

# ChatBot-Survey

lumengyi

December 2018

## 1 概述

聊天机器人 (Chatbot)，也被称为对话引擎或者对话系统，是目前的热点之一。

从对话目的上分为任务型（特定领域）对话与非任务型/闲聊型（开放领域）对话，从模型技术上分为基于检索技术模型与生成式模型，从问题的分为长（多轮）对话与短对话（单轮）。

### 1.1 模型分类

#### 1.1.1 检索技术模型 VS 生成式模型

基于检索技术的模型较为简单，主要是根据用户的输入和上下文内容，使用了知识库（存储了事先定义好的回复内容）和一些启发式方法来得到一个合适的回复。启发式方法简单的有基于规则的表达式匹配，复杂的有一些机器学习里的分类器。这些系统不能够生成任何新的内容，只是从一个固定的数据集中找到合适的内容作为回复。

生成式模型则更加复杂，它不依赖于预定义好的回复内容，而是通过抓取 (Scratch) 的方法生成新的回复内容。生成式模型典型的有基于机器翻译模型的，与传统机器翻译模型不同的是，生成式模型的任务不是将一句话翻译成其他语言的一句话，而是将用户的输入 [翻译] 为一个回答 (response)。

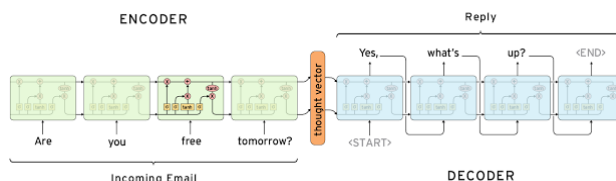


图 1: 检索技术模型 VS 生成式模型.

## 1.2 问题分类

### 1.2.1 短对话 VS 长对话

直观上处理长对话内容将更加困难，这是因为你需要在当前对话的情境下知道之前的对话说过什么。如果是一问一答的形式，技术上这将简单的多。通常对于客服对话而言，长对话更加常见，一次对话中往往会伴随着多个关联问题。2.2 开放域 VS 特定领域

面向开放域的聊天机器人技术面临更多困难，这是因为会话可能涉及的面太广，没有一个清晰的目标和意图。在一些社交网站例如 Twitter 和 Reddit 上的会话是属于开放域的，会话涉及的主题多种多样，需要的知识量也将非常巨大。

面向特定领域的相关技术则相对简单一些，这是因为特定领域给会话的主题进行了限制，目标和意图也更加清晰，典型的例子有客服系统助手和购物助手。这些系统通常是为了完成某些特定任务，尽管用户在该系统中也能够问些其他方面的东西，但是系统并不会给出相应的回复。

## 1.3 章节安排

本文首先简要介绍了智能聊天机器人的分类和研究背景，对市场上的模型平台进行比较，然后，给出了一些参考的开源框架，并总结了搭建智能问答系统的框架流程，最后，介绍了几篇后续可参考的论文。

## 2 市场上的模型平台

### 2.1 百度的 DuerOS

**DuerOS** DuerOS 开放平台主要面向企业级用户及个人开发者，提供对话式操作系统能力的输出（DuerOS 智能设备开放平台）及输入（DuerOS 技能开放平台）。

### 2.2 百度的 UNIT

**UNIT (Understanding and Interaction Technology)**，即理解与交互技术，它是建立在百度多年积累的自然语言处理与对话技术以及大数据的基础上，面向第三方开发者提供的对话系统开发平台。

文档全，有使用教程。

### 2.3 微软的 Bot Framework

**微软 Bot Framework** 是一个用于搭建、链接、测试、和部署智能机器人的平台。使用微软 Bot 框架来构建机器人，可与网站、SMS, Skype, Slack, Messenger, Office 365 邮件等进行交互。Bot Builder SDK 允许构建对话框。认知服务使机器人以更人性化的方式看到，听到，解释和交互。连接器将让机器人在最流行的对话体验中以完全保真的方式说话，无需额外的编码。可使用各种工具来开发机器人，如 QnA Maker 和语言理解。是一个应用层的框架，结合 LUIS (Language Understanding Intelligent Service)

使用该框架较多，有比较多的参考资料，感觉构建起来不是特别方便。

### 2.4 RasaCore

**RasaCore**是一个用来处理对话流程的框架，基于 python 实现。

根据搜索结果来看使用该框架的人还可以，资料较全，有官方文档，构建起来比较容易上手。

### 2.5 Pandorabots (chatbots.io)

**Pandorabots** 支持多种语言，支持包括给 Java, Node.js, Python, Ruby, PHP, Go 等语言。支持 AIML 的解析，上传，提供平台服务，根据机器人

的访问次数和数量计费。基于 aiml (AIML 是一种为了匹配模式和确定响应而进行规则定义的 XML 格式)，资料较少。

## 2.6 Facebook 的 wit.ai

**wit.ai**是 Facebook 推出的用于将自然语言转化为可处理指令的 API 平台，可免费使用，需翻墙。英文文档 <https://wit.ai/docs/recipes>。

## 2.7 Google 的 api.ai

**api.ai**是一款聊天机器人 API，功能类似 Wit.ai，也包含意图和实例的概念。不完全免费，需翻墙。

## 2.8 Motion.ai

**Motion.ai** 能够开发面向多个平台的 bot，包括 Facebook、Slack、SMS 短信、电子邮件、web 网站等等。用户只需要“拼装”Motion AI 的 bot 商店里的现成模块就能迅速实现各种功能，例如客户服务 bot，能够链接 Zendesk 等软件，扫描企业知识库自动响应客户问题。

参考资料较少。

## 2.9 世纪佳缘的 yigeai

一个 **ai** (一个 AI - 让对话无处不在)，是由世纪佳缘开发的免费创建虚拟对话机器人 (Bot) 应用的在线平台。利用一个 AI，开发者甚至产品和运营人员都可以轻松地开发聊天/对话机器人应用，而不需要具备机器学习与自然语言处理等相关知识。

## 2.10 海知智能的 ruyiai

**ruyi.ai** 个性化聊天机器人开放技术平台，面向模板进行机器学习的知识库问答系统。开放对接第三方 bot 平台。

## 3 可参考的开源框架

### 3.1 chatterbot

**chatterbot** 基于 Python，看上去比较完整的一个项目，安装简单，说明比较完善，有提供训练的材料，但以英文为主，并好像用于基于开放领域的多，基于特定领域（任务型）效果不太好。

### 3.2 DeepPavlov

**DeepPavlov** 是一个开源的对话 AI 库，建立在 TensorFlow 和 Keras 上。它为研究者提供：

- 用于实现和测试他们自己的对话模型并随后将模型共享的框架；
- 一系列预定义的 NLP 模型/对话系统组件（机器学习/深度学习/规则系统）和流程模板；
- 对话模型的基准测试环境和对相关数据的系统性评估。

并为 AI 应用开发者提供：

- 建立对话软件的框架；
- 将应用与对应基础建设（通讯、技术支持软件等）相集成的工具。

文档比较详细。

### 3.3 DeepQA

**DeepQA** 基于采用 seq2seq 结构实现的一个开源模型。

### 3.4 Seq2seqChatbotQA

**Seq2seqChatbotQA**

### 3.5 Seq2seqChatbotQA

**QA-Snake** 是一个采用 Python 编写的基于多搜索引擎和 AIML 技术的问答机器人，

### 3.6 搭建流程框架

有三个重要模块：提问处理模块、检索模块、答案抽取模块

\* 提问处理模块

查询关键词生成：查询关键词生成，从提问中提取出关键的几个关键词。

答案类型确定：确定你的提问属于哪一类的，根据答案类型确定后面所需的技术方案。

\* 检索模块

检索模块跟搜索引擎比较像，就是根据查询关键词所信息检索，返回句子或段落。

\* 答案抽取模块

答案抽取模块，通过分析和推理从检索出的句子或段落里抽取和提问一致的实体，再根据概率最大对候选答案排序，从多个结果中选出一个来。

## 4 相关论文

1. Salman Mohammed, Peng Shi, and Jimmy Lin. Strong Baselines for Simple Question Answering over Knowledge Graphs with and without Neural Networks.ACL2018.

2. (KDD 2018 tutorial) End-to-end Goal-Oriented Question Answering Systems

3. DocChat: An Information Retrieval Approach for Chatbot Engines Using Unstructured Documents. MicroSoft

## 5 Reference

- 《Deep Learning For Chatbots, Part 1 - Introduction》
- 《聊天机器人中的深度学习导读》
- 中文的一个聊天机器人教程
- 用 AIML 开发人工智能聊天机器人
- 如何基于 rasa 搭建一个中文对话系统

- [手把手教你利用微软的 Bot Framework, LUIS, QnA Maker 做一个简单的对话机器人](#)
- [DeepQA 近距离观察](#)
- [Chatbot Research 5 - 基于深度学习的检索聊天机器人](#)
- [如何设计一个基于检索模型的对话机器人](#)
- [Eric, 基于多搜索引擎的自动问答机器人](#)
- [DeepPavlov: 一个训练对话系统和聊天机器人的开源库](#)
- [如何开发一个特定领域的自动问答机器人 \(Chat Bot\)?](#)