



The Experiment Report of Machine Learning

SCHOOL: SCHOOL OF SOFTWARE ENGINEERING

SUBJECT: SOFTWARE ENGINEERING

Author:
Denzeshuai Suyue Huangjingchao

Supervisor:
Qingyao Wu

Student ID: 201530611425 and
201530612774 and
201530611722

Grade:
Undergraduate

December 9, 2017

Face classification based on AdaBoost algorithm

Abstract—

I. INTRODUCTION

A deep understanding of the principles of Adaboost, A basic approach to face detection. Learn to use Adaboost to solve the problem of face classification and integrate theory and practical engineering. Experience the complete process of machine learning

II. METHODS AND THEORY

AdaBoost is the best representation of the Boosting method. The method reduces the weight of the sample to each instance, and increases the weight of the error samples, so that the classifier is improved step by step during the iteration process. Finally, all the classifiers are linearly combined to get the final classifier.

III. EXPERIMENT

A. Dataset

1. Read data set. Read the picture, turn all the pictures into the gray scale of the size of 24*24, the number and proportion of the positive and negative samples of the data set are not limited, and the form of the data set is not limited.

2. Processing data set and extracting NPD features. Use the method of NPDFeature class in feature.py to extract features. (Hint: because the time of preprocessing data set is relatively long, we can use the dump () function in pickle library to save the preprocessed feature data to the cache, then we can read the feature data with load () function.

3. Cut the dataset into a training set and a validation set

B. Implementation

Write all the functions of the AdaboostClassifier according to the reserved interface in the ensemble.py. The following is the idea of the fit () method in the AdaboostClassifier class:

1. Initializes the weights ω of the training set, and each training sample is given the same weight value.
 2. Training a base classifier. The base classifier can use DecisionTreeClassifier in the sklearn.tree library.
 3. The classification error rate ϵ of the base classifier on the training set.
 4. According to the classification error rate ϵ , the parameters α are calculated.
 5. Update the weight ω of the training set.
 6. Iterate the steps of the above 4.2-4.6, and the number of iterations is the number of the base classifier.
- Using the method in AdaboostClassifier to predict and calculate the accuracy on the validation set

PARAMETERS

Decision tree depth	3
Weak classifier number	12

Prediction results (best results): 0.953333333333

IV. CONCLUSION

AdaBoost is an iterative algorithm that adds a new weak classifier to each round until a certain predetermined error rate is small enough. Each training sample is given a weight, indicating that it is selected by a classifier to get the probability of the training set. If a sample point has been accurately classified, the probability of being selected in the next training set will be reduced. On the contrary, if a sample point is not accurately classified, its weight will be improved.

In the concrete implementation, the weights of each sample are equal at the beginning. For the K iteration operation, we choose the sample points according to these weights, and then train the classifier. Then this classifier is used to improve the weight of the samples which are misclassified and reduce the weight of the samples that are correctly classified.