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: Ruolin Liang and
Zequan Zeng and Hainan Huang

Supervisor:
Qingyao Wu

Student ID：201721046029
201720144917 201720145150

Grade:
Graduate

December 19, 2017

Face Classification Based on AdaBoost Algorithm

Abstract—Face detection is the most important part of face analysis, which solve the problem whether face is exists. AdaBoost algorithm is a fast face detection algorithm proposed in 1995,which is the field of face detection Milestone-based advancements. This algorithm adaptively adjust the assumed error rate based on weak learning feedback. Among our experiments, AdaBoost algorithm gets the good performance.

I. INTRODUCTION

Boosting, also called Augmented Learning or Boosting, is an important integrated learning technique that can enhance prediction accuracy by a weak learner. It provides an effective new idea and method for learning algorithm design. One of the most successful applications is the AdaBoost algorithm. AdaBoost is Adaptive Boosting. Its adaptivity lies in that: the weight of a sample whose former basic classifier is misclassified increases, while the weight of a correctly classified sample decreases, and again used to train the next basic classifier. At the same time, a new weak classifier is added in each round of iteration until a pre-determined small enough error rate is reached or pre-specified maximum number of iterations is reached to determine the final strong classifier.

Adaboost algorithm can be briefly described as three steps:

1) The weight distribution D1 of training data is initialized. There are N training sample data, each training sample is given the same weight at the very beginning: w1=1/N
2) Train weak classifier. If a training sample point is classified right by the weak classifier, its corresponding weight to be reduced in the construction of the next traning set. If a training sample points are misclassified, then its weight should be increased. The updated set of samples is used to train the next classifier, and the entire training process is iteratively carried out.
3) The weak classifiers that is obtained by each training are combined into a strong classifier. After the training process of each weak classifier is finished, the weight of the weak classifier with small classification error rate is increased, which plays a important role in the final classification and reduces the weak classifier with large classification error rate. The weight plays a smaller role in the final classification.

II. METHODS AND THEORY

Given a training data set

$$T = \{(x_1, y_1), \ldots (x_n, y_n)\}, y_i \in \{1, -1\}$$

The purpose of AdaBoost is to learn a range of soft classifiers or basic classifiers from the

training number, and then combine these weak classifiers into a strong classifier.

AdaBoost algorithm is as follows:

1) The weight distribution of the initial training data. Each training sample is initially given the same weight:1/N

$$D_1 = (W_{11}, W_{12}, .. W_{1I} .., W_{1N}),$$

$$W_{1i} = \frac{1}{N}, i = 1, ... N$$

2) multiple iterations:

a.Use a training set with weight distribution to get the basic classifier.

$$G_m(x): \chi \to \{-1. +1\}$$

b. Error rate:

$$e_m = P(G_m(x_i) \neq y_i) = \sum_{i=1}^{N} w_{mi} I(G_m(x_i) \neq y_i)$$

c.Computing coefficient:

$$\alpha_m = \frac{1}{2} \ln \frac{1 - e_m}{e_m}$$

d. Update the weight distribution of the training data set.

$$D_m = (w_{m+1}, w_{m+2} ... w_{m+1,i} ... w_{m+1,N})$$

$$w_{m+1,i} = \frac{w_{mi}}{Z_m} \exp\left(-\alpha_m y_i G_m(x_i)\right),$$

i=1,2,3

$$Z_m = \sum_{i=1}^{N} w_{mi} \exp(-\alpha_m y_i \, G_m(x_i))$$

3) Combine weak classifier:

$$f(x) = \sum_{m=1}^{M} \alpha_m G_M(x)$$

Final classifier:

$$G(x) = \text{sign}\left(f(x)\right) = \text{sign}\left(\sum_{m=1}^{M} \alpha_m G_m(x)\right)$$

III. EXPERIMENT

1. Data Set

We use 500 face features and 500 non-face features and divide 330 features as test features and 670 features as train features.

2. Experimental steps

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.

Processing data set data to extract NPD features. Extract features using the NPDFeature class in feature.py. (Tip: Because the time of the pretreatment is relatively long, it can be pretreated with pickle function library dump () save the data in the cache, then may be used load () function reads the characteristic data from cache.)

The data set is divisded into training set and calidation set, this experiment does not divide the test set.

Write all Adaboost Classifier functions based on the reserved interface in ensemble.py. The following is the guide of fit function in the Adaboost Classifier class:

2.1 Initialize training set weights , each training sample is given the same weight.

2.2Training a base classifier , which can be sklearn.tree library DecisionTreeClassifier (note that the training time you need to pass the weight as a parameter).

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

2.4 Calculate the parameter according to the classification error rate .

2.5 Update training set weights .

2.6 Repeat steps 2.2-2.6 above for iteration, the number of iterations is based on the number of classifiers.

3. Experiment Result

We compare the performance of the Face Classification on the weak classifier and the Adaboost Algorithm which use ten weak classifier.

	precision	recall	f1-score	support
class-1	0.85	0.90	0.87	163
class1	0.90	0.84	0.87	167
avg / total	0.87	0.87	0.87	330

Fig.1

From Fig.1, it is shown that the precision of the weak classifier is 87% and the recall of the weak classifier is 87% and the F1 of the weak classifier is 87%.Through these datas, we can find that the NPD features of images are the good features.

	precision	recall	f1-score	support
class-1	0.96	0.98	0.97	163
class1	0.98	0.96	0.97	167
avg / total	0.97	0.97	0.97	330

Fig.2

From Fig.2, it is shown that the precision of Adaboost Algorithm is 97% and the recall of Adaboost Algorithm is 97% and the F1 of Adaboost Algorithm is 97%. Through these datas , we can find that Adaboost Algorithm improve the performance of Face Classification. We can also find that the Adaboost Algorithm has the good performance.

IV. CONCLUSION

Among our experiences, it is shown that the NPD features of images are the good features.Compare with the weak classifier, the Adaboost Algorithm can improve the performance of Face Classification and has the good performance.