

电话营销场景意图识别介绍

田兴: 2023/4/27

意图理解

举个例子:

意图: '查收费方式'. 用户句子例句:

- (1) 服务费怎么收?
- (2) 要钱吗?
- (3) 不要利息吗?
- (4) 月利息多少?

可以认为, 以上句子分别包含实体:

- (1) 服务费
- (2) <无实体>
- (3) 利息
- (4) 月利息

因此, 只要得到以下形式, 就基本可以判定用户句子的意思:

- (1) intent: '查收费方式', entity: 服务费;
- (2) intent: '查收费方式', entity: <无实体>;
- (3) intent: '查收费方式', entity: 利息;
- (4) intent: '查收费方式', entity: 月利息;

所以: '意图标签' + '实体标签' 就可以比较好地表示用户的意思.

意图识别构建过程

意图识别就相当于句子分类. 构建意图识别能力的过程如下:

(1) 根据经验总结意图标签体系.

即: 分哪些类别.

- 哪些用户想表达的意图需要做区分?
- 哪些用户的意思可以根据 '意图标签' + '实体标签' 联合区分?

(2) 根据标签体系编写正则表达式.

因为:

- 直接标注足够多的数据成本太高.
 - 有些句子用一个关键词就可以区分了, 全部数据采用人工标注, 太浪费.
- 另外:
- 如果发现模型将一个句子分类错误, 则需要用正则来马上纠正这个错误.
 - 项目早期, 甚至连原始数据都没有. 因此, 需要正则表达式来保障项目运行起来.

所以, 可以通过正则表达式来起到数据预标注的作用.

(3) 根据正则表达式做句子预标注.

当项目运行一段时间后:

- 正则表达式已具备一定的准确性.
- 正则表达式之间常常会冲突 (一个句子可以匹配多个正则), 导致准确率提升困难.
- 正则表达式越来越多, 维护困难.

因此, 根据正则表达式对句子做预标注.

(4)根据预标注数据做模型训练.

- 假设, 预标注的数据预期可以达到 70% 准确率. 则训练的模型的准确率估计不会超过 70%.
- 可以对模型识别的标签与正则识别的标签不一致的样本进行人工标注. 这大大减少了标注量.

根据标签体系编写正则

目前, 汉语, 英语, 日语, 越南语, 已经基本上确定了 **标签体系**.

- 我们会在平台上把 **标签体系** 先建立好.
- 我们会提供已有标注数据的 excel.

训练师需要:

- 多看 excel, 弄清楚已有哪些意图标签. 根据意图标签对应的句子, 理解这个标签的含义.
- 对于标签理解不清楚的, 具体的句子应该属于哪个标签, 应找我们讨论.
- 最好不要自行增删标签. 如果认为标签体系不合理, 不好用, 应该找我们讨论.
- 即使是不同语言, 标签体系也应该是具有很大的相似性的, 对不清楚的, 要多沟通, 确保大家对标签的理解是一致的, 不要搞到最后, 每个人标准都不一样.
- 要多维护, 各场景中要多引用. 要理解正则表达式的重要性, 把它良好地管理起来, 让每个标签下的正则越来越准.
- 各语言的意图库应有一个主要负责维护的训练师. 并输出文档, 内容包括: 常见词对照翻译, 语言介绍, 词干(构词)介绍.

汉语的标签体系介绍:

其它答复:

别担心, 做自我介绍, 不是本人, 改天再谈, 会按时处理, 会帮忙转告, 价格太高, 考虑一下, 请求谅解, 时间推迟, 无所谓, 太少太低

否定答复:

否定(姓名), 否定(不方便), 否定(不可以), 否定(不能), 否定(不清楚), 否定(不是), 否定(不想要), 否定(不需要), 否定(不用了), 否定(不知道), 否定(错误), 否定(没时间), 否定(没兴趣), 否定(没有), 否定(取消)

实体词:

实体(时间), 实体(金额), 实体(地址), 实体(人名), 实体(时长), 实体(组织)

常见疑问:

疑问(原因), 查公司地址, 查自我介绍, 查操作流程, 查公司介绍, 查联系方式, 查收费方式, 查物品信息, 查详细信息, 查优惠政策, 疑问(地址), 疑问(能否), 疑问(时间), 疑问(时长), 疑问(数值), 疑问(姓名)

常见质疑:

不信任, 答非所问, 打错电话, 号码来源, 骚扰电话, 是否机器人, 投诉警告, 质疑来电号码, 质疑隐私安全,

拒答话题:

政治敏感

无关领域:

无关领域

状态(已经):

已经购买, 已经记录, 已经完成

聊天用语:

结束用语, 礼貌用语, 问意图, 我在, 污言秽语, 语气词, 赞美用语, 招呼用语, 祝福用语, 祝贺用语,

肯定答复:

肯定(好了), 肯定(没问题), 肯定(嗯嗯), 肯定(是的), 肯定(需要), 肯定(有), 肯定(正确), 肯定(知道了), 肯定(好的), 肯定(可以)

资金状态:

资金充足, 资金困难

通话指令:

加快速度, 你还在吗, 请等一等, 请讲, 请讲重点, 听不清楚, 听我说话, 未能理解, 信号不好, 要求复述, 用户正忙, 重复一次, 转人工客服, 语音信箱

标签介绍

1. 通话指令:

主要是控制通话的一些触发意图.

2. 常见质疑:

营销电话经常遇到的用户质疑.

3. 答非所问 和 无关领域:

答非所问: 是指用户说机器人答非所问, 没有理解到用户的意思.

无关领域: 是用户答非所问, 比如对方是噪音识别出文字, 小孩子说话, 背景音, ASR识别错误, 等等.

4. 骚扰电话, 投诉警告, 否定(没时间) 和 用户正忙:

骚扰电话, 例句:

不要再打了吧.
怎么一直一直打.
你银行能不打电话了吗.
呃叫他不要来打扰我了.

投诉警告, 例句:

再问我就要投诉了
再打电话给我, 我就投诉你
我报警哦

否定(没时间), 例句:

哦哦, 没有时间.
我7点不行谢谢了.
也没时间了不要了.
我现在没时间, 谢谢.

用户正忙, 例句:

再在有点忙。
现在开车啊。
我，我要开车。

5. 听不清楚, 未能理解 和 要求复述:

这三个意图在平台配置时往往都是配到 用户不清楚, 触发机器人重复话术.

听不清楚, 例句:

听不见听不见。
太吵了，听不到。
不好意思，刚刚没听清。
抱歉，我没有听到声音。
你说什么我都没听到。

未能理解, 例句:

我不明白。
是什么意思啊。
你说啥。
你说说什么。
我不明白唉。

要求复述, 例句:

上一句是什么？再说一遍。
整个重复了一遍。
行前面说什么？
哎，你再说一遍听不懂。