South China University of Technology

# The Experiment Report of Machine Learning

## SCHOOL: SCHOOL OF SOFTWARE ENGINEERING

## SUBJECT: SOFTWARE ENGINEERING

Author:Zou Shuai  and Song Dingkun and Hong Shaojia

Supervisor:
Qingyao Wu

Student ID：201720145167 and 201720145006 and 201720144993

Grade:
PostGraduate

December 19, 2017

# Face Classification Based on AdaBoost Algorithm

**Abstract—The experiment will show you an attempt to use Adaboost algorithm to integrate multiple weak classifiers for face classification. We can know the basic progress of face classification and learn to use AdaBoost to solve the face classification problem.**

## I. INTRODUCTION

The experiments are based on python3 and use the following packages which include sklearn, numpy, jupyter and matplotlib, pickle, PIL and cv2. In the progress of integration, we choose the decision tree as the weak classifier. We set the max depth of the decision tree to 1 in order to make it as a weak classifier. We will train a strong classifier using AdaBoost algorithm on the training set and use it to validate on our validate set.

## II. METHODS AND THEORY

We read the dataset and convert them into a size of 24*24 grayscale. The features of the faces are extracted by using NPD. We first train a base classifier $G(x)$ according to the weight $w$ that we set. Then we calculate the classification error rate $err$ of the base classifier on the training set

$$err = P(G(x_i) \mathbin{!=} y_i) = \sum_1^n w_i * I(G(x_i) \mathbin{!=} y_i)$$

and get the weight $\alpha$ of this base classifier.

$$\alpha = \frac{1}{2}\log\left[(1 - err) / err\right]$$

Besides, we should update the weight $w$ of training set and repeat the above steps.

$$w_i = \frac{w_i}{z} * \exp(-\alpha * y_i * G(x_i))$$

where z is a normalize parameter:

$$z = \sum_1^n w_i * \exp(-\alpha * y_i * G(x_i))$$

The number of iterations is based on the number of classifiers. In each iteration, we should check whether the integrated classifier is good enough to do classification on the training set. If it is good enough and we should stop the iteration and use the final integrated classifier $f(x)$

$$f(x) = sign(\alpha_1 G_1(x) + \alpha_2 G_2(x) + \cdots + \alpha_i G_i(x))$$

where i stands for the number of base classifier on our validate set.

## III. EXPERIMENT

### A. Dataset

The experiment provides 1000 pictures of which 500 are human face RGB images and the other 500 is a non-face RGB images. We convert all of them into a size of 24*24 grayscale and then divide them into training set and validate set. The training set has the first 250 of 500 human face images and the first 250 of 500 non-face images. The validate set get the rest of the pictures. So the size of training set and the validate set are both 500 and the proportion of the positive and negative samples are both 1:1. We get the corresponding label y according to the dataset.

### B. Implementation

We first read our training set and use the class NPDFeature in feature.py to get the features of the images. We save the data in the cache using pickle function library dump() and use load() function to get the cache as the time of the pretreatment is relatively long. After that, we use fit() in the ensemble.py to train an integrated classifier. In the end , we can get precision,recall,f1-score,support as a report.txt after using the integrated classifier on the validate set. The result is shown below:

```
              precision    recall  f1-score   support

       -1.0       0.96      0.92      0.94       250
        1.0       0.92      0.96      0.94       250

avg / total       0.94      0.94      0.94       500
```

## IV. CONCLUSION

The experiment is designed to let us see the power of AdaBoost Algorithm that it can turn a list of weak classifiers to a strong classifier. Besides, we can get familiar with the basic method of face detection. The tip that we can store the data pretreated in the cache is beneficial and we can use the method in our future experiments.