# 聊天机器人研究进展及趋势

**摘要**

**关键词**

**引言**

聊天机器人是目前非常热的一个人工智能方向。聊天机器人通过学习自然语言来与人类进行对话，属于问答系统的范畴[1]。21世纪互联网的爆发对聊天机器人创造了更大的发展空间，研究者们从互联网上获得了海量的聊天数据，同时计算机硬件能力的提升也为新一代的深度学习算法提供了更好的平台。深度学习技术在自然语言处理NLP、图像识别等领域中已经取得了显著的成就，在自然语言处理方面如信息抽取、自动摘要、语义分析、词性标注已经超过了传统方法，也就说明构建聊天机器人时可以通过深度学习技术对上下文内容进行分析生成一个满意的回复[ 胡艺钟]。在工业界，很多国内外顶级的互联网公司也纷纷推出了自己的产品，2010年苹果公司推出了Siri聊天机器人，通过理解用户的语音输入，可以帮助用户预定闹钟、播放音乐、查天气等信息，甚至帮助用户预订电影票。2012 年，京东自主开发完成了聊天机器人JIMI，能够向用户提供二十四小时的服务，服务内容涵盖电子商务的各个环节，如售前咨询、售后咨询等。2014年，微软公司在中国发布了Cortana和“小冰”，Cortana类似于Siri属于个人助理范畴。“小冰”注重聊天的情感联系，并把智商（IQ）和情商（EQ）考虑在内，以适当的人际关系反应来在长时间交互中吸引用户。通过对聊天过程中语境与语义的理解，“小冰”在一定程度上实现了超越简单人机问答的智能交互，能够使用更加类似人的交流方式[从Eliza到小冰]。2015年，阿里巴巴推出的聊天机器人产品阿里小蜜，它与 JIMI功能很相似，都是向用户提供购物环节中涉及的咨询类对话服务。同年，百度公司也推出了自己聊天机器人度秘。2017年小米公司也发布了自己的AI智能音响“小爱同学”，它支持语音交互可以播放音乐、电台点播、[小说](https://baike.baidu.com/item/%E5%B0%8F%E8%AF%B4/45851" \t "/Users/apple/Documents\\x/_blank)、脱口秀，教育学习、儿童类多种有声读物内容，它还可以控制小米电视、扫地机器人、空气净化器等一系列生态链设备。

同年，阿里巴巴也推出了“天猫精灵”，他也小爱同学非常类似，可以提供感情陪护等一系列生活服务[杨磊],丰富了人们的日常生活。

聊天机器人潜在的巨大商业价值和社会价值使之日益成为工业界和学术界的研究热点[ 杨磊]。

## 1 国内外研究进展

图灵于 1936 年提出图灵机（Turing Machine）这一模型，为今后的计算机制造提供了理论基础，因而也被冠以“人工智能之父”的称号。1950 年他发表《计算机器与智能》一文，并在文中提出了机器具有智能的定义，这一标准被人们称为图灵测试[2]。图灵测试给聊天机器人种下了一颗种子。从上世纪60s到上世纪末，聊天机器人大约经历了三个重要的历史时期。

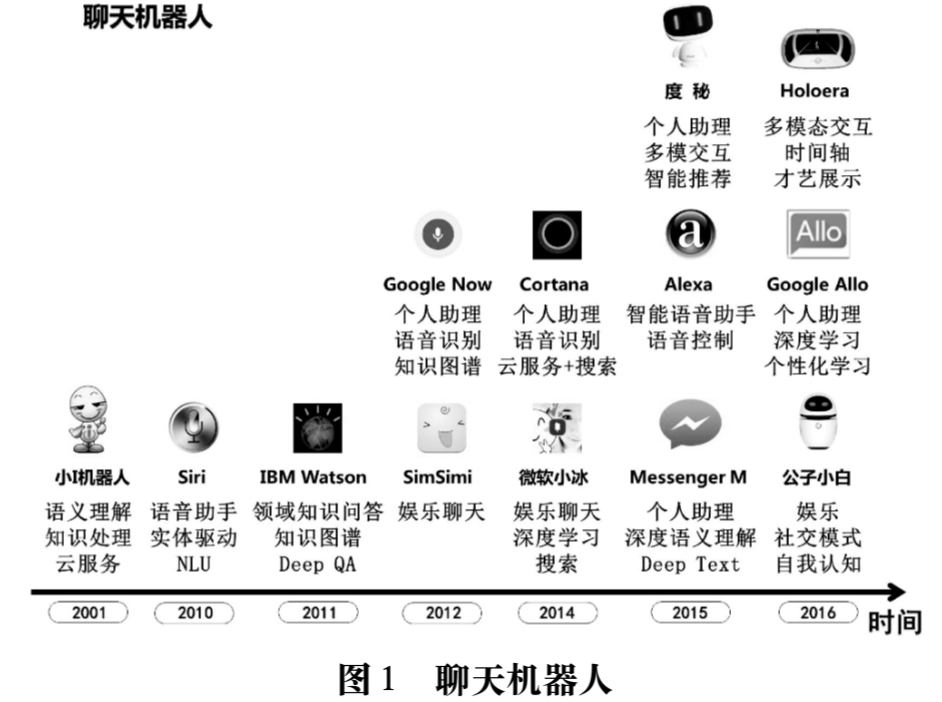
      第一个时期，1966年麻省理工学院（MIT）的约瑟夫·魏泽鲍姆（Joseph Weizenbaum）开发的聊天机器人ELIZA[3] 为世界第一款聊天机器人，它的主要用途是辅助医疗，用于在临床治疗中模仿心理医生。ELIZA的实现技术为关键词匹配及人工编写的回复规则，它并不理解对话内容且知识范围有限，只能在特定领域内使用。

      第二个时期，1988年加州大学伯克利分校（UC Berkeley）的罗伯特·威林斯基（Robert Wilensky）等人开发了名为UC（UNIX Consultant）[4]的聊天机器人系统。UC是一款帮助用户学习怎样使用UNIX操作系统的聊天机器人。它具备了分析用户的语言、确定用户操作的目标、给出解决用户需求的规划、决定需要与用户沟通的内容、以英语生成最终的对话内容以及根据用户对UNIX系统的熟悉程度进行建模的功能。

      第三个时期，1995年受到ELIZA聊天机器人的启发，理查德·华勒斯（Richard S. Wallace）博士开发了ALICE系统。随着ALICE一同发布的人工智能标记语言AIML（Artificial Intelligence Markup Language）被广泛应用在商业客服的开发中。尽管ALICE采用的是启发式模板匹配的对话策略，但是它仍然被认为是同类型聊天机器人中性能最好的系统之一。

这些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[9]。

在工业界，从市面上产品化的发展历程如图1 。



在学术界，随着近年深度学习的发展，早期AIML构建的检索式聊天机器人构建知识库时存在着大量人力资源和时间成本，为了解决此类问题Ryan[ ]等人使用双编码长短记忆网络实现了基于深度学习的检索式聊天机器人并取得了显著的效果。对于构建特定领域的聊天机器人时需要使用大量的人工特征提取等工作，Serban[]等人采用了基于循环神经网络的方法，省去了特征提取工作中的人工参与并取得了不错的效果。Google Vinyal[7]等人首次提出将序列到序列的模型用于对话系统，这是最早将机器翻译中的 sequence to sequence模型在对话系统中进行应用，该模型在机器翻译领域中将源句进行编码 encoder，对应目标句进行解码 decoder 从而完成翻译任务，Vinya 将该模型应用于对话系统中，将问题句作为 encoder，回复句作为 decoder ，从而进行对话生成任务。而后 Bahdanau[8]等人提出一种注意力模型 attention 机制，提高了翻译的准确性。Guo[9]等人引入深度强化学习的思想，将 DeepMind团队提出的DQN模型用于文本生成过程，他们的模型仍是沿用 encoder-decoder 框架，但是在 decoding 的过程中利用 Deep Q-Network 进行迭代探索生成结果。Li 等人[10]提出了一个强化学习框架，模拟两个 agent 让其自动对话，训练神经网络 seq2seq 模型，然后将 Encoder-Decoder 模型和强化学习整合，定义强化学习的 reward 保证对话的未来奖励最大，从而使对话轮数增加。该作者还对比了有监督的 seq2seq 加attention 模型和基于最大互信息作为 reward 训练生成过程的模型。针对生成式模型的“安全回复”问题，Li[ ]等人在传统的编码器-解码器模型基础上，通过最大化互信息的优化目标来最大化生成回复的概率以解决聊天机器人领域中常出现的“安全回复”问题。

在生成模型的训练过程中，除了传统的有监督模型训练，复旦大学的 Yu[16]等人提出了序列问题上的半监督学习算法，该论文提出了应用 GAN 对抗训练的方式结合强化学习算法进行文本生成，解决了传统 GAN 在 NLP 等序列问题上不能进行梯度更新的问题，为后面的 GAN 在 NLP 上的应用提供了思路。

对于蒙古语聊天机器人研究还尚未起步，主要原因还是低资源语料。

## 2聊天机器人系统框架

聊天机器人可以帮助人们进行指定操作，这就为人们省去了不少的操作与时间。你可以对他下达指令，机器人则像个下属一样为我们提供服务。这大大节省了人们的时间成本，而且机器的记忆要比人类的记忆好很多，在机器人不确定问题如何回复的时候还可以利用搜索引擎查找相关答案提供给用户。这类机器人满足了人们对于未来生活的幻想，而且着实能够提高人们的生产效率[曹东岩]。

一个完整的聊天机器人整体框架主要包含 3个功能模块，如图2所示 。其中，对话管理模块是整个系统的核心，对于系统智能程度的高低评估取决于对话管理模块生成的答复。

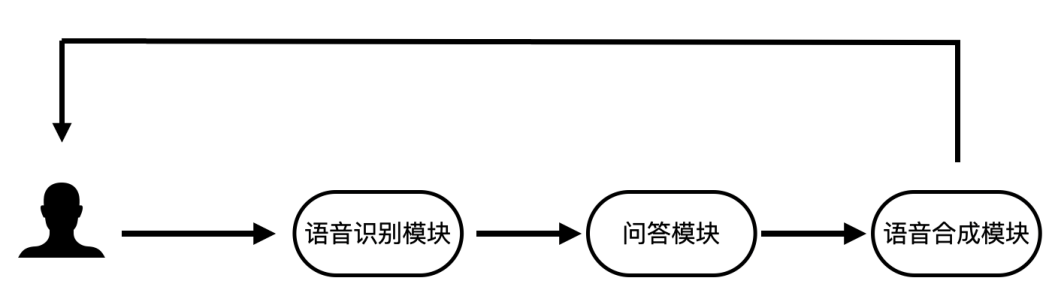


图 2 问答系统整体框架

2.1语音识别（ASR）------

2.2问答模块-----

Word embedding-----计算机并不能直接理解人们的语言，人们需要将其转化为计算机能理解的编码，在我们的任务中通常都是文本信息，所以这里需要将文本进行向量化。

模型选择-------深度学习领域的模型众多，但是目前常用的几个模型有 RNN、CNN 等模

型，RNN 是循环神经网络的简称，直观意义上理解，RNN 模型更适用于我们的

文本处理领域，当然也并非绝对，现在也有一些 CNN 模型在短文本分类等领域

表现优异，但是通常来说大多数 NLP 问题上 RNN 效果显著。

2.3语音合成（TTS）-----

## 3 聊天机器人分类及实现方法

从功能上可以分为任务型和闲聊型聊天机器人，如苹果siri、微软Cortana、百度度秘等被称为个人助理，因为他们的任务是完成人们设定的指定任务，所以都属于任务性聊天机器人。而微软的小冰和小米公司的小爱等是属于闲聊型聊天机器人，因为此类机器人可以与用户无限制范围的交互。

从技术实现方方面又可以分为检索式和生成式。

检索式聊天机器人多数为任务型的特定领域服务。检索式聊天机器人需要事先准备对话库（知识库），聊天系统接收到用户输入句子后，通过在对话库中以搜索匹配的方式进行应答内容提取，很明显这种方式对对话库要求很高，需要对话库足够大，能够尽量多地匹配用户问句，否则会经常出现找不到合适回答内容的情形，因为在真实场景下用户的输入范围非常广泛，但是它的好处是回答质量高，因为对话库中的内容都是真实的对话数据，表达比较自然，不会出现无法错误，因此很适合用在特定领域的聊天机器人。基于检索式的聊天机器人虽然可以保证回复语句的流畅性，但是受限于语料本身，很多内容在构建的数据库中没有答案，此时聊天机器人便无法进行回答[9]。

生成式聊天机器人多数为闲聊型的开放领域服务。它采取技术思路与检索式不同，使用时下最为流行的深度学习技术从大量的对话语料中学出一套规则，在接收到用户输入句子后，自动生成回复内容，这种形式的机器人的好处是可能覆盖任意话题的用户问句，但是缺点是生成的回复质量很可能会存在问题，比如可能存在语句不通顺，存在句法错误等看上去比较低级的错误，无法保证每次的生成效果。

3.1检索式聊天机器人

基于检索的短文本对话是将对话语料中的数据定义为索引信息并保存在信

息检索系统中，通过对已有索引进行文本检索，得到符合当前主题的候选回复，

再通过构建回复匹配模型来对候选回复与问句进行匹配度计算，以此作为候选

回复排序的标准，得到最佳回复[胡艺钟]。

检索式聊天机器人需要事先准备对话库，聊天系统接收到用户输入句子后，通过在对话库中以搜索匹配的方式进行应答内容提取。

Li 等人[11]认为聊天双方对于对话状态都有贡献，他们提出了让聊天机器人主动识别对话僵局的方法，继而主动提出问题以延长对话，

Li X, Mou L, Yan R, et al. Stalematebreaker: A proactive content-introducing approach to automatic human-computer conversation[J]. arXiv preprint arXiv:1604.04358, 2016.

北航和微软的团队[12]提出了聊天机器人利用多轮对话语料在每一轮的回复中结合

对话历史信息的语义表示进行选择，他们的聊天机器人也是检索式的，并且该团队公开了他们构建的豆瓣多轮对话数据集。

Wu Y, Wu W, Zhou M, et al. Sequential Match Network: A New Architecture for

Multi-turn Response Selection in Retrieval-based Chatbots[J]. arXiv preprint

arXiv:1612.01627, 2016.

3.2生成式聊天机器人

Google Vinyal[ ]等人提出从序列到序列的模型用于对话生成，这是最早将机器翻译中的 sequence to sequence模型在对话生成中进行应用。该模型在机器翻译领域中将源句进行编码 encoder，对应目标句进行解码 decoder 从而完成翻译任务，Vinya 将该模型应用于对话系统中，将问题句作为 encoder，回复句作为 decoder ，从而进行对话生成任务。

目前对于开放领域生成式聊天机器人技术而言，多数技术采用了Encoder-Decoder框架，所以本节首先描述Encoder-Decoder框架技术原理。

Encoder-Decoder框架可以看作是一种文本处理领域的研究模式，应用场景异常广泛，不仅仅可以用在对话机器人领域，还可以应用在机器翻译、文本摘要、句法分析等各种场合。图1是文本处理领域里常用的Encoder-Decoder框架最抽象的一种表示。

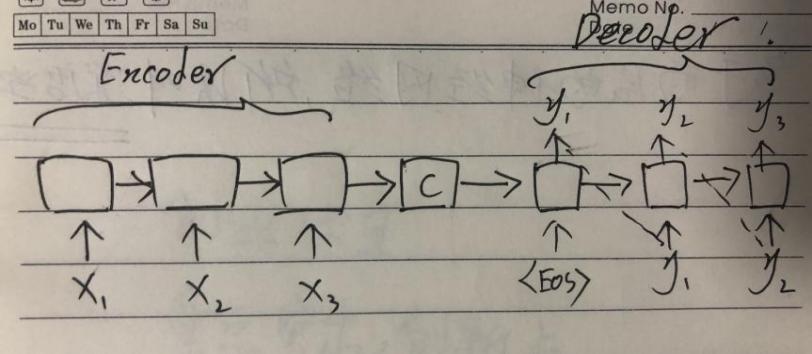


图 1 seq2seq 模型的基本结构

对于聊天机器人来说，完全可以使用上述的Encoder-Decoder框架来解决技术问题。具体而言，对应的<X,Y>中，X指的是用户输入语句，一般称作Message，而Y一般指的是聊天机器人的应答语句，一般称作Response。其含义是当用户输入Message后，经过Encoder-Decoder框架计算，首先由Encoder对Message进行语义编码，形成中间语义表示C，Decoder根据中间语义表示C生成了聊天机器人的应答Response。这样，用户反复输入不同的Message，聊天机器人每次都形成新的应答Response，形成了一个实际的对话系统。

神经网络在训练的时候容易产生过拟合，而且训练耗费大量的时间，为解决这个问题，Hiton 等[ ]提出 Drop out 机制。该机制定义为：在神经网络的训练过程中，按照一定几率将某些神经网络单元暂时从网络中丢弃，增强了模型的泛化能力。

Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: A simple way to prevent neural networks from overfitting[J]. The Journal of Machine Learning Research, 2014, 15(1): 1929-1958.

在实际实现聊天系统的时候，一般Encoder和Decoder都采用RNN模型，RNN模型对于文本这种线性序列来说是最常用的深度学习模型，RNN的改进模型LSTM以及GRU模型也是经常使用的模型，对于句子比较长的情形，LSTM和GRU模型效果要明显优于RNN模型。尽管如此，当句子长度超过30以后，LSTM模型的效果会急剧下降，一般此时会引入Attention模型，2014年Bahdanau[7]等人提出了一种Attention机制，并取得了显著效果。具体框架如图2。我们注意到在原本的 seq2seq 模型中，解码器接收到的信息仅仅是编码器encoder 的最后一个输出向量 C，然后 RNN 网络随着时间的增加很可能原本的输入信息会丢失，这就造成了解码器最终效果不理想。Attention 机制能够很好的解决该问题。Google 研究人员在 Attention 机制基础上构建了 Multi－Head的Attention 结构[8].

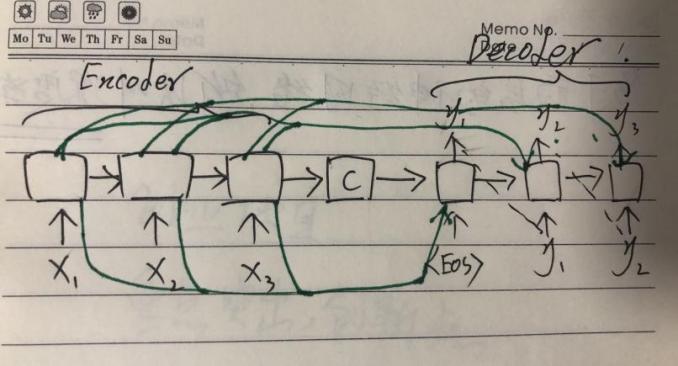


图 2 引入Attention机制后的seq2seq模型

百度的 Mou[14]等人提出基于关键词的预测，将关键词信息引入 decoder 部分，从而指导生成含有这些关键词的句子，解决了传统生成模型容易产生无信息的回复等问题。

在传统的 seq2seq 模型中，我们的输出直接选择概率最大的值，也就是所谓

的极大似然估计，因为这样比较符合语言模型，但是有时候概率最大的值往往都是训练语料中出现概率最大的词，李纪为[ ]等人把该类问题命名为“安全回复”问题。

Beam search:

## 4 聊天机器人技术评测及评测指标

CPS

4.1 BLEU

4.2 句子相似度

**5展望未来**

最完美的机器人还是兼具任务型与闲聊型两者特点的：当我们需要个人助理的时候它能够为我们助力；当我们需要陪伴的时候，它有无限的时间与精力倾听我们的心声。---基于检索式和生成式模型的深度融合。

目前聊天机器人仍然存在着多模态（文本、视频 、音频、图片）信息利用不足，情绪表达能力弱、模型融合机制缺失等问题。

让聊天机器人拥有人类一样的思维，使得对话内容更加贴切。

低语料条件下的聊天机器人构建。----少数民族语言

目前对聊天机器人生成的对话内容评价方法还尚未成熟，有待提升。

深入了解用户的目的、需求和历史信息。所谓的听懂人话。

[1]王浩畅,李斌.聊天机器人系统研究进展[J].计算机应用与软件,2018,35(12):1-6+89.

[2] A. M. Turing. Computing Machinery and Intelligence. 1950, 59(236):433-460.

[3] Weizenbaum J. ELIZA—a computer program for the study of natural language communication between man and machine[J]. Communications of the ACM, 1966, 9(1): 36-45.

[4] Wilensky R, Chin D N, Luria M, et al. The Berkeley UNIX consultant project[J]. Computational Linguistics, 1988, 14(4): 35-84.

[8] BAHDANAU D，CHO K，BENGIO Y． Neural Machine Translation by Jointly Learning to Align and Translate［J］． Computer Science，2014( 3) : 1－15．

[9] ASHISH V，NOAM S，NIKI P，et al．Attention is all you need［C］．Arxiv，2017: 1－15．

[9]曹东岩. 基于强化学习的开放领域聊天机器人对话生成算法[D].哈尔滨工业大学,2017.

[ ]Vinyals O, Le Q. A neural conversational model[J]. arXiv preprint arXiv:1506.05869, 2015.