任务型对话系统研究综述

赵阳洋 王振宇 王佩 杨添 张睿 尹凯

(华南理工大学软件学院 广州市 510000)

摘 要 人机对话技术作为人工智能领域的重要研究内容，它是人与机器的一种新型交互方式，受到学术界和工业界的广泛关注。近些年来，得益于深度学习技术在自然语言领域的突破性进展，极大地促进了人机对话技术的发展。将深度学习融入人机对话系统技术中，不但使得端到端的方法成为可能，而且提取出的特征向量非常有效几乎完全取代了人工特征。本文首先回顾了人机对话系统的发展历程，介绍了人机对话系统的两种类型，任务型对话系统和非任务型对话系统。其次，本文从理论模型、研究进展、可用性及存在的问题与挑战等角度深度剖析了任务型对话系统的两种方法，管道方法和端到端方法。重点分析深度学习技术和强化学习技术的具有代表性的前沿算法，并与传统方法进行对比。最后，对任务型人机对话系统目前的评估方法和存在的问题进行总结，并展望了任务型对话系统的未来研究方向。

关键词 对话系统；任务型对话系统；深度学习；强化学习；管道方法；端到端方法
中图法分类号 TP391

A Survey on Task-Oriented Dialogue Systems

ZHAO Yang-Yang WANG Zhen-Yu WANG Pei YANG Tian ZHANG Rui YIN Kai

(Software Engineering, South China University of Technology, Guangzhou 5100000)

Abstract The human-machine dialogue system is the core technology in the field of artificial intelligence. It is a new way of harmonious human-computer interaction, whose purpose is to provide useful information and help for users by communicating with humans in a natural and fluid language. In recent years, the breakthrough progress in deep learning technology has greatly promoted the development of human-machine dialogue technology. Therefore, human-machine dialogue technology has made substantial progress in various areas, including virtual personal assistants, entertainment, emotional chaperone, and conversational recommendation. In this paper, we first systematically describe the development process of human-machine dialogue system, and divide the human-machine dialogue system into two types according to different application scenarios, namely task-oriented dialogue systems and non-task-oriented dialogue systems. As a key branch of human machine dialogue system, task-oriented dialogue systems provide a convenient interface to help users complete tasks and have been used in a variety of applications. The non-task-oriented dialogue systems, also called known as chat robot, are different from the task-oriented dialogue systems. There are used in open field scenarios and can handle a wide variety of problems, relying on various information and ontology in the real world to solve those

本课题得到广东省科技计划项目重大专项(No.2015B01013100)、广州市科技计划项目产业技术重大攻关计划(No.201802010025)、广州市高校创新创业教育平台建设重点项目(No.2019PT103)资助. 赵阳洋，女，1995年生，博士研究生，计算机学会(CCF)会员（94136G）,主要研究领域为对话系统、强化学习和深度强化学习.E-mail: yyang.zhao@qq.com.王振宇(通信作者)，男，1966年生，博士，教授，主要研究领域为自然语言处理、对话系统和深度强化学习.E-mail: wangzy@scut.edu.cn.王佩，女，1993生，硕士研究生，主要研究方向为任务型对话系统、对话管理。Email: sewangpei@mail.scut.edu.cn.杨添，男，1994年生，博士研究生，主要研究领域为任务型对话系统. E-mail: 547941595@qq.com.张睿，男，1993年生，博士研究生，主要研究领域为对话系统、情感对话生成. E-mail: z.rui16@mail.scut.edu.cn.尹凯，男，1995年生，硕士研究生，主要研究领域为任务型对话系统. E-mail: sekaiyin@mail.scut.edu.cn.

problems. Secondly, from the perspectives of theoretical model, research progress, usability, problems and limitations, we analyze two main methods of task-oriented dialogue systems deeply, including pipeline method and end-to-end method. For the pipeline method, there are three modules to be introduced: Natural Language Understanding (NLU), Dialogue Management (DM), and Natural Language Generation (NLG). In the last part of this section, we summarize the advantages and disadvantages of the pipeline model. Among them, we divide the three tasks of the NLU module into text classification problems and sequence labeling problems. We describe the technical development of these two types of problems respectively, and finally summarize the current latest technological developments and trends of NLU. In the DM section, we divide the DM into dialogue state tracking (DST) task and dialogue policy learning (DPL) task. And we not only analyze and compare the representative algorithms for deep learning and reinforcement learning for DST and DP, but also describe the difficulties of the three challenging scenarios faced in the DPL process. In the end of DM section, we summarize the latest solutions and ideas to difficulties of the DM model in the task-oriented dialogue system. For the NLG module, we focus on the analysis of the relevant traditional methods and the deep learning techniques, and compare and analyze the two technologies. For the end-to-end method, there are two frameworks to build an end-to-end dialogue system, including a framework based on supervised learning and a framework to optimizing end-to-end dialog systems using reinforcement learning. At the end of this section, we summarize the advantages and limitations of the end-to-end method. Finally, we summarize the current issues that limit the development of task-oriented dialogue systems, and we discuss some of the potential trends in the future development of task-oriented dialogue systems, including low-resource task-oriented dialogue system, task-oriented dialogue system with domain adaptability and task-oriented dialogue system with domain knowledge and common sense, enhancing the understanding and reasoning ability of system.

Key words dialogue system; task-oriented dialogue system; deep learning; reinforcement learning; pipeline method; end-to-end method

1 引言

人机交互（Human Computer Interaction，HCI）[1]作为信息时代人类与计算机之间信息交流的基础技术，受到学术界和工业界的广泛关注。人机对话（Human-Machine Dialogue）是人机交互技术的核心领域，旨在最大限度地模仿人与人之间的对话方式，使得人类能够用更自然的方式与机器进行交流。

对话系统的发展历程可以归纳为三个阶段：基于符号规则和模板的对话系统；基于统计机器学习的对话系统和基于数据驱动的深度学习的对话系统。以 Eliza[2]为首的第一代对话系统，主要依赖专家制定的人工语法规则和本体设计。这种方法易理解，但是由于其全部使用符号规则和模板需要消耗大量的人力和物力，导致跨领域的扩展性严重不足。第二代对话系统不需要人工设计规则和模板，通过统计机器学习方法降低对话系统的手工复杂性。这种方法具有弱学习能力，但是解释性差、不易修补漏洞，仍旧难以扩大规模。近年来，第三代对话系统是目前研究的主流，使用深度学习取代浅层学习，使端到端的学习变得可行。2014 年以来，得益于网络数据的海量增长和深度学习等技术的长足进步，对话系统也因此突破传统方法的瓶颈，获得新的发展前景[3]。

对话系统的迅猛发展也引起工业界的广泛关注，以微软小冰为代表的聊天机器人使得人机对话技术更具实用价值和商业价值，但仍在自然性、逻辑性和流畅性等方面和人类有一定的差距。在任务型对话系统领域，苹果 Siri、亚马逊 Echo、微软敦煌小冰等开始步入应用阶段，帮助用户便捷处理复杂任务，减轻了人工负担。但是，这些系统的实现离不开大量手工定制规则模板，工程量巨大，缺乏通用性和可移植性，技术方法有待进一步探索。目前，任务型对话系统逐渐应用在在各行各业，使得“人机”交互方式不再是简单的输入设备和触摸屏，而是具有个性化的自然语言交互。

随着人机对话的迅猛发展，涌现出许多从不同角度描述人机对话系统的相关研究综述。文献[3]从认知技术角度概述任务型人机对话，文献[4]重点

阐述了对话生成模块中的众多深度学习模型，其对话生成的方法没有区分对话系统类型，既有适用于任务型对话系统的方法，又有适用于聊天机器人的方法。随着深度学习逐渐融入到人机对话系统技术中，京东数据科学团队编写了一个非常全面的对话系统综述[5]，从任务型对话系统和非任务型对话系统分别概述其主要的深度学习模型，内容涉及比较全面。但只列举了主要方法，较少进行横向比较，并且未覆盖注意力机制、预训练神经网络、端到端方法、多任务学习和迁移学习等研究热点，不足以让参考文章的研究人员了解掌握最新研究进展。深度学习、强化学习和知识图谱等技术的不断进步极大地促进了对话系统的出现和发展，新的技术不断涌现，急需对新的技术和热点进行归纳和梳理的综述文献。本文对任务型对话系统进行了更全面的调研，将最新的研究进展和热点进行归纳和梳理，希望对对话系统的进一步研究具有指导意义，具有以下贡献：

- 按类型对对话系统进行了划分和对比，帮助读者清晰理解和区分任务型对话系统和非任务型对话系统。然后从理论模型、研究进展、可用性及存在的问题与挑战等角度深度剖析了任务型对话系统的两种方法，重点分析深度学习技术和强化学习技术中具有代表性的前沿算法，并与传统方法进行对比。同时，对每一部分进行归纳总结。对于对话策略部分，不但列举了当前较新的代表性方法，还对其在具有挑战的场景进行重点分析，并针对每一个场景对当前已有的方法进行了详细剖析。
- 系统地归纳了任务型对话系统的评估方法，包括现有的标准测试集等，有助于读者更广泛地了解任务型对话系统的性能评估标准。
- 从低资源启动、域适应能力和引入领域知识和常识三个方面，分析和讨论了未来的研究方向。本文希望通过揭示任务型对话系统的近期的进展和热点，帮助研究人员选择任务型对话系统的未来方向。

第 2 节描述了对话系统的类型划分；第 3 节综述了现有的任务型对话系统方法，并对现有的方法进行分析对比，最后对现有方法的问题与挑战进行概括；第 4 节总结了任务型对话系统的评估方法；第 5 节讨论任务型对话系统的未来研究方向；最后给出总结与展望。

2 对话系统类型

根据不同的应用场景将对话系统分为两种类型：任务型对话系统（task-oriented dialogue systems）和非任务型对话系统(non-task-oriented dialogue systems)，非任务型对话系统又称闲聊机器人(chat bots)，其对比情况如表 1。

任务型对话系统面向垂直领域，目的是使用尽可能少的对话轮数帮助用户完成预定任务或动作，例如预定机票、酒店和餐馆等。大多数任务型对话系统对话数据规模较小，难以通过大量数据进行模型训练，前期需用手工制定的规则解决冷启动问题，这使得对话系统的构建变得昂贵和耗时，限制了对话系统在其他领域的使用。

非任务型对话系统面向开放领域，要求其回复具有一致性、多样化和个性化。由于话题自由，因此对系统的知识要求极高。实际的非任务型对话系统容易产生“安全回复”问题，如“我不知道”,“我也是”,“好的”等，使得聊天机器人的大多数答案趋近相同。同时，聊天是一个连续交互的过程，句子的语义需要结合对话上下文才能确定。但目前非任务型对话系统的语料大多是从社交网络爬虫所得，缺乏多轮对话相关的上下文语料，导致非任务型对话系统难以保持上下文信息的一致性。因此，非任务型对话系统离实际应用尚有差距。

表 1 任务型对话系统和非任务型对话系统对比表

	任务型对话系统	非任务型对话系统
目的	完成任务或动作	闲聊
领域	特定域（垂直域）	开放域
对话轮数评估	越少越好	越多越好
应用场景	虚拟个人助理	娱乐、情感陪护、营销沟通
典型系统	Siri、Cortana、敦煌小冰、度秘	小冰

3 任务型对话系统方法

任务型对话系统主要的有两种方法：管道方法（Pipeline Method）和端到端方法（End-to-End Method）。本节将从理论模型、研究进展、可用性及问题与挑战等方面分别介绍这两种方法。

3.1 管道方法

一个完整的任务型对话系统的管道方法主要包括 5 部分：自动语音识别（Automatic Speech Recognition，ASR）、自然语言理解（Natural Language Understanding，NLU）、对话管理（Dialogue Management，DM）、自然语言生成（Natural Language Generation，NLG）、语音合成（Text To Speech，TTS）。其中，对话管理包括对话状态跟踪（Dialogue State Tracking，DST）和对话策略（Dialogue Police，DP）。管道方法框架图如图 1 所示（本文不对语音识别和语音合成模块进行介绍）。

本文将概述 NLU、DM 和 NLG 三个模块的技术发展过程，并将其与现有技术进行分析对比，最后对其技术的难点和发展趋势进行总结。

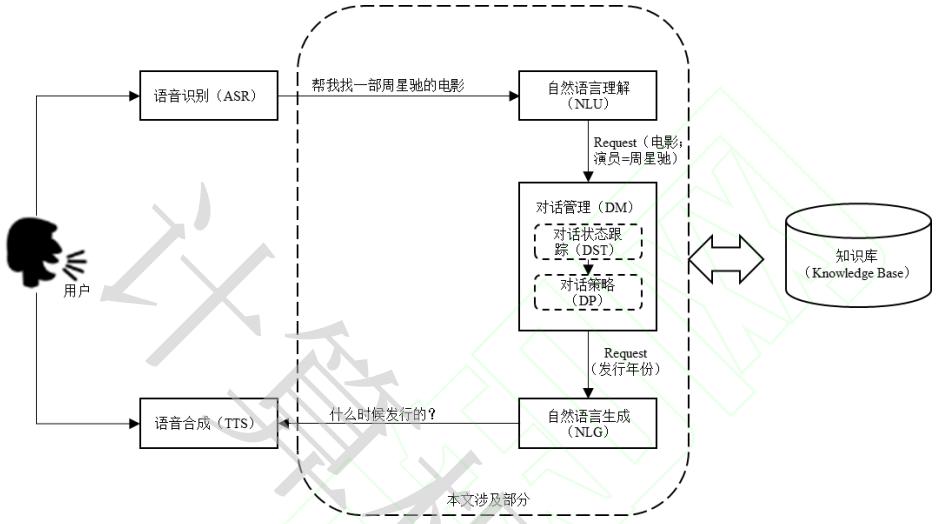

图 1 任务型对话系统的管道方法框架图

3.1.1 自然语言理解

自然语言理解的目的是将用户的输入映射到预先根据不同场景定义的语义槽中，通常包括三个任务：领域检测、意图识别和语义槽填充。自然语言理解应尽可能完整、清晰和准确地将用户输入转化为计算机能够理解的形式。NLU 模块的准确性对对话系统的质量有很大影响。表 2 是用户预订酒店时自然语言理解例子，表 3 列举了 NLU 的代表性的方法，并进行对比。

表 2 用户预定酒店时自然语言理解示例

用户输入	预订	一间	华工酒店	的	房间
语义槽	/	数量	酒店名	/	/
意图检测	预订酒店				
领域识别	酒店类				

领域识别和意图检测都属于一个文本分类任务，它根据当前用户的输入推断出用户的意图和涉及的领域，用户的意图和涉及的领域来自预定义的候选集。其技术经历了从统计学习到深度学习的转变，最早的传统统计学习模型有：支持向量机（Support Vector Machines，SVM）[6-7]、朴素贝叶斯（Naive Bayesian）[8]、k 近邻（K-Nearest Neighbor，KNN）[9]等。其中朴素贝叶斯的独立假设条件较为苛刻，在实际应用中往往是不成立的[10-11]。KNN 算法具有较强的一致性结果，其性能在很大程度上取决用于识别最近邻的距离度量，并且不擅长处理高维度数据[12-13]。Joachims 等人[14]提出使用 SVM 解决分类问题，在面对高维空间和过拟合时，具有较强的稳健性，但是其性能的优劣过于依赖对参数和核函数的选取。深度学习兴起后，多分类的神经网

表 3 NLU 的代表性方法对比

任务描述	方法名称	优点	缺点	适用场景
文本分类任务	K 近邻（KNN）	简单、有效；重新训练代价较低；	可解释性差；计算量较大	适于类域交叉或重叠较多的数据集的场景
	支持向量机（SVM）	可以解决高维、非线性问题；不容易过拟合	性能优劣过于依赖对参数和核函数的选取	适用小规模含有标注的数据集的场景

	卷积神经网络 （CNN）	训练速度快； 局部提取特征能力强	卷积核大小难以固定，若过小容易丢失数据，过大则参数空间很大	适用于大规模含有标注的短文本分类数据集的场景
	循环卷积神经网络 （RCNN）	结合了单词之间的上下文信息，可以更好的得到文本中的长依赖关系	存在数据稀疏问题； 难以确定窗口大小	适用于大规模含有标注的数据集的场景
序列标注任务	条件随机场 （CRF）	可以容纳任意上下文信息 特征设计灵活	训练代价大，复杂度高； 难以扩展	适用于搜索空间较小的场景
	长短期记忆网络 （LSTM）	能够捕捉长远的上下文信息； 具备神经网络的拟合非线性能力	无法对依赖关系建模； 训练数据不够导致过拟合严重； 没有彻底解决梯度消失问题	适用于搜索空间非常大的场景
	双向长短期记忆网络结合条件随机场 (Bi-LSTM+CRF)	结合了 LSTM 和 CRF 两者的优势，既可以考虑上下文，又可以考虑依赖关系信息； 有效利地利用过去和未来的信息	在实验过程中加入很多技巧，且实验证明这些技巧对结果提升非常有效，因此并没有完全摆脱人工特征的构造	适用于大规模含有标注的数据集的场景
领域拓展任务	零次学习模型 （Zero-Shot Learning）	领域拓展性强； 减少了标注成本； 训练时间快	领域漂移导致难以正确分类； 枢纽点影响正确率的计算； 语义间隔导致学习样本在特征空间到语义空间的映射困难	适用于含少量标注的数据集的场景
联合建模任务	基于注意力机制的循环神经网络 （Attention-Based RNN）	意图识别和槽填充联合建模，简化了 NLU 模块； 在基准 ATIS 数据集上，单领域训练，意图识别错误率低（1.79%），槽填充 F1 分数高（95.98）； 对于输入序列具有区分度	无法并行化处理，导致模型训练时间较长； 不能记忆太前或者太后的内容，因为存在梯度爆炸或消失	适用于大规模较多任务的数据集的场景
	具有长短期记忆网络门的双向循环神经网络 (Bi-directional RNN-LSTM）	领域检测、意图识别和槽填充联合建模，简化 NLU 模块； 在基准 ATIS 数据集上，多领域训练，意图识别错误率低（13.4%），槽填充 F1 分数高（89.4），意图识别准确率高（94.6%）； 领域拓展性强； 能够捕捉长远的上下文信息	无法并行化处理，导致模型训练时间较长； 网络结构比较复杂，门多，对效率有影响	适用于大规模较多领域或任务的数据集的场景

络方法已经取得不俗的成绩，比传统的统计学习模型的准确度更高且效率也更高[15-17]。文献[18-22]运用卷积神经网络模型（Convolutional Neural Networks，CNN）将查询向量表示为分类特征，Pooling 层解决了可变长度的句子的输入问题，但由于其网络视野受限，不能表达长距离的依赖关系。Zhou [23]和 Lee[24]等人提出利用循环神经网络（Recurrent Neural Networks，RNN）解决 CNN 的视野受限问题，结合 CNN 和 RNN 的各自的架构优势，不但可以抓取任意长度的序列，分析长句之间的关系，而且可以进行时间和空间扩展，同时具有记忆功能。

语义槽填充与领域识别、意图检测不同，其本质上属于序列标注问题，旨在识别句子中的语义槽和其对应的值。通常使用的线性统计方法包括条件随机场（Conditional Random Fields，CRF）[25-26]、隐马尔可夫模型（Hidden Markov Models，HMM）[27]、最大熵马尔可夫模型（Maximum Entropy Markov Models，MEMM）[28]等。线性统计方法主要依靠手工制定任务资源，开发成本高[29]。相比之下，RNN[30-31]模型及其变体，如长短期记忆网络（Long Short-Term Memory，LSTM）[32-33]、门控制循环单元（Gated Recurrent Unit，GRU）[34]使用分布式表示作为输入取代手工制定的特征，减小了开

发成本且绝对误差更小。RNN 的变体例如长短记忆网络通过三个门结构（输入门、遗忘门和输出门）可以学习长距离的依赖关系，得到了广泛应用[23]。在许多情况下，用户的输入通常为短文本或口语缩略语，缺乏任务所需的必要信息或表达含糊不清，除了可以直接利用历史消息和上下文语境信息[35]来提高 NLU 的准确性，还可以通过建立上下文信息的内存网络[36]来帮助序列标注任务。研究表明，考虑上下文语境信息、历史信息的方法可以获得更高的准确度[24]。另外，由于基于 RNN 的模型存在输出独立性问题，将深度学习与序列标注常用方法相结合成为了目前序列标注任务的主流模型，例如具有条件随机场层的双向长短记忆网络（Bi-LSTM+CRF）[37-40]，不但可以有效地利用过去和未来的信息，还可以使用句子级标记信息。

在 NLU 中，通常会对领域识别、意图检测和语义槽填充分别进行建模，但为了简化 NLU 模块，利用任务与任务之间的相关性，将 NLU 模块的三个任务进行联合建模。例如文献[41,42]提出了意图检测和语义槽填充联合模型，对两个任务同时建模进行训练。该联合模型方法不仅更简单，而且可以在实验中得到最好的 F1 值。而基于神经网络结构的序列标注方法，保持了深度学习技术的优势，无需人工定义特征，只需词向量和字向量就能达到标准水平，加入高质量的词典特征能够进一步提升识别的准确度。此外，还可以将 NLU 模块与 DM 模块或与 DST、DP 模块联合建模，详情见 3.2 端到端方法。

NLU 模块作为对话系统的重要基石，其准确性对对话系统的质量有很大影响。因此 NLU 模块作为对话系统的典型任务，受到国内外研究人员的广泛关注，发展速度迅猛。基于此，将 NLU 模块面临的主要挑战与近期发展趋势进行归纳总结：

（1）NLU 模块面临的主要挑战

① 模型的领域拓展性问题，即当用户目标发生变化或话题切换时，NLU 将现有的模型快速准确地移植到其他领域是非常困难的。虽然零次学习模型[17,43]能够应用于槽填充任务中，利用共享的潜在空间，以实现槽值的跨领域可重用性。但模型中领域漂移、语义间隔等问题需要解决。

② 在实际应用中，尽管使用深度学习模型表现优于传统方法，但深度学习模型的训练需要的大量的标注语料，成为了 NLU 领域的一个重要瓶颈。对于少量标注数据集问题，迁移学习能够利用其他模型来帮助新模型在低数据体系下的训练，而半监督的方法让智能体不依赖外界交互，通过利用未标注数据提升学习性能。因此，迁移学习[44]、半监督学习[45-46]和无监督学习是 NLU 未来研究的重点。

③ 目前，NLU 模块对于语言的理解还停留在语义表层，在复杂任务下结合知识进行逻辑推理能力还比较弱[47]。例如，用户说“帮我买一张春节回家的机票，要在春节前三天内。”面对该任务，简单的基于语义表层的槽填充不能得到正确的槽值，需要结合常识日期并通过推理才能完成该任务。为此，一些基于常识知识图谱的对话模型被陆续提出[48-50]来帮助机器理解比字面上含义更深层次的理解。但这些结合文本、对话记录、常识知识图谱的方法往往只用了单一三元组，忽略了一个子图的整体语义，导致得到的信息不够丰富。因此，使 NLU 模块在复杂情况下的推理能力有待于进一步研究与探索。

④ 用户对话自由度高，不能总是输入精确的指令，口语句式参差多变，没有明确、准确的规范句式，具有特征稀疏性、实时性和不规则性等现象，且不同的用户有不同的语言习惯，加上 ASR 模块出现的错误，都是影响 NLU 模块准确性的重要因素。为避免这些问题，一种常用的方法是利用 N 最优假设列表（N-best hypothese list）来输出最有可能的 N 个当前的用户对话状态[51]。通常情况下，这些最佳假设之间仅仅只有少量不同的词，而且很多都是短功能的词，例如语气词、冠词、停用词等，一些概率较低的词往往会被这 N 最佳假设列表忽略掉。因此，针对话语和口语的不确定性进行有效建模，使得对话系统有良好的鲁棒性是 NLU 模型的一个值得探索的课题[5]。

（2）NLU 模块的近期发展趋势

① 面向更复杂，更接近真实应用场景的数据集[52-54]。斯坦福大学在原有的 SQuAD1.0 数据集的基础上进行扩充，发布了 Question Answering Dataset（SQuAD）2.0 数据集，加入了“不可回答的问题”，进一步考验机器阅读理解能力[55-56]。对于评测多轮对话的阅读理解技术，斯坦福大学提出了 Conversational Question Answering（CoQA）数据集，目前，在 CoQA 上最高 F1 指标分数的研究成果是由微软发布的 SDNet 模型（79.3），比之前最先进的模型（65.1）高 1.6%，与人类表现（88.8%）相比，还有足够的改进空间。此数据集较 SQuSD2.0 数据集而言更加困难，主要表现在它的标准答案不

再是依赖于篇章的某个连续片段，且多轮对话的指代消解问题是另一项新挑战[57]。

② 结合上下文进行推断，对特定领域问题融合外部知识和常识解决。目前 NLU 的研究受益于大规模数据集，而不利用任何现有的知识，还处于语言表层阶段，不能在复杂情况下结合推理进行深入理解。将特定领域完整常识知识编码为键值对（key-value），并整合。在推断之前，将这些知识与上下文表示结合，进一步提升自然语言理解的综合准确度[58-59]。利用外部知识图谱，如 Freebase 和 Wikipedia 来改善自然语言理解的准确性[60-62]。大多数前期的工作都利用语言知识和知识库作为神经网络输入的附加特征，然后进行学习。例如，Liu 等人[63]和 Chen 等人[64]提出了利用语义树中编码的语言知识进行语言理解的方法，在各个领域实现了更好的语言理解能力，但这类方法容易导致错误传播的问题。Chen 等人[58]提出知识引导的结构注意网络（K-SAN）方法来避免该问题，利用先验知识作为指导，以结合非平面拓扑结构，并为不同的特定任务的子结构学习适当的注意力。其中，结构化信息可以从小的训练数据中捕获。因此，模型具有更好的泛化和鲁棒性。虽然如何引入外部知识已有许多方法，但如何应用常识尚需深入研究。

③ 采用预训练语言模型。利用预训练语言模型不但可以在缓解数据稀疏的问题，而且可以根据上下文动态生成文本表示，在向量表示中包含更多文本语义层次信息。例如，ELMo 模型（Embeddings from Language Models）[65]，采用大隐层维度（1024）训练双向语言模型（BiLM）。该模型可以很容易地添加到现有模型中，并显著改善六个具有挑战性的 NLP 问题。谷歌 AI 团队提出 BERT（Bidirectional Encoder Representation from Transformers）模型[66]，采用 Transformer[67]的编码模块作为语言模型，抛弃了 RNN 和 CNN 等结构，完全采用注意力机制（Attention）进行输入和输出之间关系的计算。在机器阅读理解顶级水平测试 SQuAD1.1 的 11 种不同 NLP 任务中通过预训练和精调得出最好的结果。最近，百度提出了知识增强的语义表示模型 ERNIE（Enhanced Representation From Knowledge Integration），其效果在各类 NLP 任务上全面超过 BERT，例如在语言推理、语义相似度、命名实体识别、情感分析和问答匹配等①。上述模型都消耗资源较多，因此在计算资源充足的情况下，结合预训练模型，可以给下游任务带来一定程度的改善。

④ 多任务学习和迁移学习。大多数 NLU 任务假设任务不相关，忽略了不同任务之间的联系和差异，将问题拆分为独立的单任务进行学习。多任务学习把多个相关的任务放在一起学习，通过共享因素或共享表示将多个任务联系在一起。例如使用神经网络实现权重共享，对多个任务同时进行训练[68]，利用不同任务之间共享的有用信息提高所有任务的学习速率和完成质量。对于目标域中训练数据不充足的 NLU 任务，迁移学习具有非常重要的意义。

3.1.2 对话管理

对话管理是对话系统的“大脑”，控制着整个对话系统的流程。DM 的输入是自然语言理解的三元组输出，并需要考虑历史对话信息和上下文的语境等信息进行全面地分析，决定系统要采取的相应的动作，其中包括追问、澄清和确认等。DM 的任务主要有：对话状态跟踪和生成对话策略。表 4 列举了 DM 的代表性方法，并进行对比。

①https://github.com/PaddlePaddle/LARK/tree/develop/ERNIE

表 4 DM 的代表性方法对比

任务描述	方法类型	方法名称	优点	缺点	适用场景
对话状态跟踪任务	基于人工规则的模型	有限状态机（FSM）	无需训练数据； 将领域知识编码到规则中比较简单	ASR 和 NLU 中的识别错误没有机会得到纠正，也无法追踪多种状态； 对于复杂状态的无法手工制定更新机制，缺少灵活性；	适用于无训练数据集的场景，即冷启动场景
		N-Best 列表	无需训练数据； 很容易将领域知识编码到规则中； 适当缓解 ASR 和 NLU 的错误识别；	相关参数需要人工编写制定，无法根据数据分布进行学习	适用于无训练数据集的场景，即冷启动场景
	生成式模型	贝叶斯网络	追踪到的状态的准确性高于基于人工规则的方法； 无需人工构建对话管理机制，具有较好的鲁棒性； 可以建模所有状态及状态转移的可能性；	仅仅可以建模简单的依赖关系； 忽略对话历史中的有用信息； 进行了不必要的独立假设；	适用于状态和动作空间较小的数据集的场景
		部分可观测马尔可夫模型（POMDP）	追踪到的状态的准确性高于基于人工规则的方法； 无需人工构建对话管理机制，具有较好的鲁棒性； 具有良好的数据驱动型	忽略对话历史中的有用信息； 进行了不必要的独立假设； 需要精确建模所有特征之间的依赖关系，而完整的建模和优化计算代价巨大	适用于状态和动作空间较小的数据集的场景
	判别式模型	循环神经网络（RNN）	模型从大量数据中学习用户行为，无需人工构建对话管理机制； 善于利用对话历史中的潜在信息特征； 可以建模任意长度的依赖关系	需要大量的标注训练数据； 由于可能存在梯度消失或梯度爆炸问题，很难去训练	适用于大规模含有标注的数据集的场景
		长短期记忆网络（LSTM）	模型从大量数据中学习用户行为，无需人工构建对话管理机制； 善于利用对话历史中的潜在信息特征； 可以建模任意长度的依赖关系； 可以缓解梯度消失或梯度爆炸问题	需要大量的标注训练数据；	适用于大规模含有标注的数据集的场景
	神经信念跟踪模型（NBT）		领域自适应能力强； 不需要大规模的标注训练数据和手工制定规则	语言限制：英语； 无法处理没有见过的值	适用于多领域对话场景
对话策略学习任务	监督学习	基于规则的模型	具体领域内效果较好	需要专家手工设计对话策略规则；可拓展性差； 不能从失败中学习	适用于无训练数据集的场景，即冷启动
	模仿学习	DAgger 算法	能够解决传统强化学习难以解决的多步决策问题	模仿成本较高，需要专家提供的策略覆盖最优结果； 采用深度网络的学习方式需要大量训练数据； 面对复杂困难的行为，很难达到好的效果	适用于多步决策、含有大规模数据集和行为简单的场景
	传统强化学习	马尔可夫决策过程（MDP）	逐渐摆脱专家手工设计决策规则；	面对过多的状态或动作空间，很难进行高效的探索；	适用于手工特征提取、状态低维且完全可观测的领域

深度强化学习	深度 Q 学习（DQN）	不需要大规模数据集； 收敛速度比传统强化学习快； 算法具有通用性； 其有效性得到验证（在预订电影票的任务中，epoch=300，任务成功率 0.7867，平均对话轮数 13.91，人工评价任务成功率 0.222）	奖励设计困难； 不能避免局部最优； 无法应用于连续动作控制； 奖励稀疏且行动空间很大时，对话策略往往会失败； 过高估计问题	适用于无大规模数据集的场景
	深度 Dyna-Q（DDQ）	只需要少量的真实用户交互数据； 使用世界模型模仿真实用户响应，接近用户行为，具有普通模拟器没有的人类语言的复杂性，进而对智能体进行优化； 其有效性得到验证（在预订电影票的任务中，epoch=300，任务成功率 0.6000，平均对话轮数 16.04，人工评价任务成功率 0.326）	DDQ 的有效性取决于计划阶段生成的模拟数据的质量； 没有好的方法来评估产生的模拟数据的世界模型的质量，从而限制 DDQ 在实际任务的适用性	适用于复杂对话和只有少量的真实用户交互数据的场景
	判别式深度 Dyna-Q（D3Q）	只需要少量的真实用户交互数据； 在 DDQ 的基础上，利用判别器区控制模拟数据的质量，这使得对话策略能进行有效且稳健的学习； 其效果超过 DDQ（在预订电影票的任务中，epoch=300，任务成功率 0.7400，平均对话轮数 13.81，人工评价任务成功率 0.440）	训练判别器需要达到纳什均衡； 可能存在训练不稳定、梯度消失、模式崩溃等问题	
	在线学习框架	无需大规模标注语料和构建用户模拟器来进行训练，就能应对复杂对话场景； 主动学习用于限制反馈请求，减轻对话策略学习中的用户反馈的噪声；	系统不断地与真实用户交互以学习对话策略，这样的在线学习过程非常耗时且昂贵； 奖励函数只关注对话任务是否成功，模型过于简单	适用于复杂对话和较小的含标注的数据集
	子目标发现网络（SDN）	使用子目标减轻了奖励的稀疏性； 不需要人工参与定义子目标； 对于复合对话任务具有适用性；	结构比较复杂，难以复现； 学习 SDN 和 HRL 是 2 个分开的过程，结合起来效果可能更好	适用于复合对话任务场景
	BBQ 网络（BBQ）	领域自适应能力强； 能够缓解 ASR 和 NLU 识别的错误； 解决了当前策略中 Q 值和环境中探索不足的部分动态的不确定性	探索成本过高	

（1）对话状态跟踪

对话状态是一种将 t 时刻的对话表示为可供系统选择下一时刻动作信息的数据结构，可以看作每个槽值的取值分布情况。DST 以当前的动作 u_n、前 n-1 轮的对话状态和相应的系统动作作为输入，输出其对当前对话状态S_t的估计。对话策略的选择依赖于 DST 估计的对话状态S_t，因此 DST 至关重要。同时，DST 也非常具有挑战性，因为 ASR 和 NLU 模块的识别往往会出错，可能导致对话系统无法准确理解用户语义。所以，ASR 和 NLU 模块通常输出 N-best 列表，DST 通过多轮对话不断修改和完善来修正 ASR 和 NLU 识别的错误。例如，表 5 是关于用户询问餐厅时 DST 的例子。

DST 主要分为三类方法：基于人工规则、基于生成式模型和基于判别模式模型。

基于人工规则的方法，如有限状态机（Finite State Machine，FSM）需要人工预先定义好所有的状态和状态转移的条件，使用分数或概率最高的

NLU 模块解析结果进行状态更新[69]。例如麻省理工的 MIT JUPITER 天气信息系统，利用人工预先编写的对话控制表中的状态变量进行状态更新[70]。1996 年，Pulman 等人[71]发现出跟踪多个对话状态的好处，Wang[72]和 Sun[73] 等人随后提出了可以计算整个 ASR 和 NLU 的 N-Best 列表分数的方法，从而修正 ASR 和 NLU 模块识别的错误。目前，大多数商业应用中的对话系统都使用基于人工规则的状态更新方法来选择最有可能的结果。该方法不需要训练集，且很容易将领域的先验知识编码到规则中，与其对应的是其相关参数需要人工制定且无法自学习，ASR 和 NLU 模块的识别错误没有机会得以纠正[74]。这种限制促进了生成式模型和判别式模型的发展。

生成式模型是从训练数据中学习相关联合概率密度分布，计算出所有对话状态的条件概率分布作为预测模型。统计学学习算法将对话过程映射为一个统计模型，并引入强化学习算法来计算对话状

表 5 用户询问餐厅时 DST 的示例

DST 的输入			DST 的输出	
系统行为/用户回复	NLU 输出		对话状态假设	置信度分布
系统：我有什么能帮助您吗？	Inform(price=便宜)	0.2	price=便宜	
用户：查询一下附近的意大利	Inform(food=意大利)	0.1	food=意大利	
餐厅	Null()	0.7	none	
系统：不好意思，你刚说什么？	Inform(food=泰菜)	0.2	price=便宜	
用户：额…一个意大利餐厅	Inform(food=意大利)	0.5	food=意大利	
	Null()	0.3	food=泰菜	
			none	
系统：意大利餐厅对吗？	Inform(food=意大利)	0.9	price=便宜	
用户：是的！	Null()	0.1	food=意大利	
			food=泰菜	
			none	

态的条件概率分布，例如贝叶斯网络、部分可观测马尔可夫模型（POMDP）[75-76]等。虽然生成式模型的效果优于基于人工规则的方法，且该方法可以自动进行数据训练，减少了人工成本[79-80]。但是生成式模型无法从 ASR、NLU 等模块 挖掘大量潜在信息特征，也无法精确建模特征之间的依赖关系。此外，生成式模型进行了不必要的独立假设，在实际应用中假设往往过于理想。

目前，基于判别式模型展现出更为有利的优势，它把 DST 当作分类任务，结合深度学习等方法进行自动特征提取，从而对对话状态进行精准建 [77-78]。与生成式模型相比，判别式模型善于从 ASR、NLU 等模块提取重要特征，直接学习后验分布从而对模型进行优化。最早的判别式对话跟踪利用手写规则定义对话状态，利用逻辑回归进行多分类，估计每类特征对应的权重[81]。除了手写规则定义对话状态，还可以结合深度学习例如深度神经网络[82] 将对话历史信息抽象成一个固定维的特征向量用于训练分类器，如最大熵模型（Maximum entropy models，MEM）[84]、网络排序（Web-style Ranking）[76]等模型将所有历史信息抽象成一个固定维的特征向量用于训练分类器。

除此之外，将马尔可夫模型（Markov Model，MM）[85]、CRF [25,86-89]、RNN [90]等模型将提取的特征用于序列建模。这些方法都能在一定程度上弥补生成式模型的缺点，但大量的标注工作加大了模型应用的难度。

最近，引入了信念跟踪的深度学习 [82]，通过学习权重和使用滑动窗口的方式，解决使用单个神经网络在任意数量的可能值上输出一系列概率分布的任务，该方法的能够容易地移植到新的领域。另一种可以解决多领域移植性的模型是多领域对话状态跟踪模型[83]，它利用领域外的数据来初始化目标领域的信念跟踪模型，即使域内数据量很少，都能改善信念跟踪的目标准确率。目前大多数方法难以拓展到更大、更复杂的对话域，2017 年，Mrkšić 等人[91]提出了神经信念跟踪模型（Neural Belief Tracker，NBT），通过建立代表性学习的最新进展来克服这些问题。它以最后一轮系统的输出、用户的话语和候选槽值对作为输入，三项输入相互作用进行上下文建模和语义解码，以确定用户是否明确表达了与输入槽值对匹配的意图。最后上下文建模

和语义解码向量经过 softmax 层产生最终预测。2018 年，Lei 等人[92]提出了一种基于单序列到序列（Single Sequence-to-Sequence）模型的框架——Sequicity 框架，将对话状态的不同表示称为信念跨度（Belief spans），这种信念跨度使得面向任务型对话系统能够在单序列到序列的模型中通过监督或强化学习进行优化。它具有良好的扩展性，显著降低了参数数量和训练时间，与传统的管道方法相比，极大的简化了系统设计和优化过程。

（2）对话策略

对话策略根据 DST 估计的对话状态S_t，通过预设的候选动作集，选择系统动作或策略a_n。DP 性能的优劣决定着人机对话系统的成败。DP 模型可以通过监督学习、强化学习和模仿学习得到[93]。

监督学习需要专家手工设计对话策略规则，通过上一步生成的动作进行监督学习。由于 DP 的性能受特定域的特性、语音识别的鲁棒性、任务的复杂程度等影响，因此手工设计对话策略规则比较困难，而且难以拓展到其他领域。这使得强化学习逐渐代替专家手工设计一系列复杂的决策规则[94]。强化学习是通过一个马尔可夫决策过程（Markov Decision Process，MDP），寻找最优策略的过程。MDP 可以描述为五元组$(S，A，P，R，\gamma)$：

S：表示所有可能状态（States）的集合，即状态集；

A：针对每个状态，做出动作（Actions）的集合，即动作集；

P：表示各个状态之间的转移概率，例如$P_{s,a}^{s'}$表示在状态s下采取动作a之后转移到状态s′的概率；

R：表示各个状态之间的转换获得的对应回报，即奖励函数（Reward Function）。每个状态对应一个值，或者一个状态-动作对（State-Action）对应一个奖励值，例如$R_{s,a}$表示状态s下采取动作a获得的回报；

γ：表示为折扣因子，用来计算累计奖励。取值范围是 0~1。一般随着时间的延长作用越来越小，表明越远的奖励对当前的贡献越少。

DP 需要基于目前状态S_t和可能的动作来选择最高累计奖励的动作。该过程仅需要定义奖励函数，例如：预订餐厅的对话中，用户成功预订则获得正奖励值，反之则获得负奖励值。

传统的强化学习需要在较多训练数据的情况下，需要计算整个行动轨迹获得的整体回报来寻找最高回报对应的最优策略，才能获得较好的结果。因此在序列多步决策问题中，强化学习需要频繁地试错，来获得稀疏的奖励，这种“随机”方式的不但搜索空间非常巨大，而且前期收敛速度非常慢。模仿学习（Imitation Learning）能够很好的解决多步决策问题[95-98]。模仿学习的原理是通过给智能体提供先验知识，从而学习、模仿人类行为。先验知识提供m个专家的决策样本$\{\pi_1,\pi_2,\dots,\pi_m\}$，每个样本定义为一个状态 s 和动作$a$的行动轨迹：

$$\pi=\left(s_1^i,a_1^i,s_2^i,a_2^i,\dots,s_{n^{i+1}}^i\right) \quad (1)$$

将所有[状态-动作]对抽取出来构造新的集合D：

$$D=\{(s_1,a_1),(s_2,a_2),(s_3,a_3),\dots\} \quad (2)$$

把集合D中的状态视为训练数据中的特征，动作视为训练数据中的标签，通过回归连续的动作和分类离散的动作，来得到最优的策略模型。模仿学习需要专家提供较多数据或提供的数据覆盖最优结果，需要花费大量的时间和精力，Kim 等人[99]提出了 APID 算法来解决该问题，该方法使用专家的示范数据定义线性约束，来引导近似策略迭代（API）所执行的优化。其关键思想是利用一个交互数据集$\mathcal{D}_{RL}=(S_i,A_i)_{i=1}^n$分别对应一个专家示例集合$\mathcal{D}_E=(S_i,A_i^*)_{i=1}^m$，即一个包含 n 个示例的[状态-动作]对样本分别对应于一个包含 m 个示例的[状态-示范动作]对，并为动作值函数最优策略增加一个变量，允许偶尔违反约束条件，以保证编码专家的次优性。因此，最后得到的是一个有约束的优化问题。AggreVATe 算法[100]也可以最小化专家成本，该算法对 Dagger 算法[98]进行了扩展，利用已有数据信息进行交互式学习，实现模仿和强化学习的技术统一。然而模仿学习不仅需要专家提供数据，而且对于学习的行为较为复杂的情况，难以提供相关行为数据。因此，模仿学习还需要进一步探索。

面对过多的状态或动作空间，强化学习很难进行高效的探索，使用传统强化学习的应用往往仅限于手工特征提取、状态低维且完全可观测的领域。为此，深度强化学习的方法被提出，通过使用深度学习作为实际应用的基础，大大加快了强化学习模型的收敛速度。Mnih 等人[101-102]将 CNN 与传统 RL 中的 Q 学习（Q-Learning）[103]算法相结合，提出了深度 Q 网络（Deep Q-Network, DQN）模型。其中 Q-学习算法分成如下两步：

① 计算当前状态行动下的价值目标函数：

$$\Delta q(s,a)=r(s')+\gamma max_{a'}q^{T-1}(s',a') \quad (3)$$

② 网络模型的更新：

$$q^T(s,a) = q^{T-1} + \frac{1}{N}[\Delta q(s,a) - q^{T-1}(s,a)] \quad (4)$$

Q-学习算法存在着一些隐患：由于数据存在不稳定性，导致迭代产生波动，难以得到平稳模型；仅适用于求解小规模、离散空间问题，对于规模大或连续空间的问题则不再有效。为此，Deepmind[101-102]提出的 DQN 结合了 Q 学习的价值估计方法和深度模型较强的拟合效果，同时结合了记忆重播（Memory Replay）机制和目标网络（Target Network）结构，在 Atari 游戏上获得了不错的效果。DQN 对 Q 学习进行完善：

① DQN 利用了 CNN 来非线性逼近值函数。

用公式表示值函数为：$q(s,a;\theta)$，当网络结构确定时，θ代表值函数。

② DQN 利用记忆重播训练强化学习过程。将采样的样本进行存储并且进行随机的采样，打破了序列样本之间的关联性。在对网络更新时对经验进行回顾，观察到的转换被存放一段时间，并且均匀地从记忆库采样来更新网络，从而实现从高维度的感知输入直接学习策略的端到端的强化学习。

③ 对于时间差分算法的 TD 偏差，DQN 设置了目标网络来单独处理。使用梯度下降法更新值函数：

$$\theta_{t+1} = \theta_t + \alpha[r + \gamma max_{a'} q(s',a';\theta^-) - q(s,a;\theta)]\nabla q(s,a;\theta) \quad (5)$$

其中，$r + \gamma max_{a'} q(s',a';\theta^-)$为 TD 目标，TD 目标的网络表示为$\theta^-$，计算值函数逼近的网络表示为θ。

在某些环境中，Q-学习和 DQN 基于目标值网络的参数，每次都选择下一状态中最大 Q 值作为所对应的动作，会导致过高估计的问题，产生正的偏差。为此，双 Q-学习方法（Double Q-Learning）[104]被提出。双 Q-学习算法将动作的选择和评估进行解耦，使用两个 Q 函数，一个 Q 函数用于动作的选择，另一个用于 Q 函数的估计。双 Q-学习的 TD 目标公式为：

$$Y_t^{DoubleQ} = \gamma Q(s_{t+1}, argmaxQ(s_{t+1},a;\theta_t),\theta_t^-) + r \quad (6)$$

其中，θ_t表示的是目标网络的参数，即网络在更新前的一个参数备份，简化了网络代替之前的网络部分。另外，还有很多对 DQN 提出改进的算法，例如基于优势学习的深度 Q 网络[105]、基于优先级采样的深度 Q 网络[106]和动态跳帧的深度 Q 网络[107]算法等。

对话策略学习在复杂的对话系统、复合任务型对话和多领域对话中仍然面临着巨大的挑战：

① 复杂的对话系统

对于复杂的对话系统，难以指定一个好的策略作为先验知识。因此，强化学习智能体（agent）①通过与未知环境交互在线学习策略成为一种常用方法[108-110]。智能体必须在新的状态下尝试新的操作，以便发现更好的策略。因此必须在选择已有最大化累计奖励的策略和发现新的更好的替代策略之间进行良好的权衡[111]，通常称为利用（Exploitation）和探索（Exploration）。通过积极与用户互动，以发现一个更好的策略，然而在实际应用中，使用用户的反馈作为奖励信号是不可靠的，并且收集成本高（该方法对应于图 2 的 a 图）。

常见的一种解决方法是使用用户模拟器来产生大量的模拟数据以供深度学习训练。然而用户模拟器通常与真实的人类用户不够近似，其设计的偏差会降低智能体的性能（该方法对应于图 2 的 b 图）。为此，受 Dyna-Q 框架[112]和深度 Dyna-Q（Deep Dyna-Q，DDQ）框架[113]（该方法对应于图 2 的 c 图）的启发，Su 等人[114]在 DDQ 的技术上改进，提出判别式深度 Dyna-Q（Discriminative Deep Dyna-Q，D3Q）（该方法对应于图 2 的 d 图）方法来提高 DDQ 的有效性和稳健性。它将 Dyna-Q 算法整合到任务型对话策略学习的整合规划中。为了避免 DDQ 对模拟数据的高度依赖，在 D3Q 中建立基于 RNN 的鉴别器以区分模拟体验和实际用户体验，控制训练数据的质量。D3Q 的训练过程包括四个阶段，如图 2 的 d 图所示：

a) 指导强化学习：智能体与真实用户互动，收集真实体验和改善对话策略。

对话策略通过最小化均方差损失函数$\nabla_{\theta Q}\mathcal{L}(\theta_Q)$优化，$y_i$是目标值函数：

$$\mathcal{L}(\theta_Q) = \mathbb{E}_{(s,a,r,s')\sim B^u}\left[\left(y_i - Q(s,a;\theta_Q)\right)^2\right] \quad (7)$$

$$y_i = r + \gamma max_{a'} Q'(s',a';\theta_{Q'}) \quad (8)$$

b) 世界模型学习：利用真实经验学习和提炼世界模型。用B_u存储的采样组(s,a,r,s')作为训练数据，在世界模型中输入当前的对话状态s和上一轮的动

① 整个强化学习系统由智能体（Agent）、状态（State）、奖赏（Reward）、动作（Action）和环境（Environment）五部分组成。其中智能体是整个强化学习系统核心。它能够感知环境的状态，并且根据环境提供的奖励信号，通过学习选择一个合适的动作，来最大化长期的奖励值。

作a，生成用户的回复动作o、奖励r和一个终止信号t：

$$h = tanh(W_h(s,a) + b_h) \quad (9)$$

$$r = W_r h + b_r \quad (10)$$

$$o = softmax(W_a h + b_a) \quad (11)$$

$$t = sigmoid(W_t h + b_t) \quad (12)$$

c) 鉴别器学习：学习和完善鉴别器以区分模拟经验与实际经验。

使用世界模型G生成的模拟数据和收集的真实数据x训练鉴别器D。使用数量为 m 的样本训练目标函数为：

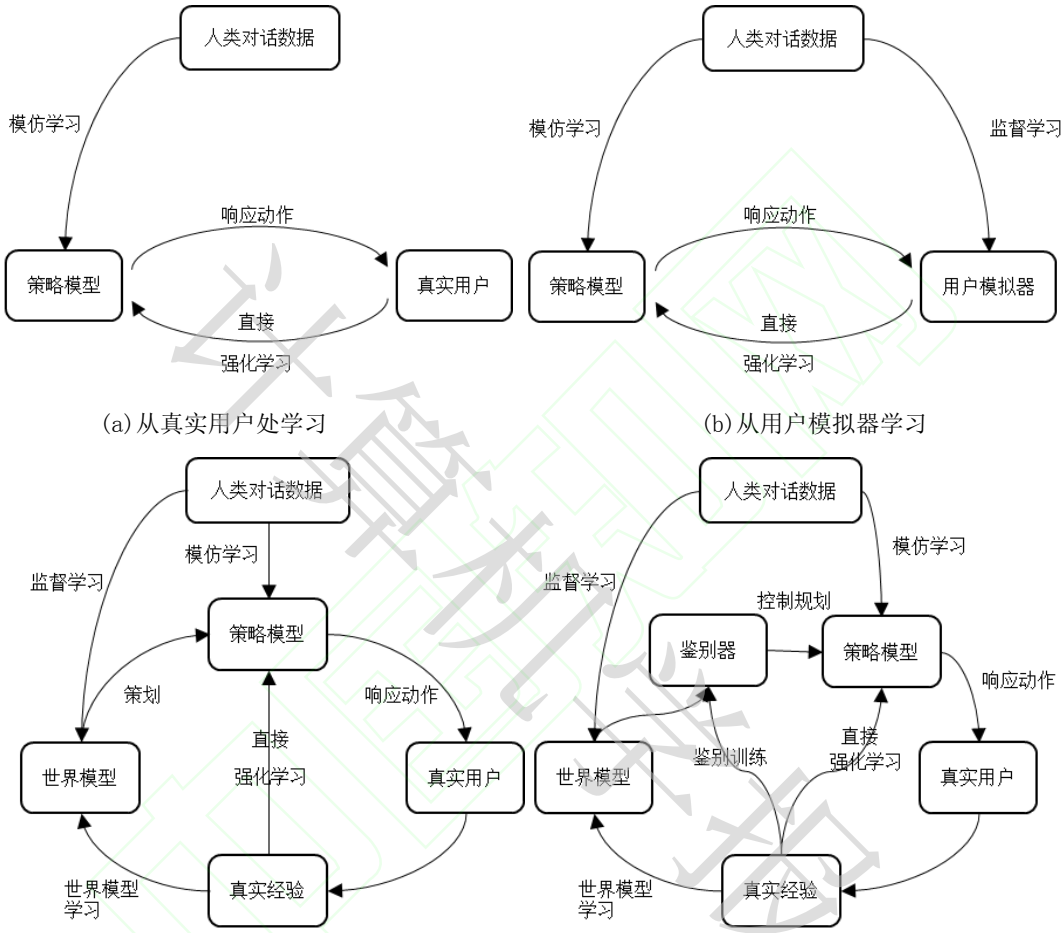

图 2 利用强化学习方法学习任务型对话策略的四种方法

$$\frac{1}{m}\sum_{i=1}^{m}\left[\log D\left(x^{(i)}\right) + \log\left(1 - D\left(G(.)^{(i)}\right)\right)\right] \quad (13)$$

受控规划：智能体使用由世界模型和鉴别器生成的高质量仿真语料来改善对话策略。

其研究结果表明，D3Q 通过控制用于规划的用户模拟器的质量，使得对话策略学习能够进行有效、稳健的学习，且 D3Q 也在领域拓展中具有有效性和鲁棒性。图 2 为利用强化学习方法学习任务型对话策略的四种方法。

另一种解决方法是通过在线学习，减少用户反馈噪声。Su 等人[115]提出一个在线学习框架，通过主动学习和高斯过程模型共同训练对话策略与奖励模型，使用 RNN 编码器-解码器以无监督的方式生成连续空间对话表示。该框架主要包含三个系统组件：对话策略、对话向量创建和基于用户反馈的奖励建模，如图 3 所示。嵌入函数的模型结构在图 3 的左侧，以对话轮数为单位级别从对话中提取情节特征f_t作为编码器的输入。编码器是一个双向长短期记忆网络（Bi-directional Long Short-Term Memory Network，Bi-LSTM）[32,116]，使用前向隐藏序列$\overrightarrow{h}_{1:T}$和后向隐藏序列$\overleftarrow{h}_{1:T}$迭代所有的特征f_t，t = 1, …, T：

$$\overrightarrow{\mathrm{h}}_t = LSTM(f_t, \overrightarrow{\mathrm{h}}_{1:t-1}) \quad (14)$$

$$\overleftarrow{\mathrm{h}}_t = LSTM(f_t, \overleftarrow{\mathrm{h}}_{1:t+1}) \quad (15)$$

解码器输出对话表示d，作为奖励模型的输入：

$$\mathrm{d} = \frac{1}{T}\sum_{t=1}^{T}\left|\overrightarrow{\mathrm{h}}_t; \overleftarrow{\mathrm{h}}_t\right| \quad (16)$$

奖励模型R(.)被描述为一个高斯过程，对每一个输入点，估计任务成功率和不确定性$\mathrm{p}(y|d,\mathcal{D})$，其中$\mathcal{D}$池中包含先前分类好的对话。基于这种不确定性，R(.)会决定是否咨询用户获取反馈，然后返回一个强化信号去更新对话策略。对话策略采用 GP-

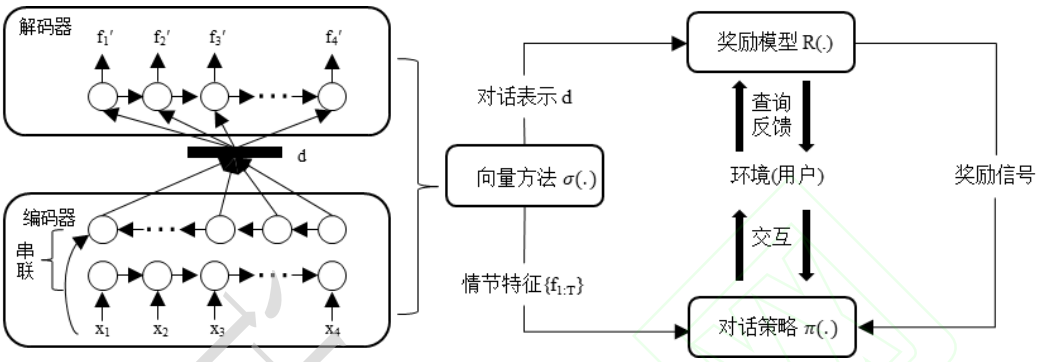

图 3 使用深度编码器–解码器网络学习对话策略的示意图

SARSA 算法[116]训练，该算法能够从最小数量的对话样本中引导稀疏值函数的估计。每次对话的质量由其累计奖励函数定义，每次奖励是小的负奖励（-1）还是最终奖励 0 或 20 取决于R(.)对任务成功的估计。该框架能够显著降低数据注释成本，并减少对话策略学习中的噪声用户反馈。

② 复合任务型对话

除了领域的拓展和复杂任务对话的问题，对话策略学习还在解决复合任务型对话的槽约束问题[117]等方面得到应用。例如制定旅行计划任务，在其中的预订航班和预定酒店两个子任务中，一个自然约束是航班的抵达时间应早于酒店登记入住的时间。在复合型任务对话中，对话策略学习非常具有挑战性：对话策略要处理很多槽，不但要处理子任务对应的槽，还要处理由所有子任务槽组成的复合任务的槽。由于槽约束，这些子任务无法独立解决，因此复合任务考虑的状态空间要比子任务状态空间的集合大得多。同时，由于复合任务对话通常需要更多轮数才能完成。因此，复合任务对话奖励具有稀疏性和有延迟性，使得策略优化更加困难。

为了解决上述问题，Peng [117]、Cuay áhuitl [118]和 Budzianowski [119]等人都使用分层强化学习（Hierarchical Reinforcement Learning，HRL）来分解复合任务，结果表明使用子任务减轻了奖励性，但对话策略学习更有效。然而，这些子任务需要人为定义，不仅需要大量的领域知识，还非常耗时。由于在许多情况下通常无法正确定义子任务，这限制了该方法在实践中的实用性。该局限性推动了有关复合任务型对话的自动学习定义子任务工作的发展，例如 Tang 等人[120]提出子目标发现网络（Subgoal Discovery Network，SDN），该方法包括两部分：

a) 为强化学习智能体自动发现有用的子任务。作者将子目标发现定义为状态轨迹分割问题。例如，对于状态轨迹$\mathrm{s} = (s_0, \ldots, s_T)$，将s分割为子轨迹$\mathrm{s}_{i:t} = (s_i, \ldots, s_t)$的可能性为：

$$\mathrm{L}_m(s_{0:t}) = \begin{cases} \sum_{i=0}^{t-1} L_{m-1}(s_{0:i})p(s_{i:t}|s_{0:i}), & m > 0, \\ I[t=0], & m = 0. \end{cases} \quad (17)$$

b) 将基于目标发现网络发现的子任务用于分层强化学习。作者采用分层强化学习方法学习了两个 Q 函数，分别θ_e和θ_i由参数化：在子目标中进行选择的高级$\mathrm{Q}^*(s,g;\theta_e)$函数和选择原始动作来完成子目标的低级$\mathrm{Q}^*(s,a,g;\theta_i)$函数。通过将子目标视为暂时的扩展动作，使用 Q 学习算法学习高级 Q 函数：

$$\theta_e \leftarrow \theta_e + \alpha \cdot (q - Q(s_t, g; \theta_e)) \cdot \nabla_{\theta_e} Q(s_t, g; \theta_e) \quad (18)$$

$$\mathrm{q} = \sum_{t=t_0}^{t_1-1} \gamma^{t-t_0} r_t^e + \gamma^{t_1-t_0} max_{g' \in \mathcal{G}} Q(s_{t_1}, g'; \theta_e) \quad (19)$$

低级 Q 函数以类似的方式学习，并遵循标准的 Q 学习更新，除了使用子目标g的内在奖励:

$$\theta_i \leftarrow \theta_i + \alpha \cdot (q_t - Q(s_t, a_t, g; \theta_e)) \cdot \nabla_{\theta_i} Q(s_t, a_t, g; \theta_i) \quad (20)$$

$$q = r_t^i + \gamma max_{a' \in A} Q(s_{t+1}, a', g; \theta_i) \quad (21)$$

除了利用对话结束时的外部奖励，分层强化学习还利用了子任务结束时所引发的内部奖励。在分层策略学习中，内在奖励与外部奖励相结合可以帮助强化学习智能体避免不必要的子任务切换，尽快完成任务。设 L 为对话最大轮数，K 为子任务的数量，外部奖励与内部奖励机制如下：

外部奖励：在对话结束时候，智能体对于成功的对话获得 $2L$ 外部奖励分数，失败则获得-L的外部奖励分数。对于每多加一轮对话，智能体获得-1的外部奖励分数，以此来使得在更短的轮数下完成任务。

内部奖励：当一个子任务结束时，智能体对于成功完成的子任务获得 $2L/K$ 内部奖励分数，失败则获得-1 的内部奖励分数。对于每多加一轮对话，智能体获得-1 的内部奖励分数，以此来使得在更短的轮数下完成任务。

将模型在模拟用户和真实用户上实验，结果表明强化学习智能体使用基于目标发现网络发现的子任务的任务成功率高于使用人工定义的子任务的任务成功率，且目标发现网络发现的子任务通常是人类可以理解的。分层强化学习与子目标发现网络的工作流程如图 4 所示。

除了分解复合任务，还可以在空间上分解复杂的策略。例如 Casanueva 等人[121]提出 Feudal 对话策略方法（Feudal Dialogue Policy，FDP），它将强化学习策略分为两步。第一步，策略决定采用独立槽还是依赖槽，然后进行动作子集的选取。第二步，抽取每个槽的子策略的状态以考虑与该槽相关的特征，并从先前选择的子集中进行具体策略动作选取。该方法不但可以减少空间探索步骤，达到优化强化学习的目的，而且不需要对奖励函数进行任何修改，且层次化结构完全由系统结构化数据库表示指定，不需要额外设计，大大减少了设计工作量。

图 4 分层强化学习与子目标发现网络的工作流程示意图

③ 多领域对话

任务型对话系统与用户的交谈可能涉及多个领域，因此任务型对话系统逐渐从限定域向开放域发展。多领域对话与复合任务型对话不同，多领域对话对应于不同域的子任务是单独的，没有跨任务的槽约束。但对话系统的领域拓展带来了许多挑战。其原因是多领域对话需要跟踪更多域的槽值和更大的对话状态空间，因此直接使用强化学习学习 DP 是低效的。早期的工作重点是使用来自不同域的数据来适应特定域[122]。在贝叶斯委员会机器（Bayesian Committee Machine，BCM）[123]的启发下，Gašić 等人[124]提出一个策略委员会模型进行有效的多域策略学习。Cuay áhuitl 等人[125]提出了另一种多域对话策略学习方法，称为 NDQN（Network of DQN，NDQN）。其主要思想是训练 DQN，其中每个 DQN 被训练用一个特定域子对话。使用一个元策略控制这些 DQN 的切换，也可以使用强化学习进行优化。

对话系统还可以通过自动添加意图或槽值以使系统更加通用[126]。因此，领域拓展的问题使得高效的探索更具有挑战性：强化学习的智能体通过明确用户意图和明确槽值的不确定性，以及避免探索到已经学习到的策略来提高探索的效率。Lipton 等人[127]提出了 BBQ 网络（Bayesian-By-Backprop Networks），BBQ 网络是 DQN 的一种变体，保留了 DQN 权重ω的近似后验分布q，其中权重的计算公式如下：

$$\omega = (\omega_1, \omega_2, \ldots, \omega_d) \quad (21)$$

为了计算方便，q是一个具有对角协方差的多元高斯分布，同时被θ参数化，参数θ计算公式如下：

$$\theta = \{(\mu_i, \rho_i)\}_{i=1}^{d} \quad (22)$$

每个权重ω_i从一个高斯后验分布$\mathcal{N}(\mu_i, \sigma_i^2)$中抽样，为了保证所有的$\sigma_i$保留严格的积极性，通过激活函数 softplus 函数：

$$\sigma_i = log(1 + exp(\rho_i)) \quad (23)$$

将σ_i参数化，并给$\mathcal{D}$维权重向量ω一个变分参数$\theta = \{(\mu_i, \rho_i)\}_{i=1}^{D}$。通过最小化变分自由能来学习这

些参数[128]，变分近似$q(\omega|\theta)$和后验$p(\omega|\mathcal{D})$之间的KL 散度（Kullback-Leibler divergence，KLD）：

$$\theta^* = argmin_\theta KL[q(\omega|\theta)\|p(\omega|\mathcal{D})] = argmin_\theta\{KL[q(\omega|\theta)\|p(\omega)] - \boldsymbol{E}_{q(\omega|\theta)}[\log p(\mathcal{D}|\omega)]\} \quad (24)$$

该方法通过汤普森抽样（Thompson-Sampling）[129-131]探索不同的策略路径，并从一个贝叶斯神经网络[132]中抽取蒙特卡罗样本。使用 DQN 来优化策略，方法包含变分信息最大化探索（Variational Information Maximizing Exploration，VIME）[132]的内在奖励。其中，贝叶斯方法的使用解决了 Q 值的不确定性，变分信息最大化探索解决了未知环境的不确定性。这些方法协同组合，在领域扩展任务中，其任务成功率高于目前最好的方法，且 BBQ 网络的学习速度比ε－贪心算法、玻尔兹曼算法、自主抽样法和基于内部奖励的方法等常用的探索策略快得多。

DM 模块在对话系统中充当“大脑”的角色，其设计的优劣决定了对话的效果和用户满意度，是影响对话系统的性能的重要因素。基于此，将 DM 模块面临的主要挑战与近期发展趋势进行归纳总结：

（1）DM 模块面临的主要挑战

① 多个场景的切换与恢复。绝大多数的真实对话涉及多个场景，实现一个能够在多领域的对话场景下决策的DM模块对于对话系统在真实对话场景中的应用具有重大意义。现有的数据驱动的 DM 模型，通过人工制定多个领域之间的规则来实现不同场景的切换。除此之外，还可以将 DM 模块分为领域相关的部分和领域无关的部分，例如 Gašić 等人[122,124] 提出使用分层结构来训练通用的 DP，然后可以在有效的情况下进行细化。类似地，Casanerva 等人[133]探索出将已知场景的数据迁移到新场景的方法，Wang 等人[134]提出了一个与领域无关的摘要空间，允许一个域上训练的策略被转移到其他域。但这些方法仅适用于具有清晰的结构，大部分构件能重复使用、支持二次开发的对话系统，对于结构模糊、领域性强的对话系统则不适用。

② 多轮对话的容错性。由于语言识别或语言理解的错误导致DM模块做出不好甚至错误的决策是影响 DM 模块性能的一项重要因素。POMDP[135]等统计模型目前已被实验证明了能比基于人工规则的方法和 MDP[136]更有效的应对噪音环境及语言和语言理解等问题，产生具有良好鲁棒性的 DM 策略[137-138]。但由于需要大量的训练数据，且仅适用于较小规模的场景等问题限制了该方法在真实场景的应用，因而该方法目前还滞留在理论研究的阶段。如何使 DM 模型具有良好的容错性从而解除在真实对话场景中限制，需要进一步关注和探讨。

③ 超出领域的话语的处理。在对话系统中，虽然用户可能对于对话系统所在的垂直领域比较熟悉，一些超出领域（out-of-domian，OOD）的话语， 例如问好、个人表态等仍然可能会被用户使用。尽管任务型对话系统只需要完成面向垂直领域的任务，但是若在面对 OOD 话语时能进行很好的处理，而不只是对用户进行提醒，这会大大提升用户满意度[139]。Lane 等人[140]虽然使用基于 SVM 的主题分类方法检测原话语是否不在领域内。如果属于 OOD 话语，系统仅提示用户不在对话系统所知范围，不能提高体验。Wang[141]和 Huang[142]等人分别提出了对超出领域话语的协处理方法和对话行为识别方法。以上两种方法可以缓解 OOD 话语带来的对话语料的词汇覆盖不足的影响，但依旧需要对对话过程中的 OOD 话语进行人工标注，增加了模型的人工负担。Ren[149]等人提出使用 StateNet 模型，该模型通过不同槽之间共享参数，使得模型参数不随着槽数目的增加而增加，因而对话状态跟踪能够适应领域本体的动态变化。但对于不能在预训练的词向量中找到新槽值的向量表示和槽值不可数的情况，该方法并不可行。

（2）DM 模块的近期发展趋势

① 高效的混合 DM 模型。针对相对复杂的实际的对话任务，对话系统往往需要兼顾任务型对话系统和非任务对话系统的任务。例如，家用机器人不但需要完成家庭服务等任务也需要兼备闲聊对话的功能。面对此需求，工业界类似于小冰[143]、阿里小蜜[144]等对话系统也逐渐往此方向发展。因此，设计一个高效的混合 DM 模型是人机对话发展的必然趋势。

② 多模态人机 DM 方法。目前的人机对话系统，例如 google Now、Alexas、思必驰、出门问问等，都存在一个普遍的缺点：人机对话总是仅仅关注语义信息[145-146]。而在实际的人机交互过程，不仅仅会产生语义信息，还会产生更多额外的具有价值的信息，例如用户动作、用户语气和用户画像等多模态信息，这些信息会对系统的回应产生重要的意义。因此，基于多模态的人机 DM 方法也将成为未来人机交互的研究重点。如何在 DM 层有效融合

和协调对话背景、用户动作、语音和语气等多模态信息，提升用户满意度和对话效果，是一个值得关注和探索的课题。

③ 富有情感的 DM 模型。现今，情感计算在情感识别、人性对话生成等方向取得了丰硕的研究成果 [147]，但基于此的智能机器人大部分是集中在用户的自然语言、动作或表情等简单的层面，没有基于语音层面的情感交互，导致了目前大多数的智能机器人的情感识别和表达能力有限。由于缺乏有效的情感交互策略，现有的智能机器人距离真实人类之间的和谐、自然的交互尚有一定的差距[148]。因此，如何设计一个富有情感的 DM 模型，生成具有情感信息的反馈，从而构建一个不但能完成用户目标而且能与用户进行和谐、自然的情感交互的智能对话系统，需要进一步探索。

3.1.3 自然语言生成

自然语言生成的主要任务是将 DM 模块输出的抽象表达转换为句法合法、语义准确的自然语言句子。一个好的应答语句应该具有上下文的连贯性、回复内容的精准性、可读性和多样性[150-151]。

NLG 的方法可以话分为：基于规则模板/句子规划的方法、基于语言模型的方法和基于深度学习的方法。基于深度学习的模型还多处于研究阶段，实际应用中还是多采用基于规则模板的方法。表 6 列举了 NLG 的代表性方法，并进行分析对比。

基于模板的方法[152]需要人工设定对话场景，并根据每个对话场景设计对话模板，这些模板的某些成分是固定的，而另一部分需要根据 DM 模块的输出填充模板。例如，使用一个简单的模板对电影票预订领域的相关问题生成回复：

表 6 NLG 的代表性方法对比

任务描述	方法类型	方法名称	优点	缺点	适用场景
自然语言生成任务	基于模板或句子规划的方法	基于模板的方法	无需训练数据； 简单，领域内回复精准	依赖于模板的质量； 无法建模复杂的语言结构； 需要人工编写模板，可移植性和可拓展性差	适用于无训练数据集的场景，即冷启动
		基于句子规划的方法	无需训练数据； 可以建模复杂的语言结构	需要大量的领域知识； 难以产生比基于人工模板方法更高质量的结果	适用于无训练数据集的场景，即冷启动
	基于统计语言模型的方法	基于类的语言模型	规则简单，易于实现和理解	计算效率低下； 依赖规则，扩展性差	适用于大规模的数据集的场景
		基于短语的方法	高效、准确率高	需要很多语义对齐处理，拓展性差	适用于大规模的数据集的场景
	基于神经网络的方法	序列到序列模型（Seq2Seq）	数据驱动，省去语言理解等过程； 表现形式灵活	需要大量的语料支持； 不能考虑上下文信息，导致回复上下文无关	适于大规模数据集和聊天机器人
		序列到序列模型引入上下文	数据驱动，省去语言理解等过程； 表现形式灵活； 生成的回复会考虑上下文信息； 相对于Seq2Seq模型在自动指标和人类成对偏好测试中效果显着改善	需要大量的语料支持； 无法平行化处理，导致模型训练时间较长	适于大规模数据集和有上下文的多轮对话
		动态神经网络（DMN）	生成的回复不但会考虑上下文，而且能结合背景知识生成响应； 具有检索和推理功能	需要大量的语料支持； 模型性能依赖于注意力机制的效果； 模型输入模块采用单向的 GRU，只能记住前向的上下文，无法获得后向上下文信息； 无法记忆过于远的信息；	适于大规模数据集和有上下文、背景知识的多轮对话或问答系统
		常识知识感知	首次在基于神经网络的对话生成中，	依赖于常识知识图谱的质量；	适于已经构建常识

会话模型（CCM）	尝试使用大规模常识知识，使得模型能够更好地理解对话，从而给出更合适、丰富的回复； 利用静态和动态图注意力机制代替知识三元组，更好地解读对话中实体的语义； 不受限于小规模的、领域的特定知识库	需要构建大规模的常识知识图谱才能完成模型	知识图谱的对话
基于对抗生成网络（GAN）的可控文本生成模型	不需要大量的训练数据； 解决了基于VAE和GAN生成文本的随机性和不可控性	句子长度太短 仅应用到文本生成，应用在其他 NLP 问题上，比如对话系统等，则需要进一步实验，以及对模型架构的更新	适于小规模数据集的文本生成任务
基于迁移学习的对话生成	可以解决目标领域数据不足的问题； 可以解决跨语言、个性化等问题；	不能处理未见过的槽值	适于多领域的对话系统

[主演人 1]、[主演人 2]、[…]主演的[电影名称]电影将于[放映日期]的[放映时间]点在[影院名称]进行放映

该模板中，[**]部分需要根据 DM 模块的输出进行填充。这种方法简单、回复精准，但是其输出质量完全取决于模板集，即使在相对简单的领域，也需要大量的人工标注和模板编写。需要在创建和维护模板的时间和精力以及输出的话语的多样性和质量之间做不可避免的权衡。因此使用基于模板的方法难以维护，且可移植性差，需要逐个场景去扩展。

基于句子规划的方法[153-154]的效果与基于模板的方法接近。基于句子规划的方法将 NLG 拆分为三个模块：内容规划、句子规划、表层生成。其过程如图 5 所示，将输入的语义符号映射为类似句法树的中间形式的表示，如句子规划树（Sentence Planning Tree，SPT）。然后通过表层实现把这些中间形式的结构转换为最终的回复。基于句子规划的方法可以建模复杂的语言结构，同样需要大量的领域知识，并且难以产生比基于人工模板方法更高质量的结果。

随着大数据技术和语言模型日益成熟，海量的数据和其他先进领域中应用语言模型逐步应用于对话生成的研究。基于类的语言模型[155]将基于句子规划的方法进行改进：对于内容规划模块，构建话语类、词类的集合，计算每个类的概率，决定哪些类应该包含在话语中；对于表面实现模块，使用 n-gram 语言模型随机生成每一个对话。从该方法生成的文本在正确性、流畅度有明显提高，且规则简单，容易理解，该方法是的瓶颈在于这些类的集合的创建过于复杂，且需要计算集合中每一个类的概率，因此计算效率低。上述的方法都难以摆脱手工制定模板的缺陷，限制了它们应用于新领域或新产品的可拓展性。基于短语的方法[156]也使用了语言模型，但不需要手工制定规则，比基于类的语言模型方法更高效、准确率也更高。由于实现短语依赖于控制该短语的语法结构，并且需要很多语义对齐处理，因此该方法也难以拓展。

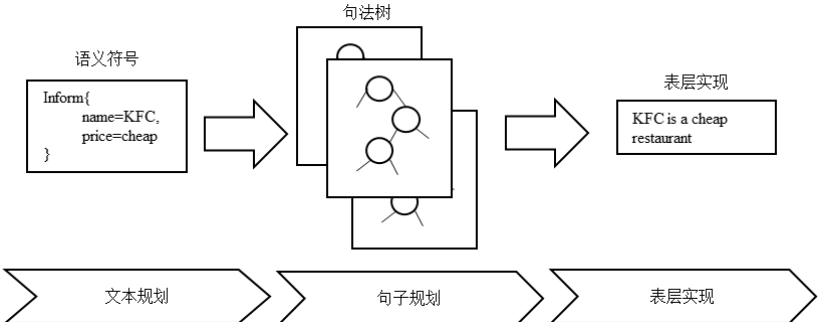

图 5 基于句子规划的自然语言生成过程示意图

NLG 模块的研究借助着深度学习的突破得到了巨大的助力推动。深度神经网络可以从海量的数据源中归纳、抽取特征和知识来学习，从而避免人工提取特征带来的复杂性和繁重问题。目前，基于

深度学习的 NLG 模型普遍以编码器-解码器（Encoder-Decoder）作为基础框架[157-158]，其框架图如图 6 所示。大多研究工作是对编码器-解码器的各个部分进行不断改进，如目标函数[159]、编码器[160-161]和解码器[162]。早期 Vinyals 等人[157]提出使用序列到序列（Sequence-to-Sequence，Seq2Seq）模型来生成简单的对话。将该模型应用于 IT 解答数据集和含有大量噪音的电影字幕数据集上，能够克服一些一般大型数据集的噪声，并从中提取知识和特征，可以执行简单形式的常识推理。基于 Seq2Seq 的生成模型虽然能够解决训练语料中未预设的问题，产生更加灵活多变的响应，但是其训练需要大规模的语料，且仅仅依靠上一句进行回复，没有考虑上下文语境。因此文献[160-161]将上下文信息引入编码器，在解码生成语句时给定信息重新输入模型参与计算来帮助解码器生成更好的回答内容。Dušek 等人 [163]发现，在对话中，说话者受对方之前话语的影响，并倾向于对方的说话方式、重用词汇以及句法结构，这种潜意识可以促进对话的顺利进行。因此，提出使用上下文感知器适应用户的说话方式和提供更多上下文准确且无重复的响应。虽然引入了上下文信息有助于提高对话的顺利进行，但是在引入上下文信息的同时，也会引入对对话没有意义的内容，从而影响生成回复的质量。因此 Kumar 等人[164]提出利用动态神经网络（Dynamic memory network，DMN）处理输入序列和背景知识，形成情景记忆模块（Episodic Memory Module），并生成相关答案。该方法不仅考虑了上下文信息，还考虑了背景知识，能够识别对对话有意义的内容，并将其激活应用到解码器中生成更好的回复。然而，上述模型在大多数情况下无法对用户进行适当的信息性响应。Zhou 等人[165]采用基于 LSTM 的编码器-解码器结构来结合问题信息，语义槽值和对话动作类型来生成更具有信息性的答案。在此基础上，他们提出常识知识感知会话模型（Commonsense knowledge aware conversational model，CCM）[48]，通过使用大规模常识知识来帮助理解背景信息，然后利用这些知识促进自然语言理解和生成，以解决由于不具备常识知识和对话背景而造成的回复不一致性或无关性等问题。另外，还有一些研究包括：基于对抗生成网络（Generative Adversarial Network，GAN）的文本可控的对话生成解决自然文本的离散性的问题，学习不可解释下的潜在表征，并生成具有指定属性的句子[166]、基于迁移学习的对话生成解决目标领域数据不足的问题，同时可以满足跨语言、回复个性化等[167-169]。

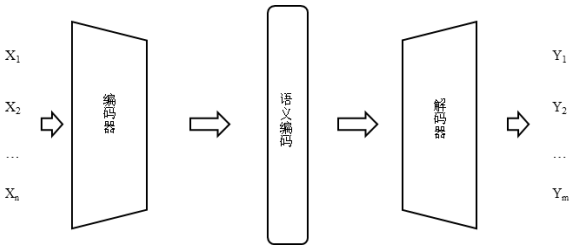

图 6 编码器-解码器框架图

NLG 模块是对话系统中关键的组成部分，创建一个结构良好且类似人的自然语言字符串对对话的可用性和感知质量都有着重要影响。虽然大数据技术、语言模型和深度神经网络的引入对对话生成技术有了很大的推动力，但实际对话场景的应用仍然存在僵硬和黑盒等不足。基于此，将 NLG 模块面临的主要挑战与近期发展趋势进行归纳总结：

（1）NLG 模块面临的挑战

① 目前的 NLG 模型表型还不够好，单纯使用模板的模型[153,155,170]过于僵硬，但使用神经网络的模型[169,171-172]又有不可控的风险。由于流畅的输出需要遵循许多语法规则，甚至覆盖大量可能的含义，因此目前工业界的对话系统的常规的方法仍然依赖于全面使用手工制作调整规则模板[173]。如何在不耗费人力物力的情况下，生成自然、生动和合适的回复需要进一步研究和探索。

② NLG 技术的评估需要完善。目前 NLG 的评估主要有：人工测评回复的适当性、流利性、可读性、多样性和利用 BLEU、METEOR、ROUGE 分数自动评估[174]。其中人工测评效率低下且过于主观，自动评估 NLG 技术的质量仍然是一个悬而未决的问题，再加上任务的复杂性对 NLG 评估也提出了重大挑战，特别是自动评估[175]。因此目前的 NLG 的评估并不成熟，制约着 NLG 的发展。实现 NLG 技术的标准化评估将实现多年来计划的共享任务评估活动（Shared-task evaluation campaigns，STEC）[176]。

③ 创建个性化回复困难。目前的人机对话系统虽然能够通过修改数据库保证其身份的一致性[177-178]，但是响应大都一成不变，不符合真实人人之间交互的灵活多变性。同时，没有提供给研究人员训练每个对话都受说话人信息的影响的公开数据集[179]。为了使得对话系统更加逼近真实人类交互行为，应根据用户的年龄、个性、心情、兴趣和个性等标签生成个性化响应。例如，餐厅预订系统

应该基于用户的固定属性比如饮食偏好、价格范围倾向等提出建议[180]。Halliday 等人[181]认为机器人使用的语言应该受到用户某些特征（年龄、性别等）影响。实现一个能够根据用户特征或属性创建不同个性化回复的 NLG 模块是目前的一个巨大的挑战。

（2）NLG 模块的发展趋势

① 基于检索模型和生成模型相结合。通常情况下，基于检索的模型无法处理过长的语句，而生成模型可能产生不一致或无意义的回复[159,161]。对于一个任务，对当前用户的输入进行语义分析，通过关键词匹配、排序等方法与数据库之间进行模糊匹配，得到一个候选回复列表，然后使用基于注意力的 Seq2Seq 模型重新评估候选应答内容。如果最高候选应答的分数高于某个阈值，将作为应答；其他问题将由基于生成的模型提供答案。该方法使用基于注意力的 Seq2Seq 模型来优化基于检索和生成模型的联合结果，比其他公开的对话系统生成的对话更加丰富和恰当。因此，将基于检索和生成模型相结合的方法用于对话生成模块，有助于提高对话系统的效果和用户满意度[182]。

② 有监督的端到端生成。基于规则或基于模板的方法[183-184]具有鲁棒性和充分性，但频繁重复相同的内容、笨拙的输出形式使得基于规则的生成模型非常繁琐。此外，该方法不易扩展到大型开放域系统[109,185]。基于语料库的方法[120,186]能够从数据中学习直接使系统能够更自然地模拟人类响应，消除对预定义规则的依赖，并使系统更易于扩建和扩展到其他领域，但现有的方法在训练数据效率、准确性和准确性方法存在缺陷。构建有监督的端到端生成模型，利用深度神经网络从海量的数据源中归纳、抽取特征和知识，从而避免人工提取特征带来的复杂性和繁重问题 [187-188]。

③ 融合任务相关的背景知识。常识性知识对于建立有效的交互行为非常重要，因为社交共享的常识知识是人们在谈话想要了解和使用的背景知识[189-191]。用于对话生成的神经模型 [123,192]倾向于产生通用响应，在大多数情况下无法对其进行适当的和信息性的响应，因为在不具备常识知识和对话背景时，只是从历史上下文中进行自然语言理解和生成是困难的。如果模型可以访问并利用大规模的知识常识和背景知识就可以理解对话，从而更恰当地响应。Hanet[193]、Ghazvininejadet[194]和 Zhu[195]等人已经尝试在对话生成中引入外部知识。但外部知识大都是以非结构化文本和面向垂直域的三元组形式进行表示和存储的。因此它们无法表示图形链接实体和关系的语义。如何引入能够表示图形链接实体和关系语义的常识知识和背景知识来帮助模型理解对话，从而形成更具信息性和适当性的响应，需要进一步关注与讨论。

④ 多样化的安全回复。当使用 Seq2Seq 模型用对话生成时会出现很多问题，最重要的问题之一是这些模型倾向于产生不具有意义的通用回复[160,165,196-197]。到目前为止，该通用响应问题已经吸引了越来越多的研究人员的注意。例如，Li 等人[159]使用最大互信息（Maximum Mutual Information, MMI）来重建传统的 Seq2Seq，这种方法使得生成的回复具有多样性，但容易产生不和语法的输出。他们进一步提出了一种快速多样化的解决方法[198]，它修改了波束搜索（Beam Training）以将有意义的响应排列到更高的位置。这些解决通用响应问题的工作[199-202]大都是为了在 Seq2Seq 的编码或解码模块添加额外的优化术语，因此也使得训练或预测更加困难。如何在不使训练或预测变得更困难的情况下，解决对话生成通用响应的问题，使得对话具有多样性是对话生成模块值得探索的课题。

3.1.4 管道方法总结

管道方法一般分别建立 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，然后按照顺序将这些模块连接起来。这种方法简单清楚，各个模块任务明确，并且可以分开研究，各自解决各自的问题[47]。

但是管道方法的问题也很明显：

（1）领域相关性强。针对每个领域都需要人工设计语义槽、动作空间和决策，导致系统的设计和领域非常相关，难以扩展到新的领域[203]。

（2）模块之间独立。各个模块之间相互独立，需要为每个模块提供大量的领域相关的标注数据。

（3）模块处理相互依赖。上游模块的错误会级联到下游模块，下游模块的反馈难以传到上游模块，使其很难识别错误来源。例如 DM 的决策出现错误，其原因可能是语言理解发生了错误，也可能是语音识别的错误。并且，由于一个模块的输入依赖于另一个模块的输出，当将一个模块调整到新环境或更新数据，其他所有模块为保证全局最优要进行相对调整。语义槽和特征也可能发生相应改变，而这个过程需要耗费大量的人力[7]。

3.2 端到端方法

深度学习的飞速发展促进了端到端方法在任

务型对话系统的应用，使得端到端的任务型对话系统成为可能。端到端方法将管道方法中的三个模块或部分用统一的端到端方法代替，根据用户的输入，直接生成相应的回复或响应模块的输出，以整体的端到端方法为例，其框架如图 7 所示。端到端任务型对话系统可以由以下框架构建：基于监督学习的框架、基于强化学习的框架和混合框架。表 7 列举了关于端到端方法的代表性方法，并进行对比：

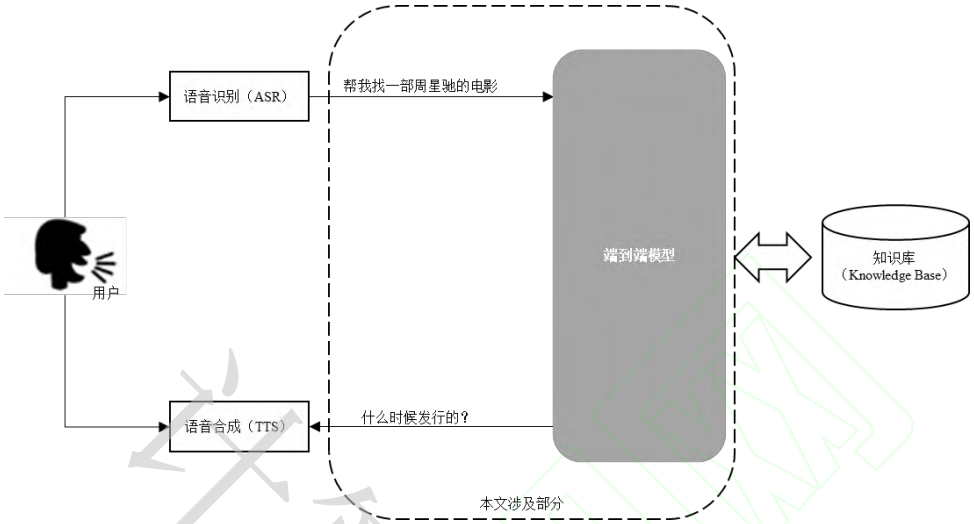

图 7 任务型对话系统的整体端到端方法框架图

表 7 端到端方法的代表性方法对比

任务描述	方法类型	方法名称	优点	缺点	适用场景
端到端方法	基于监督学习框架	基于内存的端到端模型	直接的端到端的训练； 轻量且易于使用； 具有可重复性和可解释性； 解决了未登录词问题	需要大量训练数据； 易丢失词序关系，进而丢失语义信息； 难以集合知识库；迁移性差	适于大规模数据集的场景
		基于神经网络的端到端可训练模型	只需要较少的人为干预； 可以较好的结合知识库； 解决了序列到序列模型难以应用在任务型的问题	需要大量训练数据； 由于缺乏有效的探索而无法得到一个好的策略的问题； 需要在 DST 模块预定义领域先验知识中的槽值，并且将该模块单独训练； 需人工定义对话领域相关的知识库进行检索的动作空间	适于大规模数据集且具有领域知识库的场景
		基于注意力的具有拷贝机制的序列到序列模型	无需明确建模对话状态信息； 解决了未登录词问题； 提升检索精确实体的能力	需要大量的训练数据； 生成句子具有重复问题； 没有明确地从底层知识库中获取信息，仅依赖于对话历史来进行系统响应生成； 难以完成外部数据检索且难以生成信息精准的回复； 需人工定义对话领域相关的知识库进行检索的动作空间	适于大规模数据集的场景
		一个端到端的可训练的键值检索网络	可以直接从底层知识库中提取有用信息； 无需对信念或内部跟踪器进行明	需要大量的训练数据； 知识库的规则过于严格,包含的信息不如原始文件充分；	适于大规模数据集且含有知识库的场景

		确的建模	不能处理联合知识属性； 没有推理能力； 需人工定义对话领域相关的知识库进行检索的动作空间	
基于强化学习框架	深度循环Q网络（DRQN）	能够联合优化 NLU、DP 和 DST，效果超过标准的监督学习； 有效地包含各种类型的标记数据； 模型具有记忆能力； 适用于任何数据库	可拓展性不足； 不能很好的处理非精确条件下的用户输入	适于大规模的数据集的场景
	端到端的 KB-InfoBot	允许端到端训练； 能够搜索知识库，无需编写复杂的查询； 用户输入里存在的不确定性得到了解决	该方法在真实用户进行测试时可能存在过拟合的问题； 需要动态对话语料与强化学习进行交互； 数据库很大时，需要很大的计算量，不易扩展	适于大规模的数据集且含有知识库的场景
监督学习结合深度学习或强化学习框架	混合编码网络（HCN）	可以同时用监督学习和强化学习的训练方式训练； 仅需要少量对话样本就可以自动学习对话状态，避免了人工编码对话状态	很多模块需要较多的人为干预，例如 NLU、DST、数据库和制定动作掩码等	适于小规模数据集的场景
	基于神经网络的端到端框架	具有很强的鲁棒性，能够缓解识别错误和不确定性等噪声； 允许用户在对话过程中发起对话，对话更灵活； 保证再现性	训练的过程依赖于特定数据集	适于小规模数据集的场景

3.2.1 基于监督学习的框架

基于监督学习的框架需要收集许多对话样本，然后将其用于训练对话系统的多个组成部分，以最大限度地提高预测准确性[203-206]。Bordes 等人[203]利用基于内存网络在问答系统领域展现出来的优势，将基于内存的端到端模型应用于任务型对话系统。作者利用内存网络对历史对话建模，通过一个注意力机制在内存网络中查找合适的信息，并提出了在相似词嵌入后拼接关于词的说明向量，解决了未登录词问题。该方法是一种直接的端到端的训练，但容易失去词序关系，进而丢失语义信息，且难以结合知识库。同时该模型是基于模板检索的对话生成，模板需要人工制定，因此迁移性差。文献[204]提出一个基于神经网络的端到端可训练模型，该方法将对话系统视为对话历史记录和当前数据里搜

索结果进行扩展的序列映射问题，不但既可以较好的结合知识库，只需要较少的人为干预，又解决了面向特定域的对话系统难以应用 Seq2Seq 的问题。Wen 等人[204]认为面向特定域的对话系统一直难以应用 Seq2Seq 模型解决方案的关键点在于无法将大量的领域信息建模到模型中，这些领域知识包括历史信息、用户身份信息、业务信息等。于是提出将具体的领域信息和历史信息加入到模型中，并通过将对话中的槽值转换为槽值的表示，降低了对训练数据的需求，也避免了 Seq2Seq 模型难以应用在任务型对话系统的问题。但该模型需要在 DST 模块预定义领域先验知识中的槽值，并且将该模块单独训练。如前文所述，文献[203]和[204]都属于将整个对话系统的过程视为从历史的对话和当前用户输入到系统生成响应的映射的方法，这类方法不但需要大规模的训练预料，而且可能会存在由于缺乏有效的探索而无法得到一个好的策略的问题。Eric 等人[207]提出基于注意力的具有拷贝机制的序列到序列模型解决了 Seq2Seq 模型难以应用在任务型对话系统的问题。该方法与文献[204]不同，通过结合 Seq2Seq 模型与基于注意力的复制机制两者的优点，允许模型提取和使用历史对话中相关的实体或内容，而无需明确建模对话状态信息或知识库槽值跟踪器，保留了完整的端到端可训练性。在他们的

工作中，生成的词有两个来源，一个是数据库，另一个是从原文中拷贝，生成的回复既简单又有效。之后他们把这个工作进一步延伸，提出一个端到端的可训练的键值检索网络（Key-value Retrieval Network）[205]，能够更好的对数据库进行查询。该网络可以在数据库添加基于注意力的键值检索机制，可以学习从数据库中提取分数最高的三元组回答作为最能够回答当前问题的知识。先前的方法[204-205,207]在进行对话时，要对对话状态进行表示和跟踪，这需要人工定义对话领域相关的知识库进行检索的动作空间，耗时又耗力。虽然已有使用 Seq2Seq 的模型来研究构建面向任务的对话系统[205,207]，但他们的方法还在初步阶段，在任务完成或响应生成方面表现不佳。而 Lei 等人[92]引入可以跟踪每轮对话中信念状态的文本跨度——信念跟踪器（Belief span，bspan），实现了单个 Seq2Seq 模型进行整体优化的方法，而无需人工定义对话领域相关的知识库进行检索的动作空间，命名为 Sequicity。其实现过程称为两阶段复制网（Two Stage CopyNet，TSCP），将任务型的对话问题分解为 bspan 和机器响应的生成，然后将问题转化为序列优化问题。Sequicity 采用单一序列到序列模型，实现了真实的端到端的能力。与先前的方法相比，TSCP 参数更少且训练速度更快，在大规模数据训练的基础上其任务完成度、语言质量和效率都显著优于最先进的方法[204,208-209]，包括涉及未登录词（Out of Vocabulary，OOV）的情景。除此之外，Wen 等人[210]提出使用向量表示对话状态进行隐式对话状态表示，向量每一维代表需要自己学习的槽。这种方法可以直接查询数据库，使得查询更加准确，需人工定义对话领域相关的知识库进行检索的动作空间，同时也更利于对话回复的生成。

3.2.2 基于强化学习的框架

虽然基于监督学习的方法可以产生不错的结果，但是该方法需要昂贵的训练数据，而且不允许对话系统探索不同的策略，这些策略可能比基于监督学习的方法生成回应效果更好。受此启发，使用基于强化学习来构建端到端对话系统的框架[211-212]被提出。Zhao 等人[211]首先提出一种端到端的强深度化学习框架，基于长短记忆网络的深度 Q 网络的变体，即深度循环 Q 网络（Deep Recurrent Q-Networks，DRQN），它将用户话语作为输入，输出语义系统动作，并学习压缩用户话语序列以推断对话的内部状态。同时，该模型把数据库的查询模板作为系统可选择的动作，对话策略可以输出对数据库进行相关搜索的动作，其搜索的结果是模型下一轮的输入，以便模型输出搜索结果。与传统方法相比，该方法能够联合优化 NLU、DST 和 DP 模块，超过标准的监督学习，同时该方法可以适用于任何的数据库。由于收敛需要大量的样本和需要与外部知识库交互，因此该方法可拓展性不足，且不能很好的处理非精确条件下的用户输入。由于任务型对话系统通常需要结合外部的数据库查询的信息，传统的方法是使用符号化的查询去通过属性来检索知识库，但这种符号化的操作会破坏端到端训练的函数的可微性。Dhingra 等人[212]构造了一个端到端的 KB-InfoBot，使用“软查询”的后验分布来表示用户对知识库中实体的兴趣度，并将这种软检索和强化学习结合，替换掉符号化的查询，帮助用户从知识库中获取信息，从而利用用户反馈进行端到端的训练。它能够帮助用户在不编写复杂的查询语句的情况下搜索知识库，同时，用户输入里存在的不确定性也得到了解决。

3.2.3 混合框架

端到端任务型对话系统还可以是由监督学习结合深度学习或强化学习的框架，即混合框架，进行构建。例如，结合监督学习和深度学习模式进行优化的端到端方法[213-215]。William 等人[213-214]提出混合编码网络（Hybrid Code Networks，HCN），通过使用 LSTM 来避免状态跟踪的繁琐步骤，以联合优化 DST 和 DP 模块。该方法的处理流程是首先输入用户会话，提取实体。实体结合历史对话，将其映射为向量，作为 RNN 的输入。RNN 为对话管理，结合一些领域知识、动作掩码（action mask）和模板，输出决策。根据 RNN 输出的决策，基于模板生成响应的过程。HCN 除了学习 RNN，也允许开发者通过软件编程引入领域知识模块和动作模板。RNN 之后经过密连接层和 softmax 层，将输出的动作模板的数量与动作掩码点乘并进行规范化，就得到了下一步的动作。最后，在“实体输出”模块替换上具体实体值，组织回答句子并控制分支根据行为类别不同调用 API 或者返回文本应答。该方法结合了传统的监督学习和强化学习模式进行端到端的训练，只需要更少的训练数据，就取得与传统端到端方法等同甚至更优成绩并可以进行自动学习对话状态。但是很多模块需要较多的人为干预，比如需要人为干预对话状态跟踪、数据库模块，制定动作掩码等。Li 等人[215]也使用监督学习和强化学

习的技术以端到端的方式训练完成任务型对话系统，通过结合用户模拟器来模拟用户议程进行训练，以此形成一个较完整的训练系统。该方法完整流程是用户模拟器输出查询给 NLU 模块作为输入，NLU 模块输出用 IOB（in-out-begin）格式表示的语义槽和意图。DM 模块利用 NLU 模块输出的信息输出动作。该系统具有更强的鲁棒性，能够根据强化学习在未知的情况下自动选择用户的动作，以允许在真实世界任务完成方案中与用户进行自然交互。同时允许用户在对话过程中发起对话，使得用户可以更加灵活的与系统进行交互，并演示了如何使用特定任务集并模拟用户以端到端的方式评估强化学习的智能体从而保证了系统的可重复性。该方法使用用户模拟器[216]在一定程度上解决了[203]和[204]训练数据难获得的问题，但其在训练的过程也较依赖于特定数据集。

3.2.4 端到端方法总结

为解决传统管道方法的问题，出现了一些端到端的任务型对话系统的尝试研究，端到端的系统一般使用一个 Seq2Seq 模型，根据用户的输入，直接生成相应的回复，具有结构简单、便于移植等优点，同时解决了标注数据不足的问题。由于任务型对话的特殊性，使用简单的 Seq2Seq 模型无法生成时效性、地点相关等回复，所以还需要辅以相应的知识库和常识库。知识库是用于知识管理的一种特殊数据库，但由于数据库需要通过符号查询进行检索，而一般的编码器和解码器在中间层不提供符号化表示，这种数据库衔接的不可微分特性对端到端方法的对话系统结合数据库提高了难度。为此，常见的做法是把数据库查询的模板作为系统可选择的动作，DP 在必要的时候输出对数据库进行搜索的动作，DP 可以通过监督学习或强化学习进行学习和优化[211,213]。这种方法能够适用于任何数据库或 API，但难以解决用户模糊需求的任务。另一种方法利用计算概率分布的过程代替数据库的衔接，既保证了过程可微分性，又能满足用户需求[212]。当面对很大的数据库时，概率分布计算的效率成了巨大的问题，因此利用解码器直接生成数据库的查询[203]的方法非常有效，但由于解码器的生成过程不能保证查询的格式正确，并且不能够满足任务型对话系统的逻辑推理和语义理解需求。因此，一些最新的研究集中于如何基于知识库进行序列到序列的回复生成。这也是符号计算与神经计算相结合的一个典型案例。由于常识数量巨大，非任务型对话系统已有利用外部存储模块来帮助系统更接近地模仿人类对话的模型[49]，但目前进行实验的问题还限定在开放域，因此对于任务型对话系统的端到端方发引入大规模常识还是一个巨大挑战。

4 任务型对话系统评估方法

随着任务型对话系统的发展，与其对应的评价方法也逐渐成为对话系统的一个重要课题[217]。一个面向任务的对话系统的各个子模块的评测指标如表格 8 所示：

表 8 任务型对话系统的各个子模块的评测指标

子模块名	评测指标
自然语言理解	分类问题、准确率、召回率和 F-score
对话状态跟踪	假设准确率、平均排序倒数、L2 范数、平均概率、ROC 表现、等误差率和正确接受率 5/10/20
对话策略	任务完成率、平均对话轮数
自然语言生成	目前该模块的主流实现技术为基于模板的方法，因此暂不做测评

虽然对话系统中的各个组成部分通常可以根据如表 8 所示的明确指标进行优化和评估，但评估整个对话系统需要更全面的指标[218-220]。目前任务型对话系统常用的评估方法包括基于语料库的方法[221-223]、基于用户模拟的方法[224-226]、基于人工评价的方法[81,227-228]和多种方式相结合的混合方法[115]等。本文将任务型对话系统任务常见的测评数据及相关信息归纳成表 9。

研究发现，任务型对话系统评估的关键性指标是：对话系统任务的成功率，即能够成功解决用户问题的对话（例如，购买正确的电影票、找到合适的餐馆等）[100]和对话产生的成本，即对话的轮次[218]。然而当对话系统应用于真实对话场景时，对话任务的完成率的界定非常模糊，因此基于语料库的对话评估模型成为了一个活跃的研究方向。大量实践表明[221-222]，训练语料的质量和对话系统生成内容的质量紧密相关，但是高质量的训练语料的专业的、完整的标注需要繁重的工作量。因此，用户模拟器作为一种廉价的替代品被越来越多的研究人员接受，用户模拟器试图模仿真实用户在对话中所作的事情，这样既可以减少人工消耗，也可以产生更多可用的训练数据。

用户模拟是一种廉价有效的评估方法，通过模拟不同领域下的人类的交互行为，有效地在大范围

内进行测试和评价[224-226]。虽然在用户模拟方面已经做了大量的研究工作，但构建类似于人类的模拟器仍然是一项具有挑战性的任务[229]。Dhigra 等人[212]在实践中经常观察到，在为真实用户提供服务时，用户模拟器由于过拟合不能达到很好的效果。因此，用户模拟和真实人类的差距是基于用户模拟的对话评估模型的最大限制。

由于模拟用户与人类用户之间存在差异，因此通常需要人类用户对对话系统进行测试，以便更加可靠的评估其指标。一种方法是通过招募测试人员，在特定的领域下从预定义的用户目标集合中采样一个用户目标与对话系统进行交互，根据一些重要度量标准，例如任务完成度和每个对话的平均轮次对系统进行评分[230-231]。这种方法比基于模拟用户评估模型效果更好，但招募测评人员非常昂贵且耗时，仅适用于资源雄厚的实验室。另一种方法是通过真实用户进行测试[232-233]，相比招募人员得到的指标更可靠，且实际用户基数更大，从而导致评估更大灵活性，该方法最主要的缺点是存在潜在的用户体验负面风险。

实际上，采用多种方式相结合的混合方法对对话系统评估是合理的，例如 Shah 等人[234]使用监督和强化学习从模拟用户获得的数据和在线经验中训练，然后利用真实用户对 DP 进行验证或微调。这种利用真实用户提高用户模拟与人工评价拟合程度的混合评估方法既不会产生巨大开销，也能够缩小用户模拟和真实用户的差距，但如果对实验集合没有很好的监控，真实用户的评价就不能非常完整地表现出对话的效果和特点。

因此，任务型对话系统评估方法还需以人工评价结果为目标不断地进行探索。

表 9 任务型对话系统任务常见的评测数据及其相关信息汇总表①

数据集	介绍	链接（下载&论文）	单轮或多轮	数据集大小	测评方式
MultiWOZ 2.0	1. 由 EMNLP 2018 最佳论文提出 2. 包含 7 个领域：医院、报警、酒店、餐厅、出租车、火车 3. 人-人对话的数据集	Download: http://dialogue.mi.eng.cam.ac.uk/index.php/corpus/ Paper: https://arxiv.org/pdf/1810.00278.pdf	多	包含 10438 组对话： 单领域对话平均对话轮数为 8.93 多领域为 15.39 轮 共计 115434 轮对话	客观 + 人工
Medical DS	1. 数据集来自中国在线医疗保健社区的儿科 2. 面向任务型自动诊断病情对话系统	Download: http://www.sdspeople.fudan.edu.cn/zywei/data/acl2018-mds.zip Paper: http://www.sdspeople.fudan.edu.cn/zywei/paper/liu-acl2018.pdf	多	4 类疾病 67 种症状	客观 + 人工
Snips	1. 由 Snips 收集用于模型评估 2. 包含 7 个任务：天气、播放音乐、搜索、添加到列表、预定和电影等 3. 广泛用于 SLU 模块研究	Download: https://github.com/snipsco/nlu-benchmark/tree/master/2017-06-custom-intent-engines	单	7 个意图 72 个槽值 训练集：13084 测试集：7007	客观 + 人工
MIT Restaurant Corpus	1. 餐厅 2. 广泛用于 SLU 模块研究	Download: https://groups.csail.mit.edu/sls/downloads/restaurant/	单	训练集：6894 验证集：766 测试集：1521	客观 + 人工
MIT Movie Corpus	1. 电影 2. Eng 语料库是简单的查询，Trivia 语料库是复杂的查询	Download: https://groups.csail.mit.edu/sls/downloads/movie/	单	Eng 语料库： 训练集：8798 验证集：977 Trivia 语料库： 训练集：7035 验证集：781	客观 + 人工

① https://github.com/AtmaHou/Task-Oriented-Dialogue-Dataset-Survey

	3. 广泛用于 SLU 模块研究			测试集：2443 测试集：1953	
ATIS	1. 航空旅行信息系统数据集 2. 广泛用于 SLU 模块研究	Download: 1.https://github.com/AtmaHou/Bi-LSTM_PosTagger/tree/master/data 2.https://github.com/yvchen/JointSLU/tree/master/data	单	21 个意图 120 个槽值 训练集：4478 测试集：893	客观 + 人工
Microsoft Dialogue Challenge	1. 包含三个领域：电影票预定、餐厅预定和出租车预定 2. 在每个域中具有内置用户模拟器用于训练和评估	Paper： https://arxiv.org/pdf/1807.11125.pdf	多	按意图、槽值和对话组数对三个领域的数据集大小进行描述： 电影票预定：11 、29、 2890 出租车预定：11、29、3094 餐厅预定：11 、30、4103	客观 + 人工
CamRest676	1. 预定餐厅 2. 人-人对话的数据集	Download: https://www.repository.cam.ac.uk/handle/1810/260970 Paper: https://arxiv.org/abs/1604.04562	多	一共 676 组对话， 一共 1500 轮对话， 训练集：验证集：测试集=3:1:1	客观 + 人工
Human-human goal oriented dataset	1. Maluuba 租用旅行预定数据集 2. 人-人对话的数据集	Download: https://datasets.maluuba.com/Frames/dl Paper: https://arxiv.org/abs/1706.01690 https://1drv.ms/b/s!Aqj1OvgfsHB7dsg42yp2BzDUK6U	多	一共 1369 组对话， 一共 19986 轮对话， 平均用户满意度（1-5）：4.58	客观 + 人工
Dialog bAbI task data	1. 在餐厅预定餐桌 2. 广泛用于评估端到端方法的对话系统	Download: https://research.fb.com/downloads/babi/ Paper: http://arxiv.org/abs/1605.07683	多	训练集：1000 验证集：1000 测试集：1000	客观 + 人工
Stanford Dialog Dataset	1. 斯坦福 NLP 组发布的关于汽车自动驾驶智能体的数据集 2. 包括三个任务：规划时间表、天气和导航 3. 人-人对话的数据集	Download: http://nlp.stanford.edu/projects/kvret/kvret_dataset_public.zip Paper: https://arxiv.org/abs/1705.0541	多	训练集：2425 验证集：302 测试集：301 每组对话的平均轮数为 5.25	客观 + 人工
DSTC-2	1. 餐厅预定数据集 2. 人-机对话的数据集	http://camdial.org/~mh521/dstc/	多	训练集：1612 验证集：506 测试集：1117	客观 + 人工
DSTC-4	1. 查询旅游信息 2. 人-人对话数据	不公开	多	训练集：20 个对话 测试集：15 个对话	
Movie Booking Dataset	1. 预定电影对话数据 2. 人-人对话数据	Download: https://github.com/MiuLab/TC-Bot#data Paper: TC-bot	多	280 轮对话 平均每个对话为 11 轮	客观 + 人工

TOP sematic parsing	1. 导航和活动 2. 人-机对话数据 3. 广泛应用于自然语言理解	Download: http://fb.me/semanticparsingdialog Paper: https://arxiv.org/pdf/1810.07942.pdf	单	训练集：31279 句 开发集：4462 句 测试集：9042 句	客观 + 人工

5 任务型对话系统未来研究方向

传统的任务型对话系统通常使用手工规则模板或浅层机器学习模型来单独优化模块。近期，深度学习和强化学习逐渐被用于全面优化对话系统，帮助对话系统在不断变化的环境中自动优化系统，以便系统能够有效地适应不同的任务、领域和用户行为。虽然对话系统近年来有较大的发展，已经能帮助用户完成简单的任务，但在通用性、深度理解等方面还面临着许多挑战。任务型对话系统未来发展的一些潜在趋势：

（1）低资源启动

任务型对话系统的成果往往依赖于大量高质量的语料作为训练数据，然而对话数据通常是异构的。例如聊天数据很多，但面向任务的对话数据集非常小。特定领域的对话数据的收集和标注是需要耗费大量的人力。未来的研究需要解决利用低资源启动对话系统的挑战。

对这类问题，目前的主要解决方法有：直接增加标注数据、引入领域知识、半监督学习、无监督学习、主动学习、多任务学习、数据增强和迁移学习等。其中，为现有的有限训练数据增加标注的效果最明显，但标注数据仍然是一个繁重的任务。因此，目前常用的方法是利用自动标注来帮助构建领域知识，结合领域知识来使得数据包含信息的能力更强以达到节约成本的目的，但领域知识的构建过程依旧需要消耗大量的人力物力。无监督学习和半监督学习可以更大程度的利用无标注数据帮助训练。虽然半监督或无监督学习在一定程度上避免了数据资源的浪费，但半监督学习需要做未标记数据于已存在的标注类别相联系的基本假设，如平滑假设、流形假设等，当实际情况不满足于这些基本假设，例如当面对复杂的数据时，比较难准确地还原数据的流形，无监督的样本不仅不能改进这些模型的性能，反而起到恶化的作用。并且，无监督学习需要的数据量是监督学习的百倍，甚至更多，而特定域的对话数据集收集非常困难。另外，主动学习通过发现并抽取价值更高的样本，即对实验效果影响最大的样本或信息量大的样本，请求人工标注，从而最大限度地减少了人工标注的工作量。但该方法仍然避免不了人工标注的繁重工作。数据增强技术[235] 可以利用目前不足量的数据产生更多具有同等价值的数据弥补训练数据的不足和利用非对话语料数据来增加训练数据 [199]。除此之外，多任务学习和迁移学习还可以通过使用不同语言、任务、领域或模型的信息作为对话系统训练数据的一部分[236]。目前，面对冷启动或数据不足的情况时，迁移学习逐步地被广泛应用。除了可以将其他领域的可复用信息迁移到目标领域的知识图谱，迁移学习还可以将其他模型的含有可复用特征所属的网络层次特征迁移到目标网络来帮助目标网络的训练等。然而，迁移学习在实际应用可能会在将源领域中含有噪音的信息迁移到目标领域，损伤目标领域的性能，因此源领域和目标领域的实际效果并不等价。未来任务型对话系统的发展必然需要解决低资源训练的难题。

（2）域适应能力

如何以更低的开发成本覆盖更多的领域和场景是任务型对话系统的关键问题之一，快速更新对话智能体以处理不断变化的环境非常重要。目前的任务型对话系统针对每一个领域都需要手工制定模板导致领域拓展性不足。未来的研究需要解决任务型对话系统领域拓展时遇到的挑战。

针对这类问题，目前的主要解决方法有：迁移学习、多任务学习、零次学习和单例学习等。对于多领域之间的移植问题，越来越多的研究选择迁移学习作为首选。迁移学习[237]通过利用源领域中的可复用的知识，并将其移植到目标领域作为目标领域训练数据的一部分来帮助学习和训练目标领域。其中零次学习[36]和单例学习是迁移学习的特例，也被经常用于多领域之间的快速迁移任务。两者的主要区别为标记样本的数量，零次学习为没有标记样本而单例学习为只有一个标记样本。对于快速出现的新领域，迁移学习能够快速迁移和应用，体现时效性的优势，但知识的迁移只有在“适当”的情况下才有可能，而“适当”的概念很难被量化需要经验来帮助确定，因此，迁移学习并不是适用于所有新领域的问题。多任务学习通过挖掘不同任务之间的联系和区别，共享源任务和目标任务之间的相同

参数或两者之间的共有数据特征来使得目标任务可以从不同的源任务中学习到相关的知识。虽然多任务学习通过学习不同任务之间的联系和差异来提高每个任务的学习效率和质量，但是其效果受到任务差异和数据分布带来影响，同时模型增加了参数量所以需要更大的数据量来训练模型，并且模型更复杂不利于在真实环境中实际部署使用。

（3）领域知识和常识的引入

随着互联网技术的不断发展，人们对人机交互的期望越来越高，在深度学习框架中融合语言理解和推理能力已经成为任务型对话系统乃至自然语言处理的一个重要研究课题。

在深度学习框架中融合语言理解能力和推理能力的重要方法是引入领域知识和常识，因为真实人与人之间的交互需要相关领域的知识储备，仅仅依靠对话文本包含的信息无法准确地理解用户输入和恰当地回复用户。而在实际对话中，还需要对信息进行推理并回答，常识知识的引入可以使得对话系统对于用户的话语更深入的理解，从而更贴近真实人类和谐、自然的交互方式。其中，领域知识主要包括维基百科和知识图谱两大类，知识图谱辅以维基百科知识有可能解决很多复杂的实际问题。现实世界中的大量的问题可以抽象成图模型，即节点和连边的集合，例如知识图谱可以将多源异构的数据汇聚到一起，并与多个传统人工智能领域进行融合解决真实问题。关于知识图谱的分析、推理研究，难免要与图数据的挖掘相结合，并借鉴相关的图神经网络的算法模型。虽然高质量的知识图谱对于上层应用具有很大提升效果，但高质量的知识图谱构建需要具有完备的知识、准确的知识表示、常识知识的支撑等，这就使得知识的获取、实体边界的识别、知识的形式化表示、知识的融合和知识的应用等技术面临巨大的挑战。同时，知识图谱还应该不断产生新的知识反哺知识库，这种动态的知识图谱的构建也是知识图谱的一项重要挑战。高效地融合领域知识不但要克服上述的挑战，而且还要一整套知识工程的方法，因此如何引入领域知识和构建高质量的知识图谱是未来对话系统的一项关键技术挑战。基于领域知识的对话系统比较常见，然而常识知识的收集是一个严峻的挑战。因为常识知识通常没有明确存储在现在的知识库中。目前已经研发了一些新的数据集和图谱来促进常识推理的研究，例如 Winograd Schema Challenge（WSC）[238]和可选择的替代选择（COPA）[239]，事理图谱，但是如何将常识应用在人机对话中尚无深入的研究。

6 结论与展望

本文主要从最新研究进展和热点方面对任务型对话系统的两种方法：管道方法和端到端方法，进行梳理和总结。在传统方法的基础上，以知识图谱、深度学习和强化学习等技术的最近研究进展为重点，以预训练神经网络、端到端方法、多任务学习和迁移学习等热点为方向，深度剖析任务型对话系统的相关技术。最后，总结了任务型对话系统的评估方法和测试集，并指出未来研究方向。

人机对话系统有广阔的应用前景，也在很多实际问题中发挥了作用。微软首席执行官萨提亚·纳德拉在微软 Build 2016 开发者大会上提出“对话即平台”（Conversation as a Platform，CaaP）④的概念，他认为对话是给人工智能带来颠覆式的下一代革命，更是推动人工智能领域改革的一剂兴奋剂。手机助手、问答系统、智能聊天机器人等对话系统如雨后春笋般的涌现，人机对话系统业的蓬勃发展展现了人机对话广泛的应用价值。但目前的对话系统的相关技术还处于发展阶段，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，但其通用性较差。而在低资源驱动、多领域切换、知识推理和复杂任务等场景下，对话的效果和用户的体验不具有稳定性。因此，任务型对话系统受限于场景的难易程度，还有很大进步空间。随着人工智能技术的不断发展，未来，人机对话系统会作新型人机交互界面，使自动化的机器可以取代很多烦琐、重复的人类劳动，改变人类的生活方式。

参 考 文 献

[1] Cao Jun-Kuo, Chen Guo-Lian. Human-machine dialogue system. Huaxin Building, No. 288 Jinjiacun South End of Wanshou Road: Publishing House of Eletronics Industry, 2017 (in Chinese)
(曹均阔, 陈国莲. 人机对话系统. 北京市万寿路南口金家村 288 号华信大厦: 电子工业出版社, 2017)

[2] Weizenbaum J. Eliza--a computer program for the study of natural language communication between man and machine. Communications of the ACM, 1983, 26(1): 23-28

④https://www.msra.cn/zh-cn/news/executivebylines/tech-bylines-nlp

[3] Yu Kai, Chen Lu, Chen Bo, et al. Cognitive technology in task-oriented dialogue systems-concepts, advances and future. Chinese Journal of Computers, 2015, 38(12): 2333-2348 (in Chinese)
(俞凯, 陈露, 陈博, 等. 任务型人机对话系统中的认知技术-概念, 进展及其未来. 计算机学报, 2015, 38(12): 2333-2348)

[4] Jia Xi-bin, Li Rang, Hu Chang-jian, Chen Jun-cheng. Review of intelligent dialogue system. Journal of Beijing University of Technology, 2017, 43(9): 1344-1356(in Chinese)
(贾熹滨, 李让, 胡长建, 陈军成. 智能对话系统研究综述. 北京工业大学学报, 2017, 43(9): 1344-1356)

[5] Chen H, Liu X, Yin D, et al. A survey on dialogue systems: Recent advances and new frontiers. ACM SIGKDD Explorations Newsletter, 2017, 19(2): 25-35

[6] Joachims T. Text categorization with support vector machines: Learning with many relevant features//Proceedings of the Machine Learning: ECML-98. Berlin, Germany, 1998: 137-142

[7] Hearst M A, Dumais S T, Osuna E, et al. Support vector machines. IEEE Intelligent Systems and their applications, 1998, 13(4): 18-28

[8] McCallum A, Nigam K. A comparison of event models for naive bayes text classification//Proceedings of the AAAI-98 workshop on learning for text categorization. Madison, USA, 1998: 41-48

[9] Cover T, Hart P. Nearest neighbor pattern classification. IEEE transactions on information theory, 1967, 13(1): 21-27

[10] Lewis D.D. Naive (bayes) at forty: The independence assumption in information retrieval//Proceedings of the European conference on machine learning. Chemnitz, Germany, 1998: 4-15

[11] Zhang H. The optimality of naive bayes//Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference. Miami Beach, USA, 2004: 562-567

[12] Yang Y, Liu X. A re-examination of text categorization methods//Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval. New York, USA, 1999: 42-49

[13] Weinberger K Q, Blitzer J, Saul L K. Distance metric learning for large margin nearest neighbor classification. Journal of Machine Learning Research, 2009, 10(Feb): 207-244

[14] Joachims T. Transductive inference for text classification using support vector machines//Proceedings of the Sixteenth International Conference on Machine Learning. Bled, Slovenia, 1999: 200-209

[15] Deng L, Tur G, He X, Hakkani-Tur D. Use of kernel deep convex networks and end-to-end learning for spoken language understanding//Proceeding of the IEEE Workshop on Spoken Language Technologies. Miami, USA, 2012: 210-215

[16] Dauphin Y N, Tur G, Hakkani-Tur D, et al. Zero-shot learning for semantic utterance classification. arXiv preprint arXiv:1401.0509, 2013

[17] Bapna A, Tur G, Hakkani-Tur D, Heck L.P. Towards zero-shot frame semantic parsing for domain scaling//Proceeding of the 18th Annual Conference of the International Speech Communication Association. Stockholm, Sweden, 2017: 2476-2480

[18] Hashemi H.B, Asiaee A, Kraft R. Query intent detection using convolutional neural networks//Proceeding of the International Conference on Web Search and Data Mining Workshop on Query Understanding. San Francisco, USA, 2016

[19] Huang P S, He X, Gao J, et al. Learning deep structured semantic models for web search using clickthrough data//Proceedings of the 22nd ACM international conference on Conference on information & knowledge management. San Francisco, USA, 2013: 2333-2338

[20] Shen Y, He X, Gao J, et al. Learning semantic representations using convolutional neural networks for web search//Proceedings of the 23rd International Conference on World Wide Web. Seoul, Republic of Korea, 2014: 373-374

[21] Kim Y. Convolutional neural networks for sentence classification// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar, 2014: 1746–1751

[22] Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. Baltimore, Maryland, 2014: 655–665

[23] Zhou C, Sun C, Liu Z, et al. A c-lstm neural network for text classification. arXiv preprint arXiv:1511.08630, 2015

[24] Lee J Y, Dernoncourt F. Sequential short-text classification with recurrent and convolutional neural networks//Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. San Diego, California, 2016: 515–520

[25] LAFFERTY J D, MCCALLUM A, PEREIRA F C N. Conditional random fields: Probabilistic models for segmenting and labeling sequence data//Proceedings of the Eighteenth International Conference on Machine Learning. Williamstown, USA, 2001: 282-289

[26] Peng F, McCallum A. Information extraction from research papers using conditional random fields. Information processing & management, 2006, 42(4): 963-979

[27] Rabiner L R. A tutorial on hidden markov models and selected applications in speech recognition//Proceedings of the IEEE. Scottsdale, USA, 1989: 257-286

[28] SUNG L, GUAN Y, WANG X L, et al. A maximum entropy markov model for chunking//Proceedings of the International Conference on Machine Learning and Cybernetics. Guangzhou, China, 2005: 3761-3765

[29] Ma X, Xia F. Unsupervised dependency parsing with transferring distribution via parallel guidance and entropy regularization//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. Baltimore, Maryland, 2014: 1337-1348

[30] Mesnil G, Dauphin Y, Yao K, et al. Using recurrent neural networks for slot filling in spoken language understanding. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2015, 23(3): 530-539

[31] Mesnil G, Dauphin Y, Yao K, et al. Using recurrent neural networks for slot filling in spoken language understanding. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2015, 23(3): 530-539

[32] Hochreiter S, Schmidhuber J. Long short-term memory. Neural computation, 1997, 9(8): 1735-1780

[33] Gers, Felix A., Jürgen Schmidhuber, and Fred Cummins. Learning to forget: Continual prediction with lstm. Neural Computation, 1999, 12(10): 2451-2471

[34] Cho K, Van Merriënboer B, Bahdanau D, et al. On the properties of neural machine translation: encoder-decoder approaches//Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation. Doha, Qatar, 2014: 103–111

[35] Hori C, Hori T, Watanabe S, et al. Context sensitive spoken language understanding using role dependent lstm layers//Proceedings of the Machine Learning for SLU Interaction NIPS 2015 Workshop. Montreal, Canada, 2015: 3236-3240

[36] Chen Y N, Hakkani-Tür D, Tür G, et al. End-to-end memory networks with knowledge carryover for multi-turn spoken language understanding//Proceedings of the Interspeech 2016, 17th Annual Conference of the International Speech Communication Association. San Francisco, USA, 2016: 3245-3249

[37] Ma X, Hovy E. End-to-end sequence labeling via bi-directional lstm-cnns-crf//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany, 2016: 1064–1074

[38] Huang Z, Xu W, Yu K. Bidirectional lstm-crf models for sequence tagging. arXiv preprint arXiv:1508.01991, 2015

[39] Lample G, Ballesteros M, Subramanian S, et al. Neural architectures for named entity recognition //Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. San Diego, California, 2016: 260–270

[40] Hakkani-T¨ur, D., T¨ur, G., Celikyilmaz, A., Chen, Y.-N., Gao, J., Deng, L., and Wang, Y.-Y. Multi-domain joint semantic frame parsing using bi-directional rnn-lstm//Proceedings of the 17th Annual Conference of the International Speech Communication Association. San Francisco, USA, 2016: 715–719

[41] Guo D, Tur G, Yih W, et al. Joint semantic utterance classification and slot filling with recursive neural networks//Proceedings of the Spoken Language Technology Workshop. South Lake Tahoe, USA, 2014: 554-559

[42] Xu P, Sarikaya R. Convolutional neural network based triangular crf for joint intent detection and slot filling//Proceedings of the Automatic Speech Recognition and Understanding. Olomouc, Czech Republic, 2013:78-83

[43] Lee S, Jha R. Zero-shot adaptive transfer for conversational language understanding. arXiv preprint arXiv:1808.10059, 2018

[44] Genevay A, Laroche R. Transfer learning for user adaptation in spoken dialogue systems//Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems. Singapore, 2016: 975-983

[45] Wu W L, Lu R Z, Duan J Y, et al. Spoken language understanding using weakly supervised learning. Computer speech & language, 2010, 24(2): 358-382

[46] Wu W L, Lu R Z, Duan J Y, et al. A weakly supervised learning approach for spoken language understanding//Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics. Sydney, Australia, 2006: 199-207

[47] Wang Xiao-Jie. Reflections on the human-machine dialogue system. ZTE TECHNOLOGY JOURNAL, 2017, 23(4): 47-50 (in Chinese)
(王小捷. 关于人机对话系统的思考. ZTE TECHNOLOGY JOURNAL, 2017, 23(4): 47-50)

[48] Zhou H, Young T, Huang M, et al. Commonsense knowledge aware conversation generation with graph attention//Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. Stockholm, Sweden, 2018: 4623-4629

[49] Young T, Cambria E, Chaturvedi I, et al. Augmenting end-to-end dialogue systems with commonsense knowledge//Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence. New Orleans, USA, 2018: 4970-4977

[50] Liu S, Chen H, Ren Z, et al. Knowledge diffusion for neural dialogue generation//Proceedings of the 56th Annual Meeting of the

Association for Computational Linguistics. Melbourne, Australia 2018: 1489-1498

[51] Lee C , Jung S , Kim K , et al. Hybrid approach to robust dialog management using agenda and dialog examples. Computer Speech and Language, 2010, 24(4): 609-631

[52] Choi E, He H, Iyyer M, et al. Quac: Question answering in context// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium, 2018: 2174–2184

[53] Yagcioglu S, Erdem A, Erdem E, et al. Recipeqa: A challenge dataset for multimodal comprehension of cooking recipes//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium, 2018: 1358–1368

[54] Yang Z, Qi P, Zhang S, et al. Hotpotqa: A dataset for diverse, explainable multi-hop question answering//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium, 2018: 2369–2380

[55] Rajpurkar P, Jia R, Liang P. Know what you don't know: Unanswerable questions for squad//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne, Australia, 2018: 784–789

[56] Rajpurkar P, Zhang J, Lopyrev K, et al. Squad: 100,000+ questions for machine comprehension of text//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, Texas, 2016: 2383–2392

[57] Reddy S, Chen D, Manning C D. Coqa: A conversational question answering challenge. arXiv preprint arXiv:1808.07042, 2018

[58] Mihaylov T, Frank A. Knowledgeable reader: enhancing cloze-style reading comprehension with external commonsense knowledge// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne, Australia, 2018: 821–832

[59] Wang H, Zhang F, Xie X, et al. Dkn: Deep knowledge-aware network for news recommendation. arXiv preprint arXiv:1801.08284, 2018

[60] Heck L, Hakkani-Tür D, Tur G. Leveraging knowledge graphs for web-scale unsupervised semantic parsing//Proceedings of the Interspeech. Lyon, France, 2013: 1594-1598

[61] Ma Y, Crook P A, Sarikaya R, et al. Knowledge graph inference for spoken dialog systems//Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). South Brisbane, Australia, 2015: 5346-5350

[62] Chen Y N, Hakkani-Tür D, Tur G. Deriving local relational surface forms from dependency-based entity embeddings for unsupervised spoken language understanding//Proceedings of the 2014 IEEE Spoken Language Technology Workshop (SLT). South Lake Tahoe, USA, 2014: 242-247

[63] Liu J, Pasupat P, Wang Y, et al. Query understanding enhanced by hierarchical parsing structures//Proceedings of the 2013 IEEE Workshop on Automatic Speech Recognition and Understanding. Olomouc, Czech Republic, 2013: 72-77

[64] Chen Y N, Wang W Y, Gershman A, et al. Matrix factorization with knowledge graph propagation for unsupervised spoken language understanding//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. Beijing, China, 2015: 483-494

[65] Peters M E, Neumann M, Iyyer M, et al. Deep contextualized word representations//Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. New Orleans, Louisiana, 2018: 2227–2237

[66] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018

[67] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need//Proceedings of the Advances in Neural Information Processing Systems. Long Beach, USA, 2017: 5998-6008

[68] Collobert R, Weston J. A unified architecture for natural language processing: Deep neural networks with multitask learning//Proceedings of the 25th international conference on Machine learning. Helsinki, Finland, 2008: 160-167

[69] Goddeau D, Meng H, Polifroni J, et al. A form-based dialogue manager for spoken language applications//Proceedings of the 4th International Conference on Spoken Language Processing. Philadelphia, USA, 1996: 701-704

[70] Zue V, Seneff S, Glass J R, et al. Juplter: A telephone-based conversational interface for weather information. IEEE Transactions on speech and audio processing, 2000, 8(1): 85-96

[71] Pulman S G. Conversational games ,belief revision and bayesian networks//Proceedings of the 7th Computational Linguistics in the Netherlands meeting. Nijmegen, Nederland, 1997: 1-19

[72] Wang Z, Lemon O. A simple and generic belief tracking mechanism for the dialog state tracking challenge: on the believability of observed information//Proceedings of the SIGDIAL 2013 Conference. SUPELEC, France, 2013: 423-432

[73] Sun K, Chen L, Zhu S, et al. A generalized rule based tracker for dialogue state tracking//Proceedings of the Spoken Language Technology Workshop. South Lake Tahoe, USA, 2014: 330-335

[74] Williams J D. Web-style ranking and slu combination for dialog state

tracking//Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue. Philadelphia, USA, 2014: 282-291

[75] Bai Zhan-Sheng, Lan Lan, Peng Jia-Hong, et al. Comparative study of control models in dialogue systems. Journal of Zhengzhou University (Science Edition), 2006, 38 (4): 112-116 (in Chinese)
(拜战胜, 蓝岚, 彭佳红, 等. 对话系统中控制模型的比较研究. 郑州大学学报(理学版), 2006, 38(4): 112-116)

[76] DeVault D, Stone M. Managing ambiguities across utterances in dialogue//Proceedings of the 11th Workshop on the Semantics and Pragmatics of Dialogue. Trento, Italy, 2007: 49-56

[77] Williams J, Raux A, Henderson M. The dialog state tracking challenge series: A review. Dialogue & Discourse, 2016, 7(3): 4-33

[78] Henderson M. Machine learning for dialog state tracking: A review//Proceedings of the First International Workshop on Machine Learning in Spoken Language Processing. Aizu, Japan, 2015

[79] Young S, Gašić M, Keizer S, et al. The hidden information state model: A practical framework for pomdp-based spoken dialogue management. Computer Speech & Language, 2010, 24(2): 150-174

[80] Thomson B, Young S. Bayesian update of dialogue state: A pomdp framework for spoken dialogue systems. Computer Speech & Language, 2010, 24(4): 562-588

[81] Bohus D, Rudnicky A. A k-hypotheses+ other belief updating model//Proceedings of the AAAI Workshop on Statistical and Empirical Methods in Spoken Dialogue Systems. Boston, USA, 2006: 62

[82] Henderson M, Thomson B, Young S. Deep neural network approach for the dialog state tracking challenge//Proceedings of the SIGDIAL 2013 Conference. SUPELEC, France, 2013: 467-471

[83] Mrkšić N, Séaghdha D O, Thomson B, et al. Multi-domain dialog state tracking using recurrent neural networks//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. Beijing, China, 2015: 794–799

[84] Williams J. Multi-domain learning and generalization in dialog state tracking//Proceedings of the SIGDIAL 2013 Conference. SUPELEC, France, 2013: 433-441

[85] Ren H, Xu W, Yan Y. Markovian discriminative modeling for dialog state tracking//Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue. Philadelphia, USA, 2014: 327-331

[86] Ren H, Xu W, Zhang Y, et al. Dialog state tracking using conditional random fields//Proceedings of the SIGDIAL 2013 Conference. SUPELEC, France, 2013: 457-461

[87] Lee S, Eskenazi M. Recipe for building robust spoken dialog state trackers: Dialog state tracking challenge system description//Proceedings of the SIGDIAL 2013 Conference. SUPELEC, France, 2013: 414-422

[88] Kim S, Banchs R E. Sequential labeling for tracking dynamic dialog states//Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue. Philadelphia, USA, 2014: 332

[89] Ma Y, Fosler-Lussier E. A discriminative sequence model for dialog state tracking using user goal change detection//Proceedings of the Spoken Language Technology Workshop. South Lake Tahoe, USA, 2014: 318-323

[90] Henderson M, Thomson B, Young S. Word-based dialog state tracking with recurrent neural networks//Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue. Philadelphia, USA, 2014: 292-299

[91] Mrkšić N, Séaghdha D O, Wen T H, et al. Neural belief tracker: Data-driven dialogue state tracking//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver, Canada, 2017: 1777–1788

[92] Lei W, Jin X, Kan M Y, et al. Sequicity: Simplifying task-oriented dialogue systems with single sequence-to-sequence architectures//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne, Australia, 2018: 1437-1447

[93] Cuayáhuitl H, Keizer S, Lemon O. Strategic dialogue management via deep reinforcement learning. arXiv preprint arXiv:1511.08099, 2015

[94] Lemon, O., & Rieserr, V. Reinforcement learning for adaptive dialogue systems—tutorial//Proceedings of the European Association for Computational Linguistics. Athens, Greece, 2009

[95] Ross S, Bagnell D. Efficient reductions for imitation learning//Proceedings of the thirteenth international conference on artificial intelligence and statistics. Chia Laguna Resort, Italy, 2010: 661-668

[96] Ross, S.; Gordon, G. J. and Bagnell, J. A. No-regret reductions for imitation learning and structured prediction//Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. Ft. Lauderdale, USA, 2011

[97] Judah K, Fern A, Dietterich T G. Active imitation learning via reduction to iid active learning//Proceedings of the Conference on Uncertainty in Artificial Intelligence. Catalina Island, USA, 2012: 428-437

[98] He H, Eisner J, Daume H. Imitation learning by

coaching//Proceedings of the Advances in Neural Information Processing Systems. Lake Tahoe, USA, 2012: 3149-3157

[99] Kim B, Farahmand A, Pineau J, et al. Learning from limited demonstrations//Proceedings of the Advances in Neural Information Processing Systems. Lake Tahoe, USA, 2013: 2859-2867

[100] Ross S, Bagnell J A. Reinforcement and imitation learning via interactive no-regret learning. arXiv preprint arXiv:1406.5979, 2014

[101] Mnih V, Kavukcuoglu K, Silver D, et al. Playing atari with deep reinforcement learning. arXiv preprint arXiv:1312.5602, 2013

[102] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning. Nature, 2015, 518(7540): 529

[103] Watkins C J C H. Learning from delayed rewards [Ph.D. dissertation]. King's College, Cambridge, 1989

[104] Van Hasselt H, Guez A, Silver D. Deep reinforcement learning with double q-learning//Proceedings of the AAAI Conference on Artificial Intelligence. Phoenix, USA, 2016: 2094-2100

[105] Bellemare M G, Ostrovski G, Guez A, et al. Increasing the action gap: New operators for reinforcement learning//Proceedings of the AAAI Conference on Artificial Intelligence. Phoenix, USA, 2016: 1476-1483

[106] Schaul T, Quan J, Antonoglou I, Silver D. Prioritized experience replay//Proceedings of the 4th International Conference on Learning Representations. San Juan, Puerto Rico, 2016:322-355

[107] Lakshminarayanan A S, Sharma S, Ravindran B. Dynamic frame skip deep q network//Proceedings of the Workshops at the International Joint Conference on Artificial Intelligence. New York, USA, 2016

[108] Singh S P, Kearns M J, Litman D J, et al. Reinforcement learning for spoken dialogue systems//Proceedings of the Advances in Neural Information Processing Systems. Denver, USA, 1999: 956-962

[109] Gašić M, Jurčíček F, Keizer S, et al. Gaussian processes for fast policy optimisation of pomdp-based dialogue managers//Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue. Association for Computational Linguistics. Tokyo, Japan, 2010: 201-204

[110] Fatemi M, Asri L E, Schulz H, et al. Policy networks with two-stage training for dialogue systems//Proceedings of the SIGDIAL 2016 Conference. Los Angeles, USA, 2016: 101-110

[111] Sutton R S, Barto A G. Reinforcement learning: An introduction. Cambridge: MIT press, 2018

[112] Sutton R S. Integrated architectures for learning, planning, and reacting based on approximating dynamic programming//Proceedings of the Seventh International Conference on Machine Learning. Austin, USA, 1990: 216-224

[113] Peng B, Li X, Gao J, et al. Integrating planning for task-completion dialogue policy learning//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne, Australia, 2018: 2182-2192

[114] Su S Y, Li X, Gao J, et al. Discriminative deep dyna-q: Robust planning for dialogue policy learning//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium, 2018: 3813-3823

[115] Su P H, Gasic M, Mrksic N, et al. On-line active reward learning for policy optimisation in spoken dialogue systems//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany, 2016: 2431-2441

[116] Graves A, Jaitly N, Mohamed A. Hybrid speech recognition with deep bidirectional lstm//Proceedings of the 2013 IEEE workshop on automatic speech recognition and understanding. Olomouc, Czech Republic, 2013: 273-278

[117] Peng B, Li X, Li L, et al. Composite task-completion dialogue policy learning via hierarchical deep reinforcement learning//Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen, Denmark, 2017: 2231-2240

[118] Cuayáhuitl H, Renals S, Lemon O, et al. Evaluation of a hierarchical reinforcement learning spoken dialogue system. Computer Speech & Language, 2010, 24(2): 395-429

[119] Budzianowski P, Ultes S, Su P H, et al. Sub-domain modelling for dialogue management with hierarchical reinforcement learning//Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue. Saarbrücken, Germany, 2017: 86-92

[120] Tang D, Li X, Gao J, et al. Subgoal discovery for hierarchical dialogue policy learning//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium, 2018: 2298-2309

[121] Casanueva I, Budzianowski P, Su P H, et al. Feudal reinforcement learning for dialogue management in large domains//Proceedings of the 16th annual conference of the north American chapter of the association for computational linguistics: human language technologies. New Orleans, USA, 2018: 714-719

[122] Gašić M, Kim D, Tsiakoulis P, et al. Distributed dialogue policies for multi-domain statistical dialogue management//Proceedings of the Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on. South Brisbane. Queensland, Australia, 2015: 5371-5375

[123] Tresp V. A bayesian committee machine. Neural computation, 2000, 12(11): 2719-2741

[124] Gašić M, Mrkšić N, Su P, et al. Policy committee for adaptation in multi-domain spoken dialogue systems//Proceedings of the

Automatic Speech Recognition and Understanding, 2015 IEEE Workshop on. Scottsdale, USA, 2015: 806-812

[125] Cuayáhuitl H, Yu S, Williamson A, et al. Deep reinforcement learning for multi-domain dialogue systems//Proceedings of the NIPS Workshop on Deep Reinforcement Learning. Barcelona, Spain, 2016

[126] Gašić M, Kim D, Tsiakoulis P, et al. Incremental on-line adaptation of pomdp-based dialogue managers to extended domains//Proceedings of the Fifteenth Annual Conference of the International Speech Communication Association. Singapore, 2014: 140-144

[127] Lipton Z C, Li X, Gao J, et al. Efficient dialogue policy learning with bbq-networks. arXiv preprint arXiv:1608.05081, 2016

[128] Hinton G E, Van Camp D. Keeping the neural networks simple by minimizing the description length of the weights//Proceedings of the sixth annual conference on Computational learning theory. Santa Cruz, USA, 1993: 5-13

[129] Thompson W R. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. Biometrika, 1933, 25(3/4): 285-294

[130] Chapelle O, Li L. An empirical evaluation of thompson sampling//Proceedings of the Advances in neural information processing systems. Granada, Spain, 2011: 2249-2257

[131] Russo D J, Van Roy B, Kazerouni A, et al. A tutorial on thompson sampling. Foundations and Trends® in Machine Learning, 2018, 11(1): 1-96

[132] Blundell C, Cornebise J, Kavukcuoglu K, et al. Weight uncertainty in neural networks//Proceedings of the 32nd International Conference on Machine Learning. Lille, France, 2015: 1613-1622

[133] Casanueva I, Hain T, Christensen H, et al. Knowledge transfer between speakers for personalised dialogue management//Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue. Prague, Czech Republic, 2015: 12-21

[134] Wang Z, Wen T H, Su P H, et al. Learning domain-independent dialogue policies via ontology parameterisation//Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue. Prague, Czech Republic, 2015: 412–416

[135] Goddeau D, Pineau J. Fast reinforcement learning of dialog strategies//Proceedings of the 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Istanbul, Turkey, 2000: 1233-1236

[136] Roy N. Spoken dialogue management using probabilistic reasoning//Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics. Hong Kong, China, 2000

[137] Zhang Bo, Cai Qing-Sheng, Guo Bai-Ning. Pomdp model and solution of spoken dialogue system. Journal of Computer Research and Development, 2002(02):217-224 (in Chinese)
(张波, 蔡庆生, 郭百宁. 口语对话系统的 POMDP 模型及求解 计算机研究与发展, 2002(02):217-224)

[138] Williams J D, Young S. Partially observable markov decision processes for spoken dialog systems. Computer Speech and Language, 2007, 21(2):393-422

[139] Ameixa D, Coheur L, Fialho P, et al. Luke, i am your father: Dealing with out-of-domain requests by using movies subtitles//Proceedings of the International Conference on Intelligent Virtual Agents. Springer, Cham, 2014: 13-21

[140] Lane I, Kawahara T, Matsui T, et al. Out-of-domain utterance detection using classification confidences of multiple topics. IEEE Transactions on Audio, Speech and Language Processing, 2007, 15(1): 150-161

[141] Wang Jun-Dong, Huang Pei-Jie, Lin Xian-Mao, et al. A coprocessor for out-of-domian utterances in domain specific spoken dialogue systems. Journal of Chinese Information Processing, 2015, 29(5): 194-204 (in Chinese)
(王俊东, 黄沛杰, 林仙茂, 等. 限定领域口语对话系统中超出领域话语的协处理方法. 中文信息学报, 2015, 29(5): 194-204)

[142] Wang Jun-Dong, Huang Pei-Jie, Ke Zi-Xuan, et al. Dialogue behavior recognition for out-of-domain utterances in domain specific spoken dialogue systems. Journal of Chinese Information Processing, 2016, 30(6): 182-189 (in Chinese)
(黄沛杰, 王俊东, 柯子烜, 等. 限定领域口语对话系统中超出领域话语的对话行为识别. 中文信息学报, 2016, 30(6): 182-189)

[143] Zhou L, Gao J, Li D, et al. The design and implementation of xiaoice, an empathetic social chatbot. arXiv preprint arXiv:1812.08989, 2018

[144] Li F L, Qiu M, Chen H, et al. Alime assist: an intelligent assistant for creating an innovative e-commerce experience//Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. Singapore, 2017: 2495-2498

[145] Zhu Kun-Hong. Research and implementation of multimodal teaching dialogue system. Beijing: Beijing University of Posts and Telecommunications, 2018 (in Chinese)
(朱坤鸿. 多模态教学对话系统研究与实现. 北京: 北京邮电大学, 2018.)

[146] Yi Lian. Research and implementation of multimodal dialog management system. Beijing: Beijing University of Posts and Telecommunications, 2014 (in Chinese)

(易炼. 多模态对话管理系统的研究与实现. 北京: 北京邮电大学, 2014)

[147] Luo Sen-Lin, Pan Li-Min. Emotion computing theory and technology. Systems Engineering and Electronics, 2003, 25(7) (in Chinese)
(罗森林, 潘丽敏. 情感计算理论与技术. 系统工程与电子技术, 2003, 25(7))

[148] Fu Xiao-Lan. Conducting emotional computing research, building a harmonious electronic society. Bulletin of Chinese Academy of Sciences, 2008, 23(5): 453-457 (in Chinese)
(傅小兰. 开展情感计算研究 构建和谐电子社会. 中国科学院院刊, 2008, 23(5):453-457)

[149] Liliang R, Kaige X, Lu C, Kai Y. Towards universal dialogue state tracking//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium, 2018: 2780-2786

[150] Lemon O. Learning what to say and how to say it: joint optimisation of spoken dialogue management and natural language generation. Computer Speech & Language, 2011, 25(2): 210-221

[151] Salazar V L, Cabeza E M E, Peña J L C, et al. A case based reasoning model for multilingual language generation in dialogues. Expert Systems with Applications, 2012, 39(8): 7330-7337

[152] Baptist L, Seneff S. Genesis-ii: A versatile system for language generation in conversational system applications//Proceedings of the Sixth International Conference on Spoken Language Processing. Beijing, China, 2000: 271-274

[153] Stent A, Prasad R, Walker M. Trainable sentence planning for complex information presentation in spoken dialog systems//Proceedings of the 42nd annual meeting on association for computational linguistics. Association for Computational Linguistics. Barcelona, Spain, 2004: 79

[154] Walker M A, Rambow O C, Rogati M. Training a sentence planner for spoken dialogue using boosting. Computer Speech & Language, 2002, 16(3-4): 409-433

[155] Oh A H, Rudnicky A I. Stochastic language generation for spoken dialogue systems//Proceedings of the 2000 ANLP/NAACL Workshop on Conversational systems-Volume 3. Stroudsburg, USA, 2000: 27-32

[156] Mairesse F, Gašić M, Jurčíček F, et al. Phrase-based statistical language generation using graphical models and active learning//Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics. Uppsala, Sweden, 2010: 1552-1561

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

[158] Shang L, Lu Z, Li H. Neural responding machine for short-text conversation//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing. Beijing, China, 2015: 1577-1586

[159] Li J, Galley M, Brockett C, et al. A diversity-promoting objective function for neural conversation models//Proceedings of the 15th Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. California, America, 2016: 110-119

[160] Sordoni A, Galley M, Auli M, et al. A neural network approach to context-sensitive generation of conversational responses//Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Denver, America, 2015: 196-205

[161] Serban I V, Sordoni A, Bengio Y, et al. Building end-to-end dialogue systems using generative hierarchical neural network models//Proceedings of the Thirtieth Conference on Artificial Intelligence. Phoenix, USA, 2016: 3776-3784

[162] Li J, Galley M, Brockett C, et al. A persona-based neural conversation model//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany, 2016: 994-1003

[163] Dušek O, Jurčíček F. A context-aware natural language generator for dialogue systems// Proceedings of the SIGDIAL 2016 Conference. Los Angeles, USA, 2016: 185-190

[164] Kumar A, Irsoy O, Ondruska P, et al. Ask me anything: Dynamic memory networks for natural language processing//Proceedings of the International Conference on Machine Learning. New York City, USA, 2016: 1378-1387

[165] Zhou H, Huang M. Context-aware natural language generation for spoken dialogue systems//Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. Osaka, Japan, 2016: 2032-2041

[166] Hu Z, Yang Z, Liang X, et al. Toward controlled generation of text//Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia, 2017: 1587-1596

[167] Wen T H, Heidel A, Lee H, et al. Recurrent neural network based language model personalization by social network crowdsourcing//Proceedings of the 14th Annual Conference of the International Speech Communication Association. Lyon, France, 2013: 2703-2707

[168] Shi Y, Larson M, Jonker C M. Recurrent neural network language model adaptation with curriculum learning. Computer Speech & Language, 2015, 33(1): 136-154

[169] Wen T H, Gasic M, Mrksic N, et al. Multi-domain neural network language generation for spoken dialogue systems//Proceedings of NAACL-HLT 2016. San Diego, USA, 2016:120-129

[170] Mirkovic D, Cavedon L, Purver M, et al. Dialogue management using scripts and combined confidence scores, USA, 2011-3-8

[171] Wen T H, Gasic M, Kim D, et al. Stochastic language generation in dialogue using recurrent neural networks with convolutional sentence reranking//Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue. Prague, Czech Republic, 2015: 275–284

[172] Wen T H, Gasic M, Mrksic N, et al. Semantically conditioned lstm-based natural language generation for spoken dialogue systems. //Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal, 2015: 1711–1721

[173] Tran V K, Nguyen L M. Natural language generation for spoken dialogue system using rnn encoder-decoder networks//Proceedings of the 21st Conference on Computational Natural Language Learning. Vancouver, Canada, 2017: 442–451

[174] Gatt A, Krahmer E. Survey of the state of the art in natural language generation: Core tasks, applications and evaluation. Journal of Artificial Intelligence Research, 2018, 61: 65-170

[175] Liu C W, Lowe R, Serban I V, et al. How not to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation//Proceedings of the 2016 conference on empirical methods in natural language processing. Austin, Texas, 2016: 2122-2132

[176] Rus V, Cai Z, Graesser A C. Evaluation in natural language generation: The question generation task//Proceedings of the workshop on shared tasks and comparative evaluation in natural language generation. Arlington, USA, 2007: 20-21

[177] Luo L, Huang W, Zeng Q, et al. Learning personalized end-to-end goal-oriented dialog//Proceedings of the thirty-third AAAI conference on artificial intelligence. Honolulu, USA, 2019

[178] Zhang S, Dinan E, Urbanek J, et al. Personalizing dialogue agents: I have a dog, do you have pets too?//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne, Australia, 2018: 2204-2213

[179] Serban I V, Lowe R, Henderson P, et al. A survey of available corpora for building data-driven dialogue systems. Dialogue & Discourse, 2018, 9(1): 1-49

[180] Joshi C K, Mi F, Faltings B. Personalization in goal-oriented dialog//Proceedings of the 31st conference on neural information processing systems. Long Beach, USA, 2017

[181] Halliday M A K. The linguistic sciences and language teaching. London: Longmans, 1964

[182] Qiu M, Li F L, Wang S, et al. Alime chat: A sequence to sequence and rerank based chatbot engine//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver, Canada, 2017: 498-503

[183] Cheyer A, Guzzoni D. Method and apparatus for building an intelligent automated assistant, USA, 2014-3-18

[184] Mirkovic D, Cavedon L. Dialogue management using scripts, USA, 2011-10-18

[185] Henderson M, Thomson B, Young S. Robust dialog state tracking using delexicalised recurrent neural networks and unsupervised adaptation//Proceedings of the 2014 IEEE Spoken Language Technology Workshop (SLT). South Lake Tahoe, USA, 2014: 360-365

[186] Mairesse F, Young S. Stochastic language generation in dialogue using factored language models. Computational Linguistics, 2014, 40(4): 763-799

[187] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks//Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, Canada, 2014: 3104-3112

[188] Graves Alex. Supervised sequence labelling with recurrent neural networks. Berlin: Springer, 2012

[189] Minsky Marvin. Society of mind: A response to four reviews. Artificial Intelligence, 1991, 48(3):371–396

[190] Marková I, Linell P, Grossen M, et al. Dialogue in focus groups: Exploring socially shared knowledge. Sheffield: Equinox publishing, 2007

[191] Souto P C N. Creating knowledge with and from the differences: The required dialogicality and dialogical competences. RAI Revista de Administração e Inovação, 2015, 12(2): 60-89

[192] Ritter A, Cherry C, Dolan W B. Data-driven response generation in social media//Proceedings of the conference on empirical methods in natural language processing. Association for Computational Linguistics. Edinburgh, UK, 2011: 583-593

[193] Han S, Bang J, Ryu S, et al. Exploiting knowledge base to generate responses for natural language dialog listening agents//Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue. Prague, Czech Republic, 2015: 129-133

[194] Ghazvininejad M, Brockett C, Chang M W, et al. A knowledge-grounded neural conversation model//Proceedings of the

thirty-second AAAI conference on artificial intelligence. New Orleans, USA, 2018: 5110-5117

[195] Zhu W, Mo K, Zhang Y, et al. Flexible end-to-end dialogue system for knowledge grounded conversation. arXiv preprint arXiv:1709.04264, 2017

[196] Kannan A, Kurach K, Ravi S, et al. Smart reply: Automated response suggestion for email//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, USA, 2016: 955-964

[197] Mou L, Song Y, Yan R, et al. Sequence to backward and forward sequences: A content-introducing approach to generative short-text conversation//Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. Osaka, Japan, 2016: 3349-3359

[198] Li J, Monroe W, Jurafsky D. A simple, fast diverse decoding algorithm for neural generation. arXiv preprint arXiv:1611.08562, 2016

[199] Vijayakumar A K, Cogswell M, Selvaraju R R, et al. Diverse beam search: Decoding diverse solutions from neural sequence models. arXiv preprint arXiv:1610.02424, 2016

[200] Li J, Jurafsky D. Mutual information and diverse decoding improve neural machine translation. arXiv preprint arXiv:1601.00372, 2016

[201] Li J, Monroe W, Ritter A, et al. Deep reinforcement learning for dialogue generation//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, Texas, 2016: 1192-1202

[202] Lison P, Bibauw S. Not all dialogues are created equal: Instance weighting for neural conversational models//Proceedings of the SIGDIAL 2017 Conference. Saarbrucken, Germany, 2017: 384-394

[203] Bordes A, Boureau Y L, Weston J. Learning end-to-end goal-oriented dialog. arXiv preprint arXiv, 2016: 1605.07683

[204] Wen T H, Vandyke D, Mrksic N, et al. A network-based end-to-end trainable task-oriented dialogue system//Proceedings of the 15th Conference of the European Chapter of the Association for Computational. Valencia, Spain, 2017: 438-449

[205] Eric M, Manning C D. Key-value retrieval networks for task-oriented dialogue//Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue. Saarbrücken, Germany, 2017: 37-49

[206] Yang X, Chen Y N, Hakkani-Tür D, et al. End-to-end joint learning of natural language understanding and dialogue manager//Proceedings of the International Conference on Acoustics, Speech and Signal Processing. New Orleans, USA, 2017: 5690-5694

[207] Eric M, Manning C D. A copy-augmented sequence-to-sequence architecture gives good performance on task-oriented dialogue//Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics. Valencia, Spain, 2017: 468-473

[208] Wen T H, Gasic M, Mrksic N, et al. Conditional generation and snapshot learning in neural dialogue systems//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, Texas, 2016:2153-2162

[209] Wen T H, Miao Y, Blunsom P, et al. Latent intention dialogue models//Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia, 2017: 3732—3741

[210] Wen H, Liu Y, Che W, et al. Sequence-to-sequence learning for task-oriented dialogue with dialogue state representation//Proceedings of the 27th International Conference on Computational Linguistics. Santa Fe, USA, 2018: 3781-3792

[211] Zhao T, Eskenazi M. Towards end-to-end learning for dialog state tracking and management using deep reinforcement learning//Proceedings of the SIGDIAL 2016 Conference. Los Angeles, USA, 2016: 1-10

[212] Dhingra B, Li L, Li X, et al. Towards end-to-end reinforcement learning of dialogue agents for information access//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver, Canada, 2017: 484-495

[213] Williams J D, Zweig G. End-to-end lstm-based dialog control optimized with supervised and reinforcement learning. arXiv preprint arXiv:1606.01269, 2016

[214] Williams J D, Asadi K, Zweig G. Hybrid code networks: Practical and efficient end-to-end dialog control with supervised and reinforcement learning//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver, Canada, 2017: 665-677

[215] Li X, Chen Y N, Li L, et al. End-to-end task-completion neural dialogue systems//Proceedings of the Eighth International Joint Conference on Natural Language Processing. Taipei, China, 2017: 733-743

[216] Li X, Lipton Z C, Dhingra B, et al. A user simulator for task-completion dialogues. arXiv preprint arXiv:1612.05688, 2016

[217] Zhang Wei-Nan, Zhang Yang-Zi, Liu Ting. Survey of evaluation methods for dialogue systems. SCIENTIA SINICA Informationis, 2017, 47(08): 953-966 (in Chinese)
(张伟男, 张杨子, 刘挺. 对话系统评价方法综述. 中国科学: 信息科学, 2017, 47(08): 953-966)

[218] Walker M A, Litman D J, Kamm C A, et al. Paradise: A framework for evaluating spoken dialogue agents//Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics.

Madrid, Spain, 1997: 271-280

[219] Walker M A, Litman D J, Kamm C A, et al. Evaluating spoken dialogue agents with paradise: Two case studies. Computer Speech & Language, 1998, 12(4): 317-347

[220] Hartikainen M, Salonen E P, Turunen M. Subjective evaluation of spoken dialogue systems using ser vqual method//Proceedings of the Eighth International Conference on Spoken Language Processing. Jeju Island, Korea, 2004: 2273-2276

[221] Yang Z, Levow G A, Meng H. Predicting user satisfaction in spoken dialog system evaluation with collaborative filtering. IEEE Journal of Selected Topics in Signal Processing, 2012, 6(8): 971-981

[222] El Asri L, Laroche R, Pietquin O. Task completion transfer learning for reward inference//Proceedings of the Workshops at the Twenty-Eighth AAAI Conference on Artificial Intelligence. Québec, Canada, 2014: 38–43

[223] Ultes S, Kraus M, Schmitt A, et al. Quality-adaptive spoken dialogue initiative selection and implications on reward modelling//Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue. Prague, Czech Republic, 2015: 374-383

[224] Watanabe T, Araki M, Doshita S. Evaluating dialogue strategies under communication errors using computer-to-computer simulation. IEICE transactions on information and systems, 1998, 81(9): 1025-1033

[225] Ai H, Weng F. User simulation as testing for spoken dialog systems//Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue. Columbus, USA, 2008: 164-171

[226] Schatzmann J. Statistical user and error modelling for spoken dialogue systems [Ph.D. dissertation]. University of Cambridge, UK, 2008

[227] Henderson J, Lemon O, Georgila K. Hybrid reinforcement/supervised learning for dialogue policies from communicator data//Proceedings of the IJCAI workshop on knowledge and reasoning in practical dialogue systems. Edinburgh, UK, 2005: 68-75

[228] Gašić M, Lefevre F, Jurčíček F, et al. Back-off action selection in summary space-based pomdp dialogue systems//Proceedings of the 2009 IEEE Workshop on Automatic Speech Recognition & Understanding. Merano, Italy, 2009: 456-461

[229] Pietquin O, Hastie H. A survey on metrics for the evaluation of user simulations. The knowledge engineering review, 2013, 28(1): 59-73

[230] Singh S, Litman D, Kearns M, et al. Optimizing dialogue management with reinforcement learning: Experiments with the njfun system. Journal of Artificial Intelligence Research, 2002, 16: 105-133

[231] Young S, Breslin C, Gašić M, et al. Evaluation of statistical pomdp-based dialogue systems in noisy environments//Proceedings of the Alexander Rudnicky, Antoine Raux, Ian Lane, Teruhisa Misu. Situated Dialog in Speech-Based Human-Computer Interaction. Berlin: Springer, 2016: 3-14

[232] Black A W, Burger S, Conkie A, et al. Spoken dialog challenge 2010: Comparison of live and control test results//Proceedings of the SIGDIAL 2011 Conference. Association for Computational Linguistics. Portland, Oregon, 2011: 2-7

[233] Hofmann K, Li L, Radlinski F. Online evaluation for information retrieval. Foundations and Trends® in Information Retrieval, 2016, 10(1): 1-117

[234] Shah P, Hakkani-Tur D, Liu B, et al. Bootstrapping a neural conversational agent with dialogue self-play, crowdsourcing and on-line reinforcement learning//Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. New Orleans, USA, 2018: 41-51

[235] Zhao T, Lu A, Lee K, et al. Generative encoder-decoder models for task-oriented spoken dialog systems with chatting capability//Proceedings of the SIGDIAL 2017 Conference. Saarbrucken, Germany, 2017: 27-36

[236] Luan Y, Brockett C, Dolan B, et al. Multi-task learning for speaker-role adaptation in neural conversation models//Proceedings of the 8th International Joint Conference on Natural Language Processing. Taipei, China, 2017:605-614

[237] Lipton Z, Li X, Gao J, et al. Bbq-networks: Efficient exploration in deep reinforcement learning for task-oriented dialogue systems//Proceedings of the thirty-second AAAI conference on artificial intelligence. New Orleans, USA, 2018: 5237-5244

[238] Morgenstern L, Ortiz Jr C L. The winograd schema challenge: Evaluating progress in commonsense reasoning//Proceedings of the Twenty-Ninth Conference on Artificial Intelligence. Austin, USA, 2015: 4024-4026

[239] Roemmele M, Bejan C A, Gordon A S. Choice of plausible alternatives: An evaluation of commonsense causal reasoning//Proceedings of the AAAI Spring Symposium: Logical Formalizations of Commonsense Reasoning. Stanford, USA, 2011

ZHAO Yang-Yang, born in 1995, Ph.D. candidate. Her main research interests include dialogue systems，reinforcement learning and deep reinforcement learning.

WANG Zhen-Yu, born in 1966, Ph.D., professor, Ph.D. supervisor. His main research interests include natural language processing，dialogue systems and deep reinforcement learning.

Wang Pei, born in 1993, M. S. candidate. Her main research interest is task-oriented dialogue system and dialogue management.

YANG Tian, born in 1994, Ph.D. candidate. His main research interest is task-oriented dialogue system.

ZHANG Rui, born in 1993, Ph.D. candidate. His main research interests include dialogue systems and emotional dialogue generation.

YIN Kai, born in 1995, M. S. candidate. His main research interest is task-oriented dialogue system.

Background

As the core field of artificial intelligence, the dialogue system aims to use the techniques of Automatic Speech Recognition (ASR), Natural Language Understanding (NLU), Dialogue Management (DM), Natural Language Generation (NLG)and Text To Speech (TTS) to imitate the way of humans talk and realize the information exchange between humans and computers. In the past few decades, great progress has been made in natural language processing technology and voice technology at home and abroad, which has greatly promoted the development of human-machine dialogue systems. As a result, there have been many human-machine dialogue systems with practical significance at home and abroad. At present, foreign human-machine dialogue system research has entered a relatively mature stage. In China, it was not until the 1990s that the study of human-machine dialogue system began to be attempted due to Chinese has the characteristics of language complexity, more and freely positioned function words, and flexible semantic transformation. The domestic human-machine dialogue system mostly analyzes the mechanism of the ALICE system, rewrites the structure of the ALICE system, and writes the knowledge base using the AIML language. In recent years, domestic human-machine dialogue system research has made considerable progress, such as the Benben robot of Harbin Institute of Technology and the Qrobot of Tencent, which can achieve basic dialogue.

This paper introduces the development process and division of human-machine dialogue systems, and summarizes the frontier research methods of task-oriented dialogue system, focusing on analyzing and comparing the representative algorithms of deep learning technology and traditional algorithms. Finally, Our paper discusses the issue of limiting the development of task-based dialogue systems and highlights some future trends of task-oriented dialogue system, with the hope of providing a valuable reference in its future development.

This paper is partially supported by Science and Technology Program of Guangzhou, China (No. 201802010025), University Innovation and Entrepreneurship Education Fund Project of Guangzhou (No.2019PT103), and Natural Science Foundation of Guangdong Province, China (No.2019A1515011792). These projects aim to enrich the theory of dialogue systems and develop efficient dialogue systems to expand the power and applicability of the dialogue system in industry.