

应用研究类实验报告
模板格式

一、 基本情况

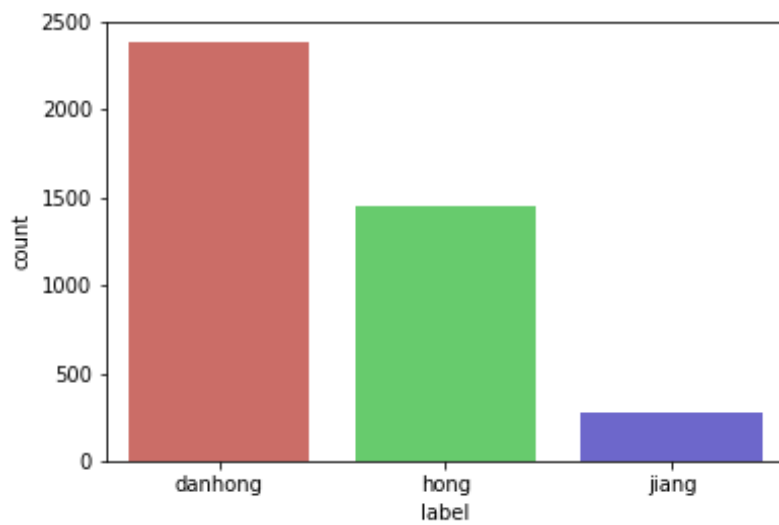
研究课题	使用不同模型对舌色训练和预测结果的比较		
课题类型			
研究人员	方圆		
开始时间	2019/1/25	预计结束时间	2019/1/28
课题方案概述：对课题总体情况的描述			
使用逻辑回归，SVM 和 XGboost 三种模型对舌色数据进行训练，并且比较不同模型下的预测结果			
研究目的 寻找出训练效果好的模型			
1) 研究现状 从准确度上看 XGboost 的效果最好			
2) 现状改善 尝试添加更多特征进行预测，例如平均亮度。			
研究内容(任务描述)			
预期目标			

进度计划

二、 技术路线

实验原理：可包括技术原理图、流程图、结构图、框架图
使用 scikit-learn 的机器学习训练模型：logisticRegression，SVM 和 XGboost 对数据进行训练，使用的特征为 1-9 列 （各种颜色的百分比）。
实验环境（硬、软）
实验数据
1) 数据描述 舌质颜色的判断 2) 数据保存路径 C:\Users\FY\Desktop\TongueColorDetectYfang\1_25_2019_Tongue Color
实验步骤

主要算法描述
<p>1) 主要算法描述</p> <p>LG, SVM 和 XGboost</p> <p>数据中 75%进行训练，25%用于测试</p> <p>具体参数：</p> <p>LogisticRegression(C=100000.0, class_weight=None, dual=False, fit_intercept=True, intercept_scaling=1, max_iter=100, multi_class='warn', n_jobs=None, penalty='l2', random_state=None, solver='lbfgs', tol=0.0001, verbose=0, warm_start=False)</p> <p>SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0, decision_function_shape='ovr', degree=3, gamma='auto_deprecated', kernel='rbf', max_iter=-1, probability=False, random_state=None, shrinking=True, tol=0.001, verbose=False)</p> <p>XGBClassifier(base_score=0.5, booster='gbtree', colsample_bylevel=1, colsample_bytree=1, gamma=0, learning_rate=0.1, max_delta_step=0, max_depth=3, min_child_weight=1, missing=None, n_estimators=100, n_jobs=1, nthread=None, objective='multi:softprob', random_state=0, reg_alpha=0, reg_lambda=1, scale_pos_weight=1, seed=None, silent=True, subsample=1)</p> <p>2) 完整代码保存路径</p> <p>C:\Users\FY\Desktop\TongueColorDetectYfang\1_25_2019_Tongue Color</p>
实验结果分析
<p>1.数据概况，总共 4124 组样本，人工判定结果(label)的样本分别如下：</p>



关于特征的具体参数:

[4123 rows x 19 columns]

danbai	int64
dan	int64
danhong	int64
hong	int64
jiang	int64
anhong	int64
honghuang	int64
qingzi	int64
danzi	int64
tip12Percent	int64
left12Percent	int64
right12Percent	int64
both12Percent	int64
middle12Percent	int64
avgBrightness	float64
peakingBrightness	float64
brightness	int64
id	int64
label	object

dtype: object

danhong 2387

hong 1454

jiang 282

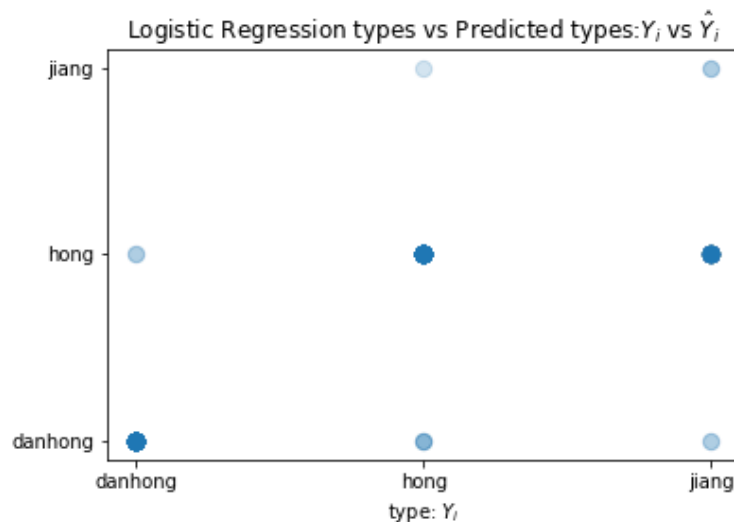
2. 机器学习结果如下:

a. 逻辑回归, logistic Regression

Accuracy of logistic Regression classifier on test set: 0.90

Accuracy of logistic Regression classifier on training set: 0.93

可以看出除了绛会被很多的错误判断为红之外, 其他颜色的预测还比较理想。

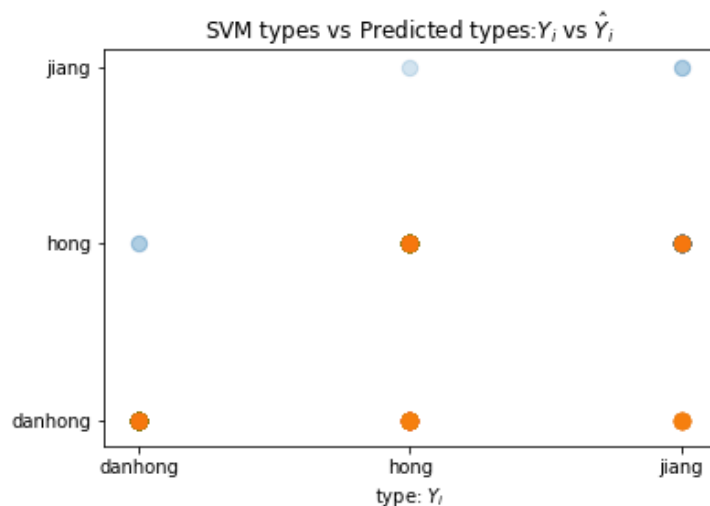


b. 支持向量机, Support vector machine

Accuracy of SVC classifier on test set: 0.79

Accuracy of SVC classifier on training set: 0.98

检测数据的准确率偏低, 训练数据有过拟合的倾向。红有多数被错误判断为淡红, 绛也多数被错误判断为大红。

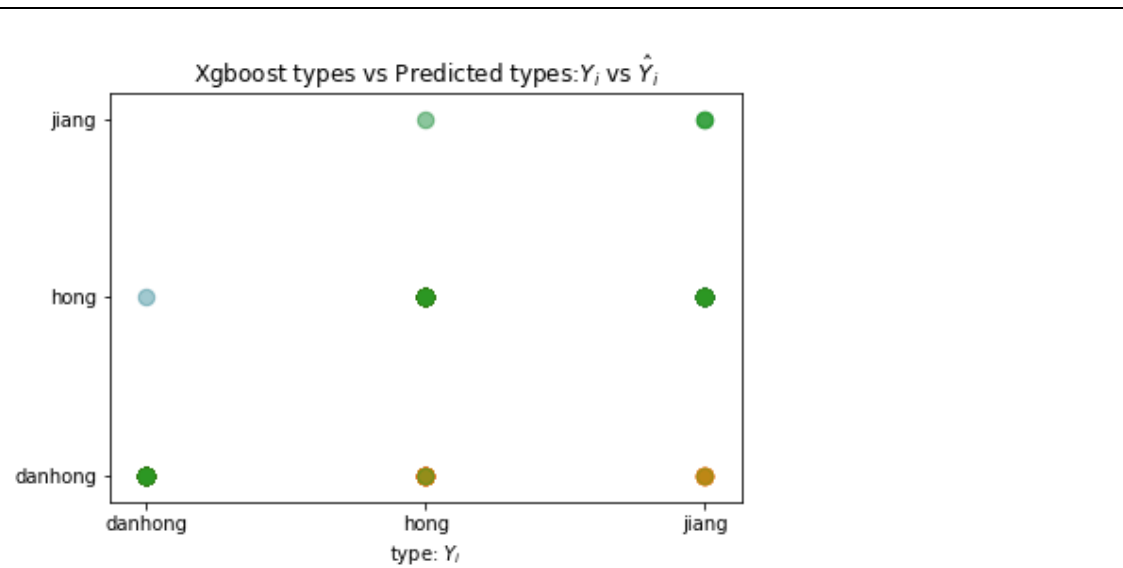


c. Xgboost

Accuracy of Xgboost classifier on test set: 0.91

Accuracy of Xgboost classifier on training set: 0.94

检测和训练的准确率是三个模型最高的, 但是绛有多数被错误判断成红。



从初步看，在这组数据中，**XGboost** 的预测效果在三种模型中是最好的。接下来如果想提高训练的准确度：**1.**增加标签为绛的样本数量。**2.** 改善或者增加特征性辅助判断。

三、 课题总结

结论
<div>1) 结论描述</div> <div>从初步看，在这组数据中，XGboost 的预测效果在三种模型中是最好的。接下来如果想提高训练的准确度：1.增加标签为绛的样本数量。2. 改善或者增加特征性辅助判断。</div> <div>2) 课题相关文档保存路径</div> <div>C:\Users\FY\Desktop\TongueColorDetectYfang\1_25_2019_Tongue Color</div>
项目中应用情况
心得、拓展设想

四、 知识参考

参考文献
没有
遇到的问题记录
没有