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:
Kaile Pan, Yongcheng Yao and
Yunxiao Zhao

Student ID：**201530612583**
201530613450
201530613757

Supervisor:
Qingyao Wu

Grade:
Undergraduate

December 21, 2017

Face Classification Based on AdaBoost Algorithm

ABSTRACT—THIS IS AN REPORT ON THE THIRD EXPERIMENT OF MACHINE LEARNING CLASS——FACE CLASSIFICATION BASED ON ADABOOST ALGORITHM

I. INTRODUCTION

Motivation of Experiment

1. Understand Adaboost further
2. Get familiar with the basic method of face detection
3. Learn to use Adaboost to solve the face classification problem,and combine the theory with the actual project
4. Experience the complete process of machine learning

Environment for Experiment

python3, at least including following python package: sklearn, numpy, matplotlib, pickle, PIL.

Time and Place

2017-12-09 9:00-12:00 AM B7-138/238

II. METHODS AND THEORY

Input: $D = \{(\mathbf{x}_1, y_1), ..., (\mathbf{x}_n, y_n)\}$, **where** $\mathbf{x}_i \in X, y_i \in \{-1, 1\}$
Initialize: Sample weight distribution $D_1 = \frac{1}{n}$
Train a base learner $h_1(\mathbf{x})$ with D_1
for m=2,3,...,M **do**
  Update the sample distribution D_m, to make the wrong predictive samples more important
  Train a new base learner $h_m(\mathbf{x})$ with D_m
end
Output: $H(\mathbf{x}) = \sum_{m=1}^{M} \alpha_m h_m(\mathbf{x})$

$$\epsilon_m = p(h_m(\mathbf{x}_i) \neq y_i) = \sum_{i=1}^{n} w_m(i) \mathbb{I}(h_m(\mathbf{x}_i) \neq y_i)$$

$$\alpha_m = \frac{1}{2} \log \frac{1 - \epsilon_m}{\epsilon_m}$$

$$z_m = \sum_{i=1}^{n} w_m(i) e^{-\alpha_m y_i h_m(\mathbf{x}_i)}$$

$$w_{m+1}(i) = \frac{w_m(i)}{z_m} e^{-\alpha_m y_i h_m(\mathbf{x}_i)}$$

III. EXPERIMENT

Dataset

This experiment provides 1000 pictures, of which 500 are human face RGB images

Implementation

max_depth=3
n_weakers_limit=10

Step

1. Read data set data. The images are supposed to 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.
2. Processing data set data to extract NPD features. Extract features using the NPDFeature class in feature.py. (Tip: Because the time of the pretreatment is relatively long, it can be pretreated with pickle function library dump () save the data in the cache, then may be used load () function reads the characteristic data from cache.)

3. The data set is divisded into training set and calidation set, this experiment does not divide the test set.
4. Write all AdaboostClassifier functions based on the reserved interface in ensemble.py. The following is the guide of fit function in the AdaboostClassifier class:

4.1 Initialize training set weights w, each training sample is given the same weight.

4.2 Training a base classifier , which can be sklearn.tree library DecisionTreeClassifier (note that the training time you need to pass the weight w as a parameter).

4.3 Calculate the classification error rate e of the base classifier on the training set.

4.4 Calculate the parameter α according to the classification error rate e.

4.5 Update training set weights w.

4.6 Repeat steps 4.2-4.6 above for iteration, the number of iterations is based on the number of classifiers.

5. 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 .

Result

Calculating the NPD table...
hit rate: 0.78

	precision	recall	f1-score	support
-1.0	0.79	0.77	0.78	100
1.0	0.77	0.79	0.78	100
avg / total	0.78	0.78	0.78	200

IV. CONCLUSION

In this experiment, we learned how to use AdaBoost to help computer to recognize a face. With the help of teacher assistant, we can manage it, though there are much space to improve.