Computer Science and Application 计算机科学与应用, 2018, 8(6), 918-929
Published Online June 2018 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2018.86102

Hans汉斯

Survey of Intelligent Chatbot Technology

Yilin Dai[1], Gongshen Liu[1,2*]

[1]School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai
[2]SJTU-Shanghai Songheng Information Content Analysis Joint Lab, Shanghai
Email: *lgshen@sjtu.edu.cn

Received: Jun. 4th, 2018; accepted: Jun. 20th, 2018; published: Jun. 28th, 2018

Abstract

As an important branch of natural language processing, intelligent chat robot is currently the hottest and most challenging research direction. It has great significance for promoting the development of human-computer interaction. This paper first briefly introduces the classification and research background of intelligent chat robots, compares the research status at home and abroad, and analyzes the advantages and disadvantages of the two implementation technologies. What's more, we also list several popular chat robots using this technology. Then, the model and evaluation method of the generative chat robot are introduced. Among them, the Encoder-decoder model, which is the foundation of many models, is introduced and analyzed in detail, and several optimization model systems which are completed on this basis are also introduced. Finally, some referenced open source frameworks available for readers to use are listed.

Keywords

Intelligent Robot, Dialogue System, Encoder-Decoder Model

智能聊天机器人的技术综述

戴怡琳[1]，刘功申[1,2*]

[1]上海交通大学电子信息与电气工程学院，上海
[2]上海交大－上海嵩恒信息内容分析技术联合实验室，上海
Email: *lgshen@sjtu.edu.cn

收稿日期：2018年6月4日；录用日期：2018年6月20日；发布日期：2018年6月28日

摘　要

智能聊天机器人作为自然语言处理的一个重要分支，是目前最火热也最具挑战的研究方向，它对于促进

*通讯作者。

文章引用: 戴怡琳, 刘功申. 智能聊天机器人的技术综述[J]. 计算机科学与应用, 2018, 8(6): 918-929.
DOI: 10.12677/csa.2018.86102

人机交互方式的发展有着重要的意义。本文首先简要介绍了智能聊天机器人的分类和研究背景，对国内外研究现状进行比较，对生成和检索两种主流的实现技术进行优缺点分析，并分别列举了几项使用该技术手段实现的聊天机器人。然后，介绍了目前较为常用的生成型聊天机器人的模型以及评估方法，其中，对作为很多模型基础的Encoder-decoder模型做了详细介绍和分析，以及在此基础上完成的几个优化模型系统。最后，给出了一些参考的开源框架以及可使用的数据以供读者使用。

关键词

智能机器人，问答系统，编码解码模型

1. 研究背景

目前市面上主要的智能聊天机器人可以分为如下两类：目标驱动型聊天机器人和无目标驱动型聊天机器人。目标驱动机器人是指机器人的服务目标或服务对象是明确的，是可以提供特殊服务的问答系统，处理特定领域的问题，即定领域的聊天机器人，比如客服机器人，订票机器人等。无目标驱动机器人是指机器人的服务对象和聊天范围不明确，可以处理的问题多种多样，解决问题时需要依赖于宇宙中的各种信息和本体，即开放领域的聊天机器人，比如娱乐聊天机器人等。

智能聊天机器人实际上是为了应对信息爆炸的今天存在的信息过载问题。具体来说，其来源是因为人们对于简单的搜索引擎仅仅返回一个网页集合的不满，而通常用户更想获得的体验是在向智能对话系统用自然语言提出一个问题之后，且智能对话系统也能够自然又通顺地回答问题，且回答内容与问题紧凑相关又答案精准。为使用者们节约了更多的时间，无需逐个浏览和仔细阅读搜索引擎返回的每个链接网址中的信息，再剔除冗余信息后才能得到期望的答案[1]。

有关于对话机器人的研究可以被追溯到 20 世纪 50 年代，当 Alan M. Turing 提出了“机器可以思考吗？”的图灵测试问题来衡量人工智能发展的程度，该领域接下来就变成了人工智能领域中一个十分有趣又具有挑战性的研究问题。

随着各种互联网公司的蓬勃发展以及各类移动终端和应用小软件的爆炸式普及，如Twitter和微博等，很多大互联网公司都在投入重金完成此领域技术的研究并陆续推出此类应用产品，比如苹果 Siri，微软 Cortana，脸书 Messenger，谷歌 Assistant 等，这些产品都让用户们在移动终端更加方便地获得需要的信息和服务，从而获得更好的用户体验。因而，人们发现智能机器人可以应用的领域十分广泛，它可以被应用到很多人机交互的领域中，比如技术问答系统，洽谈协商，电子商务，家教辅导，娱乐闲聊等[2]。

不得不说，由于人们对于智能聊天机器人不断增长的渴望和需求，人工智能在自然语言处理领域的应用变成了不论国内国外都非常热门的一个研究方向。在信息技术飞速发展，以及移动终端逐渐普及的今天，研究聊天机器人相关的技术，对于促进人工智能以及人机交互方式的发展有着十分重大的意义。

2. 国内外研究现状对比

在人工智能领域，智能聊天机器人的研究已经有了很长的历史[3]，它们都试图吸引用户不断继续聊天，通常表现为使用主导谈话的主题的手段，从而掌控谈话内容及谈话进度。但由于最初的研究受限于

计算能力和知识库，导致了所有有关人工智能的实验规模都比较狭小[4]，因此设计者们还会将谈话的内容局限在某一个特定的专家系统领域以此降低难度。

但随着 1995 信息检索技术的发展，Baidu，Google 等搜索引擎公司计算能力的飞速提升，以及 2005 年互联网业的蓬勃发展和移动终端的迅速普及，在这三方面的共同作用下，智能聊天机器人，或者说智能问答系统一下子被推到风口浪尖，研究进展也非常值得关注。

由于国外在人工智能聊天机器人及问答系统方向的研究起步较早，因而也产生了一系列比较成熟的聊天系统以供用户使用，比如苹果 Siri，微软 Cortana，脸书 Messenger，谷歌 Assistant [5]等。

这些跨平台型人工智能机器人，都借助着本公司在大数据、自然语义分析、机器学习和深度神经网络方面的技术积累，精炼形成自己的真实有趣的语料库，在不断训练的过程中通过理解对话数据中的语义和语境信息，从而实现超越一般简单人机问答的自然智能交互，为用户带来方便与乐趣。

相比国外，我国国内在智能聊天领域的投入规模和研究水平上都有着不小的差距，研究成果也并不显著。但是还是有一系列高校在此领域成绩显著，位于前列的主要有清华大学、中科院计算所、香港大学、香港中文大学和哈工大(benben [6])等。其中，高校在此领域的研究主要集中于对于自然语言处理的工具开发，比如哈尔滨工业大学的 HIT 工具(中文词法分析、句法分析和语法分析)以及台湾国防大学的 CQAS 中文问答系统(侧重于命名实体及其关系的处理)等。

3. 工程要求及分类

3.1. 实现需求

1) 语境整合

系统需要在训练过程中不断整合物理语境和语言语境来生成较为明智的回复。结合语言环境最普遍的例子就是在长对话中，人们会记录已经说过的话以及以及和对方交换过的信息。其中最普遍的方法就是将对话嵌入一个向量中，此向量还可能需要整合其它类型的语境数据，例如日期/时间、位置或者用户信息等。

2) 人格一致性与互信息

对于语义相同或者类似的输入，不论在何时输入，我们希望智能机器人会有相同的回答，比如“你叫什么名字？”和“你多大了？”等问题。这个目标看似十分容易达成，但是实际上要将固定的知识或者人格整合到模型中去是一个十分困难的研究难题。目前许多的智能聊天机器人系统可以做出语义较为合理的回答，但是却没有被进一步训练生成在语义上同样一致的回复。这一般是由于为了实验效果的增加，训练模型的数据可能来源于不同的用户而导致的。

3) 意图以及多样性

目前普遍的智能问答系统经常会生成“我不知道”或者“太好了”这样的可以适用于大部分输入问题的答案。例如，Google 的 Smart Reply 早期版本常常用“我爱你”回复一切不确定的问题。由于生成系统，特别在开放领域，没有被训练成特定意图，只是根据数据和实际训练目标或算法训练的结果，不具备一个令人满意的智能聊天机器人应该有的多样性。

3.2. 工程分类

目前该领域该方向的实现技术手段主要集中于基于规则或基于学习的方法[7]。因此，相对应的，智能聊天机器人的实现技术手段目前也分为两种：基于检索的方式[8]和基于生成的方式[9]。

1) 检索式

检索式聊天机器人是指使用了预定义回复库和某种启发方式来根据输入和语境做出合适的回复，这

种启发方式既可以像基于规则的表达式匹配一样简单，也可以像机器学习分类器一样复杂。换一句话说，在这种模式下，机器人回复的内容都处于一个对话语料库中，当其收到用户输入的句子序列后，聊天系统会在对话语料库中进行搜索匹配并提取响应的回答内容，进行输出。

该系统不要求生成任何新的文本，只是从固定的集合中挑选一种回复而已，因而这种方式要求语料库的信息尽可能的大和丰富，这样才能够更加精准地匹配用户内容，并且输出也较为高质量，因为语料库中的既定语句序列相对于生成的序列而言更加自然和真实。

该模式下的机器人使用基于规则的方式进行模型的构造，因此我们只需要完成一个模式或者样板，这样当机器人从用户端获得的问题句子在已有的模板中时，该模型就可以向用户返回一个已有的模板。理论上，任何人都可以照此方法实现一个简单的聊天机器人，但是该机器人不可能回答比较复杂的问题，其模式匹配意识是十分薄弱的。除此之外，人工地完成这些规则和模式的制定是十分耗时和耗力的。

目前，在基于规则方面的一个非常流行的智能机器人是 CleverBot，该网站提供了一个可以直接进行与机器人进行聊天的网页。

2) 生成式

生成式聊天机器人在接受到用户输入后，会采用其它技术生成一句回复，作为聊天系统的输出。这种方式并不要求非常大和精准的语料库，因为它不依赖于预定义的回复库，但生成的回复可能会出现语法错误或语句不通顺等缺点。

该模式下的机器人使用基于学习的方法进行对于对话数据和规律的学习，很好地弥补基于规则模式下实现的智能机器人的缺点，因此我们可以建造一个机器人，并让它不断地从已经存在的人与人之间的对话数据中自主得学习对话规律，并在每次收到用户问题时，自主地组织词语回答问题，这是一种十分智能的实现方法，也是目前更为火热的研究方向。

使用生成式方式并且结合机器学习的方法的优点是十分明显的：得到相较于检索式而言更加有趣多样的回答，赋有多样性，避免沉闷和无聊；端到端 End-to-end 神经网络模型的参与可以减少对于人为制定规则的依赖，提升模型在长对话数据中的性能[10]；深度学习的应用使得模型的可扩展性较强，模型本身可以和训练数据的语言互相剥离，不需要针对不同语言的数据进行数据预处理工作；可以通过扩大数据的方式持续提升模型的效果。

4. 常见技术模型

4.1. Encoder-decoder 加解密模型

在以往的研究中，我们会发现实际上智能对话系统问题可以被很好地应用到的自然语言的机器翻译框架中[11]，我们可以将用户提出的问题作文输入机器翻译模型的源序列，系统返回的答案则可以作为翻译模型的目标序列。因此，机器翻译领域相对成熟的技术与问答系统所需要的框架模型有了很好的可比性[12]，Ritter 等人借鉴了统计机器翻译的手段，使用 Twitter 上的未被结构化的对话数据集，提出了一个问答生成模型的框架[13]。

Encoder-decoder 框架目前发展较为成熟，在文本处理领域已经成为一种研究模式，可应用场景十分广泛。它除了在已有的文本摘要提取、机器翻译、词句法分析方面有很大的贡献之外，在本课题中，也可以被应用到人机对话和智能问答领域。

图 1 将 Encoder-Decoder 框架在自然语言处理领域的应用抽象为一个通用处理模型，即一个序列(或文章)转换为另外一个序列(或文章)。对于句子序列对<X,Y>，Encoder-Decoder 框架在输入源序列 X 的情况下，生成目标序列 Y，并不断改变模型参数提升这种可能性。在实际的应用中，序列 X 和序列 Y 分别由各自的单词序列构成，可以是一样或者不一样的语言：

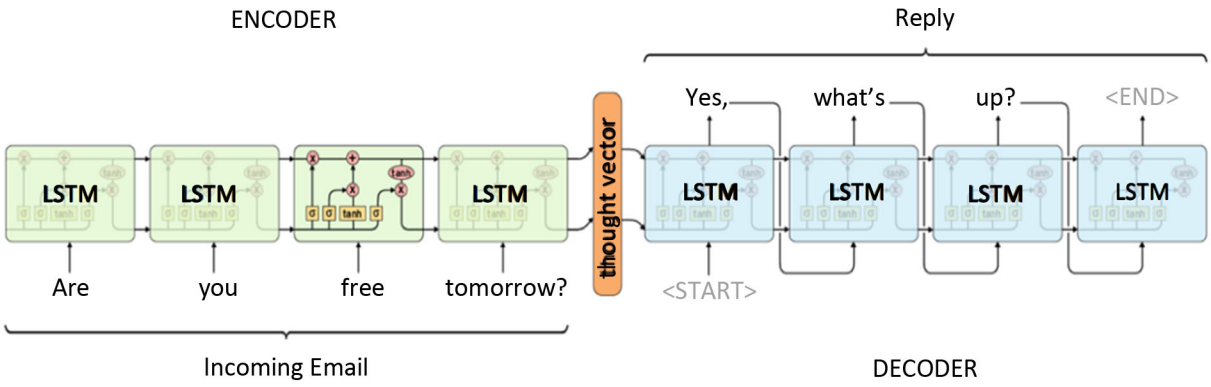

Figure 1. End2end model implemented by the E-D module
图 1. 加解密模块实现的端到端模型

$$X = \langle x_1, x_2, \cdots, x_m \rangle$$

$$Y = \langle y_1, y_2, \cdots, y_n \rangle$$

模型中神经网络(如 RNN 或 LSTM)将按照如下步骤计算此条件概率：首先，输入序列$(x_1, x_2, \cdots, x_T)$通过加密模型中的层层 LSTM 神经单元，由最后的隐藏层状态获得一个固定维度的向量表示 v；然后，根据标准的 LSTM-LM 公式，计算输出序列$(y_1, y_2, \cdots, y_{T'})$的概率，该 LSTM 的最初的隐藏层状态为输入序列$(x_1, x_2, \cdots, x_T)$的向量表示 v：

$$p(y_1, y_2, \cdots, y_{T'} \mid x_1, x_2, \cdots, x_T) = \prod_{t=1}^{T'} p(y_t \mid v, y_1, y_2, \cdots, y_{t-1})$$

在这个等式中，每个概率分布$p(y_t \mid v, y_1, y_2, \cdots, y_{t-1})$都用对于词典中的所有单词的 softmax 函数模型来表示。并且，我们要求，所有的句子都以一个特殊的句子终结符“<EOS>”来表示，这可以使得模型定义了一个关于输出序列的所有可能长度的概率分布。

Google 利用该理念实现的神经翻译系统模型，结合 LSTM 神经网络结构，实现了端到端的语言模型[14]，是现在十分主流的使用深度学习实现的智能对话系统，并提供了开源的参考架构。斯坦福大学使用该端到端模型，并在解密模型中添加注意力机制，实现了一个综合性的神经网络机器人。

4.2. Hierarchiacal Recurrent Encoder-Decoder 分级卷积加解密模型

目前研究中经常使用 Seq2seq 端到端的方式实现问答系统，但模型常常会有可能产生与问题毫不相关，意义不明，表达不准确甚至是毫无意义的安全回复，例如“我不知道”，“好的”或“我爱你”这样的答案。

对于此类问题的出现，Bengio 等人提出的一种更加复杂的模型结构——分级卷积加解密(Hierarchiacal Recurrent Encoder-Decoder)的端到端模型可以很好的解决问题。HRED 模型通过使用第二个加密模块来从之前的问句中获得更加直观的信息，从而当前输出对于之前信息的依赖性可以得到保障[15]。

在模拟对话时，我们认为 HRED 模型比标准的 RNN 模型更好是因为：上下文 RNN 会在用户之间使用一个分布式的向量来表达对话主题和内容，这对于建立一个有效的对话系统来说是非常重要的[16]；由于在序列传递过程中的计算步骤被减少，这使得于模型参数有关的目标函数的计算会更加稳定，并且有助于传播优化方法的训练信号。

4.3. Bidirectional HRED 双向分级卷积加解密模型

双向 HRED 模型的加密模块中使用一个双向的 RNN 模型，一条前向传递语句序列，另一条通过导

致输入序列反向传递。前向传递时 n 位置处的隐藏层状态包含了 n 位置处之前的信息，而反向传递时的隐藏层状态总结了 n 位置处之后的信息。为了仍然得到一个固定维度表示的上下文向量，我们可以将前后向传递的最后隐藏层状态通过直接前后相连或通过 L2 池化后相连。此种双向结构可以有效地解决短时依赖的问题，并且在其他相似的结构中也被证明是有效的[17]。

4.4. Word embedding 词嵌入

词嵌入(Word embedding)又被称为词表示(Word representation)，每个单词套用该模型后可以转换为一个实数，且每个实数对应词典中的一个特定单词[18]。它是一种用于在低维的词向量空间中用来学习深层的单词表示的技术，通过对词汇量的扩大，可以很大地提升训练速度[19]，因为会通过在词嵌入空间中非常相近的单词来共享一些信息。常用的词嵌入模型有 Word2Vec [20]，该模型是由包含了由一千多亿单词组成的 Google 新闻数据训练的，并且被证明该模型在一个非常广泛的数据集上展现出了强有力的信息。

4.5. Attention 注意力机制

Attention 结构的核心优点就是通过在模型“decoder”阶段对相关的源内容给予“关注”，从而可以在目标句子和源句子之间建立直接又简短的连接，解决机器人模型和用户之间的信息断层问题[21]。注意力机制如今作为一种事实标准，已经被有效地应用到很多其他的领域中，比如图片捕获生成，语音识别以及文字摘要等。

在传统 seq2seq 模型的解码过程中，“encoder”加密器的源序列的最后状态会被作为输入，直接传递到“decoder”解码器。直接传递固定且单一维度的隐藏状态到解码器的方法，对于简短句或中句会有较为可观的效果，却会成为较长的序列的信息瓶颈。然而，不像在 RNN 模型中将计算出来的隐藏层状态全部丢弃，注意力机制为我们提供了一种方法，可以使解码器对于源序列中的信息选择重点后进行动态记忆。也就是说，通过注意力机制，长句子的翻译质量也可以得到大幅度的提升。

注意力在每一个解码的时间步时都会进行计算，隐藏最初的 embedding 词嵌入操作和最终的 projection 投影操作后，主要包含了如下四个步骤，如图 2 所示：

1) attention weights 注意力权重的计算

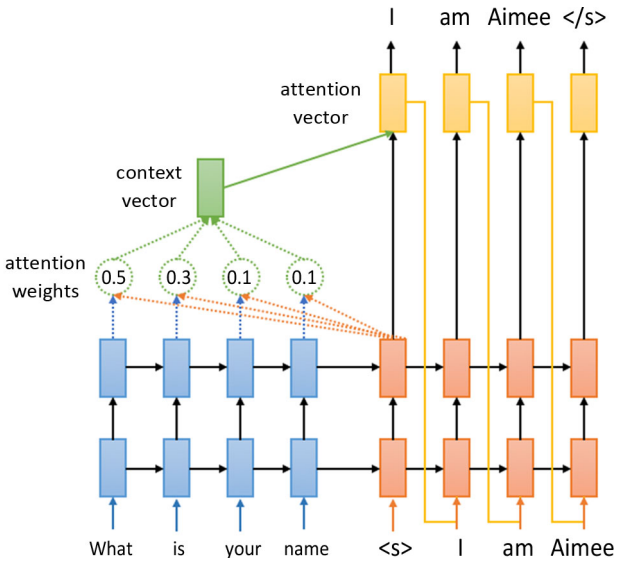

Figure 2. Calculation process of the attention mechanism
图 2. 注意力机制的计算流程

注意力权重是通过当前的目标隐藏层状态和所有的源序列的状态互相比较得出的，公式(1)表示该计算过程。

$$\alpha_{ts}=\frac{\exp\left(\text{score}\left(h_t,\bar{h}_s\right)\right)}{\sum_{s'=1}^{S}\exp\left(\text{score}\left(h_t,\bar{h}_{s'}\right)\right)} \tag{1}$$

在该公式中，score 函数会逐一比较每一个源序列的隐藏层状态 $\bar{h}_s$ 和目标序列的隐藏层状态 h_t，得到的结果将被标准化，产生一个关于源位置的分布，即 attention weights。关于 score 函数的选择有很多，它们主要的不同就在于。乘法和加法形式是目前比较流行的评分函数，如公式(2)所示：

$$\text{score}\left(h_t,\bar{h}_s\right)=\begin{cases}h_t^T W\bar{h}_s\\ v_a^T\tanh\left(W_1h_t+W_2\bar{h}_s\right)\end{cases} \tag{2}$$

2) context vector 上下文向量的计算

根据公式(2)中计算得到的注意力权重，源序列状态的权重均值(即上下文向量)的计算过程可由公式(3)表示。

$$c_t=\sum_s\alpha_{ts}\bar{h}_s \tag{3}$$

3) attention vector 最终注意力向量的生成

最终的注意力向量需要将上下文向量和当前目标序列的隐藏层状态互相结合生成，公式(4)可以表示该计算过程。

$$a_t=f\left(c_t,h_t\right)=\tanh\left(W_c\left[c_t;h_t\right]\right) \tag{4}$$

4) 模型的输入

使用公式(4)中得到的注意力向量，作为下一个时间步时初始状态输入到模型中，并用于得到模型的归一化逻辑输出和损失值。该过程与原始的 seq2seq 模型中最后一层的隐藏层状态十分相似。

5. 模型评估方法

通常情况下，用户评判一个智能聊天机器人或者问答系统的性能是否理想时，都会将该系统是否准确地完成了用户的任务作为首要的参考条件，例如，在给定的对话中是否解决客户支持问题，但是在智能问答领域并没有一个合适的标准来衡量模型的性能，一般被经常使用的用于反映对话质量的方式主要有如下几种。

1) *recall@k*

使用检索方式实现智能机器人，可以采用常用的检索型标准 recall@k，表示让这个模型从 10 个候选响应中挑出 k 个最好的响应，候选的响应中包含 1 个真实响应和 9 个干扰项噪声响应。如果模型选择出的 k 项回答中包含对应的正确响应，则该测试样本的结果将被标记为正确。自然，k 越大，那么这个任务就会更加简单。

尽管我们知道一个对于回答多分类任务的语言模型的优化并不一定能作为一个好的语言回复生成模型的好的参考标准，但是我们认为一个模型的分类能力的优化一定最终会带来生成任务的提升。

2) *perplexity*

另一个常用于估测语言模型准确率的评估方法是 perplexity 困惑度，它被定义为每个单词的平均负对数概率的指数[15]，如公式(1)所示。

$$\text{ppl}\left(\varepsilon_{test};\theta\right)=\text{e}^{-\left(\log P\left(\varepsilon_{test};\theta\right)\right)/\text{length}\left(\varepsilon_{test}\right)} \tag{5}$$

该指标可以反映“此模型对自己生成的目标序列的准确度是多少？”。更为精确的是，perplexity值可以表达的一种概念是“如果我们在每个时间步从由该语言模型计算出来的概率分布中随机挑选单词，要获得正确的答案，需要平均挑选几个单词？”。我们经常会将perplexity作为一种评估标准的参考选项，是因为perplexity值越大，就表示模型之间的差距也更容易被人眼直观感知到。

6. 公开资源

6.1. 模型框架

Chatbot 作为一场交互革命，是一个多技术融合的平台，简单来说就等于 NLU (自然语言理解)和 NLG (自然语言生成)的结合体。针对以上介绍的常用技术框架分类，介绍以下五种目前比较流行的经典框架，并从机构/作者，特点，实现原理，支持语言和流行程度五方面进行对比分析。

1) Artifiical Intelligence Markup Language

机构/作者：Dr. Richard S. Wallace。

特点：自定义的 AI 语言，作为 XML 语言的扩展，支持语言规约，开源了解析器，并支持主流的所有编程语言。

实现原理：基于检索的系统，通过 pattern 元素匹配用户问句。

支持语言：Java，Ruby，Python，C，C#和 Pascal 等。

流行程度：283,000。

2) Opendial 机构/作者：Lison, P.

特点：提供语音识别功能，有较好的澄清机制，除了在调节参数部分用到了机器学习技术之外，没有太多的机器学习和深度学习的技术。

实现原理：基于规则的系统，使用概率规则和贝叶斯网络。

支持语言：Java。

流行程度：147,000。

3) Api.ai

机构/作者：Google。

特点：提供了一个可以自己定义模板、参数和多轮对话的 AI 框架，可以很方便地使用多轮对话定义一个特定任务的聊天机器人。

实现原理：使用整个知识和数据结构的“域”。

支持语言：Java，Python，C++和 C#等。

流行程度：25,600,000。

4) Wit.ai

机构/作者：Facebook。

特点：同时提供语音识别和机器学习功能，非自动化地提供某种机器学习机制，可以理解从未见过的命令。

实现原理：“意图”和“实例”元素的使用，使用“角色”的概念在不同环境下区分实例。

支持语言：Java，Ruby，Python，C 和 Rust 等。

流行程度：351,000,000。

5) ChatterBot

机构/作者：Gunther Cox。

特点：基于检索方式的聊天机器人，但不适用于任何基于任务的对话系统。

实现原理：“意图”和“实例”元素的使用，使用“角色”的概念在不同环境下区分实例。

支持语言：Python。

流行程度：448,000。

6.2. 数据

1) Ubuntu 对话语料库

Ubuntu 对话语料库(Ubuntu Dialog Corpus)，UDC 是目前最大的公共对话数据库之一，它以一个公共 IRC 网络上的 Ubuntu 频道为基础，该频道允许大量参与者的实时交谈。该数据库可以作为训练对话系统神经网络的重要数据是由于它具备的一下四种特征：

1) 双向对话，适用于多参与者聊天，并且是最好是人－人间对话。

2) 具备大量的对话集：10^5~10^6 在人工智能的其他领域的神经网络学习是一个比较合适和典型的数据量。

3) 许多具备多个轮回的对话(大于 3 个轮回)。

4) 数据具备特定领域的目标，而不是开放的聊天机器人系统。

UDC数据集可以在官方网址进行下载，进行NLTK预处理后的训练数据由1,000,000个多轮回对话样本构成，共由7,000,000句话语和100,000,000个单词组成，其中50%是积极地(标签为1，表示该话语是对这个语境的真实响应)，50%是消极的(标签为0，表示该话语不是对这个语境的真实响应)。每个样本都由一个语境(context)，一个话语(utterance)和一个语境的响应(response)构成的。

2) Cornell 电影对话语料库

Cornell电影对话(Cornell Movie-Dialogs)数据集可以在Cornell大学CS专业的官网上进行下载，使用前需要对原始的文件进行预处理工作，通过快速搜索某些特定的模式来清理该数据集。

该语料库是从原始的电影脚本中提取出来的大量虚构的原数据对话集合，共包含了在 10,292 对电影角色之间进行的 220,579 次会话交流，涉及 617 部电影中的 9,035 个角色，合计总共 304,713 条话语。其中，每一个电影原数据都包含了：所属流派，发布年份，IMDB (Internet Movie Database 互联网电影资料库)评级，IMDB 票数以及 IMDB 评级。

3) ESL 场景对话语料库

场景对话(Scenario Conversation)数据集可以从 ESL 用于机器训练的网站进行下载，该集合中包含了图书馆，社交，购物，就餐和旅行等 25 个不同场景下的多轮对话，具体内容如表 1 所示，其中数量表示多轮对话的数量，每一个多轮对话中都包含着 5~6 轮的来回对话，合计共计 1500+条话语。

4) Reddit 社交新闻站点语料库

Reddit 社交新闻网站用于供用户在一个在线社区中，浏览因特网上的内容后做出自己的评论，也可以对他人的评论进行支持或者反对投票。类似于 Twitter 和 Weibo。抓取网站上一个月的评论就可以粗略地生成 3M 对训练样本，经过预处理后的 Reddit 数据集大约包含有 110,000 条话语作为训练数据。

7. 小结

1) 实现智能聊天机器人的方法背景介绍

本文对常用的检索型和生成性两种方式进行了详细介绍和对比，总结如表 2 和表 3 所示，并列举了一部分使用该方法生成的实例。在实际条件和资源允许的情况下，也常常会使用检索和生成互相结合的方式来实现，以追求更良好的表现性能。

2) 本文主要研究总结

Table 1. ESL scene speech database
表 1. ESL 场景对话语料库

内容	数量	内容	数量
Small talk	72	Buying a car	39
College life	72	Driving	60
Library	42	Health	78
Transferring to a university	42	Employment	72
Socializing	42	Unemployment	33
Dating	24	Travel	84
Renting an apartment	135	At a hotel	63
Taking the bus	45	Buying a house	36
Daily life	78	Selling a house	51
Shopping	42	In a new neighborhood	39
At the bank	51	Crime	42
Food	63	Voting	51

Table 2. Comparison of search and generation chatbots
表 2. 检索型与生成型聊天机器人优点对比

Pros	检索型	生成型
	生成较为精准的答案	端到端的学习过程
	生成较为通顺自然的回答	生成更为安全的回答
	更加灵活的模型系统	容易获取上下文中的信息
	评估方法更为简单	回答富有情感和可控制性

Table 3. Comparison of the retrieve-based and generative chatbots
表 3. 检索型与生成型聊天机器人优缺点对比

Cons	检索型	生成型
	满足多重匹配时回复随机	难以衡量和评估模型优劣
	需要大量问题－答案对的数据且过于依赖数据质量	无趣和不流利的回答
	较难获取上下文的信息	需要极为经验丰富的开发者

首先，关于主流的实现智能聊天机器人所采取的生成型技术框架，本文主要介绍了高度抽象的Sequence-to-sequence 端到端的模型，它可以对中间的词法分析，句法分析可以省略，并减少了对于序列的过多假设和猜想，十分高效。实现时一般使用 Encoder-decoder 加解密模型，并结合 RNN 或 LSTM 等神经元网络实现，其余模型都在该模型的基础上进行变种修改，如添加 Word embedding 词嵌入模型和 Attention 注意力机制等，侧重于解决信息传递，人格一致和回答多样性等问题。

Encoder 框架，即“加密”模块，就是使用几层神经元细胞构成的网络，按照特定的规则，对源序列 X 进行非线性编码，并转化为中间语义表示 C $\left(\mathrm{C}=\mathrm{f}\left(x_1,x_2,\cdots,x_m\right)\right)$，由于该语义表示 C 包含了之前输入的问句中的基本信息，因而又可以被称为思考向量。Decoder 框架，即“解密”模块，该模块基于中间语义表示 C 和已生成的历史单词信息，生成 i 时刻单词 y_i $\left(y_i=g\left(y_1,y_2,\cdots,y_{i-1}\right)\right)$，依次产生每个 y_i，那么整

个系统从宏观的输入和输出上而言，就是根据输入的句子序列 X，产生了目标句子序列 Y，即问与答。

其次，本文还对模型训练时使用的训练数据及评估方法进行介绍。智能问答领域不同于其他使用机器学习实现的自然语言处理问题，最为准确的评估方式应当是人的判断，但由于时间和经济条件的限制不被经常采纳。通常情况下可以用的自动模型评估方式较少，常用的为基于检索型的 Recall@k 和基于生成的 Perplexity。

最后，本文介绍了可参考和利用的五个开源模型，分别使用了不同的检索和生成型方式，并提供了可链接的网址进行学习，读者可以使用本文中提供的数据库进行训练，或根据自己的要求定义和生成聊天机器人。

基金项目

国家自然科学基金支持(编号：61772337，61472248)。

参考文献

[1] 冯升. 聊天机器人问答系统现状与发展[J]. 机器人技术与应用, 2016(4): 34-36.

[2] Gasic, M., Breslin, C., Henderson, M., Kim, D., Szummer, M., Thomson, B., Tsiakoulis, P. and Young, S. (2013) On-line Policy Optimisation of Bayesian Spoken Dialogue Systems via Human Interaction. *IEEE International Conference on Acoustics*, *Speech and Signal Processing*, Vancouver, BC, 26-31 May 2013, 8367-8371. https://doi.org/10.1109/ICASSP.2013.6639297

[3] Weizenbaum, J. (1966) Eliza: A Computer Program for the Study of Natural Language Communication between Man and Machine. *Communications of the ACM*, **9**, 36-45. https://doi.org/10.1145/365153.365168

[4] Rambow, O., Bangalore, S. and Walker, M. (2001) Natural Language Generation in Dialog Systems. *Proceedings of the First International Conference on Human Language Technology Research* (*HLT*'01), Stroudsburg, PA, USA, 18-21 March 2001, 1-4. https://doi.org/10.3115/1072133.1072207

[5] Qiu, M.H., LI, F.L., Wang, S.Y., Gao, X., Chen, Y., Zhao, W.P., Chen, H.Q., Huang, J. and Chu, W. Alime Chat: A Sequence to Sequence and Rerank Based Chatbot Engine. *Proceedings of the* 55*th Annual Meeting of the Association for Computational Linguistics* (*Vol.* 2: *Short Papers*), Vancouver, Canada, July 30-August 4 2017, 498-503.

[6] Zhang, W.-N., Liu, T., Qin, B., Zhang, Y., Che, W.X., Zhao, Y.Y. and Ding, X. (2017) Benben: A Chinese Intelligent Conversational Robot. *Proceedings of the* 55*th Annual Meeting of the Association for Computational Linguistics-System Demonstrations*, Vancouver, Canada, July 30-August 4 2017, 13-18.

[7] Williams and Young (2007) Partially Observable Markov Decision Processes for Spoken Dialogue Systems. *Computer Speech & Language*, **21**, 393-422.

[8] Ji, Z.C., Lu, Z.D. and Li, H. (2014) An Information Retrieval Approach to Short Text Conversation. arXiv preprint arXiv:1408.6988

[9] Bahdanau, D., Cho, K. and Bengio, Y. (2015) Neural Machine Translation by Jointly Learning to Align and Translate. arXiv preprint arXiv:1409.0473

[10] Li, X.J., Chen, Y.-N., Li, L.H., Gao, J.F. and Celikyilmaz, A. (2017) End-to-End Task-Completion Neural Dialogue System. arXiv preprint arXiv:1703.01008

[11] Echihabi, A. and Marcu, D. (2003) A Noisy-Channel Approach to Question Answering. *Proceedings of the* 41*st Annual Meeting on Association for Computational Linguistics-Volume* 1 (*ACL*'03), Sapporo, 7-12 July 2003, 16-23. https://doi.org/10.3115/1075096.1075099

[12] Leuski, A. and Traum, D.R. (2010) Practical Language Processing for Virtual Humans. 22*nd Annual Conference on Innovative Applications of Artificial Intelligence* (*IAAI*-10).

[13] Ritter, A., Cherry, C. and Dolan, W. (2011) Data-Driven Response Generation in Social Media. *EMNLP*, 583-593.

[14] Sutskever, I., Vinyals, O. and Le, Q.V. (2014) Sequence to Sequence Learning with Neural Networks. *Advances in Neural Information Processing Systems*, **2014**, 3104-3112.

[15] Serban, I.V., Sordoni, A., Bengio, Y., *et al*. (2016) Building End-to-End Dialogue Systems Using Generative Hierarchical Neural Network Models. *AAAI*, **16**, 3776-3784.

[16] Sordoni, A., *et al.* A Hierarchical Recurrent Encoder-Decoder for Generative Context-Aware Query Suggestion. *Proceedings of the* 24*th ACM International on Conference on Information and Knowledge Management*, 553-562.

[17] Clark, H.H. and Brennan, S.E. (1991) Grounding in Communication Perspectives on Socially Shared Cognition. American Psychological Association, Washington, DC, 127-149.

[18] Graves, A. (2013) Generating Sequences with Recurrent Neural Networks. arXiv:1308.0850v5

[19] Bengio, Y., *et al*. (2003) A Neural Probabilistic Language Model. *Journal of Machine Learning Research*, **3**, 1137-1155.

[20] Mikolov, T., *et al.* (2013) Efficient Estimation of Word Representations in Vector Space. arXiv preprint arXiv:1301.3781

[21] Park, C., Kim, K. and Kim, S. Attention-Based Dialogue Embedding for Dialogue Breakdown Detection.