# 第1章 聊天机器人概述

随着人工智能的发展，很多以前仅出现在科幻小说与科幻电影中的场景正在逐渐变成现实。我们过去对于未来世界的想象中，往往会有这样的场景，每个人都拥有一个智能助手或者智能管家，帮助主人完成许多日常工作，比如根据当天的天气信息、交通信息等，帮助安排出行计划、安排会议时间地点、帮助预定晚餐等。

这些未来世界随处可见的场景中均属于人工智能应用的主要场景，其中涉及语音识别、人机交互、自然语言处理、计算机视觉技术等技术，这些技术的集合体即在于帮助我们实现一种可以理解我们人类语言的机器，进而通过人类的语言指令帮助我们完成相应的工作任务，这种可以与人类通过语音交流的机器也就是我们要介绍的聊天机器人。近年来，关于聊天机器人的研究在学术界和工业界都得到了广泛的关注，其中聊天机器人在学术界受到青睐是由于聊天机器人是图灵测试的一种实现方式，而图灵测试是人工智能领域王冠上最璀璨的明珠，聊天机器人作为人工智能领域“最后的战场”，被认为隐含着上帝造物的秘密。而在工业界，随着各大互联网巨头纷纷推出自己的聊天机器人产品，比如微软推出了基于情感计算的聊天机器人小冰，百度推出了用于交互式搜索的聊天机器人小度，大大推动了聊天机器人产品化的发展。此外聊天机器人系统可以看作是机器人产业与“互联网+”的结合，也符合国家的科研及产业化发展的未来方向。

从本章节开始，我们将带大家走进聊天机器人的世界，我们将通过实际的项目例子带大家一步一步完成不同聊天机器人的搭建与部署，不仅帮助大家了解聊天机器人的实现原理，更帮助大家从零开始构建一个属于自己的聊天机器人。

## 1.1 聊天机器人概述

聊天机器人是一种通过自然语言模拟人类，进而与人类进行沟通对话的程序系统，可以运行在特定的平台如PC或者移动终端，也可以结合机器人设备及其他类人的硬件上来运行。聊天机器人的研究源于图灵提出的图灵测试，图灵测试是人工智能这个当下最火热的研究领域的起点，那么什么是图灵测试呢？

### 图灵测试

被誉为人工智能之父的阿兰·图灵（Alan Turing）于1950年在Mind期刊上发表了文章《机器具有智能吗？》（Computing Machinery and Intelligence），提出了著名的图灵测试（Turing Test）。



图1.1 阿兰·图灵

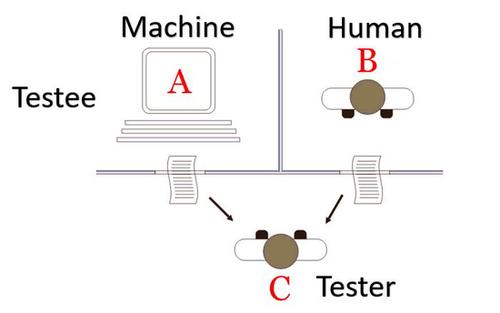


图1.2 图灵测试

图灵测试是指将被测试的机器与人分别于测试者隔离开来，在互相不接触对方的情况下，让测试者分别与二者进行问答。如果在相当长时间内，测试者无法根据这些问题判断对方是人还是计算机。那么，就可以认为这个机器具有同人相当的智力，即这台计算机是能思维的。图灵测试是促使人工智能从哲学探讨转换为科学研究的一个重要因素，引导了后来人工智能的很多研究方向，其中人与机器之间交互对话的过程也就是最早的聊天机器人的雏形。

### 1.1.2 聊天机器人的发展历程

受图灵测试的启发，早期聊天机器人是使用音频或文本进行对话的计算机程序。这类程序的设计通常模拟人类作为聊天对象的行为，并以图灵测试作为是否成功的标准。

目前公认最早的聊天机器人程序是1966年由麻省理工学院（MIT）的约瑟夫·维泽保姆（Joseph Weizenbaum）开发的ELIZA，维泽保姆的本意是设计一个可以模仿精神科医生对话的系统，通过对用户输入的语句进行模式匹配和智能短语搜索，然后基于预先设置的对话脚本与用户交流，这些脚本可以模仿一个罗杰斯学派的心理治疗师，但是实际上它并不理解对话内容。

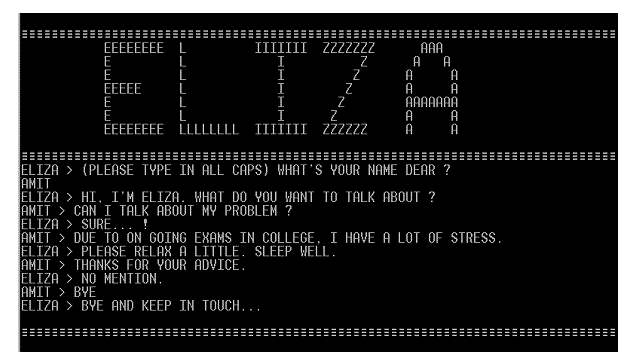


图1.3 最早的聊天机器人ELIZA

在实际对话过程中，ELIZA根据用户的输入不断提出问题，这些问题可能并不相关，但是用户会自动对其进行解读，甚至认为系统了解其真实的情况，认为回复的语句是有意义的。下图展示即为ELIZA与用户之间的一段对话：

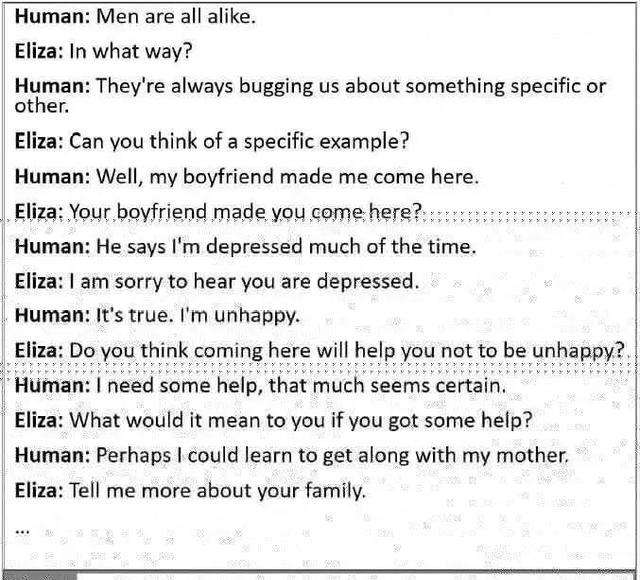


图1.4 ELIZA与用户的一段对话示例

1972年美国精神病学家肯尼斯·科尔比（Kenneth Colby）在斯坦福大学（Stanford University）使用LISP语言编写了聊天机器人PARRY。PARRY的结构类似于 Eliza，但是其会话策略更加先进严谨，同时具备更好的控制结构、语言理解能力，尤其是可模仿机器人情绪的心理模型。研究人员使用一种变体的图灵测试对PARRY进行测试，测试结果表明，其回复正确率达到了随机投票所产生的正确率。

1988年英国程序员罗洛·卡朋特（Rollo Carpenter）创建了聊天机器人Jabberwacky，Jabberwacky是第一个能够模拟人声的聊天机器人。卡朋特当时创建这个聊天机器人的目的是为了让聊天机器人能够通过图灵测试，该项目也是通过与人类互动创造人工智能聊天机器人的早期尝试。Jabberwacky在对话中使用上下文模式匹配技术返回最合适的回复语句，该项目于1997年正式上线后，通过存储用户与系统之间的对话过程，采集了许多对话数据，推动了后续的研究发展。

1992年微软的Creative Labs基于Jabberwacky的制作脚本而发明了Dr. Sbaitso，这款聊天机器人运行在MS-DOS系统上，它为聊天机器人添加了一个用户界面提供连接，使得其可以更好地模拟心理学家对于相关症状信息的问答过程。

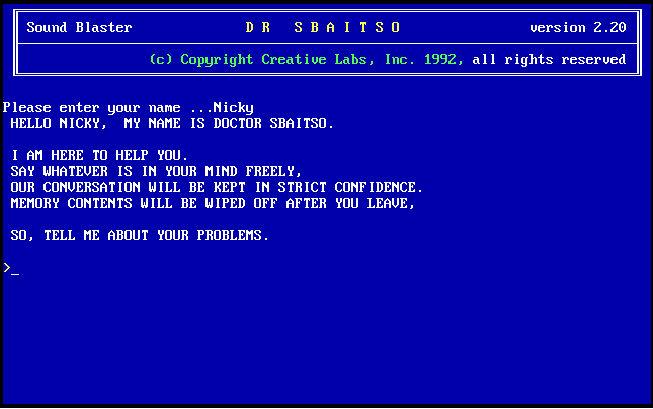


图1.5 Dr. Sbaitso聊天机器人

1995年，业界诞生了标志性的聊天机器人产品，即A.L.I.C.E（Artificial Linguistic Internet Computer Entity）。理查德·华莱士（Richard Wallace）从ELIZA中得到启发，通过对对话数据的统计分析，他发现人们在日常谈话中涉及的谈话主体不过几千个，因此华莱士估算我们大概只需四万个日常回答即可覆盖全部的日常常用对话情况，一旦将这四万个预编程序语言全部输进ALICE，那么它就可以回应95％的人们对它说的话。ALICE使用了一种称之为AIML（ Artificial Intelligence Markup Language，人工智能标记语言）的自定义AI语言，目前仍然被广泛应用在移动端虚拟助手的开发中。尽管ALICE采用的是启发式模板匹配的对话策略，但是他仍然被认为是同类型聊天机器人中性能最好的聊天机器人，分别在 2000、2001 和 2004 年三次斩获勒布纳人工智能奖（Loebner Prize）。

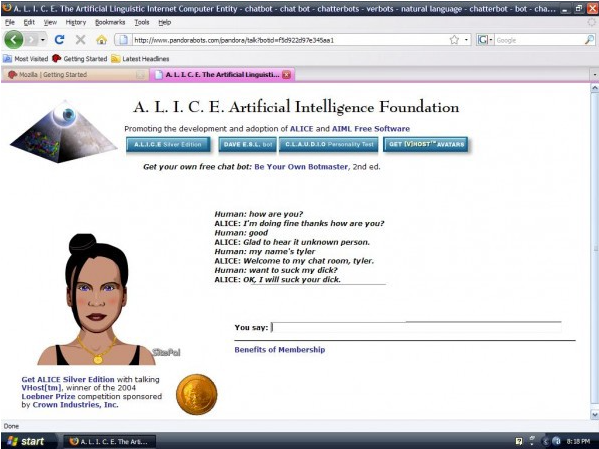


图1.6 ALICE聊天机器人

2001年，一款由ActiveBuddy开发的聊天机器人Smarterchild在全球的即时信息和SMS网络中得到了广泛的应用，使得聊天机器人第一次被应用在了即时通讯领域，它经过预编程能对用户查询给出相应的反馈，比如即时访问新闻、天气、股票信息、电影网站、黄页列表和详细的体育数据等，一定程度上类似于后来的Siri。



图1.7 Smarterchild聊天机器人

2006年，IBM开始研发能够用自然语言回答问题的Watson，结合了信息分析、自然语言处理和机器学习领域的大量技术创新，通过自然语言理解技术和处理结构化和非结构化数据的能力，具有理解自然语言的能力，同时通过对数据的分析使其具有一定的逻辑推理能力，而且像人类一样具有进行学习和认知的能力。Waston最令人印象深刻的是在2011年的Jeopardy比赛中击败了人类选手。



图1.8 Waston击败人类选手

聊天机器人的最新一次重大革新是随着苹果私人助理Siri的诞生而到来的。2010年Siri作为IPhone的一款应用程序出现，后来被集成到IOS操作系统中，可以与IOS操作系统中的应用程序进行互动交互，它的语音识别引擎由Nuance Communications提供，并使用了先进的机器学习技术，Siri除了具备聊天功能，更重要的是突破在于能够通过自然语言交互的形式实现问答、推荐、手机操作等功能，同时支持广泛的用户命令，包括执行电话操作，订餐、订票、放音乐，安排事件和提醒，处理设备设置，搜索互联网，浏览区域，查找娱乐信息等。



图1.9 Siri

继Siri之后，各大互联网巨头公司纷纷加入聊天机器人的开发领域，推出了许多款各具特色的聊天机器人，比如谷歌（Google）于2012年也发布了聊天机器人Google Now，亚马逊于2014年发布了Alexa，微软于2015年发布了Cortana，可以帮助用户设置提醒、识别自然声音以及使用Bing搜索引擎中的信息来回答用户的问题。

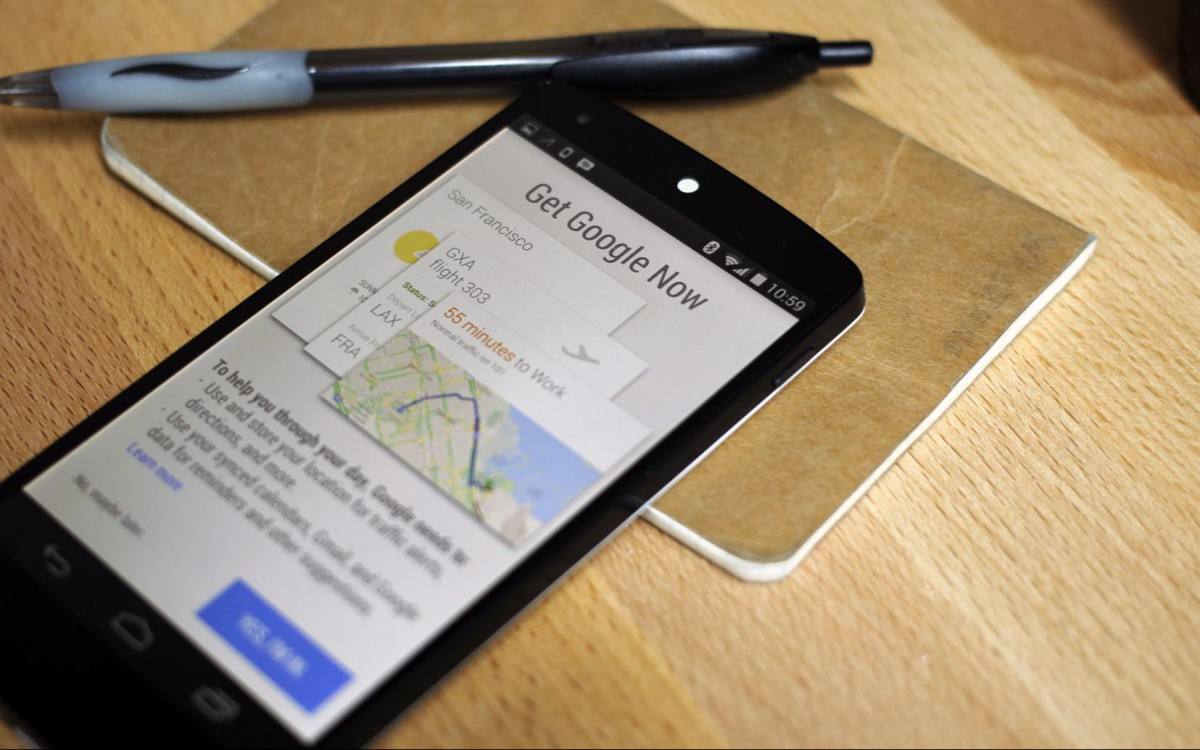


图1.10 Google Now

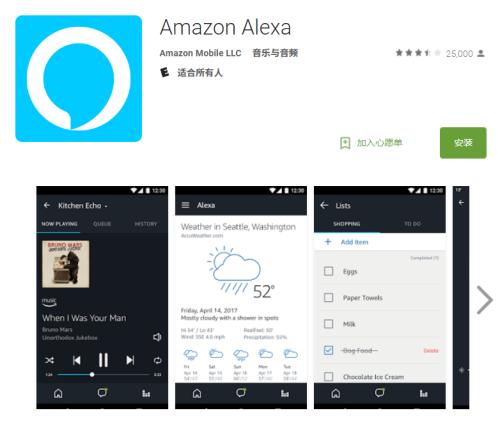


图1.11 Amazon Alexa



图1.12 Microsoft Cortana

此外国内近年来在聊天机器人领域也有快速的发展，涌现了许多优秀的聊天机器人产品，比如阿里巴巴的小蜜，微软的小冰等等。



图1.13 阿里巴巴小蜜



图1.14 微软小冰

## 1.2 聊天机器人的分类与应用场景

在继续接下来的介绍前，我们先介绍下聊天机器人中一些常用的概念及术语，这些术语将在后续的章节中频繁被使用到。

### 1.2.1 相关术语

1）意图（Intent）

当一个用户和聊天机器人进行交互时，用户使用聊天机器人想要完成的任务或者想要获取的信息一般称为用户的意图。例如当用户向聊天机器人发送“帮我定一张电影票”的信息时，那么用户希望传达给聊天机器人的意图即可以命名为“定电影票（book\_movie）”，同样当用户向聊天机器人发送“明天的天气怎么样”或者“帮我查看下天气情况”这种信息时，可以将用户的意图命名为“查询天气信息（query\_weather\_info）”。我们在实际项目中可以基于具体场景的需要定义不同的意图信息。

2）实体（Entities）

在意图中往往包含不同的命名为实体的元数据，同样在上面的例子中“帮我订一张电影票”，其中涉及的实体即为“电影”，同理，实体可以帮扩其他的东西，比如“航班”、“天气”等。我们可以在一个意图中编辑多个实体，例如当用户输入的信息为“我想订一个8寸的奶油蛋糕”，那么这里可能有两个实体，分别是物品类别为“奶油蛋糕”，大小信息为“8寸”。

3）话术（Utterances）

话术是指对同一个问题或者意图的不同表达方式，例如对于查询天气信息，用户可能有不同的表达方式，“查询天气”，“帮我查下天气信息”，“请给我看下明天的天气”，这些表达均表达了同样的意图，但是具体的句子有不同的变化，甚至可能出现句子差异性较大的情况。

### 1.2.2 聊天机器人的分类

近年来，各种聊天机器人层出不穷，对于聊天机器人基于不同的维度有不同的分类情况，常用的分类维度主要有基于应用场景的分类、基于实现方式的分类，基于交互方式的分类以及基于功能的分类方式。

1）基于应用场景分类

从应用场景的分类角度来看，聊天机器人一般可以分为在线客服聊天机器人、娱乐陪伴聊天机器人、教育领域机器人、个人助手和智能问答机器人五类。其中在线客服聊天机器人的主要功能是自动回复用户提出的与产品或服务相关的问题，在企业运营上可以极大地降低企业客服成本、同时提升用户体验。其中具有代表性的商用在线客服聊天机器人系统有京东 JIMI客服机器人、阿里小蜜等。以京东JIMI客服机器人为例，用户可以通过与JIMI聊天了解商品的具体信息、了解平台的活动信息、反馈购物中存在的问题等。另外，对于无法回答的问题，JIMI可以及时将用户转向人工客服。阿里巴巴集团在2015年7月24日发布了一款人工智能购物助理虚拟机器人，取名“阿里小蜜”，阿里小蜜基于客户需求所在的垂直领域（服务、导购、助手等），通过“智能＋人工”的方式提供良好的客户体验。

娱乐陪伴类型的聊天机器人系统的主要功能是同用户进行不限定主题的开放域对话，也就是我们常说的闲聊模式，从而起到陪伴、慰藉等作用。其应用场景集中在社交媒体、儿童陪伴及娱乐等领域。其中代表性的聊天机器人如人人网的小黄鸡和微软的“小冰”等。其中微软的“小冰”除了能够与用户进行开放主题的聊天，还能提供一些特定的服务，例如支持用户询问天气信息、回答用户关于生活常识的疑问等。

另一种具有实际意义的聊天机器人为应用于教育场景下的聊天机器人，这种机器人可以根据教育内容的不同及学生的学习状态进行个性化辅助教学。例如，通过构建交互式的语言使用环境，帮助用户学习某种语言的聊天机器人；在用户学习某项专业技能时，指导用户逐步深入地学习并掌握该技能的聊天机器人；此外还有在用户的不同年龄阶段，针对用户的学习能力可以帮助用户制定个性化的学习方案的聊天机器人，如目前流行的早教机器人等。

个人助理类聊天机器人是指可以通过语音或文字与用户进行交互，帮助用户完成个人事务，如天气查询、短信收发、定位及路线推荐、闹钟及日程提醒、订餐订票等，从而让用户可以更便捷地处理日常事务。个人助理的典型聊天机器人产品为Siri。

智能问答类聊天机器人是指可以回答用户以自然语言形式提出的事实型问题及其他需要计算和逻辑推理的复杂问题，以满足用户的信息需求并起到辅助用户决策的目的。例如回答“中华人民共和国的成立时间是什么时候”、“姚明的出生地在哪”这种事实型问题。智能问答聊天机器人的应用场景相对单一，作为聊天机器人的一个服务子模块。典型的智能问答系统包括 IBM 研发的 Watson。

2）基于实现方式分类

从实现的角度看，聊天机器人主要分为检索式聊天机器人和生成式聊天机器人。检索式聊天机器人的回答是提前定义的，对于用户输入的问题，聊天机器人使用规则搜索引擎、模式匹配或者机器学习训练好的分类器从知识库中挑选一个最佳的回复展示给用户。因此需要预先构建一个知识库，聊天机器人在知识库中以检索的方式进行应答内容提取。这种实现方式受构建的知识库影响较大，展示给用户的内容的质量严重依赖知识库的质量，若构建的知识库质量较差，则检索式聊天机器人往往会出现找不到合适回复的情况。

生成式聊天机器人采用不同的实现方式，不依赖于预先定义的回答，而是在训练聊天机器人的过程中，基于包含上下文信息的大量语料，直接生成用户回复给用户的语句。常用的模型为Encoder-Decoder模型，聊天机器人在接收到用户输入的自然语言后，将输入自动生成一句话作为对用户输入的应答，类似机器翻译的过程，但是输出的结果不是输入的其他语言形式，而是该输入相应的回复内容。生成式聊天机器人的优点是理论上可以覆盖任意话题、任意句式的用户输入，缺点是生成的应答句子的质量很可能存在问题，比如出现语句不通顺、句法错误等比较低级的错误。

当前的生产环境中，常用的聊天机器人实现方式一般都是基于检索的聊天机器人，但是随着基于深度学习的seq2seq（sequence to sequence）模型的出现，基于生成式的聊天机器人将会越来越多。

3）基于交互方式分类

按照交互方式分类，聊天机器人一般可以分为主动交互聊天机器人和被动交互聊天机器人，目前我们接触到的大部分聊天机器人都属于被动交互的范畴，即一般由用户发起对话请求，聊天机器人通过理解用户的对话意图并做出相应的响应。而主动交互聊天机器人由机器人首先发起，通过共享或推荐用户感兴趣或热点事件与用户进行交互，目前主动交互更多作为传统交互方式的一种补充，作为辅助手段使用。

4）基于功能分类

从功能角度来看，聊天机器人主要分为以闲聊为主的闲聊系统，问答系统和任务导向的对话系统这三大类。其中闲聊系统主要是应用于开放域用于娱乐陪伴的聊天机器人，如前面介绍的微软“小冰”；对于问答系统，主要是智能问答类机器人，在其中还可以继续细分为为事实型问答（Factoid QA）、常见问题集问答系统（FQA）和开放域问答系统（Open Domain QA）；对于任务导向的对话系统基本上是通过多轮交互的方式帮助用户实现特定的需求，例如定披萨或者订机票等。

我们接下来的介绍主要按照功能来分类，主要研究问答系统、对话系统及闲聊系统。

### 1.2.3 问答系统

问答系统（Question Answering，QA）由最初的搜索需求发展而来，不同于传统的信息检索系统，问答系统接收的输入为用户输入的自然语言，通过Web搜索或者链接知识库等方式，检索到用户的答案后，将其转换为准确简介的自然语言并返回给用户。

问答系统一般为“一问一答”的交互模式，因此构建问答系统一般不需要涉及对话状态管理及对话策略学习等技术，因此问答系统更侧重在对自然语言的理解层面，注重对问句进行分析，旨在获取并挖掘问句的主题词、问题此、中心动词等，目前问句分析主要采用模板批评和语义解析两种方式。

传统的信息检索系统以关键字作为输入，以文档或者结构化的数据作为输出，用户需要依赖多种检索操作来让搜索引擎“明白”其搜索意图，而问答系统将这部分工作交给机器来完成，整体是一个问题驱动的信息获取过程，因此问答系统更适用于特殊而复杂的信息需求，可以从多样化的、非结构化的信息中获取问题的答案，并对问题进行更多的语义理解。

现代的问答系统是一种融合了知识库、信息检索、机器学习、自然语言处理等技术的综合系统，基于不同的分类原则，一般可以将其分为不同的类型。按照其问题答案的数据来源和回答方式的不同，一般可以分为基于Web信息检索的问答系统（Web Question Answering，WebQA）、基于知识库的问答系统（Knowledge Based Question Answering，KBQA），以及社区问答系统（Community Question Answering，CQA）。

其中基于Web信息检索的问答系统以搜索引擎为支撑，在理解用户意图后，在全网范围内检索相关答案并返回给用户，典型的系统如早期的AnswerBus问答系统；基于知识库的问答系统则是当下应用最广泛的问答系统，该类问答系统不仅需要对复杂的自然语言输入进行语义理解，还要对相关知识库进行知识融合，并在此基础上进行一定的知识推理，典型的知识库问答系统如IBM的Watson。社区问答系统是基于社交平台的问答数据来构建的，问题的答案一般来自于社区用户提供，通过检索网络社区中语义相似的问题，将其相应答案返回给用户而实现。

基于问答领域进行分类，可以分为基于事实的问答系统、基于常见问题集的问答系统（Frequently Asked Question，FAQ）和开放域的问答系统。基于事实的问答系统是指通过学习百科知识、期刊、杂志、新闻及文学作品等内容，从这些资源中挖掘出知识，构建出“问题”、“问题类型”、“答案”，进而可以回答“诺贝尔奖获得者有那些？”等问题。基于常见问题集的问答系统，通常是面向一个垂直的领域，构建常用问答数据库，通过在问答数据库中找到与用户输入问题相似的问题，然后将该问题的答案返回给用户。开放域的问答系统通常是通过抽取海量的聊天数据，提供一个可以闲聊的服务。

除了上面介绍的这些主流问答系统类型，还有一些其他的问答系统类型，例如混合式问答系统（Hybrid QA）



图1.15 问答系统基本框架

### 1.2.4 对话系统

对话系统是一种典型的人机交互的应用，不同于问答系统的“一问一答”，对话系统往往需要人机之间多轮对话，对话内容可能涉及不同的领域，对于当前用户输入的信息，往往需要结合上下文信息，同时结合当前语境信息，给与用户相应的回复。

按照用途的不同，一般分为开放域的对话系统和面向任务的对话系统，其中开放域的对话系统主要用于闲聊的对话方式，用户往往不具有明确的目的性，在实现方式上主要基于海量FAQ的检索方式以及端到端的方式。

面向任务的对话系统即为我们当前的主流对话系统，其设计目的是帮助用户解决特定的需求，按照技术实现的方式，面向任务的对话系统可以划分为两类，一类是分模块串行的管道式（Pipeline）的方式，另一类是采用端到端的对话方式，采用数据驱动的模型，忽略中间过程。

目前主流的面向任务的对话系统实现为管道式的方式，整体架构如图所示：



图1.16 任务导向对话系统结构图

该架构主要包括语音识别模块（ASR）、自然语言理解（NLU）模块、对话状态管理模块（DST）、对话策略学习模块（DPL）、自然语言生成模块（NLG）和文本转语音模块（TTS）合成器。其中语音识别模块将用户输入的语音信息转成便于子计算机处理的文本信息，并将其传输至自然语言理解模块，自然语言理解模块的目的是理解其中的语义信息将其映射为计算机可理解的结构化语义表述，也就是识别对话领域和用户意图，然后解析用户输入信息中的语义槽信息。对话状态管理是追踪用户需求并更新管理当前的对话状态，该模块以多轮对话历史、当前的用户输入状态，结合对话上下文信息理解用户输入的含义，通常需要和对话系统背后的数据库或知识库进行交互，查询满足用户需求的实体信息。对话策略学习模块根据当前的对话状态和预定的对话策略觉得系统下一步的动作，然后自然语言生成模块将对话策略学习模块选择的系统动作转换为自然语言，通过文字转语音模块反馈给用户。

一个典型的任务导向型对话系统与用户的交互示例如下：

用户：你好（User Dialog Act：greeting）

机器：你好，我可以帮助你搜索并预订餐厅（Sys Dialog Act：greeting）

用户：我想要找一家在中央公园2.0附近的意大利餐厅（User Dialog Act：inform(地址=中央公园2.0，口味=意大利)）

机器：请问你想要在中央公园2.0附近什么价位的意大利餐厅？（Sys Dialog Act：request(价位)）

用户：普通的就好（User Dialog Act：inform(价位=普通)）

机器：我知道那里有一家柯南家的老意大利料理餐厅（Sys Dialog Act：inform(名称=柯南家的老意大利料理)）

用户：换一个（User Dialog Act：alter()）

机器：我知道那里有一家意大利油泼面餐厅（Sys Dialog Act：inform(名称=意大利油泼面)）

用户：它家的价格怎么样？（User Dialog Act：request(价位)）

机器：意大利油泼面的价位是普通（Sys Dialog Act：inform(价位=普通)）

用户：就这家吧，帮我预订（User Dialog Act：order()）

机器：请问有几个人？（Sys Dialog Act：request(人数)）

用户：单身狗（User Dialog Act：inform(人数=1)）

机器：真的就1个人吗？（Sys Dialog Act：confirm(人数=1)）

用户：是的（User Dialog Act：confirm()）

机器：已经预定成功（Sys Dialog Act：order\_success()）

用户：好的，再见（User Dialog Act：bye()）

机器：再见，欢迎下次使用（Sys Dialog Act：bye()）

对于对话行为（Dialog Acts），如果是用户发起的对话行为（User Dialog Act），那么它是一种处理后的用户意图的抽象表达，是一种形式化的意图描述。如果是系统发起的行为（Sys Dialog Act），那么该行为是根据当前用户的行为，结合上下文信息等综合考虑得出下一步所要进行的操作的抽象表达，这个抽象表达后续会送入NLG部件，生成自然语言。

通过上面的例子，一般在交互过程中，系统需要理解用户的意图，常见的不同意图有：

用户的greeting：问好

用户的inform：用户提供一个信息，例如想要的餐厅的地址

用户的request：询问一个信息，例如当前结果餐厅的电话

用户的confirm：确认信息正确（例如上一条是机器问你对不对）

用户的bye：结束对话

机器的greeting：问好，也可以是自我介绍

机器的inform：提供机器知道的信息，例如当前结果餐厅的信息

机器的request：机器必须有足够的信息才能完成任务，如果欠缺一些必须信息，例如餐厅地址、口味，则会向用户询问

机器的confirm：根用户确认信息是否正确

机器的bye：结束对话

整个对话系统，就是为了完成某个特定任务，这个任务所需要的特定条件需需要由用户提供（例如帮助买咖啡需要咖啡品种，热或冷等信息），当信息足够的时候，机器就能完成相应任务。

### 1.2.5 闲聊系统

闲聊系统是聊天机器人的另一个典型应用，针对的是用户没有特定的目的，没有具体的需求或者目的比较模糊的情况下进行的多轮人机对话，不同于问答系统和对话系统，前两者的一个期望是用户尽可能快的获取到所需要的信息，快速找到想要的回答或者快速完成任务，尽可能减少与用户的交互时间，而聊天机器人往往设计上需要尽可能的占用用户时间，尽可能的延长与用户聊天、陪伴的时间，或者尽可能的再次让用户使用。

目前，大量聊天机器人的产品均定位于闲聊系统，例如微软推出的“小冰”，早期在人人网快速爆红的“小黄鸡”均属于闲聊系统。当前很多硬件产品均集成了闲聊系统，如小米音箱、天猫精灵、叮咚音箱等。



图1.17 人人网小黄鸡闲聊机器人

基于不同的实现方式，闲聊系统主要分为基于检索的闲聊系统和基于生成的闲聊系统，基于检索的闲聊系统是指通过构建一个庞大的对话库，闲聊系统收到用户输入的信息后，在对话库中通过搜索匹配的方式进行答案内容的提取，这种方式的好处在于返回给用户的语句来语音真实的对话场景，因此可以避免出现语法错误，表达比较自然。在具体实现中一般通过结合搜索引擎如Elasticsearch或者Solr对全部对话库进行粗粒度检索，获取候选答案，然后使用匹配算法对候选答案进行排序，选取候选答案中与用户输入句子最匹配的答案返回给用户。

基于生成的闲聊系统来源于机器翻译的技术，通过将输入文本“翻译”成输出文本“回答”，也就是基于用户的输入，采用一定的技术自动身材一句话作为回答，这样做的好处是可以覆盖任意话题的用户问句，缺点是生成的句子往往存在语句不通顺，存在语法错误等问题。

## 1.3 聊天机器人的前沿技术发展

当前与聊天机器人相关的前沿技术主要集中在深度学习、强化学习及知识图谱与聊天机器人的融合，这些技术均为当前的研究热点。

### 1.3.1 深度学习与聊天机器人

深度学习的概念源于对人工神经网络的研究，属于机器学习的子领域，通过使用深度神经网络对数据实现建模的一种方法，不同于机器学习需要对数据进行特征工程的操作，深度学习将神经网络当作一个“黑盒”，直接对数据进行建模，通过大量数据对模型进行训练，进而实现对样本的预测和分类。

随着深度学习技术在图像处理、语音识别等领域取得的进展和突破，越来越多的研究人员开始将深度学习技术应用与自然语言处理领域中，且在自然语言处理任务中取得了较好的效果，由于深度学习技术强大的特征抽取与特征表示能力，近年来在聊天机器人上的应用也越来越多。

1. 意图识别：在聊天机器人的自然语言理解模块中一个重要的任务即为意图识别，尤其是在任务导向的对话系统中，如何在多轮对话中根据上下文理解用户的意图是一个重点问题也是一个难点。2018年以来，Google推出的BERT模型，采用预训练语言表示的方法，基于大型文本语料库上训练的通用模型，并将其应用于下游任务，在上下文语义理解任务中取得了较好的效果。

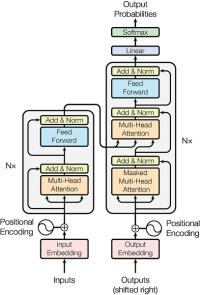


图1.18 BERT结构图

1. 生成模型：聊天机器人中的自然语言生成模块一般是采用基于规则模板的方式将文本组合为答案并返回给用户。随着深度学习的发展，Seq2Seq开始应用到基于生成模型的聊天机器人中，Seq2Seq模型也就是Encoder-Decoder模型，通过将输入转换为一个固定长度的向量，然后将该向量转码为输出向量，这种End2End的思路，通过将问题当作输入语料，答案当作输出语料来训练模型。

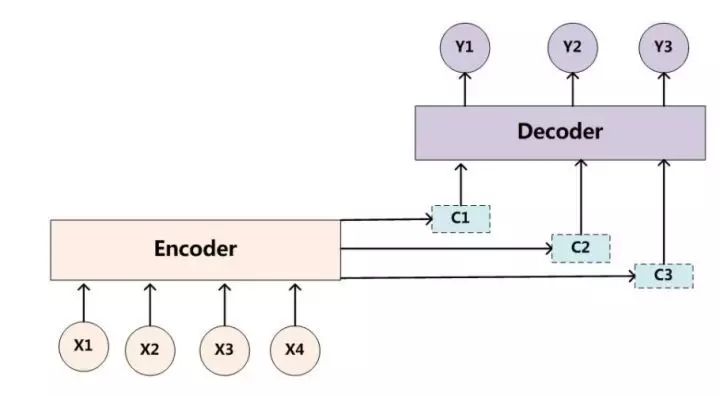


图1.19 Encoder-Decoder模型结构图

### 1.3.2 强化学习与聊天机器人

强化学习是机器学习中的一个子领域，强调如何基于环境而行动，以取得最大化的预期利益，主要来源于行为主义理论，即有机体通过与环境进行交互，在环境给与奖励或者惩罚的刺激下，逐步修正自己的行动，产生能获得最大利益的习惯性行为。  
 强化学习的整个学习过程是封闭的，也就是模型的优化并不需要认为的指导和标注数据，整个学习过程在一个时间序列中完成，在这个时间序列的每个节点，我们不去直接告诉模型应该采取哪种动作，而是让模型自己去尝试，最终通过环境给与的奖励或惩罚来修正模型，在不断的尝试中，模型会学习到采取那种动作才能获得最终的高收益。

当前在任务导向的对话系统中，对于对话管理模块，研究人员考虑结合强化学习，让对话管理模块担任强化学习的智能体，通过用户和对话系统的多轮对话交互，自动学习对话管理的策略，提高对话系统的性能。

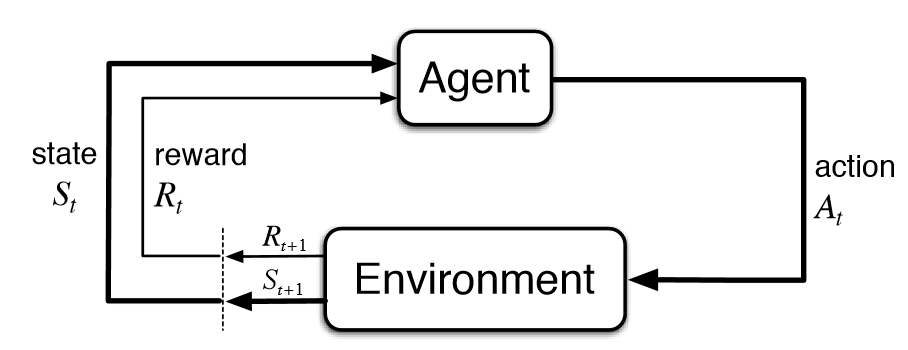


图1.20 强化学习结构图

### 1.3.3 知识图谱与聊天机器人

知识图谱是2012年由谷歌提出，为了更加准确的实现语义搜索，解决原有的基于关键字搜索无法理解字符串语义内容的局限性，知识图谱的目标是描述真实世界中存在的各种实体以及他们之间的关系。

在知识图谱中，一般采用三元组的方式表示知识，如“姚明出生于上海”，用三元组表示为（姚明，出生地，上海），这里“出生地”即为“姚明”和“上海”这两个实体之间的实体关系。

当前将知识图谱结合在问答系统中，可以很好的提高问答系统的性能。

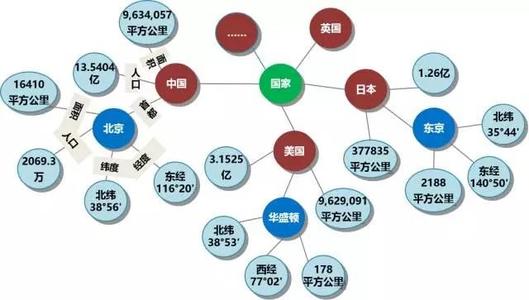


图1.21 知识图谱

## 1.4 聊天机器人的挑战

随着人工智能技术的发展，聊天机器人的生态也日趋成熟与完善，为了让聊天机器人可以更加“智能”，交互体验更加拟人化，可以更好的理解用户的真实意图，结合对话场景更好的给用户提供服务，当前聊天机器人技术还面临多重挑战。

1. 整合语境信息的问题：对于用户的输入，聊天机器人需要整合历史对话语境及物理语境，这种情况尤其在长对话过程中，系统对用户当前输入的信息进行反馈时，需要考虑前几轮对话过程中已经获取的信息，这就需要系统“记忆”整个对话过程中的全部信息。常用的方法是将获取到的对话信息都拼接到一个向量中，但在长对话上进行这样的操作是极富挑战性的。此外在对话过程中也需要整合其他类型的语境数据，例如日期／时间、位置或用户信息，例如当用户询问“明天天气怎么样？”，聊天机器人需要结合当前用户的询问时间信息及所在位置信息进行相应的回复。
2. 聊天机器人的人格一致性问题：当生成回复时，对于语义相同的输入，聊天机器人应该生成相同的回答。例如，你想在“你多大了？”和“你的年龄是多少？”上得到同样的回答。这个问题在闲聊机器人的研究中尤其重要，当前聊天机器人的研究中主要集中在让聊天机器人学习生成语义合理的回复，但是由于训练中使用的语料数据往往是基于不同用户而采集的，因此将固定的知识或者人格整合进模型变得非常困难。
3. 回复多样性的问题：为了让聊天机器人更具有智能性，也就是我们常说的“更像一个人”，我们希望聊天机器人的回复具有多样性，避免回答相同的答案，然而当前聊天机器人基于数据和训练目标的原因，往往使用像“太好了！”或“我不知道”这样的“万能回答”作为其回复，但这种无效回复会严重降低聊天机器人的交互体验。