每日监控：

1. 通过自动化脚本从主流应用商店爬取Top 1w应用名称对应的图标icon并返回 (metaid, 应用名称，应用图标icon) 字段
2. 利用已有的CNN预测所有当天爬取的应用图标icon，并将预测结果与实际的metaid进行比较
3. 如果有任何应用预测错误，发出报警并人工介入查找原因（很大可能是某个应用的图标进行了大的改变）
   1. 如果某些应用有了全新的图标，更新训练数据集并重新训练CNN模型

Yes

Supervised Learning

卷积神经网络CNN

清洗+分词

LDA

清洗+分词

清洗+分词

模型

Trace Back

and

Check

score

Trace Back

and

Check

score

判断基于图像和文本两个维度判断的得分是否有很大差异

Yes

No

输出 Output:

1. pkgname -> (metaid, score)
2. pkgname -> (“need new id”, max\_score)

算法/规则

导出如下信息并进行人工介入:

packageName, (metaid#1, score#1), (metaid#2, score#2)

Unsupervised Learning

**O ()**

Filter

（基于开发者信息等其他维度先行判断这对应用是否肯定不应合包，尽量降低计算维度）

转化成词向量进行pairwise相似度比较

余弦相似度

余弦相似度

余弦相似度

LDA

主题相似度+关键词相似度

（规则或模型）

可以尝试采用pre-train好的CNN在有标注的icon数据集上进行fine-tune

卷积神经网络CNN

Trace Back

and

Check

Trace Back

and

Check

导出如下信息并进行人工介入:

packageName, (metaid#1, score#1), (metaid#2, score#2)

No

Yes

Yes

输出 Output:

1. pkgname -> (metaid, score)
2. pkgname -> (“need new id”, max\_score)

算法/规则

算法/规则

算法/规则

判断基于图像和文本两个维度判断的得分是否有很大差异

score

score