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:
SunXingxing

Supervisor:
Qingyao Wu

Student ID：
201721045350

Grade:
Postgraduate

December 22, 2017

Face Classification Based on AdaBoost Algorithm

Abstract Boosting is an approach to machine learning based on the idea of create a highly accuarate prediction method by combining many relatively weak classifiers.though one basic classifier do not well in classifying the dataset, but the final result come to a pleasent result. The Adaboost algorithm was the first parctical boosting algorith, and remains one of the most widely used and studied with applications in various fields.In experiment three, we use the AdaBoost algorithm for face recognition.Experimental result on a large face database of 600 faces of 300 individuals show the feasibility of this method for fast face recognition.

Keywords: Face recognition, Adaboost, Decesion Tree

I. INTRODUCTION

Face recognition technology can be used in a widely range application such as identity authentication, access control.In 2017, Hangzhou's first self-service supermarket was opened for trial operation.Alibaba also used the this technology to improve shopping experience.Many researches show that AdaBoost algorithm can be goog engouh in this area.

AdaBoost algorithm has the potential of fast training .Here, we concentrate on the AdaBoost algorithm and evaluate its performance for face recognition.

II. METHODS AND THEORY

Boosting is a method to combine a collection of weak classification functions to form a stronger classifier.AdaBoost is an adaptive algorithm to boost a sequence of classifiers, in that the weights are updated dynamiclly in every iteration according to error in privous basic learner .

Given: $(x_1, y_1), \ldots, (x_m, y_m)$ where x_i X, y_i 1, +1 .
Initialize: $D_1(i) = 1/m$ for $i = 1, \ldots, m$.
For $t = 1, \ldots, T$:
Train weak learner using distribution D_t.
- Get weak hypothesis $h_t : X \rightarrow \{-1, +1\}$.
- Aim: select h_t with low weighted error:

$\varepsilon_t = \Pr_{i\sim D_t}[h_t(x_i) \neq y_i]$
Choose $\alpha_t = \frac{1}{} \ln \frac{1-\varepsilon_t}{\varepsilon_t}$
Update, for $i = 1, \ldots, m$:

$$D_{t+1}(i) = \frac{D(i)\exp(\alpha y h(x))}{Z_t}$$

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

Output the final hypothesis:
$H(x) = \text{sign } h(x)$

Fig. 1 The boosting algorithm AdaBoost.

Figure 1 is the pseudocode for AdaBoost.Here we are given m labeled training examples (x1, y1), . . . , (xm, ym) where the xi's are in some domain X , and the labels yi 1, +1 . On each interation t = 1, . . . , T , a distribution Dt is computed as in the figure over the m training examples, and a given weak learner or weak learning algorithm is applied to find a weak hypothesis X ->{1, +1 } .The final or combined hypothesis H computes the sign of a weighted combination of weak hypotheses

$$F(x) = \sum_{t=1}^{T} \alpha_t h_t(x). \qquad (1)$$

This is equivalent to saying that H is computed as a weighted majority vote of the weak hypotheses ht where each is assigned weight α t .

In AdaBoost algorithm, we should caputres imgages' features as fast as possible.

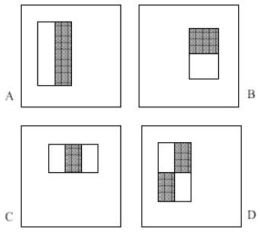

Fig 2 Rectangular filters

We local features by subtracting sum of pixels in white area from the sum of pixels in black area; 2-rectangle features (A and B), 3-rectangle feature (C) and 4-rectangle feature (D).But In a 24 x 24 patch with 4x4 detector, there are over 160,000 locations for rectangles.With too much features, we need a efficent method to compute the sum.

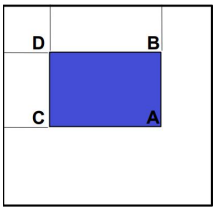

Fig 3 Compute the sum

As figure 3 shows, we let A,B,C,D be the values of the integral image at the corners of a rectangle.Then the sum of original image values within the rectangle can be computed: sum = A − B − C + D.

III. EXPERIMENT

We used 1000 faces and nonfaces mixed dataset to train our Adaboost model.

The face database is divided into two non-overlapping sets for training and testing. The training data consist of 600images: 300 images is face and the other is nonface. The remaining 400images are used for validating the model.

At first images are converted into a size of 24 * 24 grayscale, the data set label is setted as {-1, +1}.After processing data set data to extract NPD features, we start to finish
all AdaboostClassifier functions based on the reserved interface in ensemble.py.The detail steps are shown in Figure 4.

1 Initialize training set weights , each training sample is given the same weight.
2Training a base classifier , which can be sklearn.tree library DecisionTreeClassifier (note that the training time you need to pass the weight as a parameter).
3 Calculate the classification error rate of the base classifier on the training set.
4 Calculate the parameter according to the classification error rate .
5 Update training set weights .
6 Repeat steps 2 - 6 above for iteration, the number of iterations is based on the number of classifiers.

Fig 4 AdaboostClassifier steps

Finally, we 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 .

Resutl:

	precision	recall	f1-score	support
face	0.90	0.84	0.87	300
nonface	0.85	0.90	0.88	300
avg / total	0.87	0.87	0.87	600

Fig 5 TrainSet result

	precision	recall	f1-score	support
face	0.86	0.79	0.82	200
nonface	0.80	0.87	0.83	200
avg / total	0.83	0.83	0.83	400

Fig 6 Validset result

From the figuer 5 and figuer 6, we find that we do not get a high perfermance by using AdaBoost.The reasons causing this problems may:

Firstly,we divide the dataset by hand and later experiments based on divided training set and valid set, The result may be influenced by the fixed dataset.

Another reason may be the problem of over-fitting,which is serious for boosting on face data.

IV. CONCLUSION

We have evaluated the AdaBoost algorithm for face recognition. At privous part of this report, we introduce the AdaBoost algorithm and later learn how to train model in pratice.Using face recognition experiment, we have fully understand how AdaBoost works.

Besides, we do not have a good result approach AdaBoost.The most likely reason is we divide the dataset inproperly.All in all,AdaBoost is a boost method to make weak learner stronger.We still have a lot of work to improve the experiment.