工作计划完成情况（测试）

|  |  |  |
| --- | --- | --- |
| 日期 | 计划内容 | 完成情况 |
| 8.25-8.26 | 查找APT组织相关资料，了解课题背景 | 完成 |
| 8.27 | 第一次汇报，与负责老师交流，确定数据集的收集来源及范围 | 完成 |
| 8.28-8.31 | 完成数据集的收集整理，与开发人员沟通，制作测试需要的完整资料 | 完成 |
| 9.1 | 进行功能测试，向开发人员反馈测试情况 | 未完成 |

总结：

前期按时完成了课题背景的调研，发现老师所给样例数据集包含3500+个数据，足够使用，而且数据类型比较特殊，是二进制处理后的病毒文件，不易自行制作。第一周周五汇报跟老师交流确认可以使用该数据集的一部分做训练，一部分测试。

我们的处理流程为“将病毒文件提交到cuckoo虚拟环境测试→获得测试报告→对测试报告进行关键词等特征的提取→训练模型→预测结果”，其中将原始文件提交到cuckoo测试并获取报告需要较长的处理时间（大概一小时只能处理不到10个），而且cuckoo的环境配置比较复杂，所以我和闫星润一起对这部分进行了研究，将原始病毒文件一次性处理成测试报告，后续负责模型研究的成员只需拿跑好的报告文件做输入即可。目前我们已有200个左右处理好的报告，后续计划根据模型训练情况继续出一些报告，并为最终的模型检验预留一些测试数据。

由于模型开发侧成员还未交付具有模型导出读取功能的初步实现，目前每次运行都得重新训练模型，时间成本太大，因此功能测试延后，但最迟需要在本周五进行。等待期间协助杜梁楷进行数据库端的开发。