

## 图片认知分类系统设计与开发

### 异常情况说明书

队 名： 吃藕队

指导老师： 陈鹏

队 长： 尚振鸿

队 员： 于晴，潘颖慧

目录

1 输入恶意无效标签 ..... 1

    1.1 异常说明..... 1

    1.2 应对策略..... 1

2 部分图片长时间未判定 ..... 1

    2.1 异常说明..... 1

    2.2 应对策略..... 2

3 部分推送图片总被忽略 ..... 2

    3.1 异常说明..... 2

    3.2 应对策略..... 2

4 导出图片标签化结果丢失 ..... 3

    4.1 异常说明..... 3

    4.2 应对策略..... 3

## 1 输入恶意无效标签

### 1.1 异常说明

系统被别有用心的人使用，可能冒充志愿者对图片输入恶意无效的标签，企图影响最终的图片标签判定的结果，影响最终判定的准确率。

### 1.2 应对策略

#### (1) 图片识别接口首先识别

图片在分发给志愿者之前先由图片识别接口对图片进行初次识别，接口识别的结果可以作为最终标签判定的参考。

#### (2) 限定每个志愿者对于图片所打的标签数量

对于每张图片，只会分发给志愿者一次，一个志愿者只能对一张图片添加自定义的 4 个描述标签。这样可以将无效恶意标签的影响降到最低。

#### (3) 对一张图片采集适量的标签再进入最终判定流程

对于一张图片我们将会分发给不同的 25 个志愿者，最终将得到(25--200 个标签)，然后才会进入最终的标签判定结果，在保证迭代速度的前提下尽可能让每个志愿者的标签对最终结果的判定的影响降低。

#### (4) 采用"语义距离算法"判定图片最终标签

我们会采用 NLP 中的语义距离算法，使用同义词语料库，计算所有标签的相似度，从中找出相似度在一定阈值内的标签，作为图片最终的标签，该策略可以将恶意无效标签的影响降到最低。

## 2 部分图片长时间未判定

### 2.1 异常说明

我们的系统对于每一张图片设定了判定阈值 25，当一张图片被分发 25 次后，开始进入标签判定流程。对于分发系统，可能出现长时间后，有一部分图片的分发次数还没

有达到 25 次，迟迟没有进入判定流程。

## 2.2 应对策略

### (1) 统计每张图片被分发的次数

我们将会统计每张图片的分发次数，然后根据分发次数进行策略调整。

### (2) 分发次数越接近阈值越先分发

对于统计得到的图片的分发次数进行排序，图片的分发次数越接近 25，越会被系统优先分发给志愿者。这样便会加快迭代速度，尽快有图片进入标签判定流程，尽快产出结果。

## 3 部分推送图片总被忽略

### 3.1 异常说明

如果我们采用一般的图片随机分发方式，可能推送的图片不是志愿者感兴趣的类别，可能会被志愿者忽略，这样就降低了迭代的速度。

### 3.2 应对策略

#### 1) 添加常见分类

管理员可以在 web 管理端添加一些常见的图片分类，以作为兴趣选择项。

#### 2) 收集用户的兴趣

在志愿者首次登录成功之后，将给志愿者展示管理员添加的兴趣类别，志愿者选择自己感兴趣的类别，然后志愿者收到的分发的图片就是感兴趣的类别，这样便做到了定向精准推送。

#### 3) 将分类类别与图片进行关联

在将图片分发给志愿者之前，首先将图片利用图片识别接口进行初步识别，然后利用语义距离算法算出每张图片的标签与每个分类的语义距离之和，然后用最小的语义距离的分类作为图片的初步分类，这样便可以将兴趣与图片进行关联，然后根据志愿者的兴趣进行推送。

## 4 导出图片标签化结果丢失

### 4.1 异常说明

当图片标签化判定完成后，管理员将最终图片标签化的结果导出，但是，可能导出的图片标签化数据文件不小心丢失，造成标签化结果丢失。

### 4.2 应对策略

#### (1) 在数据库记录每次标签化的导出记录

当管理员选择在 web 端导出图片标签化结果时，我们将在数据库中记录管理员此次的导出记录，管理员可以在管理员端任何时间，多次重复导出标签化结果文件，即使管理员不小心将导出的数据文件丢失，也可以选择在此导出。