# 基于决策的黑盒文本对抗样本攻击方案研究与实现

## 摘要

摘要

## 绪论

绪论

### 背景和意义

背景和意义

### 研究现状

研究现状

### 主要研究内容

#### 动态灵活

重点在讲解动态生成字符和运行速度快

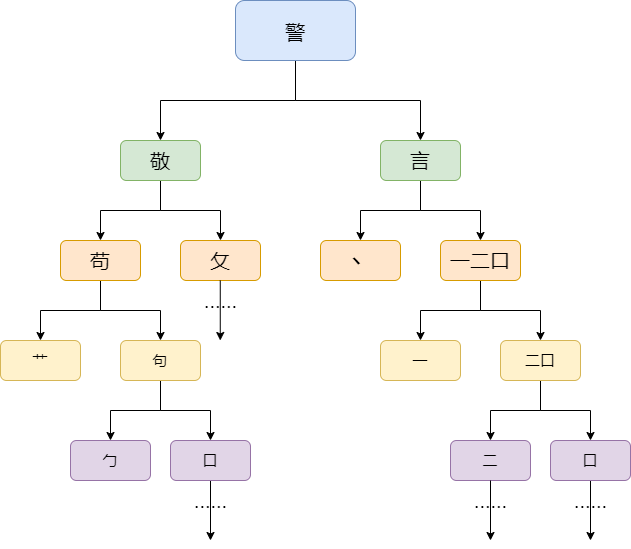
## 正文

### 以汉字形态学为核心的结构

#### 汉字递归结构拆解

##### 三元组结构 (C1,C2,结构)

##### 汉字结构二叉树



#### 汉字拆分

##### 基于汉字结构二叉树进行拆分,只选取左右部分进行拆分

对汉字进行广度优先遍历,将结构为左右或者左中右等横向可拆分的节点作为遍历节点,将其余的节点作为叶子节点,这样出队序列就是汉字的拆分序列.考虑到组合结构的复杂性,为了简便起见在遇到组合节点的时候认为汉字不可横向拆分,停止遍历.

#### 动态生成异体字

##### 基于汉字递归结构搜索.优先使用增加偏旁的汉字,其次使用汉字的部分组件.

基于汉字读半边的常见现象,将汉字加上一些偏旁对整体的影响不是很大.将原汉字作为偏旁在汉字拆分表中进行搜索.

#### 动态生成形近字

##### 以同偏旁汉字为切入口,通过递归方法计算汉字的结构加权相似度.利用动态规划的思想加速汉字相似度比对

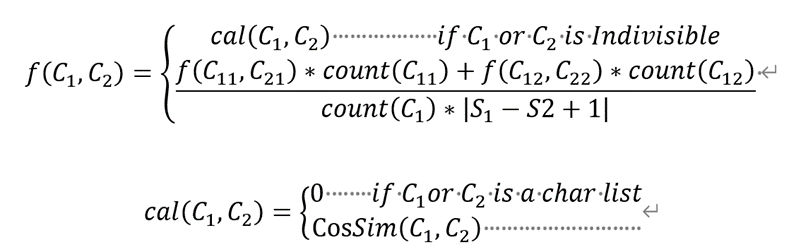
对于汉字来说,其组成部分可能是能够独立成字的偏旁(如"海",由"⺡"和"每"组成),也可能是不能独立成字的部件组合(如"侵",由"亻"和"彐 冖 又"组成).对于不能成字的部件组合,我们用一个字符元组来表示.因此,规定单字符串和字符元组为汉字C.

对于每个汉字C,都可以通过直接查询或者对其所有部件笔画数进行求和的方法求出对应的笔画数count(C).

同样的,能够通过查表获取其对应的结构S,其中字符元组的结构记为专门的组合类型.S是一个枚举类型,根据结构的相似程度依次分配了相应的数字.可以通过计算其差值作为结构偏差系数.

该算法主要是利用汉字结构二叉树进行了计算.对于要比较相似度的两个汉字,比较他们的结构是否一致.对于不一致的结构,若其中一个是独体字或者组合字则直接认为他们不相似,返回0作为相似度.否则计算结构枚举的差作为结构偏差系数,然后作为结构一致的计算.   
对于结构一致的情况如果两者都是独体字,就去直接计算他们的相似度并返回上一级.否则,分别递归计算两个子部件的相似度,然后按笔画数占比求和,乘以结构偏差系数得到两个汉字关于该节点的相似度.

计算完成后排序选取相似度最大的.



#### 字符插入攻击

##### 选择日语平假名片假名等对汉语认读影响不大的字符进行插入.

### 以 Unicode 为核心的形态学攻击

#### 同形异码字

##### 使用 Unicode 中具有类似形态但是编码不同的汉字作为替换的对象进行攻击

#### Unicode 顺序攻击

## 实验结果

### 多模型运行结果

### 和基线模型进行对比

## 总结和展望

### 优势

### 不足

### 继续努力的方向