从Google语音助手说起：人机对话的未来趋势

郑银河　黄民烈
清华大学

关键词：对话系统　语言助手　人机对话

谷歌在近期召开的2018年度开发者大会(Google I/O 2018)上展示了语音助手Google Duplex。这项新技术可以通过电话与人进行对话，从而执行特定的任务，如预订餐馆、预约发廊等。用户只需要告诉Google Duplex自己的需求，它就会像一个真人助手一样，帮用户在后台给商家打电话，并向用户反馈预约结果。谷歌首席执行官桑达尔·皮查伊(Sundar Pichai)现场展示了两段Google Duplex的通话录音，通话过程的流畅和仿真程度引起了台下观众的阵阵惊呼：几乎无法辨别对话时面对的是真人还是机器。由此可见，在一些特定任务场景，人机对话系统的性能达到了实用的程度。这项技术引起了工业界和学术界的广泛关注，媒体宣传进一步推波助澜，唤起了人们对人机对话系统新的期待。

人机对话技术的研究进展

与机器自如地交谈，一直以来都是人类的美好愿景，相关场景也经常出现在各类科幻小说与电影中。如何与机器使用自然语言展开对话，是人工智能领域的一个重要研究课题。事实上，人机对话技术被视为人工智能领域最富挑战性的任务。一个机器具有的语言交互能力甚至被当作判断这个机器是否有“智能”的标准——最早的图灵测试就是以人机对话的方式进行设定的[1]。从某种意义上说，解决了开放领域的人机对话问题就等同于通过了图灵测试。

对话系统因为符合人类自然语言的交互特性，相比传统信息获取方式有很大的优势。用户与系统可以在多轮对话过程中通过询问、澄清、确认等方式，实现复杂场景的信息获取、任务完成、情感抚慰和陪伴等需求。随着未来机器人技术的发展，服务机器人、社交机器人必然会成为智能社会新的成员，而人机对话的技术能力则直接决定了“人机”和谐相处的可能性。

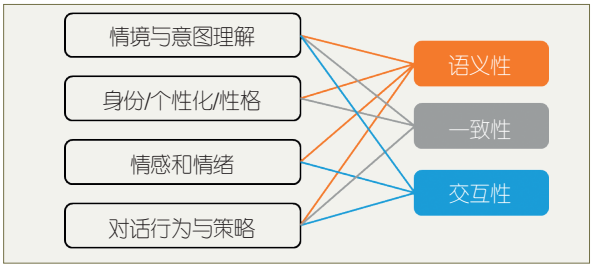


图1　人机对话技术所面临的主要问题

人机对话技术的主要挑战集中在如何解决交互过程中的语义性、一致性和交互性（见图1）。**语义性**涉及自然语言理解的根本问题，需要在交互过程中理解用户输入内容、上下文、对话背景（对话者身份、角色）与情境（情感、情绪），还常常涉及利用常识、世界知识、背景知识进行逻辑推理等。一

致性对于“类人水平”的对话机器人来说非常重要，常常体现为身份、个性化、性格等因素上的一致性。而这些因素如何体现在语言交互上，是非常前沿和具有挑战性的问题。**交互性**是对话系统的终极目标：通过有效的交互，满足用户的信息需求或情感需要，或实现特定的任务目标。因此，研究对话交互中的主动和被动行为、对话策略优化、对话状态跟踪与管理就显得非常必要。

人机对话技术从对话交互目标的类型上可大致分为三类（见表1）。第一类是以闲聊为主的人机对话。这类对话系统不以完成某个具体任务为目标，纯粹是为了与用户进行情感交流和沟通。这种人机对话技术通常应用于娱乐、情感陪护或者护理型机器人上。微软小冰就是其中的典型代表。她可以陪用户聊天解闷，甚至引导用户的情绪。这类对话系统所产生的回复不限定于某个特定领域，属于开放域的对话系统。

第二类人机对话技术主要用于任务导向的对话系统。这类系统具有明确的任务目标，通常是为了完成某一个具体的任务，如订餐馆、订票等。这样的系统可以极大地减少人力成本，因此受到了商业公司的广泛关注。这类对话系统所能处理的问题与回复通常只限定于某个特定领域或者某些内容范围。谷歌语音助手所展示的大部分功能就属于这一范畴。但受限于当前技术，使用任务型对话系统完全替代人工还不现实，因此**人机协作式**的对话系统是值得深入研究和探索的课题。

任务导向的对话系统还有一类比较特殊的分支，那就是知识问答系统。不同于常规的任务导向对话系统，知识问答系统主要关注于使用准确、简洁的自然语言来回答用户用自然语言所提出的问题，从而快速、准确地帮助人们获取信息。事实上，我们也可以将“获取信息”当作是某种“任务”。随着技术的进步，这类系统也在传统的单轮次问答模式上引入了多轮次的对话交互（如交互式的商品推荐），以应对复杂的信息获取需求。最近斯坦福大学推出的数据集CoQA，实际上就是基于给定的文档，进行对话式的问答任务。由于知识问答系统可以更便捷地帮助用户获取信息，其成为了目前人工智能领域中一个备受关注并且具有广泛发展前景的研究方向。这类人机对话技术不仅可以应用于某些特定领域，如客服、教育等，还可以应用于开放领域的知识问答。IBM Watson、搜狗公司的汪仔机器人就是这一类型的典型代表。

任务型对话系统

任务型对话系统由于其潜在的商业价值，始终是人机对话领域的研究重点。这类系统面向特定的任务，因此需要包含与任务相关的对话逻辑。这些对话逻辑又常常涉及复杂的领域知识库。关于任务导向的人机对话研究主要集中在对话意图理解、对话状态追踪、对话管理和对话生成几个方面。

对话意图理解是指从语义的角度对用户所输入的语句进行分析，属于解决**语义性**问题的重要组成部分。这一任务有时又被称为自然语言理解(NLU)，主要包含两个子任务：用户意图识别(intent understanding)和槽值检测(slot filling)。用户意图识别主要用来检测用户语句中所包含的意图。这些意图往往是在对话系统的设计阶段预定义好的，

表1 对话系统的分类[2]

	闲聊	任务执行	知识问答
目标定位	闲聊、好玩	完成任务和动作执行	知识获取
领域限制	开放领域	特定领域	特定领域/开放领域
应用场景	娱乐、情感陪护，服务机器人	个人助理、机器人预定系统	客服、教育、信息获取
参考系统	微软小冰	苹果Siri、谷歌Google Assistant、三星Bixby、Amazon Echo、百度Duer、微软小娜	IBM Watson、搜狗深度问答、Wolfram Alpha

因此用户意图识别通常被当作句子分类任务。很多现成的分类模型可以直接用来解决这个问题，如时下流行的深度学习模型。槽值检测主要用来提取用户输入语句中的槽值（也可理解为命令执行所需要的参数），其往往被视作一个序列标记任务：输入是一个词的序列，输出是每个词所对应的槽值。虽然有许多序列标注模型（比如循环神经网络）可以直接使用，但我们常常面临的挑战是当前领域可用但数据不足，因此研究小样本学习和领域迁移对于解决该任务意义重大。

对话状态追踪主要负责维护对话的历史信息，属于解决**交互性**问题的一部分。该模块会维护当前对话的状态，并随着对话的推进对其进行更新。目前大部分的对话系统将对话状态定义为完成某项任务所需的各个槽值的取值。传统方法一般依赖规则更新对话状态，基于这些方法的开发与维护耗时耗力，并且可扩展性较差。现有研究大都使用统计方法追踪槽值取值的概率分布，这些方法可以利用已有对话数据自动学习合理的对话追踪策略。一些常用模型包括最大熵模型、记忆神经网络模型等。

对话管理解决的主要问题是根据当前对话状态预测系统应该采取的动作，属于解决**交互性**问题的一部分。早期系统主要使用基于规则的方法，但是这些方法需要开发者对任务有较为深入的了解，并且手工编写模板的代价较高。这一情况随着统计学习模型的引入得到了改善，其中基于部分可观察马尔科夫决策过程 (POMDP) 的概率图模型在对话管理中取得了非常好的效果 [3]。这类模型可以通过概率推理的形式产生新的对话动作，并且通过强化学习方法进行训练。

任务导向的对话生成通常利用一个结构化的对话意图表示生成自然语言回复。传统的对话生成方法采用两个阶段的做法：首先是句子规划，也就是将输入的语义符号映射为一种语言学结构；然后是表层实现，即将这个结构转化成合适的句子。这些方法通常依赖于手工规则。近几年的新趋势是使用神经网络模型实现端到端的生成模型，通常有两种做法，从结构化意图表示到自然语句的生成 [4]，或者直接实现从自然语句到自然语句的生成。这些都是基于编码 - 解码结构的序列生成方法。

闲聊型对话系统

近些年，由于深度学习的兴起，基于大规模语料的闲聊式人机对话系统得到了快速发展。这类对话系统并不关注于某项特定的任务，其主要目标是和人们在开放领域展开对话。闲聊式对话系统的技术实现可以大致分为两类[1]：检索式模型 [5] 和生成式模型 [6]。

检索式模型的主要思路是从对话语料库中找出与输入语句最为相近的回复，这些回复通常是预先存储的数据。首先构建一个丰富的对话语料库，对于每一条输入语句，检索模型会从候选语句中挑选语义匹配度最大的语句作为其回复。检索模型的核心是其所使用的语义匹配算法。早期的研究工作大多只关注单轮对话，这些方法不能有效利用上下文信息做出准确匹配。近年来，基于多轮对话的检索式对话系统受到越来越多的关注，这些模型在选择回复的过程中不只考虑了当前的对话，还考虑了丰富的历史对话。这些上下文信息极大地提升了检索式模型的性能。同时基于深度学习的匹配算法在匹配性能上也有了较大的提高。

与检索式模型不同，生成式模型的主要目标是根据当前对话的上下文信息生成回复。回复有可能是模型在训练阶段没有见过的崭新回复。生成式模型借鉴了机器翻译的思路，随着序列到序列 (Seq2Seq) 转换模型在机器翻译中取得成功，其在对话生成模型中也得到了广泛的应用 [7]。这些模型一般采用“编码 - 解码”结构：在编码端使用一个循环神经网络将输入语句编码为一个向量表示，而在解码端使用

[1] 还有一类混合模型：同时考虑检索结果和生成模型，本文不做讨论。

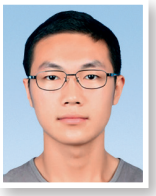
另一个循环神经网络，并采用注意力机制逐一生成回复内容[8]。

然而，当前的聊天机器人在语义性、一致性和交互性方面还存在显著不足。在语义性方面，基于深度学习的生成模型更容易生成一些无意义的万能回复，生成内容的信息量、合适性、逻辑性还存在较大不足，距离真正意义上的语义理解还有很大距离。在一致性方面，很容易在多轮交互中产生语义、身份、个性上的冲突，具有明显的“机器”特点。在交互性方面，当前的聊天机器人在情感交互、策略应对方面还存在显著不足，不能根据用户的话题、状态自适应地调整自身的策略，如话题策略、主动被动策略、情感表达策略等，还无法实现流畅、自然的人机交互。

随着技术的发展与媒体的宣传，人们更倾向于将对话系统当作是自己的生活伴侣，而不是一个只能用于执行任务的机器。为了满足用户的这一期望，需要在对话生成过程中考虑很多“类人”的特性，如具有一定的情商，能够进行情感响应和交互[9]，或在对话交互中体现个性、语言风格、性格[10]等。同时，构建一个真正“智能”的对话系统也离不开对知识的处理[11]。也就是说，一个“类人”的对话机器人应该是聪明的、有身份、有个性、有自己风格的，是一个能够从人类伙伴和环境的交互中不断学习和成长的社交个体。

总结

对话系统几乎涵盖了所有自然语言处理领域的难题，一直以来都是人工智能领域的研究重点。近年来，随着深度学习方法的崛起，相关技术不断进步，有效推动了对话系统的发展。工业界涌现出了一批适用于各种场景的对话系统，这些系统有助于人们解决很多实际问题，极大提升了人们的工作效率。然而目前对话系统在语义性、一致性、交互性方面还存在显著不足，构建“类人”水平的对话机器人还需要在情境理解，个性化、性格、风格嵌入，知识运用等多方面开展持续研究。 ■



郑银河

清华大学计算机系博士后，北京三星通信研究院工程师。主要研究方向为自然语言处理、对话系统等。

zhengyinhe1@163.com

黄民烈

CCF专业会员。清华大学计算机系副教授，博士生导师，人工智能研究所副所长。主要研究方向为深度学习、强化学习、自然语言处理等。

aihuang@tsinghua.edu.cn

参考文献

[1] Turing A. Computing machinery and intelligence[J]. *Mind*, New Series, 1950, 59(236): 433-460.

[2] 张伟男，张杨子，刘挺. 对话系统评价方法综述[J]. 中国科学：信息科学，2017, 47(8):953-966.

[3] Young S, Gašić M, Thomson B, et al. POMDP-based statistical spoken dialog systems: A review[J]. *Proceedings of the IEEE*, 2013, 101(5): 1160-1179.

[4] Wen T, Gasic M, Mrksic N. Semantically Conditioned LSTM-based natural language generation for spoken dialogue systems[C]// *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (EMNLP 2015), 2015: 1711-1721

[5] Ji Z, Lu Z, Li H. An information retrieval approach to short text conversation[OL]. (2014) arXiv preprint arXiv:1408.6988.

[6] Vinyals O, Le Q. A neural conversational model[OL]. (2015). arXiv preprint arXiv:1506.05869.

[7] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[OL]. (2014). arXiv preprint arXiv:1409.0473.

[8] Sutskever I, Vinyals O, Le Q. Sequence to sequence learning with neural networks[C]// *Advances in Neural Information Processing Systems*. 2014: 3104-3112.

[9] Zhou H, Huang M, Zhang T, et al. Emotional chatting machine: emotional conversation generation with internal

and external memory[C]// *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence* (AAAI 2018), New Orleans, Louisiana, USA.

[10]Qian Q, Huang M, Zhao H, et al. Assigning personality/identity to a chatting machine for coherent conversation generation[C]// *Proceedings of the 2018 International Joint Conference on Artificial Intelligence* (IJCAI-ECAI 2018), Stockholm, Sweden.

[11]Zhou H, Yang T, Huang M, et al. Commonsense knowledge aware conversation generation with graph attention[C]// *Proceedings of the 2018 International Joint Conference on Artificial Intelligence* (IJCAI-ECAI 2018), Stockholm, Sweden.