**应用研究类实验报告**

**模板格式**

# 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   1. 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 | | | |
|  | | | |

# 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） |
|  |
| 实现数据 |
| 1. 数据描述   舌质颜色的判断   1. 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 |
| 1. 主要算法描述   LG, SVM和XGboost  数据中75%进行训练，25%用于测试  具体参数：  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)  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)  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)   1. 完整代码保存路径   C:\Users\FY\Desktop\TongueColorDetectYfang\1\_25\_2019\_Tongue Color |
| 实验结果分析 |
| 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：    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： |
| 1. 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  可以看出来出了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。     1. 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。     1. 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。 |

# 课题总结

|  |
| --- |
| 结论 |
| 1. 结论描述   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。   1. 课题相关文档保存路径   C:\Users\FY\Desktop\TongueColorDetectYfang\1\_25\_2019\_Tongue Color |
| 项目中应用情况 |
|  |
| 心得、拓展设想 |
|  |
|  |

# 知识参考

|  |
| --- |
| 参考文献 |
| 没有 |
| 遇到的问题记录 |
| 没有 |