型的“过拟合欠泛化”问题……

=======================================================

我的联系方式：

电话：

邮箱：