

20231213技术协助需求

一、交警执法规范音频分析

1.1 背景

从交警执法记录仪录制的视频中提取音频部分，分析交警是否在执法过程中规范用语，是否言语不当。

1.2 技术壁垒



1.2.1 音频情感分析

根据论文描述，情感模型不具备泛化能力。经过我们多次实验，证实此论点。

 在情感分析领域中，音频情感分析属于小众方案，公开数据量单一并且多为专业人员根据需求演绎，因此训练出的公共模型**没有泛化能力**，需要根据实际数据集进行训练。并且在实际使用效果中，仅能够将音频状态分为三种的三分类，**不具备多种复杂情感分析的功能**。

1.2.2 语音识别文字分析

先将语音识别成文字，再根据字典检索文明用语和不当用语。

首先，同一时间多人讲话，即多人语音重叠的时候，对语音识别产生严重的影响，因此进行语音分离任务，但是在公开模型中，模型对采样率的要求使得使用的语音分离模型和语音识别模型不能够同时达到最优状态。

其次，在语音识别中，因沟通中的口语化较多，方言等问题，使得语音识别不完全准确，有误识别情况，方言识别识别准确率更低。

再者，识别出来的文本进行文本匹配过程中，因误识别导致的错别字情况会产生较多的识别率下降，后采用拼音匹配方式进行文本匹配，拼音匹配中会存在因相同拼音字母造成的误识别问题。

1.3 无法解决的问题

- 1、现在开源模型不足以满足对本项目复杂的语音识别；
- 2、考虑自行训练模型的情况，未有足够的数据进行训练，并且方言部分未有人员能够听懂。