

# 场景式与检索式问答机器人

# 目录

- 产品概述
- 场景式对话机器人
- 任务型对话机器人
  - ◆ 主动型对话机器人
  - ◆ 被动型对话机器人
- 问答型对话机器人
  - ◆ FAQ检索问答
  - ◆ FAQ知识库问题挖掘

# 产品概述

对话机器人是以智能对话系统为核心，应用于客服/营销/企业信息服务等多方场景的产品。对话机器人以文本、语音和多模态的产品形式，辅助或替代人工对话，赋能对话全流程以实现降本增效。

依据应用范围，我将它分为两种类别：

1. 行业场景式对话机器人。（淘宝客服）
2. 通用情感式聊天机器人。（微软小冰）

# 场景式对话机器人

场景式对话机器人，聚焦于具体行业领域。根据对话目的不同，主要可分为任务型和问答型对话机器人，旨在以最短的对话轮次满足用户的轮次满足用户的任务需求。比如智能客服、营销等服务场景中。



## 任务型

用户希望完成特定任务，对话机器人理解用户意图后，执行后台已对接能力，完成指定任务并回复。

举例：帮我订一张去上海的火车票。



## 问答型

用户希望得到某个问题的答案，对话机器人匹配到所需答案并回答客户。

举例：请问如何在app中绑定银行卡。

# 任务型对话机器人

任务型对话机器人根据用户定位不同，我将它分为主动型和被动型这两种情况。

- 主动型是指用户主动提问，并且能很好归纳概括自己的问题需求，系统来理解用户需求和意图。
- 被动型是指客户没有主动提问能力或者不知道对话系统能力，由系统给用户推送可以解决的问题，用户进行选择或者回答。

# 主动型对话机器人

- 自然语言理解 (NLU)

将用户的输入映射到预先根据不同场景定义的语义槽中，让机器理解语言的意思，通常包含领域检测、意图识别和语义槽填充。

- 对话管理 (DM)

考虑历史对话信息和上下文的语境进行全面分析，决定系统要采取的相应动作，主要任务有对话状态跟踪、生成对话策略。

- 自然语言生成 (NLG)

将机器输出的抽象表达转换为句法合法、语义准确的自然语言句子，以自然语言文本去表达给定的意图。

# 主动型对话机器人

## 自然语言理解 (NLU)

### ➤ 意图识别。

- 识别用户说话意图，比如订票任务中常见意图有订火车票。
- 单意图识别可以认为是一个多分类任务。
- 多意图识别可以认为是一个多标签任务。

### ➤ 语义槽填充

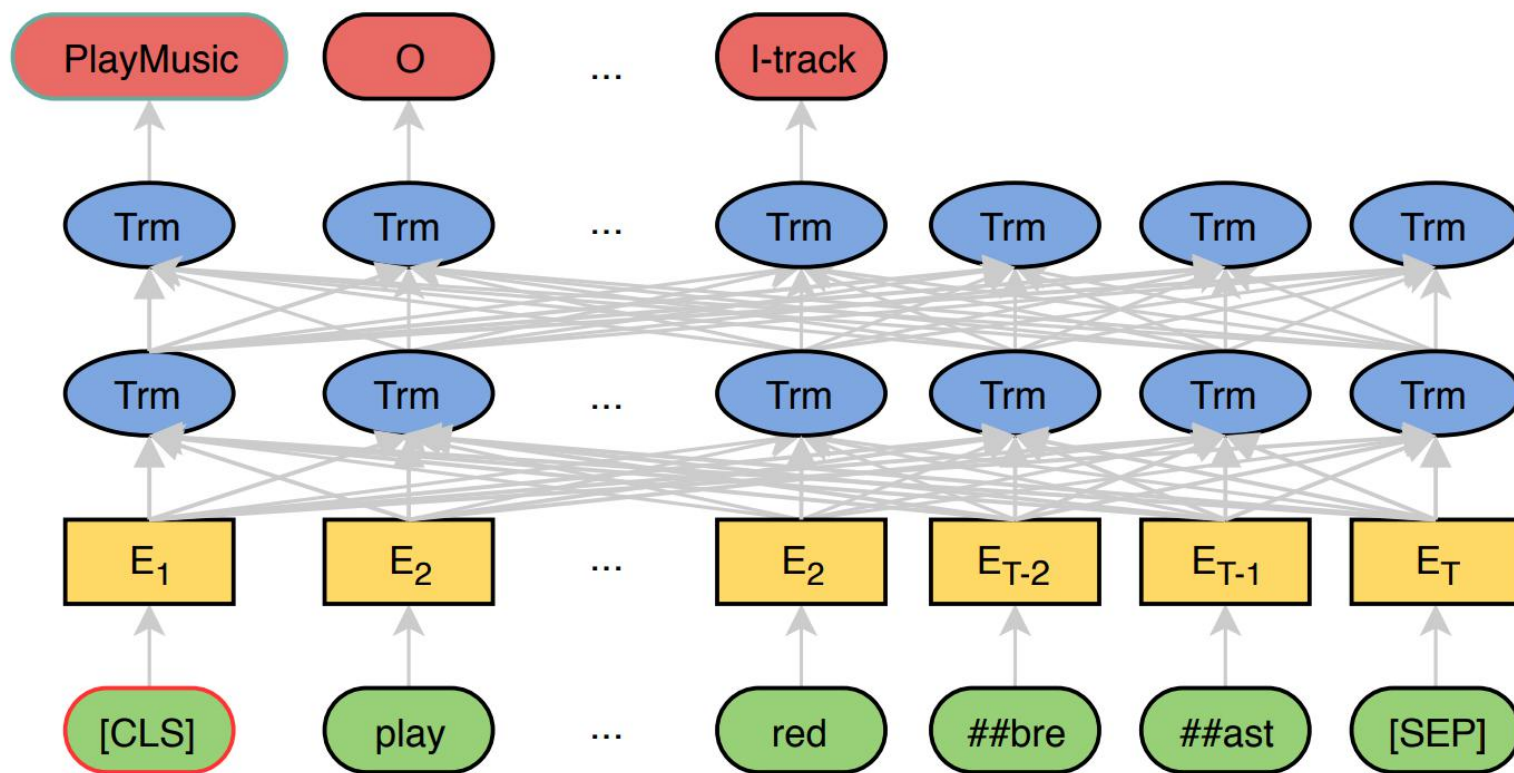
- 识别用户输入中语义槽，即任务实体对象，比如订票任务中出发时间、出发地点、目的地点等都是该任务的语义槽。
- 槽识别任务可以认为是一个序列标注任务。

举例：我想订一张去上海的火车票。

意图：订火车票。 槽位：目的地(上海)，缺失槽位：出发地、出发时间等。

# 主动型对话机器人

意图识别和语义槽识别—JointBert模型<sup>[1]</sup>



[1]. <https://arxiv.org/abs/1902.10909>. BERT for Joint Intent Classification and Slot Filling

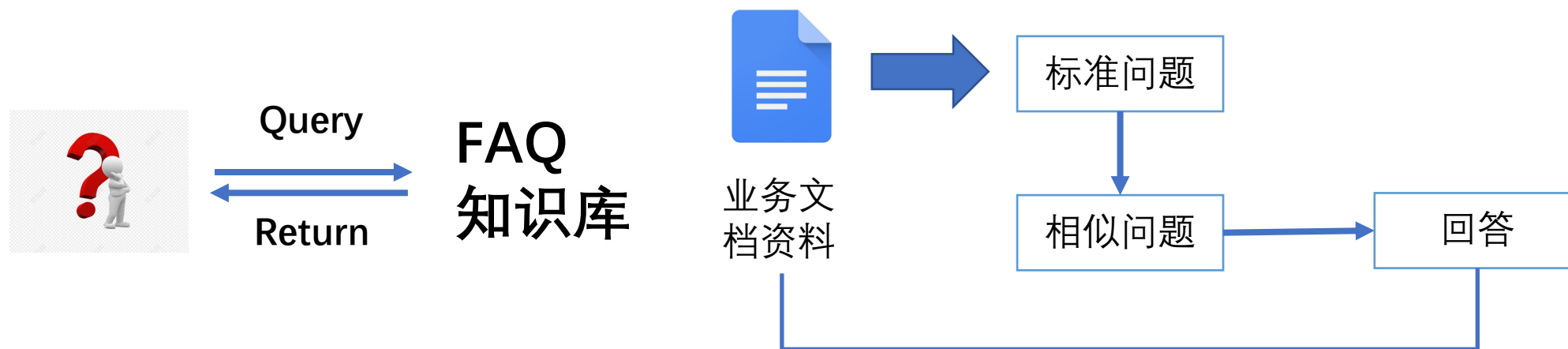


# 被动型对话机器人

- 对话系统主动给用户提供服务，供用户进行选择，并根据用户的选择来转移对话状态。
- 优点：可在有限状态之间转移，给予用户明确的回复，整个对话逻辑可控。缺点：业务逻辑固定单一，适合有限状态集的业务模式，对于客户的主动提问缺乏合适的应对。
- 可与检索式对话机器人、用户画像等相结合，设计成一套对话产品。
- 举例：猜您想要了解
  1. 今日大盘运势
  2. 今日领涨板块
  3. 今日热门板块
  4. 关注股票情况

# 问答型对话机器人

FAQ检索型问答（FAQ, Frequently Asked Questions常见问题集的问答系统），作为基础发展的问答系统，被对话机器人厂商广泛使用。FAQ问答是基于业务文档资料整理标准问题、相似问题与其相应答案。当问题输入后，系统会在知识库内查询与之相匹配的问句并输出对应问答。



# 问答型对话机器人

## 核心技术—文本匹配算法

- 传统文本匹配（信息检索中的BM25， TF-IDF等）
- 深度语义匹配（以Bert为代表的预训练语言模型）

### 1. 双塔式

经过语言模型获取语句向量表征，最后进行向量相似度计算。

### 2. 交互式

在模型内部进行token的深度交互，最终得出文本匹配度。

# 问答型对话机器人

## 挖掘常见客户问题

1. 存在大量无标签问题语料，需要从中挖掘出用户高频问题，丰富FAQ知识库。
2. Bert模型存在向量表征无区分度，对比学习中所说的模型坍塌问题。
3. 针对以上两点问题，我们使用文本对比学习方法SimCSE来训练语料表征。

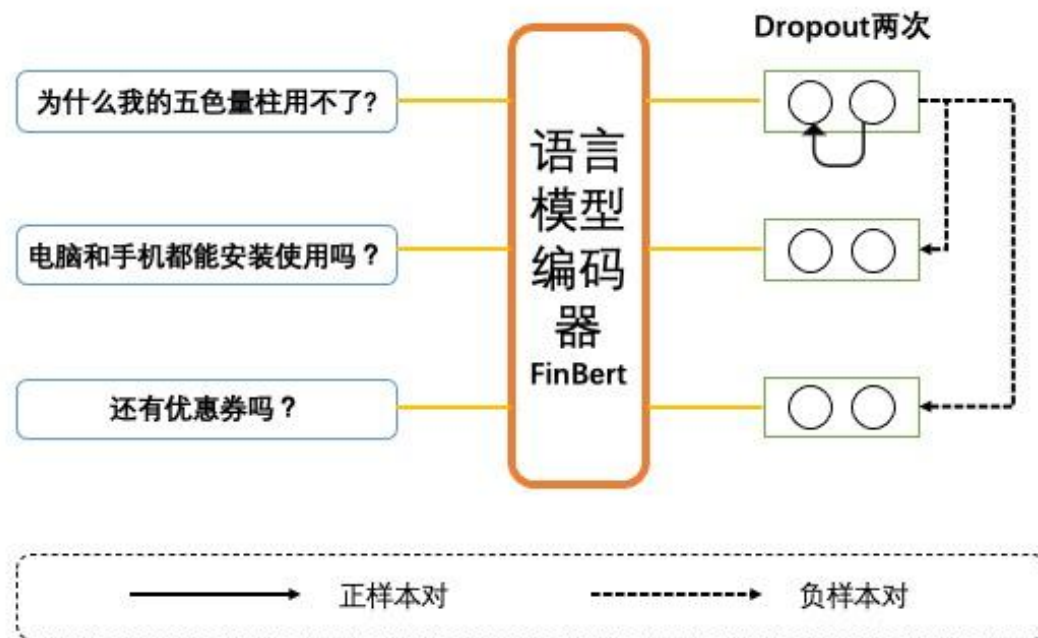
# 问答型对话机器人

## Unsupervised SimCSE

1. 使用Dropout来对文本增加噪音，从而构造一个正样本对。
2. 负样本对使用batch内的其他句子。
3. 整体效果要优于传统EDA、回译等方法。

# 问答型对话机器人

## 文档语料挖掘常见问题



感谢聆听