



AffectNet Paper Summary and Training Result

Jiaming Nie

June 13, 2018

1 AffectNet Paper 模型

AffectNet的原paper利用了三种模型对人脸表进行建模分类，具体是三种：

Table 1: AffectNet 中几种分类与回归模型

| | |
|------------------------------------|---|
| Categorical Model | 分类模型，根据不同表情图片与标签进行分类 |
| Dimensional Model | 纬度模型，基于不同表情的Arousal(情绪的强烈程度)和Valence(情绪的正负倾向),以上两个数值均为连续的。 |
| Facial Action Coding System (FACS) | 根据脸部提取的特征点，用于描述脸部表情的动作，并不直接给出表情的分类。用Action Unit来表示。AU6和AU12可能均表示高兴。 |

2 模型Baseline

2.1 Categorical Model Baseline

数据集本身并不均衡，对于不均衡的数据集，采取以下几种方式：

- Imbalanced learning
- Down-Sampling 下采样
- Up-Sampling 上采样
- Weighted-Loss 加权损失

Weighted-Loss的处理方法 在数据集中，根据样本所占的相对比例对损失函数进行修改，使得模型输出结果更贴近于真实情况。当模型对占比较少的样本预测错误时，对网络的反馈会更大，对模型的修正效果会更好。

Weighted-Loss function定义：

$$E = - \sum_{i=1}^K H_{l,i} \log(\hat{p}_i) \quad (1)$$

在公式1中，相关系数的定义如下：

- $H_{l,i}$ 第 l 行第 i 类的偏振系数(penalization factor)
- K 总的类数 (number of classes)
- \hat{p}_i 预测的softmax输出值，区间[0,1]

重新定义的Weighted-Loss Function为：

$$E = \log\left(\sum_j \exp(x_j)\right) \sum_i H_{l,i} - \sum_i H_{l,i} x_i \quad (2)$$

当矩阵 H 是单位矩阵 I 时，加权损失函数和标准的交叉损失函数相同。具体的定义如下：

$$H_{i,j} = \begin{cases} f_i/f_{min} \\ 0 \end{cases} \quad (3)$$

f_i : 第 i 个类的总数， f_{min} 是数据集个数最少类的样本数。

所用的CNN是AlexNet，并且数据在处理前，对图片进行 256×256 的crop，并在数据增强阶段进行 224×224 的random crop。

训练的参数：

- Epoch: 20
- Batch Size: 256
- Learning Rate: 0.01 (每隔10000次叠代下降10%)
- Momentum: 0.9

2.1.1 不同模型下的结果

如图1中所示，对于Original和skew-normalized下，top-1和top-2 F1 score的结果。加权损失的表现是最好的。

TABLE 7
F1-Scores of four different approaches of training AlexNet

| | Imbalanced | | | | Down-Sampling | | | | Up-Sampling | | | | Weighted-Loss | | | |
|----------|------------|-------|-------|------|---------------|------|-------|------|-------------|------|-------|------|---------------|------|-------|------|
| | Top-1 | | Top-2 | | Top-1 | | Top-2 | | Top-1 | | Top-2 | | Top-1 | | Top-2 | |
| | Orig* | Norm* | Orig | Norm | Orig | Norm | Orig | Norm | Orig | Norm | Orig | Norm | Orig | Norm | Orig | Norm |
| Neutral | 0.63 | 0.49 | 0.82 | 0.66 | 0.58 | 0.49 | 0.78 | 0.70 | 0.61 | 0.50 | 0.81 | 0.64 | 0.57 | 0.52 | 0.81 | 0.77 |
| Happy | 0.88 | 0.65 | 0.95 | 0.80 | 0.85 | 0.68 | 0.92 | 0.85 | 0.85 | 0.71 | 0.95 | 0.80 | 0.82 | 0.73 | 0.92 | 0.88 |
| Sad | 0.63 | 0.60 | 0.84 | 0.81 | 0.64 | 0.60 | 0.81 | 0.78 | 0.6 | 0.57 | 0.81 | 0.77 | 0.63 | 0.61 | 0.83 | 0.81 |
| Surprise | 0.61 | 0.64 | 0.84 | 0.86 | 0.53 | 0.63 | 0.75 | 0.83 | 0.57 | 0.66 | 0.80 | 0.81 | 0.51 | 0.63 | 0.77 | 0.86 |
| Fear | 0.52 | 0.54 | 0.78 | 0.79 | 0.54 | 0.57 | 0.80 | 0.82 | 0.56 | 0.58 | 0.75 | 0.76 | 0.56 | 0.66 | 0.79 | 0.86 |
| Disgust | 0.52 | 0.55 | 0.76 | 0.78 | 0.53 | 0.64 | 0.74 | 0.81 | 0.53 | 0.59 | 0.70 | 0.72 | 0.48 | 0.66 | 0.69 | 0.83 |
| Anger | 0.65 | 0.59 | 0.83 | 0.80 | 0.62 | 0.60 | 0.79 | 0.78 | 0.63 | 0.59 | 0.81 | 0.77 | 0.60 | 0.60 | 0.81 | 0.81 |
| Contempt | 0.08 | 0.08 | 0.49 | 0.49 | 0.22 | 0.32 | 0.60 | 0.70 | 0.15 | 0.18 | 0.42 | 0.42 | 0.27 | 0.59 | 0.58 | 0.79 |

*Orig and Norm stand for **O**riginal and skew-**N**ormalized, respectively.

Figure 1: Top 2 F1-score

直接对数据集进行训练，样本较少的表交叉较差，采取了加权损失函数之后，F1 score有所提高。

2.1.2 加权损失函数下Confusion Matrix

由图1可得出，加权损失函下，总体F-score的表现较好，测试集的Confusion Matrix的结果如下：

2.1.3 几种不同的评判标准

以下几种标准均用于二分类模型，此处是将正确分类的结果总结为第1类，错误的结果第二类，在准确率的基础上对Baseline的结果进行评测。

TABLE 8
Confusion Matrix of Weighted-Loss Approach on the Test Set

| | | Predicted | | | | | | | |
|--------|----|-----------|------|------|------|------|------|------|------|
| | | NE | HA | SA | SU | FE | DI | AN | CO |
| Actual | NE | 53.3 | 2.8 | 9.8 | 8.7 | 1.7 | 2.5 | 10.4 | 10.9 |
| | HA | 4.5 | 72.8 | 1.1 | 6.0 | 0.6 | 1.7 | 1.0 | 12.2 |
| | SA | 13.0 | 1.3 | 61.7 | 3.6 | 5.8 | 4.4 | 9.2 | 1.2 |
| | SU | 3.4 | 1.2 | 1.7 | 69.9 | 18.9 | 1.7 | 2.8 | 0.5 |
| | FE | 1.5 | 1.5 | 4.6 | 13.5 | 70.4 | 4.2 | 4.3 | 0.2 |
| | DI | 2.0 | 2.2 | 5.8 | 3.3 | 6.2 | 68.6 | 10.6 | 1.3 |
| | AN | 6.2 | 1.2 | 5.0 | 3.2 | 5.8 | 11.1 | 65.8 | 1.9 |
| | CO | 16.2 | 13.1 | 3.5 | 3.1 | 0.5 | 4.3 | 5.7 | 53.8 |

Figure 2: 测试集 Confusion Matrix, 在加权损失函数下

- F1 Score
- Cohen's Kappa 评判分类准确性, 区间[0,1]
- Krippendorfs Alpha 评判分类准确性, 区间[0,1]
- Area under the ROC curve (AUC) 区间[0,1]低于0.6说明模型效果很差
- Area under the Precision-Recall curve (AUC-PR), 区间[0,1]

TABLE 9
Evaluation Metrics and Comparison of CNN baselines, SVM and MS Cognitive on Categorical Model of Affect.

| | CNN Baselines | | | | | | | | SVM | | MS Cognitive | |
|-----------------------|---------------|------|---------------|------|-------------|------|---------------|------|------|------|--------------|------|
| | Imbalanced | | Down-Sampling | | Up-Sampling | | Weighted-Loss | | Orig | Norm | Orig | Norm |
| | Orig | Norm | Orig | Norm | Orig | Norm | Orig | Norm | | | | |
| Accuracy | 0.72 | 0.54 | 0.68 | 0.58 | 0.68 | 0.57 | 0.64 | 0.63 | 0.60 | 0.37 | 0.68 | 0.48 |
| F ₁ -Score | 0.57 | 0.52 | 0.56 | 0.57 | 0.56 | 0.55 | 0.55 | 0.62 | 0.37 | 0.31 | 0.51 | 0.45 |
| Kappa | 0.53 | 0.46 | 0.51 | 0.51 | 0.52 | 0.49 | 0.5 | 0.57 | 0.32 | 0.25 | 0.46 | 0.40 |
| Alpha | 0.52 | 0.45 | 0.51 | 0.51 | 0.51 | 0.48 | 0.5 | 0.57 | 0.31 | 0.22 | 0.46 | 0.37 |
| AUC | 0.85 | 0.80 | 0.82 | 0.85 | 0.82 | 0.84 | 0.86 | 0.86 | 0.77 | 0.70 | 0.83 | 0.77 |
| AUCPR | 0.56 | 0.55 | 0.54 | 0.57 | 0.55 | 0.56 | 0.58 | 0.64 | 0.39 | 0.37 | 0.52 | 0.50 |

Figure 3: Baseline的不同评判标准

2.2 Dimensional Model Baseline

针对表情的维度模型中, 另一种是基于Arousal(情绪的强烈程度)和Valence(情绪的正负倾向)。

AlexNet作为回归模型, 最后的全连接被替代为仅含有一个神经元的线性回归层。

输出数值的范围在[-1,1]之间。所采取的损失函数为欧式距离。损失函数定义如下:

$$E = \frac{1}{2N} \sum_{n=1}^N \|\hat{y}_n - y_n\|_2^2 \quad (4)$$

在回归模型中, 图片被剪切为256×256。回归模型的相关参数:

- Epoch: 16
- Batch Size: 256
- Learning Rate: 0.001
- Momentum: 0.9

2.2.1 评判标准

- RMSE Root Mean Square Error
- CORR Pearson's Correlation Coefficient
- Concordance Correlation Coefficient (CCC)

具体结果见于图4中。其中同时给了基于支持向量机的回归模型。AlexNet的结果优于SVR。

| | CNN (AlexNet) | | SVR | |
|-------------|---------------|---------|---------|---------|
| | Valence | Arousal | Valence | Arousal |
| RMSE | 0.394 | 0.402 | 0.494 | 0.400 |
| CORR | 0.602 | 0.539 | 0.429 | 0.360 |
| SAGR | 0.728 | 0.670 | 0.619 | 0.748 |
| CCC | 0.541 | 0.450 | 0.340 | 0.199 |

Figure 4: Baseline的不同评判标准

3 ResNet18 Training Result (Using Keras)

基于resnet模型训练结果如下(增加了Batch Normalization层，在每次卷积之后):
一些参数:

- Epoch : 30
- batch size: 16
- Learning Rate: 0.01
- L2 Regularizer: 0.0001 (1e-4)
- Optimizer: Adam

模型在训练集和验证集的结果如下:

测试集的结果:

Loss: 1.67 Accuracy: 71.2%

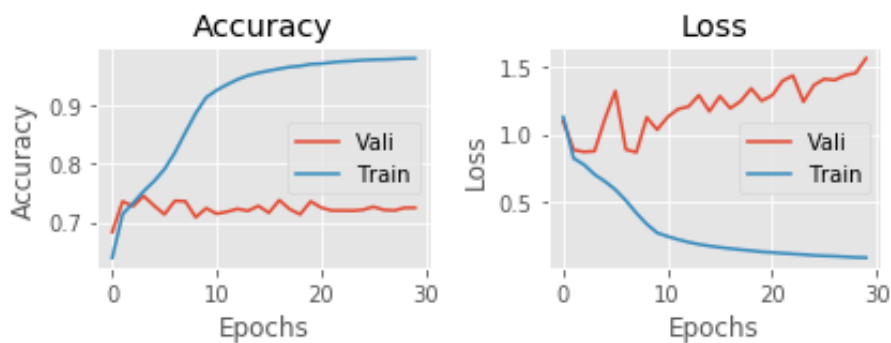


Figure 5: ReseNet34 结果，增加了BN层

4 其他结果

4.1 Paper Result

目前只查到一篇report，作者对8分类和3分类，RGB和灰度图像进行了训练。

三分类是将表情几种分类为Positive,negative和neutral. <http://lup.lub.lu.se/luur/download?func=download>

4.1.1 8分类，验证集

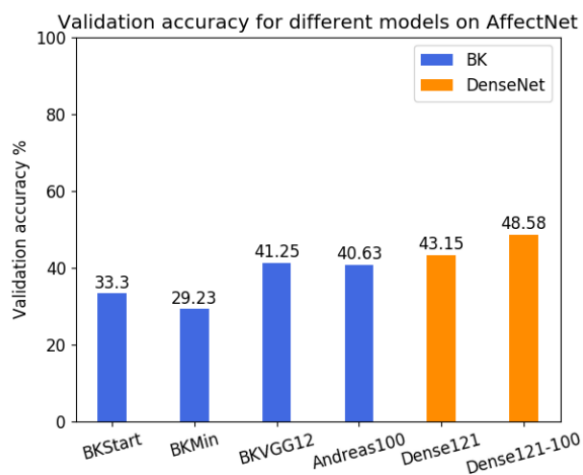


Figure 6: ReseNet34 结果，增加了BN层

4.1.2 8分类和3分类，验证集

模型为BKVGG12，结果为8分类和3分类，灰度和RGB。

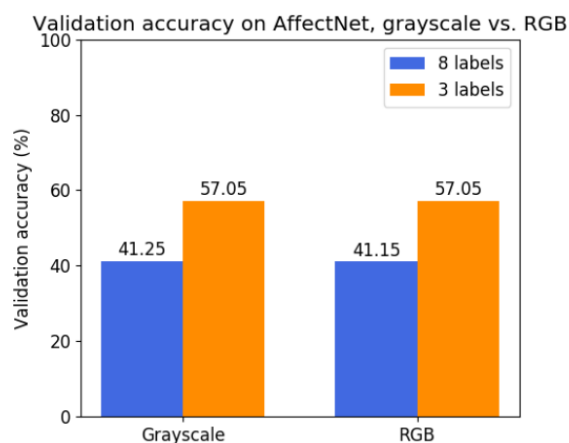


Figure 7: ReseNet34 结果，增加了BN层

4.1.3 Confusion Matrix

Confusion Matrix:

| Prediction \ Ground truth | Neutral | Happiness | Sadness | Surprise | Fear | Disgust | Anger | Contempt |
|---------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| Neutral | 0.734 | 0.15 | 0.05 | 0.02 | 0.012 | 0 | 0.034 | 0 |
| Happiness | 0.052 | 0.944 | 0 | 0.004 | 0 | 0 | 0 | 0 |
| Sadness | 0.384 | 0.094 | 0.412 | 0.014 | 0.028 | 0.002 | 0.066 | 0 |
| Surprise | 0.344 | 0.228 | 0.044 | 0.252 | 0.106 | 0.002 | 0.024 | 0 |
| Fear | 0.25 | 0.078 | 0.104 | 0.1 | 0.362 | 0.014 | 0.092 | 0 |
| Disgust | 0.252 | 0.166 | 0.106 | 0.03 | 0.016 | 0.136 | 0.294 | 0 |
| Anger | 0.356 | 0.068 | 0.062 | 0.016 | 0.024 | 0.016 | 0.458 | 0 |
| Contempt | 0.284 | 0.636 | 0.02 | 0.002 | 0. | 0.008 | 0.048 | 0.002 |

Figure 8: 3/8分类，灰度和RGB

4.2 中文网络

查到一篇博客，使用AlexNet，训练集的准确度为70%,验证集38%.

resource: <https://blog.csdn.net/ZWX2445205419/article/details/79086288>

5 总结

从不同的训练结果来看，Happiness和Neutral模型识别的能力比较高。

Disgust, Contempt和Disgust三种表情分类识别能力较弱。