The Experiment Report of Machine Learning

SCHOOL: SCHOOL OF SOFTWARE ENGINEERING

SUBJECT: SOFTWARE ENGINEERING

Author:
王煜、夏俊煊、颜常霖

Supervisor:
Qingyao Wu

Student ID：201530612927、
201530613139 and201530613368

Grade:
Undergraduate

December 22, 2017

Face Classification Based on AdaBoost Algorithm

Abstract—This experiment wants us to further understand AdaBoost and get familiar with the basic method of face detetion.

I. Introduction

AdaBoost is an iterative algorithm. The core idea of AdaBoost is to train different classifiers, ie weak classifiers, against the same training set, and then combine these weak classifiers to construct a stronger final classifier.
The algorithm itself is to change the distribution of data to determine the weight of each sample based on the correct classification of each sample in each training set and the accuracy of the last overall classification. The new data with modified weights are sent to the lower classifier for training, and then the classifiers obtained by each training are fused together to be the final decision classifier.

II. Methods and Theory

This alogrithm's input is
D={(x1,y1),(x2,y2),…,(xn,yn)},where xi ∈X,yi∈{-1,1}
First we train them in an average weight(if the number of the sample is N,then the initial weight is 1/N) to train a base learner.Then we will make the wrong predictive samples more important, that is to enlarge its weight in the next iteration and handle it in next round.After each iteration,this algorithm generates a new base learner hm(x) and its importance score α m. Weak classifier with low error rate has more weight in the final classifier

III. Experiment

A. Dataset

Thos experiment provides 1000 pictures,of which 500 are human face RGB images, stored in the ./datasets/original/face, the other 500 is non_face RGB images, stored in ./datasets/original/nonface.

B. Implementation

Initialize the weight of every sample,and we set weak classifer `turn=10#弱分类器` ,and set the AecisionTreeClassifier's max_depth to 3 `each=DecisionTreeClassifier(max_depth=3)` ,we calculate the fault and update the weight of the wrong sample.the details can be seen in our code.

C. Method

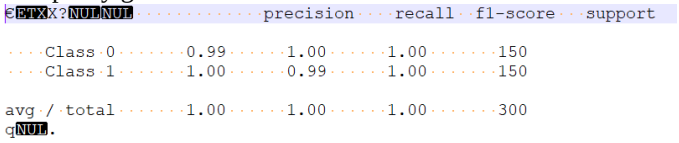
We read the pictures and processing data set data to extract NPD features,it is achieved in feature.py. then,we divide the data set into training set and calidation set.then,we use the AdaBoost algorithm to train these pictures.
(1)initialize training set weght,each sample is given the same weight.
(2)train a base classifier,which can be sklearn.tree liabraryDecisionTreeClassifier.
(3)calculate the classification error rate of the base classifier on the train set.
(4)calculate the parameter α according to the classification error rate.
(5)update training set weight .
(6)repeat the iteration.
Then ,we predict and verify the accuracy on the validation set and output the result in report.txt.
By change our max_length and other parameters,our accuracy rises from 65 percent to 99 even 100 percent!

D. Comparision

When we refer to AdaBoost, we usually think of another algorithm to solve this problem: GBDT.
AdaBoost uses misdirected data points to identify problems and adjusts the model by adjusting the weight of the misclassified data points. GDBT identifies problems by negative gradients and improves the model by calculating negative gradients.

E. Result

Finally ,the accuracy reaches 0.99 to 1.00.We think it is pretty good.



IV. Conclusion

After this experiment ,we are aware of the theory of AdaBoost algorithm. Instead of resampling, each iteration of AdaBoost changes the distribution of the sample, The sample

distribution changes depending on whether the sample is correctly classified:Always classify the correct sample with low weight,Samples that are always misclassified have high weights .
The end result is a weighted combination of weak classifiers, Weights represent the performance of the weak classifier.

In our code,we use some functions in the numpy,PIL and so on rather than using multiple for loops.When we use for loops,the speed is so slower that running is very time-wasting.We set max_depth to 3 and 10 weak classifer ,and after the adjustment to those function and parameters,we get a high accuracy 。 So we can say , AdaBoost is suitable to solve face classification problem,and it hardly overfitting.