**测试方案与测试用例设计**

**1. 概述**

本文档是《基于机器学习的APT检测》系统的测试方案。本系统使用多种随机森林分类器对病毒样本所属的APT组织和国家进行预测，以可视化界面展示，并与云数据库进行连接。因此主要测试对象分模型、界面和数据库连接功能三部分，其中后两部分可以集中进行测试。测试的目的是验证训练端人员提供的的模型的有效性，前端界面的功能运转正常且界面合理易用，为软件整体质量提供保障。

**2. 测试方法**

对于模型测试，使用训练和未训练过的测试样本作为输入运行模型，统计模型预测表现的各种指标，属于动态黑盒的测试方法。对于前端界面和功能测试，由于代码量不大，且组员在开发过程中互相阅读过代码并进行改进，这里不重复进行静态白盒测试，主要进行运行时的测试和代码Debug，属于动态白盒测试。

**3. 测试内容**

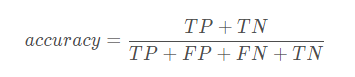
* 1. 模型评价

本组课题属于分类模型。常采用的指标有：

* 四个基础比例：
* 正确地预测为正例：TP
* 正确地预测为反例：TN
* 错误地预测为正例：FP
* 错误地预测为反例：FN
* **混淆矩阵Confusion Matrix：**以上四个比例构成的矩阵：



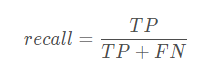
* **准确率Accuracy**：被预测正确的比例：



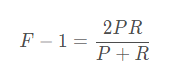
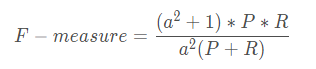
* **精确率(查准率) Precision**："正确被预测为正(TP)"占所有"实际被预测为正的(TP+FP)"的比例：



* **召回率(查全率) Recall**："正确被预测为正(TP)"占所有"应该被预测为正(TP+FN)"的比例：



* **F-measure**：查准率和查全率通常是一对矛盾的度量。F-measure就是一个二者调和的指标。如果A取1，那么F-measure实际上就是precison和recall的调和平均值，此时的F-measure就称为F1值：



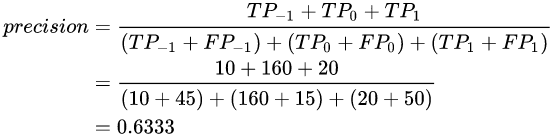
* **ROC曲线：**以模型的“False positive rate”与“True positive rate”值分别为横纵轴的曲线。曲线越往左上角贴近说明模型效果越好：



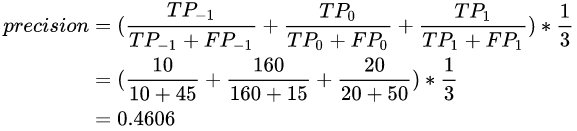
* **AUC（Area Under Curve）：**指ROC曲线下的面积是多大，取值在0与1之间。AUC = 1，代表ROC曲线在纵轴上，预测完全准确。

以上指标最初用于二分类模型的评价。而本课题是多分类的问题，所以相应的指标含义会发生一些变化，并引入适用于多分类的新指标：

* 属于类C的样本被正确分类到类C，记这一类样本数为 TP
* 不属于类C的样本（可能是A也可能是B…）被错误分类到类C，记这一类样本数为 FP
* 属于类别C的样本被错误分类到类C的其他类，记这一类样本数为 FN
* 不属于类别C的样本被正确分类到了类别C的其他类，记这一类样本数为 TN
* 假设有N个类，混淆矩阵相应变为NxN的矩阵
* **micro方法**：指把所有的类放在一起算。例如对于precision，就是把所有类的TP加和，再除以所有类的TP和FN的加和。因此micro方法下的precision和recall都等于accuracy。



* **macro方法**：先分别求出每个类的precision，再算术平均。



计划使用以上所有指标对我们的模型进行测试评价。

* 1. 功能及界面测试
* 选择本地模型及待预测文件（单/多个）进行预测，输出预测结果；
* 查看本地模型参数；
* 连接华为云数据库；
* 查询数据库中的预测数据；
* 本地预测结果上传数据库；
* 界面布局合理易用，未接触过该系统的用户可以快速理解界面，在阅读用户手册后可以完全掌握系统的所有操作。

**4. 测试用例设计**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 测试用例编号 | 测试对象 | 测试输入 | 预期输出 | 备注 |
| 1.1 | 未训练的分类器 | 所有训练样本 | 训练后的分类器及模型的表现参数 | 训练模型时使用K折交叉验证，同时完成模型训练和模型表现参数的计算 |
| 1.2 | 训练后的分类器 | 分类器的表现参数 | 评价模型是否达到可用要求 | 根据参数进行人工评价，最低要求达到课题规定的75%准确率 |
| 2.1 | APT预测程序 | 点击界面“浏览”按钮，选择单/多个病毒报告文件；在下拉框选择模型 | 对话框中显示选择的模型和所有待测文件路径，为预测做好准备 |  |
| 2.2 | APT预测程序 | 在2.1的基础上，点击开始预测按钮 | 结果显示区域展示使用的模型，预测文件名，预测概率分布，预测分类 | 选择多个待测文件时，以列表显示每一项 |
| 2.3 | APT预测程序 | 在2.1的基础上，调整阈值滑动条，点击开始预测按钮 | 观察到预测分类与2.2有所变化，更多/更少的样本被划分到其他分类 |  |
| 2.4 | APT预测程序 | 选中某个模型，点击现实模型参数按钮 | 弹出新窗口，显示该模型的算法、采样策略信息和精确率、召回率等表现参数，以及混淆矩阵和ROC曲线两张图片 |  |
| 2.5 | APT预测程序 | 输入云数据库地址，点击连接 | 左下角状态栏刷新，显示连接到数据库 |  |
| 2.6 | APT预测程序 | 选择数据库已有文件，点击查询数据库 | 显示数据库中对应文件的预测结果 |  |
| 2.7 | APT预测程序 | 在2.2的基础上，点击上传预测结果，并点击刷新 | 在数据库已有文件下拉框中出现刚上传的文件，可以查询其预测结果 |  |