



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:
Shoukai Xu and Yaofu Chen

Supervisor:
Mingkui Tan or Qingyao Wu

Student ID: 201530611449 and
201530612026
And 201530611210

Grade:
Undergraduate or Graduate

December 9, 2017

AdaBoost 算法的人脸分类

I. INTRODUCTION

- 1、深入理解 Adaboost 的原理
- 2、熟悉人脸检测的基本方法
- 3、学会利用 Adaboost 解决人脸分类问题，将理论和实际工程接轨
- 4、体验机器学习的完整过程

II. METHODS AND THEORY

- 1、读取数据集数据。读取图片，将全部图片转成大小为 24*24 的灰度图，数据集正负类样本的个数和比例不限，数据集标签形式不限。
- 2、处理数据集数据，提取 NPD 特征。使用 feature.py 中 NPDFeature 类的方法提取特征。（提示：因为预处理数据集的时间比较长，可以用 pickle 库中的 dump() 函数将预处理后的特征数据保存到缓存中，之后可以使用 load() 函数读取特征数据）
- 3、将数据集切分为训练集和验证集，本次实验不划分测试集。
- 4、根据 ensemble.py 中的预留的接口编写 AdaboostClassifier 所有函数。以下为 AdaboostClassifier 类中的 fit() 方法的思路：
 - 4.1 初始化训练集的权值，每一个训练样本被赋予相同的权值。
 - 4.2 训练一个基分类器，基分类器可以使用 sklearn.tree 库中 DecisionTreeClassifier(注意训练的时候需要将权重 作为参数传入)。
 - 4.3 计算基分类器在训练集上的分类误差率。
 - 4.4 根据分类误差率，计算参数。
 - 4.5 更新训练集的权值。
 - 4.6 重复以上 4.2-4.6 的步骤进行迭代，迭代次数为基分类器的个数。
- 5、用 AdaboostClassifier 中的方法在验证集上进行预测并计算精确率，并用 sklearn.metrics 库的 classification_report() 函数将预测结果写入 report.txt 中。
- 6、整理实验结果并完成实验报告。

实验环境：python3 包含 python 包：
sklearn, numpy, matplotlib, pickle, PIL。PyCharm
Community 集成开发环境

III. EXPERIMENT

A. 数据集

本实验提供 1000 张图片，其中 500 张是含有人脸的 RGB 图片，储存在 ./datasets/original/face 内；另外 500 张是不含有人脸的 RGB 图，储存在 ./datasets/original/nonface 内。数据集包含在示例仓库内，请自行下载并将其切分为训练集，验证集。

B. 执行

- 1.初始化：权重 w 全为 1，将全部图片转化为 24×24 的灰度图
- 2.过程：读入数据集，利用 NPDFeature 类读取 nonface 和 face 的 NPD 特征并存入缓存。从缓存中读取数据，划分训练集和测试集，利用 AdaBoostClassifier 进行模型训练和测试，将测试结果与预测结果存入 report.txt 中
- 3.结果：

support	precision	recall	f1-score
102 nonface	0.98	0.89	0.93
98 face	0.90	0.98	0.94
avg / total	0.94	0.94	0.93
200			

IV. CONCLUSION

经验与收获：

经过这次的实验，我们理解并懂得了 Adaboost 的实现原理，学会了人脸检测的基本方法，明白了人脸识别的算法并没有想象中的那么困难与复杂，初步了解了机器学习对实际工程解决实际问题的重要作用，同时，小组通过团队合作，完成实验的效率比一个人的时候效率要高很多，在现实中解决实际问题时，团队合作是必不可少的。