

# 关于人机对话系统的思考

## Thoughts on Human-Computer Dialogue System

中图分类号: TN929.5 文献标志码: A 文章编号: 1009-6868 (2017) 04-0047-004

**摘要:** 提出了一系列非常重要、影响人机对话质量的问题,包括:如何面向自然语言理解(NLU)构建对话任务分析、深度推理,如何利用语言学尤其是互动语言学研究成果构建对话管理(DM),如何有效建模人机对话中不同任务间的关约束来发展联合模型等。认为尽管人机对话系统的基础模型已取得了长足进步,但如果不能有效地解决上述问题,就不可能获得高质量的人机对话系统,自然语言处理的水平也就难以得到实质性提升。

**关键词:** 人机对话系统; NLU; DM; 自然语言生成(NLG)

**Abstract:** In this paper, a series important problems which have significant influence on the quality of human-computer dialogue are proposed, including how to normalize task analysis and introduce deep reasoning for natural language understanding (NLU), how to build linguistically, especially interact linguistically reasonable dialogue management (DM), and how to develop joint method for modeling correlations between different subtasks in dialogues. There has been lots of progresses on human-computer dialogue, but if the problems mentioned above cannot be solved, it is difficult to achieve high quality dialogue systems and important improvements in natural language processing.

**Key words:** human-computer dialogue system; NLU; DM; natural language generative(NLG)

王小捷/WANG Xiaojie

(北京邮电大学, 北京 100876)  
(Beijing University of Posts and  
Telecommunications, Beijing 100876,  
China)

- 对话任务分析的自动化、深度推理的缺乏是自然语言理解面对的重要挑战,目前相关工作较少,需要大力发展
- 对话管理应汲取会话分析、互动语言学等方面的研究成果,研究相关计算模型,才有可能构建自然和谐的人机对话系统
- 联合建模人机对话系统多个子任务的方法比传统的管道方法具有诸多的优势,深度学习提供了有效联合建模工具,已成为一个重要的技术发展趋势

人类社会的一个主要特征是人际交流,广泛深入的人际交流是促进合作,推动社会向前发展的重要动力。人际交流的主要工具是包括汉语、英语在内的各种自然语言,基于自然语言的对话是人际交流的主要方式之一。基于自然语言的对话也称为会话,可以是在两人之间进行,也可以在多人之间进行。此外,自说自话的独白也常被语言学家归入会话。

近年来,随着技术日新月异的发展,各种人造机器成为人类的亲密伙

伴,人与机器的交流变得越来越重要。为了与机器交流,让机器明白人类需要它们做什么,人类发明了很多种机器语言。开始,这些机器语言只是少数人使用,随着机器进入到普通人的日常社会生活,就需要机器能与普通人进行交流,发明越来越简单的机器语言或操作方式是一个途径,而让机器能懂得人类语言无疑是一个更自然、更令人向往的方式。人机对话系统旨在实现这种方式,满足普通人与机器进行顺畅交互的需求。

人机对话系统经历了 50 多年的发展,取得了不少的进展,目前更是面临巨大的发展机遇。如何抓住机遇,促成人机对话系统的进一步发

展,是每位研究者都需要认真思考的问题。

### 1 人机对话系统简述

20 世纪 60 年代,Weizenbaum 创建了 ELIZA 对话系统<sup>[1]</sup>,试图模仿心理治疗师与人进行对话交流;90 年代,Senef 等人开发了自动机票预订系统(ATIS)<sup>[2]</sup>,通过与用户进行多轮对话来获取用户需求(如出发地、出发日期、航班号等),以帮助用户预订机票。这两个系统分别代表了两类不同的对话系统,后者属于目标驱动型对话,前者属于非目标驱动型对话。

目标驱动型人机对话系统通过与用户进行多轮基于自然语言的对话

收稿时间: 2017-05-26  
网络出版时间: 2017-07-05

话来逐步收集与目标相关的信息,辅助用户成功获得某种服务。与 ATIS 系统类似的还有电信服务系统 HMIHY<sup>[3]</sup>、天气信息查询系统 JUPITER<sup>[4]</sup>、旅行计划制定系统 DARPA Communicator<sup>[5]</sup>等。目标驱动型的人机对话系统也称为面向任务的人机对话系统。

而以 ELIZA 系统为代表的非目标驱动型人机对话系统通常只是对用户输入进行基于自然语言的响应,而不完成特定信息服务任务,类似于闲聊似的对话,所以也称为聊天机器人。与目标驱动型系统相比,这类系统对收集信息、自主确定对话行为等方面的要求都较弱。

一个典型的任务驱动型人机对话系统的技术模块如图 1 所示<sup>[6]</sup>,系统主要包含自然语言理解(NLU)、对话管理(DM)、自然语言生成(NLG)3个部分。用户输入可以是语音,也可以是文本,如果是语音,则 NLU 部分包含语音识别模块;而系统输出为语音时,NLG 部分还需包含语音合成模块。NLU 部分进一步又分解为3个子任务:其一是领域识别,判定用户谈论的是什么领域的事情;其二是意图识别,识别用户话语的目的,比如是告知某个信息,还是确认某个信息;其三是语义分析,又称为槽填充,旨在标识用户话语中与目标有关的语义类别,比如预订机票时需要的出发地、目的地、时间等语义类别。DM 则进一步分解为对话状态追踪和动作选择两个子任务;对话状态追踪获取对话的当前状态;动作选择确定系统响应的动作。DM 模块负责控制整

个对话过程,例如:决定系统在上一轮对话中获取了用户出发地信息之后,下一步该和用户说什么。NLG 模块则把决定要说什么具体转化为自然语言,通常也进一步分解为几个子任务。

非目标驱动的对话系统大致按此结构,但是大多没有显式的意图识别、语义标注等 NLU 过程,也没有显式的 DM。

在实际应用中,目标驱动与非目标驱动的划分也并不是绝对的,常常会需要目标驱动和非目标驱动混合出现。比如:目前的个人助理系统 Apple Siri、Microsoft Cortana 等,有时可能要与用户闲聊,这时是非目标驱动的,有时可能要帮助用户订票,这时是目标驱动的。这一系统框架经历了一系列的发展,其中的各个技术部分也在飞速发展<sup>[7-9]</sup>。

## 2 自然语言理解

人机对话系统中 NLU 的主要目标是识别输入话语的领域和意图,获得任务相关的语义信息。例如:在订机票时,客户说“订1张5月10号去北京的票”,其领域是“航班信息服务”,意图是“订票”,与订票任务相关的语义信息包括作为“订票数量”的“1张”、作为“起飞时间”的“5月10号”,以及作为“目的地”的“北京”。订票数量、起飞时间和目的地称为槽,句子中提供的这些槽的具体值称为槽值。

如何准确地进行领域、意图和槽值的识别或标注是当前针对人机对话的 NLU 研究中得到关注最多的方

面。近年来人们不断设计新的模型和算法来提升语言理解的性能<sup>[9]</sup>,而另一个研究重点是当面对的对话任务发生变化时,如何快速准确地将模型推广到新的对话领域中,即模型的跨领域适应问题。

然而,笔者认为,目前的语言理解中还存在1个问题没有得到应有的重视,即:如何为 NLU 模块设定目标。从前面的例子可以看到,目前的语言理解对于给定的用户输入通过分析获得什么信息都是基于对话任务而确定的。比如:NLU 算法只会在句子中去标注订票数量、起飞时间和目的地等槽的槽值,而不会去标注“运动类型”等无关的信息。但是对话系统如何能知道它只应该去关注这些信息,而不应关注其他信息呢?目前的方法是完全由人预先基于对话任务来确定语言理解的目标,再基于这些目标来训练模型。这种方法需要预先对任务有完整的分析,建立关于任务的完整意图和槽的体系。这类似于软件工程中在构建一个软件时需要完整的需求分析一样。对于软件工程,有一套需求分析技术,而对于人机对话系统,目前还并没有这样的任务分析技术,一般的人机对话系统的任务分析还停留在经验阶段,这给人机对话系统的普及带来了困难。

进一步,在目前的 NLU 中,对于语言的理解还停留在比较浅的层次。例如:在槽标注时,大多数模型还只能在语言表层进行标注,在复杂情况下结合推理进行深入理解的能力还较弱或者基本没有。例如,客户说“订一张端午节后一天的票,如果那天预报有大雾的话就往后推一天。”这时简单地在语言表层标注时间槽已经不太可能,需要结合其他信息并通过推理得到正确的槽值。语言推理不仅需要对话语言本体的理解,其背后还需要更为丰富的关于现实世界对象、事件及其关系的知识进行支撑,并且有的知识是显式的,而有

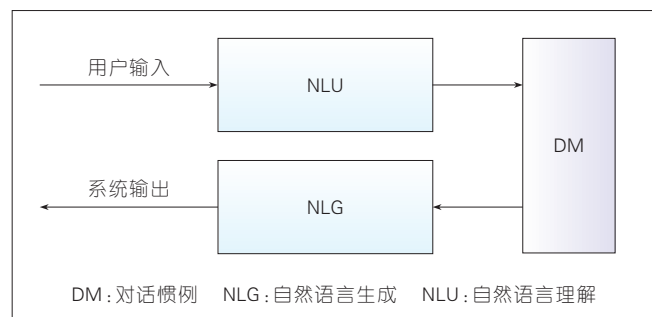


图1 人机对话系统框架

些知识是隐式的。例如,上一句中的“端午节”“下雨”等是现实世界中的对象和事件,其中,端午节具体是哪一天可以直接查询获得。但是其中还含有简介的、隐式的推理知识,比如在上句中人可能会依据如下的隐含逻辑去理解说话人:如果有大雾的话,飞机可能不能按时起飞,客户大概不希望延误,所以决定在这种情况下推迟一天,推迟的目的是为了等一个好天气。如果知道存在这样一个隐含的推理,那么作为好的客服人员,可能就要再问客户:如果后一天还报有大雾怎么办?这类知识以及要基于这类知识进行的推理将导致对语言的理解更为复杂和困难。

### 3 对话管理

DM 追踪对话状态,并基于此决定系统的响应动作。对话状态追踪和动作选择的研究已得到了广泛关注,取得了大量进展。这些研究一方面由 DM 计算模型的发展所带来<sup>[79]</sup>,例如,从马尔科夫决策过程(MDP)到部分可观马尔科夫决策过程(POMDP),从 Q-Learning 算法到深度 Q 网络(DQN);另一方面也离不开对语言学研究成果的有效借鉴,毕竟人机对话是希望建模人人对话,只有对人类对话的本质有了更深刻的理解,才有可能在人机对话的研究中取得相应的进展。这方面,已有的借鉴包括会话分析研究中话轮、相邻对等概念,它们已成为目前计算模型的建模单元,此外,言语动作等语用学的概念也已成为 DM 的决策目标。

但是,更多、更深入的语言学研究成果还没能在计算模型中得到有效体现。如格莱斯的对话合作原则、礼貌原则等还很少进入计算模型的考量范围。近 20 年来,一些新兴的关于语言的理论也还几乎没有进入计算语言学家的视野,其中就包括笔者认为对于对话建模和 DM 具有重要借鉴价值的互动语言学。

互动语言学<sup>[80]</sup>考察和研究日常真

实人类互动过程中的语言使用情况,探讨语言结构及其运用模式与互动交流是如何相互影响的,为互动参与者在真实、情景化的日常交往中完成互动与交流提供理论解释和支持。互动语言学者从功能语言学、会话分析的成果以及对真实对话数据的研究出发,对传统的语言学有直接的批评。传统的以乔姆斯基生成语言学为代表的语言学者认为以语言能力为基础的语言结构的理想形式是语言学研究的对象,语言行为是语言能力在具体使用情景中的实现。互动语言学者则认为,语言的结构和意义是在情景互动过程中产生,并随互动的发展不断变化,是说话者和接受者共同协商的产物。例如:形式语言学将句子这一语言单位作为单个说话者语言能力的产物;而互动语言学者则认为,句子的生成是说话者和接受者在交际过程中互动的产物,它不仅与说话者有关,还关系到接受者,因为说话者在互动中往往需要根据接受者的反应调整话语<sup>[81]</sup>。

这种语言结构和意义可以基于互动而不断进行建构的观点如果能用于 DM,帮助构建更好的对话结构,进行更好的对话言语行为决策,将是十分有价值的。至于如何借鉴,则还需要深入了解互动语言学的相关成果,同时综合考虑在当前科学与技术发展水平下,构建其数学模型和计算模型的可行性。

笔者在此仅举一个可能的例子:交互语言学认为语言结构是在语言交互中通过互动而产生的,那么一个好的对话行为决策所基于的对话语言分析和理解可能不应仅局限在一个说话人话轮中进行。在对话中,将对话双方的邻接话轮,即相邻对,作为一个分析单元可能更为可取。目前的模型是可以基于多轮对话信息来进行对话行为选择。但是,通常只是简单地拼接对话历史,而没有考虑对话历史中的结构信息。而如果以相邻对为基础单元,在计算模型上可

以构建关于对话的层次结构,有利于建模对话的语言结构,从而为对话行为提供更准确的决策信息。

### 4 系统构建

传统上,一般采用分而治之的方式实现如图 1 所示的人机对话系统。即:先分别构建 NLU、DM 和 NLG 等模块,这些子模块通常还要分解为更小的子任务分别建模,之后按序将这些模块连接起来。这种方案的优势是简洁清楚,各部分的任务明确,可以分别开展研究,各自解决各自的问题。

但是这种研究方案的问题也很明显。首先,各个子任务分别研究导致难以建模子任务之间的关联约束信息,而有时这种关联约束信息是非常有价值的。例如:在 NLU 模块中的意图识别和语义分析两个子任务就是紧密相关的,一个具有告知所需航班起飞时间意图的句子很可能包含时间信息,反之,包含时间信息的句子很可能是表明了一种告知时间的意图。这两个子任务的交互约束信息是可以互相帮助,共同提高彼此性能的。而如果分别建模,这种交互约束信息就不能得到充分利用了。

其次,各部分分别研究时,某个任务产生的错误结果只能用来调整这个任务本身,而不能有效地反传到前面的子任务上,而实际上后一任务的错误很可能是前面子任务错误的累计导致的。例如:在 DM 的决策出现错误时,其原因可能是状态追踪发生错误,也有可能是语言理解部分发生了错误,如果分别建模,则 DM 的错误不可能直接用于对 NLU 部分的错误进行修正。

因此,人机对话系统的一个重要发展方向是多任务的联合建模。联合可以有不同的层次,例如:NLU 模块内各个子任务的联合建模,综合利用意图识别和语义分析两个子任务之间的交互约束信息,可以获得更好的理解性能;综合 NLU 与 DM 的跨模



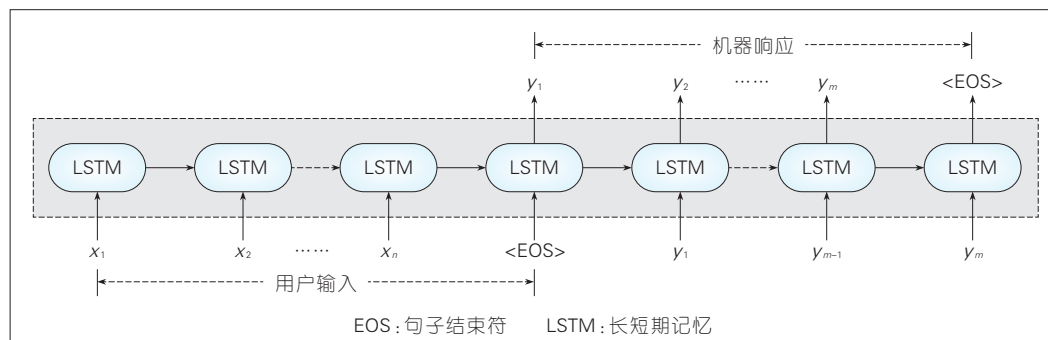
块联合建模,使得DM的决策错误不仅可以用于修正状态追踪的错误,也可以直接用于修正语言理解的错误。

另一方面,近年来深度神经网络研究的发展也为联合建模提供了重要的技术工具。2015年出现了多种基于序列到序列深层神经网络模型的非目标驱动型对话系统。图2是一个简单的例子,其中输入句子在编码后逐步解码得到机器输出的响应。初期的模型大多只是处理一问一答的应答系统,而随着输入中加入对话历史,应答系统逐步向对话系统发展。与此同时,随着当前深度强化学习的发展,面向目标驱动型对话系统也逐渐成为可能。

这种跨越NLU和DM的联合建模具有特别重要的意义,因为它将语言输入和决策(说什么)直接关联起来了,跨越了NLU任务。实际上,我们在人际对话中是难以看到人的语言理解结果是什么,而只能看到决策的结果是什么。我们目前定义的NLU任务只是我们给人机对话引擎设计的,人在对话中是否有类似的理解过程和结果其实是不清楚的,也有可能不存在或与我们定义的不一样。我们并不是要否定NLU任务定义本身,而是认为,如果未来能发展出直接关联语言输入和决策结果的好模型,那么基于对这种模型内部状态的分析,也许可以帮助我们了解在对话过程中真正需要的NLU任务是什么。

## 5 结束语

人机对话系统业已展现出广泛的应用价值,也在很多实际问题中发挥了作用。在微软Build 2016开发者大会上,微软首席执行官纳德拉提出了作为平台的对话(CaaP),将对话作为平台技术,作为整合人工智能技术



▲图2 序列到序列的应答系统

的途径,这吸引了非常广泛的研究和商业关注。

从目前的问题研究和技术水平来看,为特定任务提供信息服务的人机对话系统,在进行精心的任务定义和设计之后,是可以获得较好的服务质量的,但是其鲁棒性不能得到有效保证。处于任务边界的话语将很可能导致系统不能优雅地退出工作,而对于更为复杂的多任务、跨任务等情形,系统的鲁棒性就更难以保证了。然而,反观人类的日常人际对话,任务边界模糊在大多数情况下并不是问题,除非进入非常专业的领域,跨领域并不太影响人类的人际交互能力,鲁棒性似乎是人际对话的一个基本特质。人际交互时不懂可以马上就问,这本身也就是维持对话的一种方式,而不像在阅读时遇到不懂问题就难以继续推进阅读本身。这些迹象表明,也许目前我们还远没有触及到人机对话系统的核心问题,未来还需要更多、更深入的工作。

### 参考文献

- [1] 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. DOI:10.1145/365153.365168
- [2] SENEFF S, HIRSCHMAN L, ZUE V W. Interactive Problem Solving and Dialogue in the ATIS Domain[C]// The Workshop on Speech and Natural Language. Association for Computational Linguistics, 1991:354–359. DOI:10.3115/112405.112743
- [3] GORIN A L, RICCARDI G, WRIGHT J H. How May I Help You?[J]. Speech Communication,

1997, 23(1–2):113–127. DOI:10.1016/S0167-6393(97)00040-X

- [4] ZUE V, SENEFF S, GLASS J R, et al. JUPITER: A Telephone-Based Conversational Interface for Weather Information[J]. Speech & Audio Processing IEEE Transactions on, 2000, 8(1):85–96. DOI: 10.1109/89.817460
- [5] WALKER M, ABERDEEN J, BOL J, et al. DARPA Communicator Dialog Travel Planning Systems: The June 2000 Data Collection[J]. Veterinary Pathology, 2001, 9(3):190–201
- [6] JURAFSKY, D. MARTIN, J. H. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition[M]. New Jersey: Prentice-Hall
- [7] YOUNG S, GASIC M, THOMSON B, et al. POMDP-Based Statistical Spoken Dialog Systems: A Review[J]. Proceedings of the IEEE, 2013, 101(5):1160–1179. DOI: 10.1109/JPROC.2012.2225812
- [8] YU K, CHEN L, CHEN B, et al. Cognitive Technology in Task-Oriented Dialogue Systems: Concepts, Advances and Future[J]. Chinese Journal of Computers, 2014, 37(18): 1–17
- [9] WANG X, YUAN C. Recent Advances on Human-Computer Dialogue[J]. CaaI Transactions on Intelligence Technology, 2016, 1(4):303–312. DOI: 10.1016/j.trit.2016.12.004
- [10] 林大津, 谢朝群. 互动语言学的发展历程及其前景[J]. 现代外语, 2003, 26(4):411–418
- [11] 罗桂花. 互动语言学:语言产生于互动 互动塑造语言[N]. 中国社会科学报, 2012–10–8 (A07)

### 作者简介



王小捷, 北京邮电大学教授、博士生导师、智能科学与技术中心主任, 中国人工智能学会常务理事、自然语言理解专业委员会主任; 主要从事自然语言处理、人机对话、多模态认知计算等方面的研究; 已在相关领域发表学术论文200余篇。