



华南理工大学

South China University of Technology

The Experiment Report of Machine Learning

SCHOOL: SCHOOL OF SOFTWARE ENGINEERING

SUBJECT: SOFTWARE ENGINEERING

Author:

Zhaotao Lin and Yuhong Zhang

Supervisor:

Qingyao Wu

Student ID

201530612309 and 201530613665

Grade:

Undergraduate

December 21, 2017

FACE CLASSIFICATION BASED ON ADABOOST ALGORITHM

Abstract— This report is mainly to illustrate the experiment of face classification based on Adaboost algorithm.

I. INTRODUCTION

The Adaboost algorithm is a kind of method for classification. Its basic principle is to combine some weak classification methods and combine new and strong classification methods. The purpose of face detection is to find all the child Windows that contain faces, and separate the child window of the face from the child window of the non-human face. In the experiment, we understand Adaboost further, get familiar with the basic method of face detection. Then, we learn to use Adaboost to solve the face classification problem, and combine the theory with the actual project. Generally, we experience the complete process of machine learning.

II. METHODS AND THEORY

A. Adaboost algorithm

The entire Adaboost iteration algorithm is 3 steps: The weight distribution of the initial training data. If there are N samples, each of the training samples will initially be given the same weight: $1/N$. Train the weak classifier. In the specific training process, if a sample point has been accurately classified, then the weight of the next training set will be reduced. Conversely, if a sample point is not classified correctly, its weight is improved. The updated sample set is then used to train the next classifier, and the whole training process goes on so iteratively. Combine the weak classifier of various training into strong classifier.

B. Decision tree algorithm

Decision tree algorithm is a method to approximate discrete function values. It is a typical classification method. Firstly, the data is processed and the inductive algorithm is used to generate readable rules and decision trees, and then the new data is analyzed using the decision. Essentially, a decision tree is a process of classifying data through a series of rules. Decision tree structure can be divided into two steps. The first step is the generation of the decision tree: the process of generating the decision tree from the training sample set. In general, the training sample data set is based on the actual needs of historical and some degree of comprehensive data for data analysis and processing of data sets. The second step, the decision tree cutting technology: the decision tree pruning is decision tree of on one phase generated test, calibration and repair process, main is to use the new sample data in a data set (referred to as test data sets) check occurring in

the process of decision tree to generate preliminary rules, will that affect the prediction accuracy of equilibrium branch off.

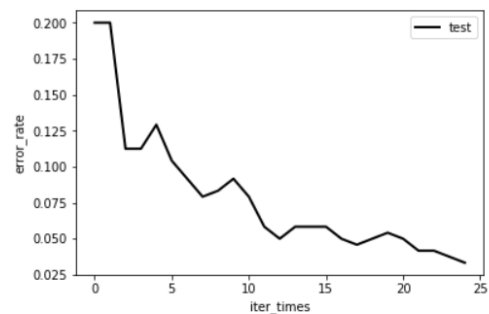
III. EXPERIMENT

A. Dataset

This experiment provides 1000 pictures, of which 500 are human face RGB images, the other 500 is a non-face RGB images.

B. Implement

- Read data set data. The images are supposed to be converted into a size of $24 * 24$ grayscale, the number and the proportion of the positive and negative samples is not limited, the data set label is not limited.
- Processing data set data to extract NPD features. Extract features using the NPDFeature class in feature.py.
- The data set is divided into training set and validation set, this experiment does not divide the test set.
- Write all AdaboostClassifier functions based on the reserved interface in ensemble.py.
- Predict and verify the accuracy on the validation set using the method in AdaboostClassifier and use classification_report () of the sklearn.metrics library function writes predicted result to report.txt .
- Organize the experiment results and complete the lab report (the lab report template will be included in the example repository).



	precision	recall	f1-score	support
nonface	0.97	0.97	0.97	120
face	0.97	0.97	0.97	120
avg / total	0.97	0.97	0.97	240

IV. CONCLUSION

In this experiment, we conducted face detection through Adaboost algorithm, which was more difficult than last time. During the experiment, the biggest difficulty was the preparation of fit functions, the adjustment of various parameters, and many small details and precautions. It was difficult to start the experiment. After discussing and borrowing their code from other groups, we finally completed the experiment.