

前言

探意机器人是以电话为媒介载体的语音交互机器人，搭载着 CRM 客户管理的功能展现给用户。探意机器人基于一知智能底层的语音识别（ASR，语义理解（NLP），语音合成（TTS）等基础设施型产品之上，搭载着对话管理，客户管理，用户画像等技术，表现为一个中台应用产品。

一、系统各组成部分功能及结构介绍

1.1 功能说明

探意智能语音机器人是一款应用于电话沟通场景的智能机器人，其具有强大逻辑理解思维，完善的语境应对策略，实现真人语音交互，语义精准识别。该产品由以下模块构成：呼叫中心、AI 机器人、客户管理、数据统计、系统管理、帮助中心。

1.2 产品系统架构

1.2.1 应用部署图

探意机器人分为资源层、中间件层、服务层与功能层四个层次，虚线为客户可能不需要的资源与能力，如下图所示（略）：

一知智能支持给客户私有化部署系统。在客户自有机房里安装服务器设备，安装中间件层或对接客户的数据库，服务层也全部安装在客户机房内，包括 ASR、TTS 和语义理解引擎（需要超高配服务器）。

客户可以使用公网访问探意 CRM 系统，也可以使用 Web API。

二、系统构造训练细节介绍

本系统的语义理解部分：

1. 语句编码模块使用了基于多层 transformer 的深度学习语义编码。此模块是基于大量语料预训练之后，参数固定的。在此基础之上根据不同的下游任务，建立模型，并进行微调。
2. 训练过程中采用了梯度下降的方法，对正确意图和预测意图之间的交叉熵损失函数进行优化。
3. 将 f1 score 作为模型效果的衡量指标，并且加入了早停的机制防止模型过拟合，保存下效果最好的模型。

一知智能作为一家研发型人工智能公司，在探意机器人上应用了很多自研的 AI 技术：

——意图学习算法

探意基于 Transfer learning 的方法，依靠参加机器阅读理解比赛 squad 等国际赛事积累的技术经验，以及公司长期积累的海量语料，针对不同的领域训练不同的模型，然后根据语义的相似度对语句进行聚类以及再学习。

该技术使得智能机器人具备再学习能力，更加智能。在智能语言话术方面，探意将未识别问题归为几大类并按占比排序。对客户来说，需要重点关注与完善的 AI 话术一目了然，省时省力又高效！

——TTS（语音合成）自主技术

探意拥有完全自主产权的 TTS 语音变量合成技术。

与行业普遍调用阿里云、百度云等 TTS 服务不同，探意应用自主研发的基于 Tacotron 的语音合成，不论在发音精确性，还是在灵活定制发声方面，都具备天生的相对优势。

——基于 NLU 技术的语义理解能力

智能机器人行业大部分产品普遍采用“正则表达式”实现语义理解，它的优势主要在于占用计算资源较少。而 NLU（自然语言理解）通过复杂的算法模型来判断说话者的意图，具有“正则表达式”无法比拟的灵活性和丰富性。

探意搭载的 NLU 算法模型在不断优化迭代后，精度已逼近 90%。通过有效整合“正则表达式 + Transformer-based 意图匹配模型”两种技术手段，使得关键词未能识别的“漏网之鱼”，通过 NLU 这张大网进一步识别，帮助 AI 准确理解客户语义，极大提升了智能客服的处理能力。

——客户性别识别

探意基于基音检测算法，通过获取通话记录的音频数据，完整科学的语音处理流程，最终实现根据基音频率的高低区分性别。

——情绪识别

传统的情绪识别是基于词典的方法，自动分析语音文本中的情绪字眼，从而识别说话者的喜、哀、怒、惧等情绪。这种方法的识别能力受限于词典质量的好坏，此外，一些本身带有歧义的词，不能简单地判定是正面或者负面的情绪。

探意利用自主研发的情绪识别模型，以及平台上丰富的聊天语料对系统进行大量训练，能够精确探测用户语言情绪，通过情绪分析识别用户的意向程度。

——早期媒体识别

探意利用传统的 MFCC 音频声学特征，结合 DNN 深度学习网络，对电话音频进行分类。根据电话开头的提示音特征，将通话情况分为空号、停机、欠费、通话中、正常等 8 个类别，方便直观地查看呼出情况。

——客户属性识别

传统的属性识别基于词典和关键词或者是规则，这种方法简单，易快速构建，但缺点是实用性差，人工维护量大。探意使用了基于实体识别的方法，在大数据预训练的词向量基础上，针对差异化的用户需求，训练不同的深度神经网络来进行属性识别。

——语境分析 辅助多轮对话

当客户提问存在多个答案的模糊问题时，传统机器人可能会引发问答冲突，需要客户主动去修改关键词、优化知识库。而探意 AI 可基于上下文语境分析客户想咨询的具体问题，给客户满意的答复。