Facial Expression Recognition

Chenqi Liu & Zhenghuan Sun & Junhao Yang
2018533249 & 2018533202 & 2018533079
shanghaitech

Abstract

Facial expression plays an important role in our daily life. In order to get a high-accuracy classifier, we try to use six different algorithms to train databases. Then this classifier can be used to divide all facial expression into "positive", "negative" and "neutral". Some explanations are proposed to interpret the experimental results of models.

1 Introduction

Facial expression is one or more motions or positions of the muscles beneath the skin of the face. According to one set of controversial theories, these movements convey the emotional state of an individual to observers.Thus facial expressions are a form of nonverbal communication which are a primary means of conveying social information between humans. Numerous studies have been conducted on automatic facial expression analysis because of its practical importance in many fields, such as social security. There are too much information in facial expression, we have different research targets to study these information.[1] What is most popular is to classify facial expression into different groups according to the meaning behind face. There are also many existing problems in this filed :

1. Compared with the target detection, the data of the task such as face recognition is insufficient and easy to be over fitted;

2. It is greatly affected by external conditions such as illumination;

3. Identity information which shows personal characters in the same class;

4. Compared with face recognition, expression is more abstract and subtle.

In this paper, we focus on the classification of the facial expression. We try to divide facial expression into three groups "negative", "positive"and "neutral" according to the meaning of these facial expression. Our target is to get a efficient classifier which can classify the new picture into one of three groups correctly.

2 Preparations

2.1 Databases

Having sufficient labeled training data that include as many variations of the populations and environments as possible is important for the design of a deep expression recognition system. After comparing different data we get from Internet, we choose to use data from "Kaggle" (one website focusing on data modeling and data analysis competition).

`https://www.kaggle.com/mahmoudima/mma-facial-expression`

It contains tree directories for training, validation, and test.Each directory contains seven subdirectories corresponding to seven facial expression categories.They are respectively "angry", "disgust", "fear", "happy", "neutral", "sad"and "surprise".

2.2 Pre-processing

Firstly, to divide pictures into three groups, we need to use a reflection to map seven labels into three labels so that we can train data in three groups.Thus, we regard "disgust", "angry", "fear", "sad" as "negative" and regard "surprise", "happy" as "positive". Also, we change our data from RGB colors to black-white so that we can speed up our algorithm.

Next, we apply standardization on our datasets to make data acceptable to some machine learning algorithm and much easier to be handled by.

```
sle = StandardScaler()
train_x_le = sle.fit_transform(train_x)
val_x_le = sle.transform(val_x)
test_x_le = sle.transform(test_x)
```

Finally, noting that variations that are irrelevant to facial expressions, such as different backgrounds, illuminations and head poses, are fairly common in unconstrained scenarios. Also, our pictures are 60×60 pixies which leads to 3600 dimensions. It's complex and time-wasting to train all data in this way. So to lower the influence of noises in picture and decrease the difficulty of training the data, we apply PCA (Principal components analysis) into data:

```
from sklearn.decomposition import PCA
pca = PCA(n_components=20)
train_x_pca = pca.fit_transform(train_x)
val_x_pca = pca.transform(val_x)
test_x_pca = pca.transform(test_x)
```

3 Methodology And Experiment

3.1 Logistic Regression

We firstly consider some simply models to train data. Logistic regression is a statistical model that in its basic form using a logistic function to model a binary dependent variable, although many more complex extensions exist. Here we use one of its extensions to train our databases:

$$\beta^{t+1} \leftarrow \beta^{t} + \alpha \frac{\partial L(\beta)}{\partial \beta} = \beta^{t} + \alpha \sum_{i=1}^{b} (p_i - y_i)\mathbf{x}_i \qquad [2]$$

Our experimental results of training are shown in below pictures. Through confusion matrix and accuracy bar graph, we can get that "negative" and "neutral" pictures are easily wrongly classified to "positive" group and "positive" pictures are classified almost successfully.

3.2 Bagging

In bagging, we randomly select n(put back after take out, n < N) samples from training data to train multiple classifiers, and the weights of each classifier are consistent.Noting that we use logistic regression as the basic classifier instead of random classifier. Finally, the final result is obtained by voting. This sampling method is called bootstrap aggregating in Statistics which is also called bagging.

Our experimental results of training are shown in below pictures. Through confusion matrix and accuracy bar graph, we can get that "negative" and "neutral" pictures are easily wrongly classified to "positive" group and "positive" pictures are classified almost successfully. It's obviously the same with Logistic Regression. The reason maybe we use logistic regression as the basic classifier to training bagging which shows bagging itself is not good at facial recognition.

3.3 Random Forest

Using Random Forest to eliminate the effects of imbalanced datasets are proved to be effective, so we consider using Random Forest.[3] Random Forest differs in only one way from bagging is : they use a modified tree learning algorithm that selects, at each candidate split in the learning process, a random subset of the features. This process is sometimes called "feature bagging". The reason for doing this is the correlation of the trees in an ordinary bootstrap sample: if one or a few features are very strong predictors for the response variable (target output), these features will be selected in many of the B trees, causing them to become correlated.

Our experimental results of training are shown in supplementary materials.Through confusion matrix and accuracy bar graph, we can get that "negative" and "neutral" pictures are easily wrongly classified to "positive" group and "positive" pictures are classified almost successfully.

3.4 Adaboost

Adaboost algorithm is an improved boosting classification algorithm. The method is to increase the weight of the error samples which are linearly combined by the first several classifiers, so that each time a new classifier is trained, it can focus on the training samples which are easy to cause classification errora. Each weak classifier uses the weighted voting mechanism instead of the average voting mechanism. The weak classifier with higher accuracy has larger weight, and the weak classifier with low accuracy has lower weight.

Our experimental results of training are shown in supplementary materials.Through confusion matrix and accuracy bar graph, we can get that "negative" and "neutral" pictures are easily wrongly classified to "positive" group and "positive" pictures are classified almost successfully. It's almost the same with Logistic Regression.

3.5 Gradient Boost

Gradient Boost differs in only one way from Adaboost is: how the two algorithms identify the shortcomings of weak learners (eg. decision trees). While the Adboost model identifies the shortcomings by using high weight data points, Gradient Boost performs the same by using gradients in the loss function. The loss function is a measure indicating how good are model's coefficients at fitting the underlying data.

Our experimental results of training are shown in supplementary materials.Through confusion matrix and accuracy bar graph, we can get that "negative" and "neutral" pictures are easily wrongly classified to "positive" group and "positive" pictures are classified almost successfully. But comparing to all algorithms before, Gradient Boost has better performance in "negative" picture recognition, but work worse in "positive" and "neutral".

3.6 Resnet[4]

A residual neural network (Resnet) is an artificial neural network (ANN) of a kind that builds on constructs known from pyramidal cells in the cerebral cortex. Residual neural networks do this by utilizing skip connections, or shortcuts to jump over some layers. Typical Resnet models are implemented with double- or triple-layer skips that contains nonlinearities (ReLU) and batch normalization in between. To balance the efficiency and effect of algorithm, we won't chose a very complex model. Here our epoch is 5, learning rate is 0.001 and depth is 3.

Our experimental results of training are showed in supplementary materials. Through confusion matrix and accuracy bar graph, we can get that Resnet performance best in all six algorithm. But in "positive" picture recognition, it will still wrongly classify some pictures into "neutral".

4 Conclusion

We want to divide facial expression into three groups "negative", "positive"and "neutral" according to the meaning of these facial expression. To realize this target, we choose 6 different model training methods to train data so that we can quickly and accurately classify a new picture into one of three models.

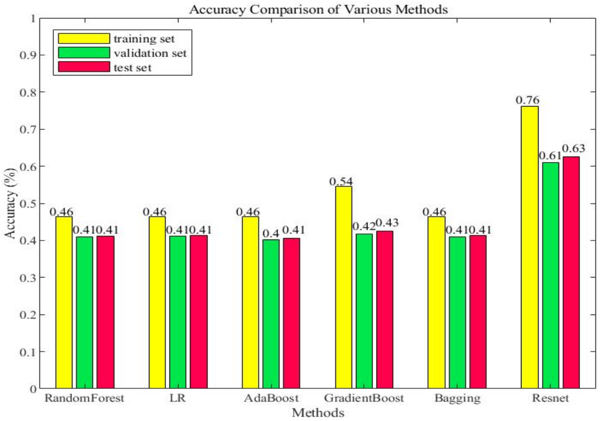

Figure 1: All methods accuracy.

Most of the datasets have a better performance in terms of accuracy on proposed residual network than that on the benchmark algorithm.[5] Noting our accuracy is not very high, only one of them is more than 50 %. After discussions and information searching, our group think the reasons are as followed.

1. Our resource pictures don't contain enough information because of low pixel. Because of the low pixel of the pictures, some details of the picture are indistinct, which leads to fails in captures of some characters by models.

2. Our resource pictures are shot at the same direction which means we can only get information at only one direction. So our model can only classify pictures at one direction instead of comparing different directions.

3. To decrease the difficulty of training the data, we choose PCA to reduce dimensions which may leads to information lost in some dimensions.

There are still many points that we can do to improve the performance of machine learning algorithms. Firstly, we can choose more vivid resource pictures. Also, noting that because people in different age ranges, cultures and genders display and interpret facial expression in different ways, an ideal facial expression datasets is expected to include abundant sample images with precise face attribute labels. So in the future we should choose clearly-classified datasets which contain pictures of people containing different cultures, genders and so on. In addition, we can increase the depth of Resnet to have a better performance in model training.

References

[1] Shan Li & Weihong Deng (2018) Deep Facial Expression Recognition: A Survey.

[2] Tommy Huang, "logistic regression", "https://medium.com/@chih.sheng.huang821".

[3] Hanwu Luo & Xiubao Pan (2019)Logistic Regression and Random Forest for Effective Imbalanced Classification .

[4] Kaiming He & Xiangyu Zhang & Shaoqing Ren & Jian Sun (2015) Deep Residual Learning for Image Recognition.

[5] Xingcheng Luo & Ruihan Shen & Jian Hu (2017) The performance of proposed deep residual learning network of images .