

面向任务的基于深度学习的多轮对话系统与技术

姚冬¹ 李舟军² 陈舒玮² 季震¹ 张锐¹ 宋磊¹ 蓝海波¹

1 国网冀北电力有限公司 北京 100053

2 北京航空航天大学计算机学院 北京 100191

(yao.dong@jibei.sgcc.com.cn)

摘要 自然语言是人类智慧的结晶,以自然语言的形式与计算机进行交互是人们长久以来的期待。随着自然语言处理技术的发展与深度学习方法的兴起,人机对话系统成为了新的研究热点。人机对话系统按照功能可以分为任务导向型对话系统、闲聊型对话系统、问答型对话系统。任务导向型对话系统是一种典型的人机对话系统,旨在帮助用户完成某些特定的任务,有着十分重要的学术意义和应用价值。文中系统地阐述了一种在实际工程应用中的任务导向型对话系统的通用框架,主要包括自然语言理解、对话管理以及自然语言生成3个部分;介绍了上述各部分所采用的经典深度学习和机器学习方法。最后,对自然语言理解任务进行了实证性的实验验证与分析,结果表明文中内容可以为任务导向型对话系统的构建提供有效指导。

关键词: 任务导向型对话系统;自然语言理解;对话管理;自然语言生成;深度学习

中图法分类号 TP391

Task-oriented Dialogue System and Technology Based on Deep Learning

YAO Dong¹, LI Zhou-jun², CHEN Shu-wei², JI Zhen¹, ZHANG Rui¹, SONG Lei¹ and LAN Hai-bo¹

1 State Grid Jibei Electric Company Limited, Beijing 100053, China

2 School of Computer Science and Engineering, Beihang University, Beijing 100191, China

Abstract Natural language is the crystallization of human wisdom, and interacting with computers in the form of natural language has long been expected. With the development of natural language processing technology and the rise of deep learning methods, human-computer dialogue systems have become a new research hotspot. Human-computer dialogue systems can be divided into task-oriented dialogue systems, chit-chat-oriented dialogue systems, and question-and-answer dialogue systems according to their functions. The task-oriented dialogue system is a typical man-machine dialogue system, which aims to help users complete certain specific tasks, and has very important academic significance and application value. This paper systematically illustrates the general framework of task-oriented dialogue systems in practical engineering applications, including natural language understanding, dialogue management, and natural language generation. Then, the classical deep learning and machine learning methods used in the above parts are introduced. Finally, the task of natural language understanding is empirically verified and analyzed. This paper can provide effective guidance for the construction of a task-oriented dialogue system.

Keywords Task-oriented dialogue system, Natural language understanding, Dialog management, Natural language generation, Deep learning

1 引言

自然语言是人类表达情感、观念的主要工具和人类思维的重要载体,而最常见的自然语言应用场景则是对话。人们希望通过自然语言控制机器,甚至实现机器与人类的对话与交流。使用自然语言作为输入输出的媒介,会使用户获得更

自然友好的人机交互体验,从而为人类生活提供便捷服务。随着近年来人工智能技术的蓬勃发展,实现基于自然语言的对话系统逐渐成为了人工智能领域的热门研究之一,受到了工业界和学术界的高度重视和广泛关注。

对话系统按照对话的目标可以分为3类:任务导向型对话系统、闲聊型对话系统以及问答型对话系统^[1]。任务导向

到稿日期:2020-06-16 返修日期:2020-07-17

基金项目:国家自然科学基金(U1636211,61672081);软件开发环境国家重点实验室课题(SKLSDE-2019ZX-17);国网人工智能技术在调控运行全过程安全管控中的应用研究项目(520101180044)

This work was supported by the National Natural Science Foundation of China(U1636211,61672081),Fund of the State Key Laboratory of Software Development Environment(SKLSDE-2019ZX-17) and Research on the Application of Artificial Intelligence in the Safety Control of the Whole Process of Regulation and Operation in State Grid(520101180044).

通信作者:李舟军(lizj@buaa.edu.cn)

型对话系统的目标是通过自然语言交互的形式帮助用户完成某个特定的任务,例如商品导购、网上订票、智能客服等;闲聊型对话系统的主要目标是通过模拟人的聊天行为,来与用户闲聊,只需提供合理的、相关的、有趣的、个性化的回复,其主要功能是提供情感陪护;问答型对话系统旨在基于特定的知识,回答用户特定的问题,以帮助用户获取其关注的信息。

本文主要聚焦于任务导向型对话系统,而为了帮助用户完成某个特定的任务,人机之间一般需要基于给定的场景,进行多个轮次的对话(简称多轮对话)。目前,市场上推出的各种智能音箱、智能客服系统、智能手机助手等均使用了多轮对话技术,包括国内的阿里巴巴、腾讯、小米和国外的谷歌、微软等公司的智能音箱产品,我们常用的网购平台如淘宝、唯品会、网易考拉中的客服导购聊天机器人,以及苹果 Siri、微软小娜、讯飞 AIUI、百度度秘等智能助手。多轮对话技术已成为实现人机交互的重要手段,具有较大的应用价值。

当前任务导向型多轮对话系统的技术方案主要有模块级联方法和端到端方法两种^[2]。多模块级联方法也称流水线式方法,通常由多个不同功能的模块级联而成,流水线式方法往往拥有更好的可维护性和可解释性,能够对系统进行样例粒度的改进,但各个模块之间的误差容易积累。端到端方法通常由一个神经网络模型构成,可以直接由用户的输入经过模型之后得到系统的输出,理论上可以提高对话系统的鲁棒性,但是往往需要更多数据进行训练,可维护性和可解释性较差。由于可维护性和可解释性的重要性,实际产品通常采用流水线式框架,这种框架是任务导向型对话系统的一种通用框架。

流水线式任务导向型多轮对话系统通常包括自然语言理解模块、对话管理模块和自然语言生成模块。意图识别和槽填充算法是自然语言理解模块的重点,其中意图识别的目的在于提取用户的实际意图,槽填充的目的在于分析自然语言中所包含的实体信息。良好的意图识别和槽填充算法能够极大地提升多轮对话系统的整体性能。对话管理模块负责维护多轮对话的状态,产生回复策略,对系统的运转同样十分重要。自然语言生成模块负责生成自然语言回复,其生成文本的流畅性、合理性对用户体验十分重要。

本文详细地介绍了流水线式任务导向型对话系统框架,梳理了任务导向型对话系统的一种通用范式,同时阐述了在实际产品中自然语言理解、对话管理、自然语言生成的常用方法。本文第2节介绍了任务导向型对话系统的研究现状,总结了自然语言理解、对话管理、自然语言生成等相关技术的发展和现有的算法;第3节对任务导向型对话系统的各个模块的主流算法进行了详细介绍和分析;第4节在 NLPCC 2018 Shared Task4 数据集上对自然语言理解的主流算法进行了实验验证和对比,并分析了槽填充和意图识别两个子任务的关系;最后总结全文。

2 任务导向型对话系统的研究现状

目前,任务型对话系统的构建方式主要有两种:多模块级联(Pipeline)的方法和端到端的方法。基于模块间级联的任务型对话系统的设计雏形来自于 Young 发表于 2000 年的论

文^[3],通常包括语音识别(ASR)、自然语言理解(NLU)、对话管理(DM)、自然语言生成(NLG)、语音合成(TTS) 5 个模块。

在多模块级联的方法中,各模块之间需要定义相应的接口,从而明确每个子模块的输入和输出。其中,语音识别模块和语音合成模块的主要功能是进行文本和语音之间的转换,与其他模块之间的耦合很小,相对独立。而自然语言理解、对话管理、自然语言生成 3 个模块的逻辑联系则更加紧密,是任务型对话系统的核心部分。

自然语言理解模块的主要作用是对用户的输入文本进行分析,将其解析为预先定义好结构和内容的语义表达式,这些语义特征通常包括句子级别的(sentence-level)和词级别的(token-level),对应于用户的意图和提及的重要属性。

随着深度学习技术的迅速发展,传统的机器学习模型逐步被深度模型所取代。对于意图识别(分类)子任务,文献[4-5]使用卷积神经网络抽取句子的向量表示,这些特征有效提升了后续分类器的性能。对于槽填充(序列标注)子任务,深度信念网络^[6]、循环神经网络^[7-8]等模型都取得了很好的效果。目前,基于双向长短期记忆模型和条件随机场的模型在许多数据集上取得了最优结果^[9]。此外,由于意图识别和槽填充两个任务在逻辑上并非是独立的,而是互相影响的,因此同时进行意图识别和槽填充的模型^[10]也是一个重要的研究方向,文献[11]使用了联合训练的方式,在一些数据集上取得了最优结果。

对话管理模块的主要功能是在多轮对话中维护对话的历史信息和当前的状态,同时根据历史和当前状态选择特定的回复策略,这个模块一般需要与外部的知识库进行交互。

为了使对话管理模块更有泛化性,带有随机性的概率模型成为了研究主流。其中,最典型的工作^[12]是使用马尔可夫决策过程对其进行建模,通过强化学习技术^[13]学习模型参数。这些工作在一定程度上解决了基于规则的模型覆盖率低、鲁棒性差的问题。Young 等^[14]进一步提出基于部分可观测马尔可夫决策过程的方法,以应对自然语言理解模块的结果不明确的情况。

自然语言生成模块的主要功能是根据对话策略和当前的对话状态生成自然语言形式的回复。

传统的自然语言生成模块多使用基于预先定义的规则,将抽象动作映射到固定的回复中。通常的回复模板是树状的结构,树节点是由模板标记语言表示的标准回复。抽象动作包括动作种类和一些具体的槽值。根据种类确定好模板后,将对应的槽值填入模板对应的位置,从而生成标准的回复^[15-16]。

为了使回复更具多样性,基于深度学习技术的方法先后被提出。Wen 等^[17]使用循环神经网络对抽象动作进行解码(decode),之后使用卷积神经网络进行重排序(rerank)。而文献[18]则使用了注意力机制(Attention Machine)对上下文进行编码(encode),之后使用长短期记忆网络进行回复生成,达到了最佳效果。

3 多轮对话系统框架及其相关技术

任务导向型对话系统框架主要由自然语言理解、对话管

理、自然语言生成 3 部分组成。图 1 给出了一种流水线式任务导向型对话系统的通用框架。现有的任务导向型对话系统产品通常包括语音识别与语音合成的功能,以支持用户与系统进行语音交互。本文主要关注面向文本的任务导向型对话系统技术,因此只对自然语言理解、对话管理和自然语言生成进行重点介绍。

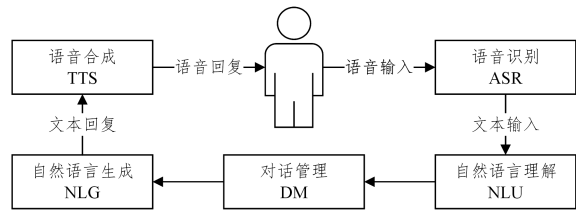


图 1 流水线式任务导向型对话系统框架

Fig. 1 Illustration of pipeline-based task-oriented dialog system framework

3.1 自然语言理解模块

自然语言理解模块主要的任务是解析用户的自然语言语句,进行用户意图识别 (User Intent Detection) 以及语义槽填充 (Semantic Slot Filling),将用户输入的自然语言信息解析为计算机能够处理的结构化数据。“意图 (Intention)”和“槽 (Slot)”通常由领域专家进行人工设计,其中用户意图是对当前对话领域的再次细分。

3.1.1 意图识别

意图指用户对话过程中所表达的真实目的,一个意图往往对应着系统能够完成的一种任务。意图识别是抽取句子级别 (sentence-level) 的特征,可以理解为一个分类问题,即一个从自然语言文本到预定义的意图列表集合的映射关系。

深度学习是目前意图识别的主流技术。图 2 给出了基于双向 LSTM 和注意力机制的意图识别模型^[19]。分词后的输入文本首先被表示成词嵌入的形式,随后由前向 LSTM 和后向 LSTM 对词嵌入表示进行进一步编码,最后将前向 LSTM 和后向 LSTM 的输出向量拼接在一起,经过注意力层后得到最终输出。

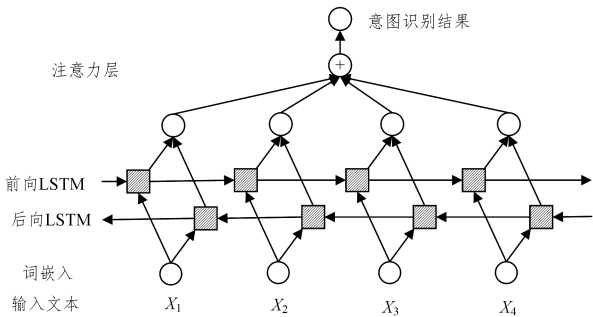


图 2 Bi-LSTM+Attention 模型

Fig. 2 Bi-LSTM with Attention mechanism

图 3 给出了基于卷积神经网络的意图识别模型——Text-CNN^[20],输入文本切分后的结果首先被转换为词向量序列,然后根据卷积核的大小将词向量序列填充到指定的长度,再进行卷积操作和最大池化操作,最后经全连接层和 Softmax 层得到最终的意图概率分布。

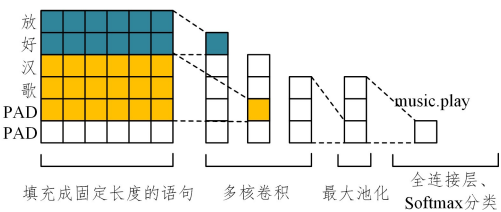


图 3 TextCNN 模型

Fig. 3 TextCNN model

3.1.2 槽填充

槽信息可以对系统的任务目标进行进一步的细化。槽填充是抽取词级别 (token-level) 的特征,通常被理解为序列标注问题,即对语句中的每个词 (token) 标注一个预先定义好范围的语义标签,在标注完成后,将语句中的词序列 (token-sequence) 映射到语义槽中。

目前常用的槽填充方法以深度学习方法为主,其中 Bi-LSTM+CRF^[21]为槽填充任务的经典模型,其模型结构如图 4 所示。分词后的输入首先在词嵌入层被转化为词向量表示,然后经过双向 LSTM 进行编码,最后经 CRF 层^[22]得到最终输出。

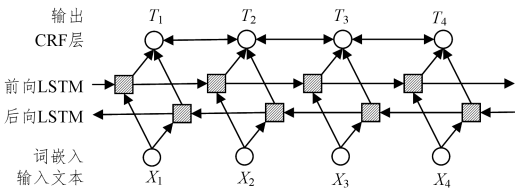


图 4 Bi-LSTM+CRF 模型

Fig. 4 Bi-LSTM with CRF layer

CRF 和 LSTM 也可以单独用于槽填充任务,即在词嵌入层之后只有 LSTM 层或 CRF 层。此外,基于卷积神经网络 (CNN) 的方法、基于序列到序列 (sequence-to-sequence)^[23]的编码-解码 (encoder-decoder) 模型以及注意力机制 (attention)^[24]等拓展的方法,均在解决序列标注问题的过程中展现出了良好的效果^[25]。

近两年,预训练技术在自然语言处理领域取得了广泛且显著的成果。特别地,BERT^[26]预训练模型使用多层双向的 Transformer^[27],创新性地提出了掩码语言模型和下一句预测的联合训练方法,在序列标注问题上大幅提高了效果。

3.1.3 联合模型

联合模型指可同时学习和完成意图识别和槽填充任务的模型。图 5 给出了一种经典的联合模型 Joint,与 Liu 等的工作类似^[28],它使用一个双向 LSTM 作为编码器来编码输入文本,使用一个解码器解码出槽填充任务的标注序列,同时直接基于编码器最终的输出来预测文本的意图。

Zhang 等^[29]提出了一种基于胶囊的神经网络模型,其动态地按协议路由方案来完成时隙填充和意图检测。通过新的路由方案使用推断的意图表示来进一步协同填充时隙。

图 5 还给出了一个输入输出样例。输入文本为“放好汉歌”,意图识别结果为“music. play”,槽填充结果为“O B-Song I-Song I-Song”。其中,music. play 表示意图为播放歌曲,槽

填充结果采用了 BIO 标注规范, Song 表示歌曲名称, 最终的歌曲名为“好汉歌”。

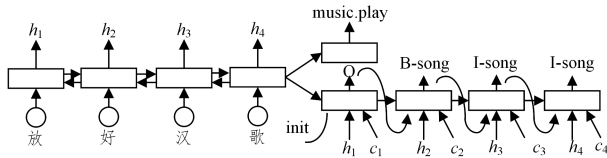


图 5 Joint 模型

Fig. 5 Joint model

3.2 对话管理模块

对话管理模块的输入为自然语言理解模块提取到的用户意图和填充好的语义槽, 该模块主要包括对话状态追踪(Dialogue State Tracking, DST)以及对话策略选择(Dialogue Policy Learning, DPL)两个主要部分。对话状态追踪部分负责维护多轮对话的状态并根据与知识库的交互结果、当前模块的输入、历史对话状态三者综合得到新的对话状态, 从而生成对话策略, 对话策略作为对话管理模块的输出, 控制回复生成模块产生对用户的回复语句。

3.2.1 对话状态追踪

对话状态追踪任务根据自然语言处理模块的输出和历史状态来决定当前的对话状态, 属于在时间序列上对状态转移进行建模的方法, 其方法主要包括基于规则的和基于统计的两种, 这两种方法在一定程度上是可以互相融合的。

该框架下基于规则的对话状态追踪方法在实际产品中较为常用。对于特定的领域, 一般需要预先定义好对话状态, 设对话的槽名共 K 个, 则第 k 个槽名包括 V_k 个槽值, 则必然有:

$$\sum_k V_k = V$$

若一共有 V 个槽值, 则有一个 V 维的向量 $\vec{v}$ 作为对话状态的表征, 其中 $\vec{v}$ 是各个槽名状态向量 $\vec{v}_k$ 的拼接。因此, 可以针对每个槽位 $\vec{v}_k$ 的转移来设定规则。对于转移规则, 一般使用基于有限状态自动机(FSM)的或者说基于规则的方法。在普遍情况下, 一个槽位的状态更新规则主要包括两个方面: 如果该槽位为排他性的槽位, 则该槽位置 1 时会将其其他槽位全部置 0; 如果该槽位为非排他性的槽位, 则该槽位置 1 时对其其他槽位无影响。根据上述规则可以给出状态转移方程, 得到各个状态的当前向量, 最后将其拼接到一起, 从而得到最终的对话状态。

基于规则的方法往往适用于冷启动和业务简单的场景, 基于统计的对话状态追踪方法则更为灵活且在大数据驱动的场景下表现更好。为了更好地描述基于统计的方法, 我们首先给出了一些形式化定义: 对于多轮对话数据, 设用户在时刻 t 的输入语句为 q_t , 系统回复为 r_t , 则用户与系统的交互过程可以表示为 $\{q_1, r_1, q_2, r_2, \dots, q_T, r_T\}$, 其中 T 为对话轮次。

对话状态追踪任务可以描述为根据当前轮的用户输入 q_t 、用户和系统的交互历史 $\{(q, \vec{s}, r)_i\}_{i=1}^{t-1}$, 来预测当前轮的对话状态 $\vec{s}_t$, 因此可以将对话状态追踪任务理解为对以下概率模型进行建模:

$$p(\vec{s}_t | q_t, (q, \vec{s}, r)_{t-1}, \dots, (q, \vec{s}, r)_1)$$

对于该条件概率的建模, 一般使用表示层、交互层、追踪层构成的层级架构。

表示层的功能是对文本进行编码, 得到文本的特征表示, 是对自然语言进行处理的第一步, 可以采用长短期记忆网络(LSTM)^[30]、门限循环单元(GRU)^[31]、预训练模型 BERT^[26]等, 其中预训练模型 BERT 可以进行较好的特征表示, 但计算量较大。

对于交互层, 可以使用多种方法, 将文本的特征转换为需要预测的对话状态。一种比较简单的方法是设置交互矩阵, 与槽值向量进行交互:

$$\vec{s} = \vec{e}MV$$

其中, $\vec{e}$ 为文本的编码结果, M 为交互矩阵, V 为槽向量矩阵。经过交互, 可以将任意维度的文本表征向量转换为目标维度的向量。

随着深度学习的发展, 深度学习方法逐渐被人们应用到对话状态跟踪任务中。Su 等^[32]提出将四元组 $\langle \text{Top User Dialogue Act}, \text{Full Belief State}, \text{System Summary Act}, \text{Turn No.} \rangle$ 构成的特征作为对话状态跟踪模型的输入, 该方法取得了很好的实验效果; 此外, Su 等还提出了经典的基于循环神经网络(RNN)和卷积神经网络(CNN)的对话状态跟踪模型。Mrksić 等^[33]提出了一个神经信念跟踪器(NBT)来预测槽值, 它根据系统历史对话动作、用户原始输入来判断一个槽值对是否出现, 这样就可以通过遍历所有槽值对来完成对话状态的更新。Chen 等^[34]提出了带有插槽连接的对话状态跟踪(DST-SC)模型, 明确考虑了跨不同域的插槽相关性, 以避免用户经常采用省略号和引用来表示其他域的槽已提及的值的现象。给定目标插槽, DST-SC 中的插槽连接机制可以推断其源插槽并直接复制源插槽值, 从而大大降低了学习和推理的难度。

3.2.2 对话策略选择

对话策略选择的基本目标是根据当前的对话状态选择合适的对话策略, 进而生成相应的回复。

与对话状态追踪任务相似, 对话策略选择部分同样分为基于规则的和基于统计的两种方法。前者的优势是性能稳定、生效快、冷启动方便、可解释性强; 后者则在数据充分的情况下能够获得更好的性能。

基于规则的方法通常是由领域专家编写相应的规则, 即在特定状态的分布上, 如何获得对话策略。基于统计的方法通常引入强化学习, 即使用支持强化学习的模型在训练过程中进行增量学习, 从而能够在实践中对模型进行更新。

Li 等^[35]提出了 Deep Q-network 来学习对话决策选择, 这个网络架构融合了当前语义框架(Current Semantic Frame)和数据库返回结果(database returned results), 取得了良好的实验结果。对话决策选择有学习缓慢、冷启动困难的问题, Su 等^[36]提出将监督学习与强化学习相结合, 先使用少量监督数据对模型进行预训练, 然后使用强化学习进行微调, 很好地解决了纯强化学习框架的弊端。

3.3 自然语言生成模块

自然语言生成模块接收对话管理模块输出的对话策略和

当前对话状态,并据此生成自然语言回复用户。

与对话管理部分类似,自然语言生成模块的主流方法包括基于规则的和基于统计的两类。基于规则的方法本质上属于模板匹配,通常使用基于模板引擎的渲染方法。基于模板引擎的方法需要在对话管理和自然语言生成模块之间定义回复策略。回复策略通常由分级的类型和相应的参数组成,分级的类型用于确定需要被渲染的模板,参数则用于渲染基础模板,生成对应的自然语言回复。一个典型的模板匹配场景如图 6 所示。

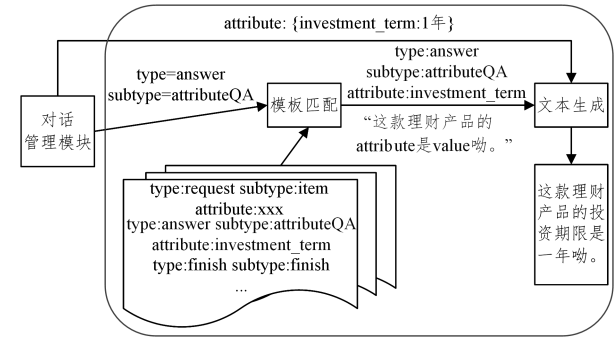


图 6 基于模板匹配的自然语言回复生成

Fig. 6 Template-based natural language response generation

基于统计的方法包括检索式和生成式两类,检索式方法主要建立一个候选回复库,根据对话管理模块输出的对话策略,在候选库中选择合适的回复,模型则对其进行建模;生成式的方法是从对话管理模块的输出直接得到目标回复。图 7 给出了基本的生成式方法的模型框架。

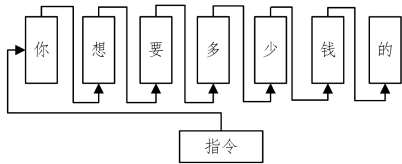


图 7 生成式自然语言回复生成

Fig. 7 Generation-based natural language response generation

近年来,基于神经网络的方法被广泛应用于自然语言生成。Wen 等^[37]使用一个正向循环神经网络生成器、一个卷积神经网络重排器与一个后向循环神经网络重排器,通过共同优化所有的子模块来产生以固定对话动作为条件的句子。Zhou 等^[38]采用基于编码器-解码器的结构,结合问题信息、语义槽值和对话动作类型来生成正确答案。它使用注意力机制来关注解码器当前解码状态的关键信息,通过编码对话动作类型,可以生成许多不同动作类型的回复答案。Li 等^[39]研究了时隙一致性,使用迭代整流网络(IRN)构建可靠的 NLG 系统,用自举算法对训练候选样本进行采样,并使用强化学习将与空位不一致相关的离散奖励纳入训练中。

4 对自然语言理解主流方法的实证性实验验证

4.1 数据集

本文使用 NLPCC 2018 Shared Task4 数据集^[40]进行实验,该数据集包含 3 个领域:音乐(music)领域、导航(navigation)领域、通话(phone)领域。其统计信息如表 1 所列。表 2 列出了训练集和测试集中的数据样例,训练集和测试集均包

括原始输入文本以及对应的意图和槽值。

表 1 数据集统计信息

Table 1 Statistics of datasets

data type	Training set	Test set
Sessions	4 705	1 177
Utterances	21 352	5 350
Average utterances persession	4.54	4.55
Average words perutterance	5.93	6.08

表 2 数据集样例

Table 2 Examples for training and test set

Example	
Training set	Text: 播放光辉岁月
	Intent: music, play
	Slots: <song> 光辉岁月 </song>
Test set	Text: 给我来一首谭咏麟的朋友
	Intent: music, play
	Slots: <singer> 谭咏麟 </singer> <song> 朋友 </song>

4.2 实验方法与参数设置

槽识别实验所使用的方法及参数设置如下。

CRF:使用腾讯 AI Lab 开源预训练词向量^[41],词向量维度为 200。

Bi-LSTM:词向量设置同上。训练批次大小(batchsize)为 25,训练轮数(epochnumber)为 50,使用 Adam 优化器,学习率为 0.001。LSTM 的隐藏单元(hiddenunits)数量为 128,dropout 概率为 0.5。

Bi-LSTM+CRF:参数设置同 Bi-LSTM。

Joint:参数设置同 Bi-LSTM 方法。

意图识别实验所使用的方法如下。

LSTM:参数设置同 Bi-LSTM 方法。

TextCNN:池化层使用最大池化,卷积核尺寸为[3,4,5],卷积核数量为 128,卷积步长为[1,1,1,1]。

Bi-LSTM+Attention:参数设置同 Bi-LSTM 方法。

Joint:同槽识别实验中的 Joint 方法。

4.3 评价指标

槽填充任务:本文使用句子级槽填充准确率(P_{slot})来评价槽填充模型。

$$P_{slot} = \frac{\text{槽填充正确的语句数}}{\text{语句总数}}$$

意图识别任务:本文使用意图识别的分类准确率(P_{intent})来评价意图识别模型。

$$P_{intent} = \frac{\text{意图识别正确的语句数}}{\text{语句总数}}$$

4.4 实验结果及分析

表 3 和表 4 分别列出了意图识别和槽填充实验的实验结果。

表 3 槽填充实验结果

Table 3 Experiment results of slot filling

Methods	P_{slot}
Bi-LSTM	0.778
CRF	0.834
Bi-LSTM-CRF	0.842
Joint	0.802

表 4 意图识别实验结果

Table 4 Experiment results of intention recognition

Methods	P_{intent}
LSTM	0.866
TextCNN	0.858
Bi-LSTM+Attention	0.865
Joint	0.874

在槽填充任务中,Bi-LSTM+CRF 达到了最好的句子级准确率;在意图识别任务中,Joint 模型达到了最好的准确率。这说明联合学习对意图识别有促进作用,但是对槽填充任务却有一定的副作用。这也表明,在实践中我们可以利用构造辅助任务的方式来提高意图识别的效果,槽填充对于意图识别来说就是一种很好的辅助任务。但意图识别对于槽填充来说并不是一种很好的辅助任务。

结束语 本文系统地介绍了基于多模块级联的任务型对话系统框架的相关技术与方法,并给出了其中自然语言理解算法部分的实验验证。在面向特定领域时,组织足够的训练数据训练模型,并根据领域特点设计对话管理和自然语言生成模块的专家规则,最终可以根据本文提供的框架实现一个任务型对话系统。

参 考 文 献

[1] ZHAO Y. Research on the Key Technologies of Customizable Knowledge-Based Chatbots [D]. Beijing: Beihang University, 2018.

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

[3] YOUNG S J. Probabilistic methods in spoken-dialogue systems [J]. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 2000, 358(1769): 1389-1402.

[4] HASHEMI H B, ASIAEE A, KRAFT R. Query intent detection using convolutional neural networks[C]// International Conference on Web Search and Data Mining, Workshop on Query Understanding. 2016.

[5] SHEN Y, HE X, GAO J, et al. Learning semantic representations using convolutional neural networks for web search[C]// Proceedings of the 23rd International Conference on World Wide Web. ACM, 2014: 373-374.

[6] DEORAS A, SARIKAYA R. Deep belief network based semantic taggers for spoken language understanding [C] // Interspeech. 2013: 2713-2717.

[7] MESNIL G, HE X, DENG L, et al. Investigation of recurrent-neural-network architectures and learning methods for spoken language understanding[C]// Interspeech. 2013: 3771-3775.

[8] YAO K, ZWEIG G, HWANG M Y, et al. Recurrent neural networks for language understanding[C]// Interspeech. 2013: 2524-2528.

[9] HOCHREITER S, SCHMIDHUBER J. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8): 1735-1780.

[10] HAKKANI-TÜR D, TÜR G, CELIKYILMAZ A, et al. Multi-Domain Joint Semantic Frame Parsing Using Bi-Directional

RNN-LSTM[C]// Interspeech. 2016: 715-719.

[11] GOO C W, GAO G, HSU Y K, et al. Slot-Gated Modeling for Joint Slot Filling and Intent Prediction[C]// Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2018: 753-757.

[12] LEVIN E, PIERACCINI R, ECKERT W. A stochastic model of human-machine interaction for learning dialog strategies[J]. IEEE Transactions on Speech and Audio Processing, 2000, 8(1): 11-23.

[13] SINGH S, LITMAN D, KEARNS M, et al. Optimizing dialogue management with reinforcement learning: Experiments with the NJFun system[J]. Journal of Artificial Intelligence Research, 2002, 16: 105-133.

[14] YOUNG S, GAŠIĆ M, KEIZER S, et al. The hidden information state model: A practical framework for POMDP-based spoken dialogue management[J]. Computer Speech & Language, 2010, 24(2): 150-174.

[15] WALKER M A, RAMBOW O C, ROGATI M. Training a sentence planner for spoken dialogue using boosting[J]. Computer Speech & Language, 2002, 16(3/4): 409-433.

[16] STENT A, PRASAD R, WALKER M. Trainable sentence planning for complex information presentation in spoken dialog systems[C]// Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2004: 79.

[17] WEN T H, GASIC M, KIM D, et al. Stochastic language generation in dialogue using recurrent neural networks with convolutional sentence reranking[J]. arXiv:1508.01755, 2015.

[18] ZHOU H, HUANG M. Context-aware natural language generation for spoken dialogue systems[C]// Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers. 2016: 2032-2041.

[19] ZHOU P, SHI W, TIAN J, et al. Attention-based bidirectional long short-term memory networks for relation classification [C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. 2016: 207-212.

[20] KIM Y. Convolutional Neural Networks for Sentence Classification[C] // Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014: 1746-1751.

[21] HUANG Z, XU W, YU K. Bidirectional LSTM-CRF models for sequence tagging[J]. arXiv:1508.01991, 2015.

[22] LAFFERTY J D, MCCALLUM A, PEREIRA F C N. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data[J]. Proceedings of Icml, 2001, 3(2): 282-289.

[23] KALCHBRENNER N, BLUNSOM P. Recurrent continuous translation models[C]// Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. 2013: 1700-1709.

[24] LUONG T, PHAM H, MANNING C D. Effective Approaches to Attention-based Neural Machine Translation[C] // Procee-

dings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015;1412-1421.

[25] ZHU S, YU K. Encoder-decoder with focus-mechanism for sequence labelling based apoken language understanding [C] // 2017 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP). IEEE, 2017;5675-5679.

[26] DEVLIN J, CHANG M W, LEE K, et al. BERT;Pre-training of Deep Bidirectional Transformers for Language Understanding [C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2019;4171-4186.

[27] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Advances in Neural Information Processing Systems. 2017;5998-6008.

[28] LIU B, LANE I. Attention-Based Recurrent Neural Network Models for Joint Intent Detection and Slot Filling[C]// Inter-speech 2016. 2016.

[29] ZHANG C W, LI Y L, DU N, et al. Joint Slot Filling and Intent Detection via Capsule Neural Networks[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019;5259-5267.

[30] HOCHREITER S, SCHMIDHUBE R, JÜRGE N. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8):1735-1780.

[31] CHO K, VAN MERRIËNBOER B, GULCEHRE C, et al. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation [C] // Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014;1724-1734.

[32] SU P H, VANDYKE D, GASIC M, et al. Learning from Real Users;Rating Dialogue Success with Nerual Networks for Reinforcement Learning in Spoken Dialogue Systems[OL]. <https://www.oalib.com/paper/4050158#.YFwoVPliras>.

[33] MRKŠIĆ N, ÓSEAGHDHA D, WEN T H, et al. Neural belief tracker;Data-driven dialogue state tracking[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver, Canada, Association for Computational Linguistics, 2017;1777-1788.

[34] Y O, CHEN M, DAI X Y, et al. Dialogue State Tracking with Explicit Slot Connection Modeling[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020;34-40.

[35] LI X, CHEN Y N, LI L, et al. End-to-End Task-Completion Neural Dialogue Systems[J]. arXiv;1703.01008.

[36] SU P H, BUDZIANOWSKI P, ULTES S, et al. Sample-efficient Actor-Critic Reinforcement Learning with Supervised Data for Dialogue Management[J]. arXiv;1707.00130.

[37] WEN T H, GASIC M, KIM D, et al. Stochastic language generation in dialogue using recurrent neural networks with convolutional sentence reranking[C]// Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue. Prague, Czech Republic, Association for Computational Linguistics, 2015;275-284.

[38] ZHOU H, HUANG M, ZHU X. Context-aware natural language generation for spoken dialogue systems[C]// COLING. 2016;2032-2041.

[39] LI Y M, YAO K S, QINL B, et al. Slot-consistent NLG for Task-oriented Dialogue Systems with Iterative Rectification Network[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020;97-106.

[40] ZHAO X, CAO Y. Overview of the NLPCC 2018 Shared Task: Spoken Language Understanding in Task-Oriented Dialog Systems[OL]. <http://tcci.ccf.org.cn/conference/2018/papers/EV33.pdf>.

[41] SONG Y, SHI S, LI J, et al. Directional Skip-Gram; Explicitly Distinguishing Left and Right Context for Word Embedding [OL]. https://www.researchgate.net/publication/325447242_Directional_Skip-Gram_Explicitly_Distinguishing_Left_and_Right_Context_for_Word_Embeddings. 2018.



YAO Dong, born in 1985, postgraduate. His main research interests include power system dispatching and control operation, application of artificial intelligence in power system dispatching and control operation.



LI Zhou-jun, born in 1963, Ph.D, professor, is a member of China Computer Federation. His main research interests include data mining, natural language processing, network and information security.