The Experiment Report of Machine Learning

SCHOOL: SCHOOL OF SOFTWARE ENGINEERING

SUBJECT: SOFTWARE ENGINEERING

Author:
ZhaoRong Likaizhe and Hongzijie

Supervisor:
Qingyao Wu

Student ID：201530613740
201530611982
201530611593

Grade:
Undergraduate

December15, 2017

I. Face Classification Based on AdaBoost Algorithm

Abstract—using Adaboost further to build a Neural Network solve the face classification problem

II. Introduction

As the most famous algorithm on behalf of the boosting family, adaboost algorithm deals with the wrong samples just like the gbdt method, but differently, it simply makes the wrong samples with a higher weight, then throws the old samples with a new weight into a new classifier, after several iteration ends, it uses the accuracy as each classifier's weight, then add them, after a sign function, the result turns to be the final classified answer.

III. Methods and Theory

The algorithm takes as input a training set $(x_1; y_1); \cdots; (x_m; y_m)$ where each x_i belongs to some *domain* or *instance space* X, and each *label* y_i is in some label set Y . to maintain a distribution or set of weights over the training set.
The most basic theoretical property of AdaBoost concerns its ability to reduce the training error.prove that the training error (the fraction of mistakes on the training set) of the final hypothesis H is at most

IV. Experiment

A data set

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

B implementation

First initialize a Adaboost classifier:Build a boosted classifier from the training set (X, y).saving the weak classifier alpha.final estimation classification matrix and fit every weak Classifier and add its prediction to the list calculcate the error rate.Calculate the total error of the training set and store it in the weak classifier,

C.**method**:

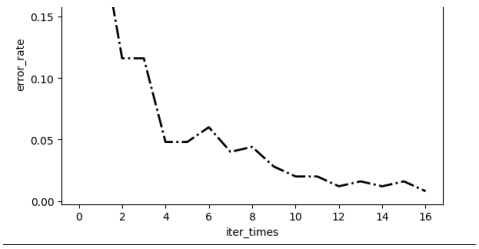
Converting the images into a size of 24 * 24 grayscale and extract NPD features.Then using a base classifier and adjust the weight and do iterations to Predict and verify the accuracy on the validation set using the method in Adaboostclassifier. To make the weak classifier really to be weak, we define the max depth of the CART, the tree used as the basic classification tree in sklearn package, to be 1, to avoid the overfitting. Since the train error becomes bigger than we use 3 or 4 even default as the max depth, it shows a better result than classifier with a higher depth on test set. In contrast, we use a simple classifier with no depth limit has a 65 percent accuracy on the test set while its accuracy on the train set is 100. It means that adaboost method do have great effect to handle with overfitting problem.

D. comparision:

To make the weak classifier really to be weak, we define the max depth of the CART, the tree used as the basic classification tree in sklearn package, to be 1, to avoid the overfitting. Since the train error becomes bigger than we use 3 or 4 even default as the max depth, it shows a better result than classifier with a higher depth on test set. In contrast, we use a simple classifier with no depth limit has a 65 percent accuracy on the test set while its accuracy on the train set is 100. It means that adaboost method do have great effect to handle with overfitting problem.

E.result:

	precision	recall	f1-score	support
nonface	0.99	0.99	0.99	118
face	0.99	0.99	0.99	132
avg / total	0.99	0.99	0.99	250



V. Conclusion

One mistake I made is that I make the label to be 0 and 1, rather than -1 and 1, it leads to a fatal error that every weight will change just because its label is zero.
Since that the adaboost method uses CART as it basic classifier, it's easy to notice that another useful boost method gbdt, use CART as its basic classifier too. Not like adaboost, gbdt tends to minimize its loss, and use negative gradient to fit the loss, each weak classifier is used to minimize the last classifier's loss, then use the weak classifier to update the strong classifier. Adjust the parameter of gbdt is not that easy, so we just use the default parameters. The accuracy is 0.94 to test set. The result is as good as adaboost, but works slower because that we didn't set parameters.

Adaboost algorithem can really worked on Face Classification.With the times of the iteractions increased using the right γ the error rate will reduce a lot.And it is simple fast to program and has no parameter to tune.