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:
Riyang Hu
Xianzhe Wu
Canguang Li

Supervisor:
Mingkui Tan

Student ID:
201530611616
201530613078
201530611937

Grade:
Undergraduate

December 21, 2017

Face Classification based on AdaBoost algorithm

***Abstract*—In this experiment we use AdaBoost algorithm and decision tree to realizing Face Classification, which check the given photo with a face in it or no, and try to imporve its accuracy.**

I. Introduction

FACE Recognization is a hotpoint in AI nowadays. We use AdaBoost algorithm with decision trees as weak learners to realizing our goal **Face-Classification** here. With enough samples, we think there should be more than 0.9 of the accuracy.

II. Methods and Theory

AdaBoost, short for Adaptive Boosting, is a machine learning meta-algorithm formulated by Yoav Freund and Robert Schapire. It can be used in conjunction with many other types of learning algorithms to improve performance, like desicion tree. The output of the other learning algorithms('weak learners') is combined into a weighted sum that represents the final output of the boost classifier.

AdaBoost with decision trees as weak learners is often referred to as the best out-of-the-box classifier. When used with decision tree learning, information gathered at each stage of the AdaBoost algorithm about the relative 'hardness' of each training sample is fed into the tree growing algorithm such that later trees tend to foucus on harder-to-classify examples.

$Algorithm's\ process$:
$\boldsymbol{Given}$:$(\boldsymbol{x}_1, y_1), ..., (\boldsymbol{x}_m, y_m)$ where $\boldsymbol{x}_i \in X, y_i \in Y$.
$\boldsymbol{Initialize}$ $D_1(i) = 1/m$.
$\boldsymbol{for}$ $t = 1, ..., T$:

- Train decision tree using distribution D_t.
- Get weak hypothesis h_t: $X \to \{-1, 1\}$ with error

$$\epsilon_t = \sum [h_t(\boldsymbol{x}_i) \neq y_i] \tag{1}$$

- Choose $\alpha_t = \frac{1}{2} ln(\frac{1-\epsilon_t}{\epsilon_t})$
- Update:

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times \begin{cases} e^{-\alpha_t} & if\ h_t(\boldsymbol{x}_i) = y_i \\ e^{\alpha_t} & if\ h_t(\boldsymbol{x}_i) \neq y_i \end{cases} \tag{2}$$

where Z_t is a normalization factor (chosen so that D_{t+1} will be a distribution).

$\boldsymbol{Output\ the\ final\ hypothesis}$:

$$H(x) = sign(\sum_{t=1}^{T} \alpha_t h_t(\boldsymbol{x})) \tag{3}$$

III. Experiments

A. Dataset

This experiment uses 250 pictures, of which 125 are human face RGB images, the other 125 is a non-face RGB images.We spilt them into training set and validation set according to the proportion of 8:2.

B. Implementation

- 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.
- 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 , 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 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.
- 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.

TABLE I
Simulation Parameters

number of base learners	5
max_depth of decision tree	3

TABLE II
Experiment results

	precision	recall	f1-score	support
non-face	0.95	0.93	0.94	150
face	0.93	0.95	0.94	150
avg/total	0.94	0.94	0.94	300

IV. Conclusion

In this experiment , we firstly established the structure that combined Adaboost algorithm with decision tree . Then , we adjusted the parameters of the structure to make the precision higher . The number of base classifiers and the maximum depth of decision tree were main factors which influenced the precision of learning process . Besides , the split between training set and test set also mattered . At last , we got the proper parameters and reached the highest precision of 0.95 .