# 台湾AIdea“AOI 瑕疵分類”挑战方案

李勇 2019年08月30日

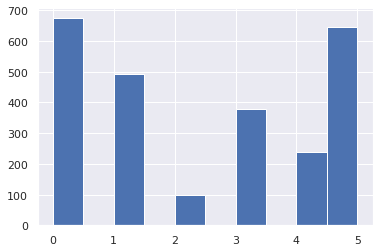
## 数据分布

### 1.1.分布数据表格

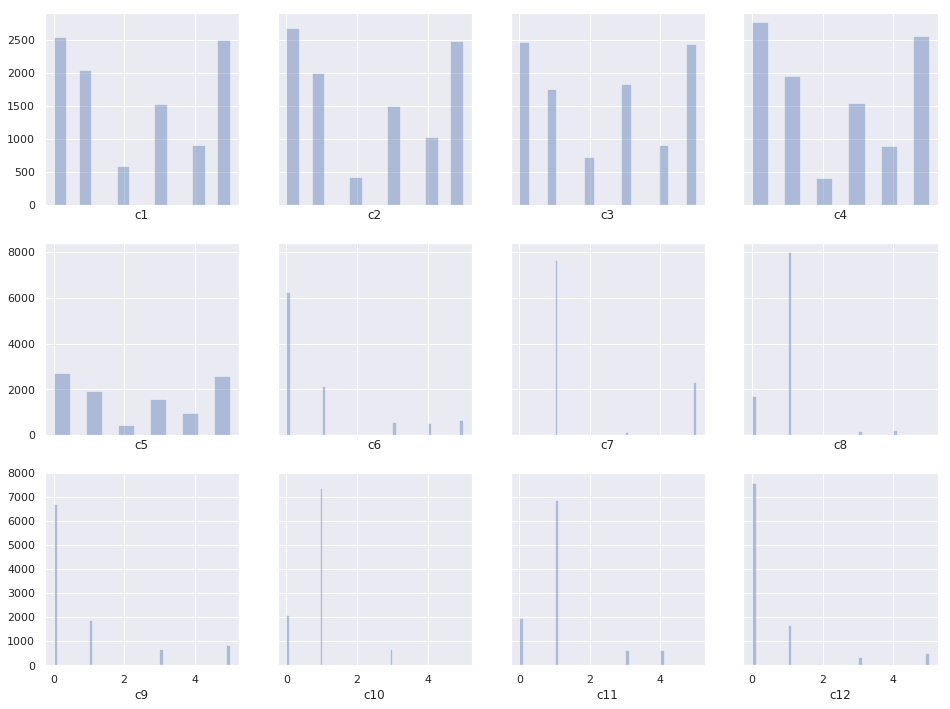
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 类 别 | 0 | 1 | 2 | 3 | 4 | 5 | 合计 |
| 训练集 | 674 | 492 | 100 | 378 | 240 | 644 | 2,528 |
| 测试集 | 2552 | 2049 | 592 | 1530 | 914 | 2505 | 10,142 |

### 1.2图表展示

训练集：



测试集：（根据98%+精度的组合分类器统计得到）



c1,c2,c3,c4为四个Classifier得到的测试数据分布。这些个数据分布反应出：正常数据（类别0）还是很容易在工业中大量获取的。

### 1.3 数据分布问题及策略

样本数量太少：

A)半监督的方式，给测试集样本打标签，参与训练；

数据分布不均匀：

A)通过copy的方式，实现样本均衡

## 网络设计

### 2.1 基础网络尝试

|  |  |  |
| --- | --- | --- |
| 基础网络 | 表现 | 说明 |
| ResNet19 | 92%左右 | 只是finetune最后fc层 |
| DenseNet | 30%左右 | Finetun很多层 |

结论是：

继续采用ResNet19，适当放宽finetune范围；

数据量不够驱动一个复杂网络有效做全部的参数update；

### 2.2 提高分类准确率

综合多个网络的预测结果，投票表决。

## 算法流程详情

|  |
| --- |
| 1.训练五个基础网络on train set； |
| 2.根据基础网络表决结果，对测试集打标签； |
| 3.混合train set和test set，通过copy实现data balance； |
| 1. 在新的合成数据集上训练5个决策网络； |
| 1. 根据决策网络给出最终的预测结果。 |