




DataFunSummit

语音交互中的无效query识别

崔世起
小米人工智能部-小爱

目录

CONTENTS

PART 01. 无效query介绍

PART 02. 非人机交互识别

PART 03. 意图不明识别

PART 04. 小结

PART 01. 无效query介绍

用户query的类型划分

	Query类型	Query示例	意图
有效query (可以结果满足)	单轮意图明确	*打开客厅空调*	智能家居
	场景意图明确	*千与千寻 (前台应用:music)*	音乐
		换一个 (上一轮：讲个笑话)	笑话
	多意图	*灰姑娘*	音乐|电台|视频
无效query (无法结果满足)	非人机交互	*你没打完呢*	无
	意图不明	*让谁先待会儿下雨*	未知

非人机交互query

定义　query不是用户给设备下发的指令

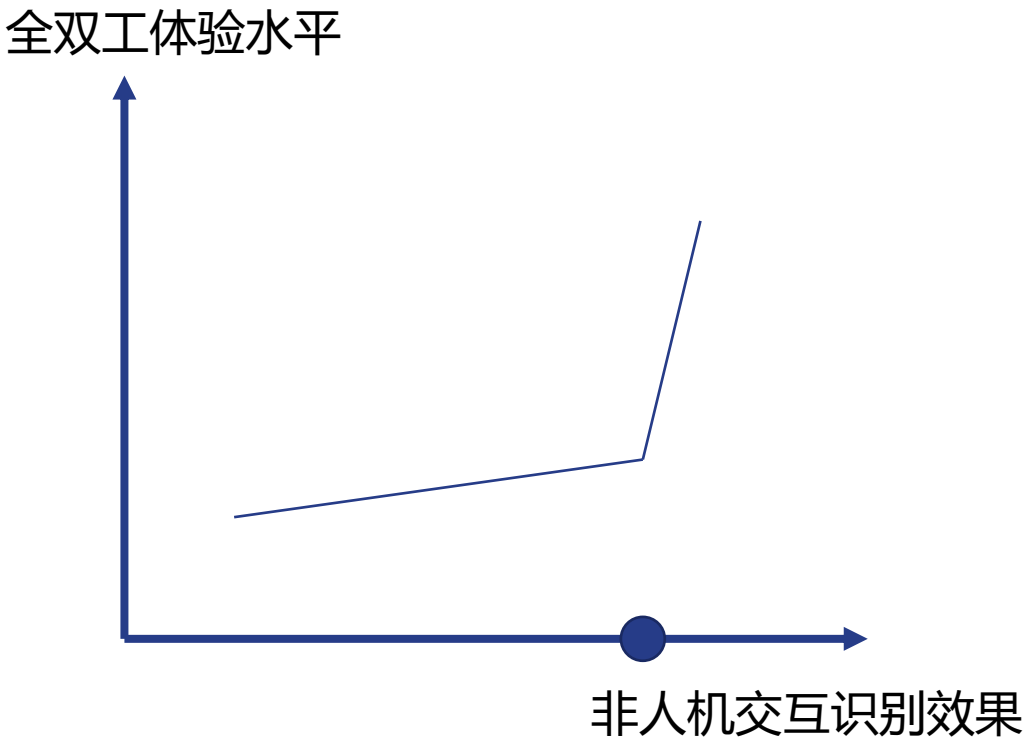
原因　误录入周围的人声

你没打完呢

那个跨年演讲不错

影响　乱响应，打扰用户

影响全双工连续对话体验



意图不明query

定义　query是人机交互指令，但根据query无法判断用户意图

类型　乱序无意义　*播放吴楚丽江穿个*

表达不完整　*帮我把手机*

Query意图比较模糊　*召唤、天空*

影响　误落入精品垂域，答非所问

非人机交互 + 意图不明　占比：5%-20%

无效query体验优化

非人机交互：识别 + 不响应

意图不明：识别 + 兜底回复或者引导澄清

PART 02. 非人机交互识别

非人机交互识别-难点

信息不完备

- 判断是否人机交互，需要多维度的信息（声音、视觉等）
- 只靠音频信息会有很大的歧义性

语音变化的多样性

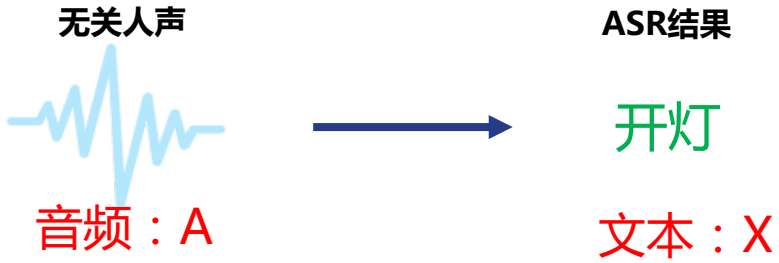
- 同一句话，由于语气、语调、语速、音色的不同，产生不同的音频
- 数据样本难以覆盖每种类型的语音

鸡尾酒会效应

- 嘈杂环境下的有效指令识别

非人机交互识别-问题建模

1. Language understanding Or Speech understanding



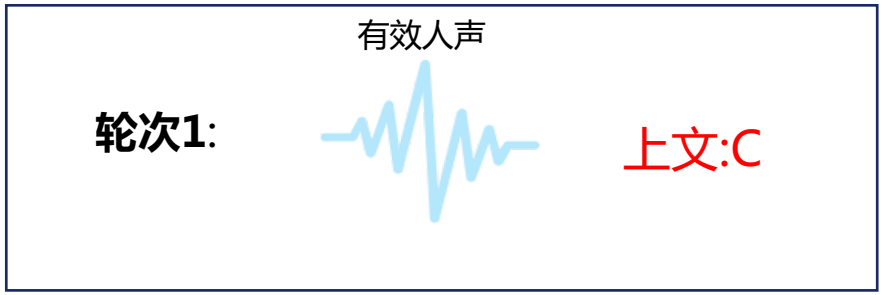
$\mathcal{P}(Y|X)$ ×

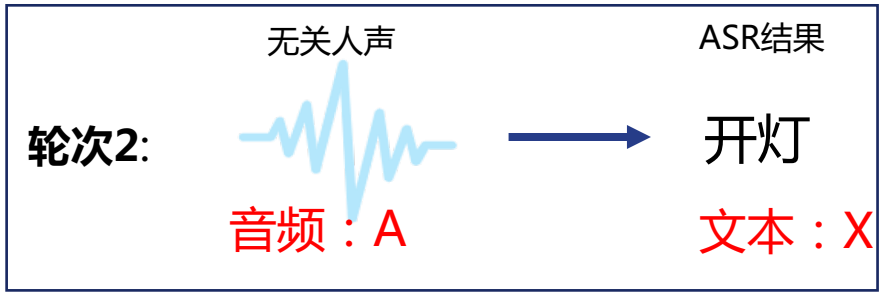
- 缺失了音频中包含的识别信号
- ASR错误传播

$\mathcal{P}(Y|A)$ ✓

非人机交互识别-问题建模

2. 是否需要依赖上文





$\mathcal{P}(Y|X)$ ✗

- 缺少音频特征
- ASR错误传播

$\mathcal{P}(Y|A)$ ✓

$\mathcal{P}(Y|X,C)$?

- 轮次之间弱依赖
- 增加上文依赖，建模难度增大

非人机交互识别-解决方案

建模

针对单轮语音的二分类

关键任务

数据集构建 + 特征和模型设计

非人机交互识别-数据集构造

非人机数据标注成本很高

听音频标注，标注10万条样本耗费100人日。

提升数据标注质量

1. 存在大量的模糊样本，需提高label一致性
   - 详细的标注规范，针对各类型音频提供示例
   - 多人标注验证

数据质量的提升 -> 模型效果的提升

非人机交互识别-数据集构造

样本挖掘

1. 提升样本的多样性
   - 随机采样
   - 正样本挖掘：基于ASR置信度、基于误唤醒检测
2. 提升样本的有效性
   - 挖掘困难样本：模型打分置信度低
   - 挖掘错误分类样本

非人机交互识别-模型分析

语音特征

频谱 优于 mfcc、fbank特征

加入通过声学信号处理获取的特征没有提升

语音Encoder

CNN -> CNN+LSTM+ATTENTION

CNN是个很强的baseline

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：https://d.book118.com/388035046024006032