# 正则模板演化项目 ——目标说明

## 目标

此正则模板演化项目旨在为交通信息短文本分类提供一种可动态演化的正则表达式匹配模板。

## 方法

使用演化算法，例如，遗传算法，对问题进行求解。

1. 编码：依据样本中的高频词和正则表达式的限制进行编码。
2. 适应度计算：给定的正则模板对各类样本进行分类，以对应类的准确率和误分类比例之差作为适应度。
3. 演化算子：依据所选择的演化算法设计相应的演化算子。

## 数据

数据包含4类样本，分别标记为1，2，3，4类，其中1类为交通事故，2类为设施故障，3类为路况信息，4类为交通热点。

各类数据的数量如下表所示。

|  |  |
| --- | --- |
| 类型 | 数量 |
| 1 | 1174 |
| 2 | 140 |
| 3 | 6201 |
| 4 | 1712 |
| 汇总 | 9227 |