

图片认知分类系统设计与开发

算法设计说明书

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 任务分配算法 1

 1.1 算法流程图 1

 1.2 算法说明 2

2 标签判定算法 3

 2.1 算法流程图 3

 2.2 算法说明 4

3 兴趣推荐算法 5

 3.1 算法流程图 5

 3.2 算法说明 6

第六届“中国软件杯”大学生软件设计大赛

1 任务分配算法

1.1 算法流程图

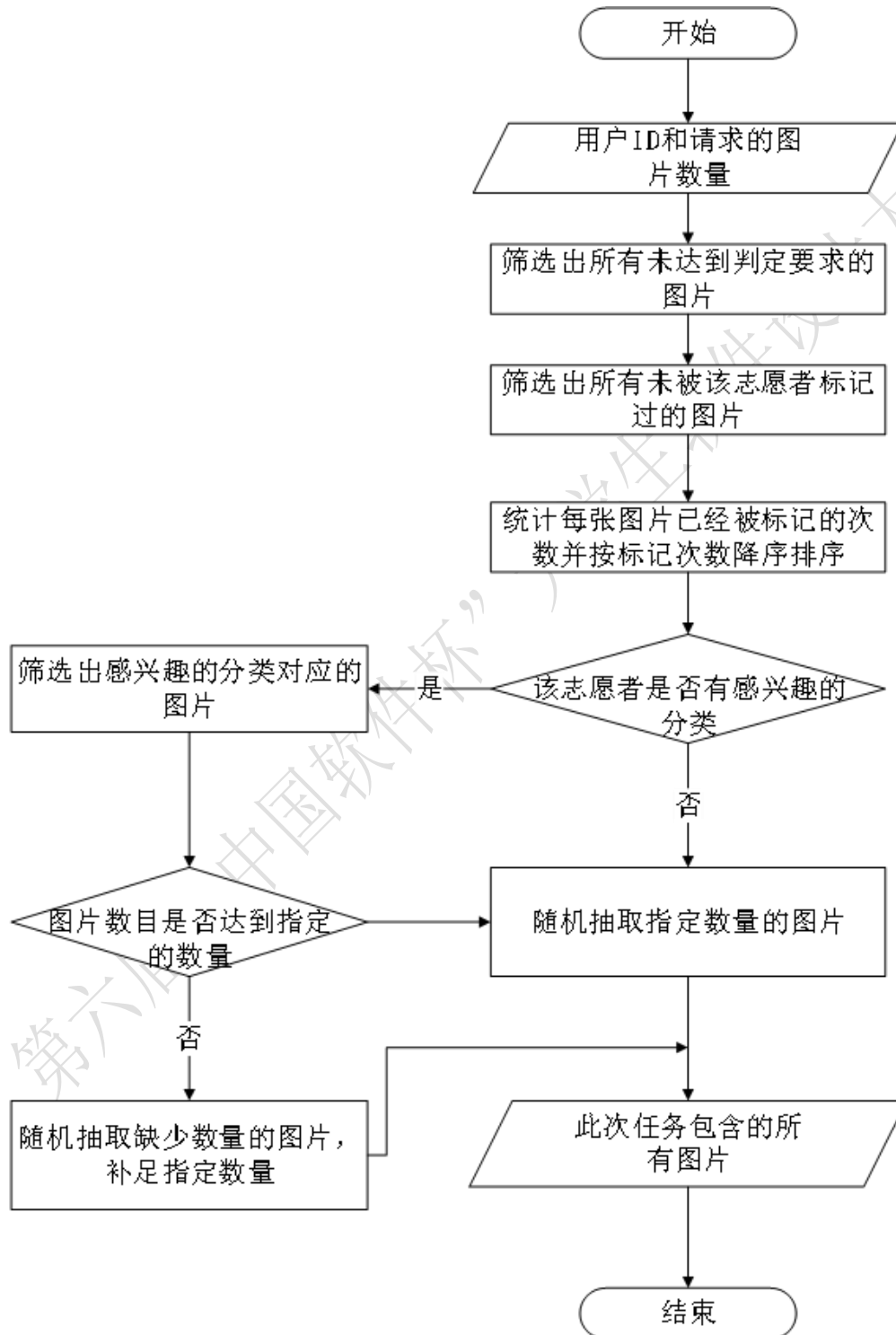


图 1.1 任务分配算法流程图

1.2 算法说明

(1) 限定每张图片只可以被每个志愿者标记一次

志愿者每次获取任务都将获取到的是全新的图片，一张图片只可以被志愿者标记一次，这在一定程度上遏制了恶意标签对最终结果的判定。

(2) 优先根据兴趣进行推送

当志愿者请求标记图片任务时，如果该志愿者有感兴趣的分类，优先推送感兴趣的分类相关的图片，如果感兴趣的类别的图片不够一次任务图片的数量，则再随机抽取图片，最终进行推送。如果该志愿者没有感兴趣的分类，则随机选择指定数量的图片推送。

(3) 接近判定阈值的图片越优先推送

对于被标记次数越接近判定阈值 25 的图片，我们越早进行推送，这样可以保证这部分图片优先被判定，加快了标记的迭代速度。

2 标签判定算法

2.1 算法流程图

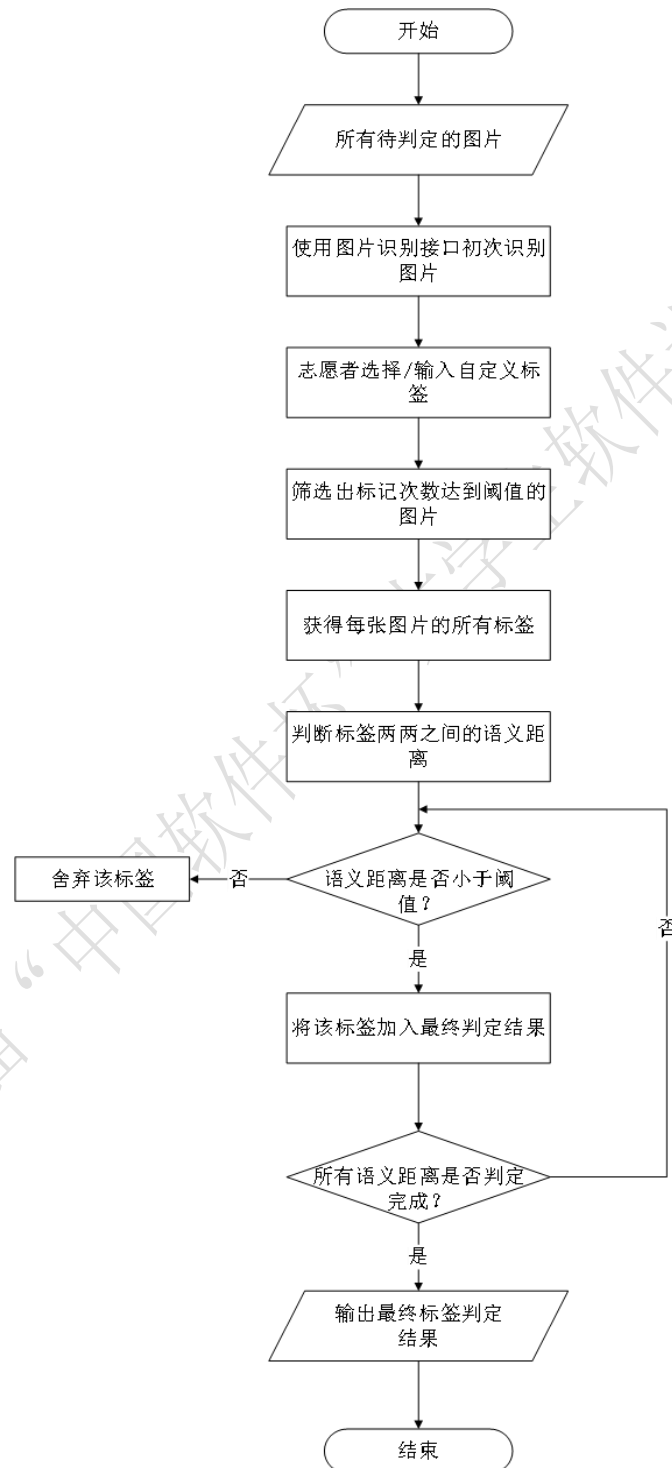


图 2.1 标签判定算法

2.2 算法说明

（1）对于所有上传的图片资源先利用识别接口进行识别

图片在分发给志愿者之前，先对于每张图片进行初步识别，得到识别接口给出的标签，这些标签将作为最终标签判定的参考标签，这样可以在一定程度上提交标签判定的准确率。

（2）志愿者可以选择备选标签和自己输入标签

对于每一张分发给志愿者的图片，我们都将接口识别提供的标签提供给志愿者进行选择，如果备选标签中有志愿者想输入的标签，则志愿者可以直接点击选择，免去了手动输入标签的麻烦。对于备选标签中没有的标签或者是图片更详细的细节，则志愿者可以手动输入标签。

（3）每天定时启动标签判定处理流程

我们使用定时任务，在每晚 12:00，对于标记次数达到 25 次的图片的标签进行判定处理，此时做图片标记任务的志愿者很少，服务器压力较小，任务处理速度快。

（4）使用语义距离算法对标签进行筛选和最终判定

对于获得的标签，我们将使用同义词词典语料库，计算标签两两之间的语义距离，然后选择语义距离在一定阈值内的标签作为最终判定结果。

3 兴趣推荐算法

3.1 算法流程图

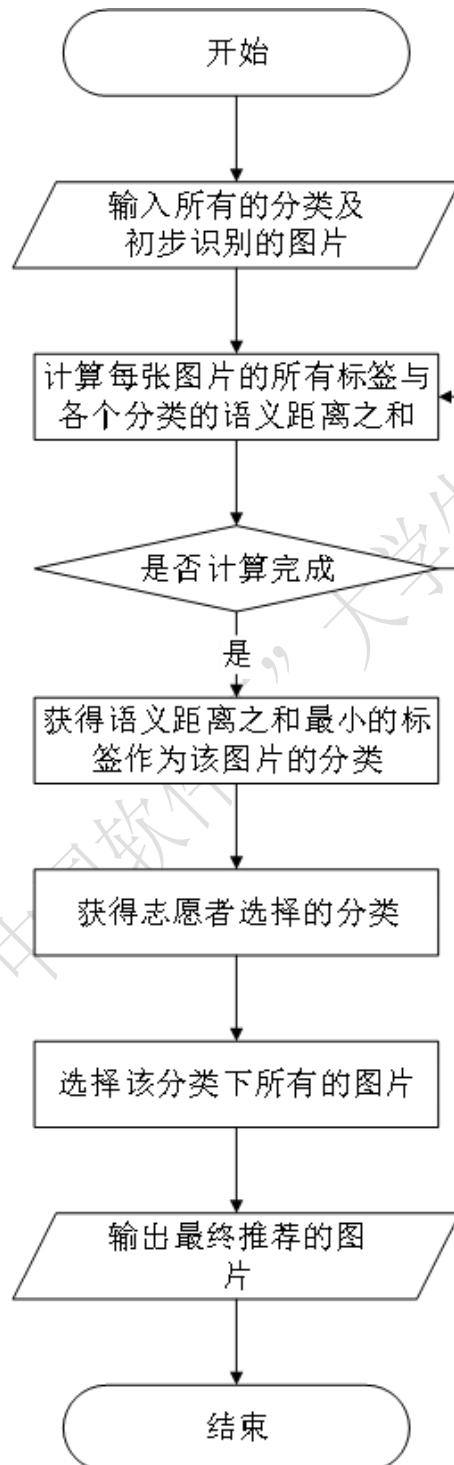


图 3.1 兴趣推荐算法

3.2 算法说明

（1）用图片识别接口初步识别

对于初次上传的图片资源，首先利用图片识别接口进行初次识别，以备后续与图片进行分类相关联。

（2）利用语义距离算法关联图片与分类

利用同义词词典语料库，计算每张图片的标签与每个分类的语义距离之和，然后选用语义距离最小的分类作为该图片的分类。

（3）基于内容的兴趣推荐算法

在志愿者初次注册登录系统时，让志愿者手动选择感兴趣的分类，根据志愿者的选择，有针对性地推送相应分类的图片给志愿者，同时志愿者可以随时修改自己选择的感兴趣的分类。