華南理工大學

South China University of Technology

The Experiment Report of Machine Learning

SCHOOL: SCHOOL OF SOFTWARE ENGINEERING

SUBJECT: SOFTWARE ENGINEERING

Author:
Juncong Huang, Yunzhe Guo and Shumin Yan

Supervisor:
Mingkui Tan

Student ID：
201530611746, 201530611524 and 201530613337

Grade:
Undergraduate

December 19, 2017

Face Classification Based on AdaBoost Algorithm

Abstract— The purpose of face classification is to identify a picture includes a human face or not. Adaboost algorithm is widely used in face classification. Adaboost is an iterative algorithm that train different classifiers for the same training set and then assemble these weakly classifiers to form a stronger classifier. In this experiment, we will use Adaboost algorithm to realize face classification.

I. INTRODUCTION

The motivations of the experiment are as follows. Firstly, understand Adaboost further. Secondly, get familiar with the basic method of face detection.Thirdly, learn to use Adaboost to solve the face classification problem, and combine the theory with the actual project. Fourthly, experience the complete process of machine learning.

II. METHODS AND THEORY

In this experiment, we use Adaboost algorithm to realize face classification. Adaboost is an iterative algorithm that train different classifiers for the same training set and then assemble these weakly classifiers to form a stronger classifier.

The Adaboost algorithm is as follow.

Input: D $= \{(x_1, y_1), \dots (x_n, y_n)\}$,where $x_i \in X, y_i \in \{-1,1\}$
Initialize: Sample distribution w_m
Base learner: L
Process:

$w_i(i) = \frac{1}{n}$
For m=1,2,…M do
$h_m(x) = L(D, w_m)$
$\epsilon_m = \sum_{i=1}^{n} w_m(i) \prod(h_m(x_i) \neq y_i)$
If $\epsilon_m > 0.5$ then
Break
End
$a_m = \frac{1}{2} log \frac{1-\epsilon_m}{\epsilon_m}$
$w_{m+1}(i) = \frac{w_m(i)}{z_m} e^{-a_m y_i h_m(X_i)}$, where i=1,2,…,n
and $z_m = \sum_{i=1}^{n} w_m(i) e^{-a_m y_i h_m(x_i)}$
End
Output:H(x) $= \sum_{m=1}^{M} a_m h_m(x)$

III. EXPERIMENT

A. Dataset

The dataset includes 1000 pictures, of which 500 are human face RGB images, the other 500 are non-face RGB images. The trainning set has 800 samples and the validation set has 200 samples.

B. Implementation

Experiment Step

1. Read data set data. Convert the images into a size of 24 * 24 grayscale. The data that is a human face RGB image is labeled "**+1**" and the data that is a non-face RGB image is labeled "**-1**".

2. Processing data set data to extract NPD features. Extract features using the NPDFeature class in feature.py. Because the time of the pretreatment is relatively long, we pretreat it with pickle function library dump () save the data in the cache, then use load () function reads the characteristic data from cache.

3. The data set is divisded into training set and validation set. In the experiment, the trainning set has **800** samples and the validation set has **200** samples.

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 ω, each training sample is given the same weight. In this experiment, we set weights as $\frac{1}{n}$ (n is the number of training set.)

4.2Training a base classifier. In the experiment we use sklearn.tree library DecisionTreeClassifier. The training time we need to pass the weight ω as a parameter.

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

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

4.5 Update training set weights ω.

4.6 Repeat steps 4.2-4.6 above for iteration, the number of iterations is based on the number of classifiers. In the experiment the number of classifiers is **10.**

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

	precision	recall	f1-score	support
face	0.90	0.79	0.84	104
nonface	0.80	0.91	0.85	96
avg / total	0.85	0.84	0.84	200

Fig. 1. Results in report.txt

IV. CONCLUSION

Through this experiment, we understand the principle of Adaboost algorithm futher and know how to implement it. We are familiar with the basic method of face detection and use Adaboost to solve the face classification problem. What's more, we combine the theory with the actual project and experience the complete process of machine learning.