

显式知识注入的任务型对话理解模型

李帅鹏¹, 王平辉¹, 孙望淳¹, 杨 阳¹, 杜友田¹, 马小科², 杜永杰³

(1. 西安交通大学智能网络与网络安全教育部重点实验室, 西安 710049; 2. 西安电子科技大学计算机科学与技术学院, 西安 710049; 3. 青岛海尔科技有限公司数字家庭网络国家工程研究中心, 青岛 266000)

摘要: 传统对话理解模型依赖对话历史识别用户意图, 由于缺乏丰富的知识信息, 对生僻或特有内容的理解能力欠佳。通过隐式编码将知识加入模型的方法将知识注入与模型训练高度绑定, 难以适应知识库的更新迭代, 也会导致知识噪声, 引入无关知识破坏原有语义。为解决上述问题, 本文提出一种显式知识注入的多任务学习对话理解模型。将知识以自然语言形式插入到对话文本中, 即插即用, 满足知识源动态发展的需要; 通过对话理解的主任务, 关联知识识别的辅助任务, 进行多任务学习, 减少知识噪声。实验结果表明, 与现有方法相比, 本文提出的模型在意图识别和语义槽填充任务上的宏 F_1 值分别提升了4.87%和2.09%。

关键词: 任务型对话理解; 知识库; 多任务学习

中图分类号: TP391 **文献标志码:** A

Task-Oriented Dialogue Understanding with Explicit Knowledge Injection

LI Shuaipeng¹, WANG Pinghui¹, SUN Wangchun¹, YANG Yang¹, DU Youtian¹,
MA Xiaoke², DU Yongjie³

(1. Ministry of Education Key Laboratory for Intelligent Networks and Network Security, Xi'an Jiaotong University, Xi'an 710049, China; 2. School of Computer Science and Technology, Xidian University, Xi'an 710049, China; 3. Digital Home Network National Engineering Laboratory, Haier Group, Qingdao 266000, China)

Abstract: Dialogue understanding aims to detect user intent given dialogue history. Due to the lack of domain knowledge, traditional dialogue understanding models fail to understand domain-specific entities. Knowledge-enhanced approaches are proposed to improve model performance with structured knowledge, where the knowledge is implicitly injected with knowledge embeddings. However, knowledge embeddings have to be updated with the update of the knowledge base, which brings extra costs. Besides, existing methods suffer from the knowledge noise and incorporate the context-irrelevant knowledge that changes the semantics of the utterance. To address the above issues, this paper proposes a multi-task learning dialogue understanding model with explicit knowledge injection (K-CAM). K-CAM injects knowledge into the model using natural language knowledge without retraining the model for updated knowledge embeddings. A multi-task learning objective of joint intent detection, slot filling, and relevant knowledge recognition is further proposed to resist the knowledge noise problem. Extensive experimental results show that the proposed model K-CAM achieves a significant improvement of 4.87% and 2.09% in macro F_1 on the intent

detection and slot filling tasks compared to other baselines.

Key words: task-oriented dialogue understanding; knowledge base; multi-task learning

引言

任务型对话系统(Task-oriented dialogue system, TOD)通过人机交互协助用户完成预约、搜索等事务,具有巨大商业价值,是当今热点研究方向之一^[1]。其中,对话理解旨在识别用户意图和对话内容,是任务型对话系统中的关键模块。如图1所示,任务型对话理解对用户当前的对话语句进行识别解析,将其转换成机器可以理解的语义框架。一般而言,该过程主要包括对话意图识别和语义槽填充两个任务^[2]。在传统方法中,这两个任务被单独建模实现^[3-4],忽略了意图识别和语义槽填充任务间的强相关性,难以充分利用所有的有监督信息。因此,一些研究工作^[5-7]提出了意图识别和语义槽填充的联合训练方式,并通过结合对话历史信息^[8]提升模型的效果。

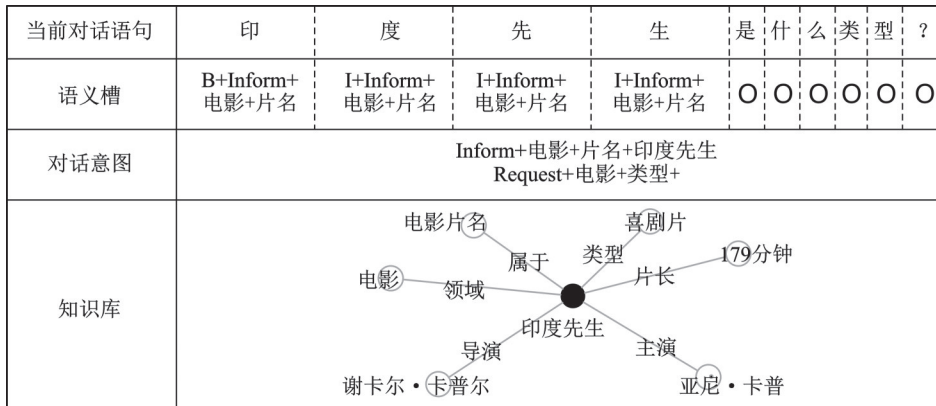


图1 对话理解任务示例

Fig.1 Example of dialogue comprehension tasks

传统任务型对话理解模型只依赖对话内容进行分析,缺乏更加丰富的知识背景信息,导致很多时候模型可能并不清楚用户在问或者说什么,尤其在垂直领域难以理解用户意图,造成对话乃至问答的效果不佳。如图1中的当前对话语句“印度先生是什么类型?”,如果不增加“(印度先生,属于,电影片名)”等领域知识对“印度先生”这一实体进行补充说明,模型就无法得知这句话是在询问一部电影的类型,将直接影响对话理解的准确性。因此,有效地融合知识信息对提升任务型对话理解的效果起到了重要的作用。为此,有研究工作^[9-11]提出结合注意力机制,隐式地将知识编码加入对话理解模型训练过程中,改善对话理解的效果。但此类方法将知识注入与模型训练高度绑定,面对新的知识,需要重新训练模型来获取知识编码,难以适应知识库的更新迭代。同时,也会导致知识噪声,引入与对话内容无关的知识破坏原有语义。

针对上述问题,本文将知识以自然语言形式插入到对话文本中,即插即用,满足知识源动态发展的需要;此外,通过对话理解的主任务,关联知识识别的辅助任务,进行多任务学习,减少知识噪声,提出一种显式知识注入的多任务学习对话理解模型(Knowledge-context augmented multi-task learning, K-CAM)。具体地,该模型分为知识注入模块、对话编码模块和多任务学习模块。首先,知识注入模块对匹配的知识三元组进行筛选,避免插入的知识三元组数量过多而破坏原始语义。受基于知识图谱的语言表示模型K-BERT^[12]的启发,将筛选后的知识三元组以自然语言的形式显式插入到原句对应实体

的位置以构建知识-句子树。其次,对话编码模块分别对知识-句子树和对话历史进行编码,并将两部分的编码向量表示进行拼接,得到融合知识和对话历史的对话文本表征。最后,多任务学习模块通过意图识别和语义槽填充的主任务,关联知识识别的辅助任务,进行多任务联合学习,区分原始语句和被插入的知识,最终实现对当前对话语义的准确捕获。本文在目前规模最大的中文任务型对话数据集 RiSAWOZ^[13]上进行了实验。结果表明,本文提出的方法与基线方法相比,在意图识别和语义槽填充任务上的宏 F_1 值分别显著提升了4.87%和2.09%。

1 相关工作

随着科学技术的发展和现实需求中训练数据的不断扩展,深度学习方法在对话理解领域被广泛应用。任务型对话理解主要包括意图识别和语义槽填充两个任务。早期研究主要是对以上两个任务单独建模。Yao等^[3-4]利用循环神经网络(Recurrent neural network, RNN)及其变体强大的序列建模能力对对话语句进行建模后,再分别完成这两个任务。此外,语义槽填充任务经常使用条件随机场(Conditional random field, CRF)^[14]与RNN相结合的方法以提升模型性能。但由于单独建模的方法阻碍了两个任务间的信息共享,一些研究工作提出了意图识别和语义槽填充的联合模型。Liu等^[5]提出一种基于注意力的序列到序列(Sequence to sequence, Seq2Seq)模型,用于联合训练。Goo等^[6]提出一种语义槽门控结构,学习意图和语义槽向量之间的关系,并通过联合优化获得更好的理解效果。Chen等^[7]提出一个基于大规模预训练模型BERT^[15]的联合意图识别和语义槽填充模型。

然而,这类对话理解模型由于缺乏相关的知识,在理解对话文本中的生僻或特有内容时常常出现困难。为此,Wang等^[9]提出一种对话理解模型,基于TransE^[16]算法获取知识表征,计算对话历史编码与知识编码的注意力,隐式融合知识。Hao等^[10]基于BERT编码知识,并通过注意力机制融合对话历史表征、外部知识表征和句子表征,用于意图识别和语义槽填充的联合训练。Wu等^[11]在此基础上,进一步使用门控机制过滤不相关的知识三元组。此类方法基于知识的隐式编码,每次更新知识库,都需要重新训练对应的知识表征,拓展性差,训练成本高。

有研究工作尝试将知识显式加入到预训练语言模型中,避免对知识的重复编码,提升预训练语言模型对特有内容的理解能力。Liu等^[12]提出一种知识增强预训练语言模型K-BERT,将原始语句中的实体所匹配的知识三元组与对应实体相连接,语句序列增广为树结构,再通过为这种树结构设置特定的位置编码,树结构的输入可以重新被作为序列输入到预训练模型中。然而,这种方法未对知识进一步筛选,过多的知识会改变原始语义。同时,显式注入知识的方式会增加原始语句的长度,导致语义槽标签改变,混淆插入知识和原始语句,带来知识噪声,因此这种方法无法被直接应用到任务型对话理解问题上。

本文从融合知识的方式入手,将知识以自然语言形式显式插入到对话文本中,实现了知识的可插拔,以适应知识库的改变。同时,基于多任务学习策略,充分利用对话理解两大任务间的共享信息,并添加辅助任务来区分被插入的知识和原始语句,减少知识噪声。

2 模型设计

本文提出的显式知识注入的多任务学习对话理解模型K-CAM包括知识注入、对话编码和多任务学习3个模块,如图2所示。其中,知识注入模块首先通过随机知识筛选限制候选知识数量,防止注入过多知识而分割原始语义,并保证了注入知识的多样性。然后,将知识以自然语言的形式插入到原始句子中,构建知识-句子树,实现对相关知识的显式融合;对话编码模块分别编码知识-句子树和对话历史,并将两部分的向量表示进行拼接,形成融合知识和对话历史的当前对话表示;多任务学习模块基于

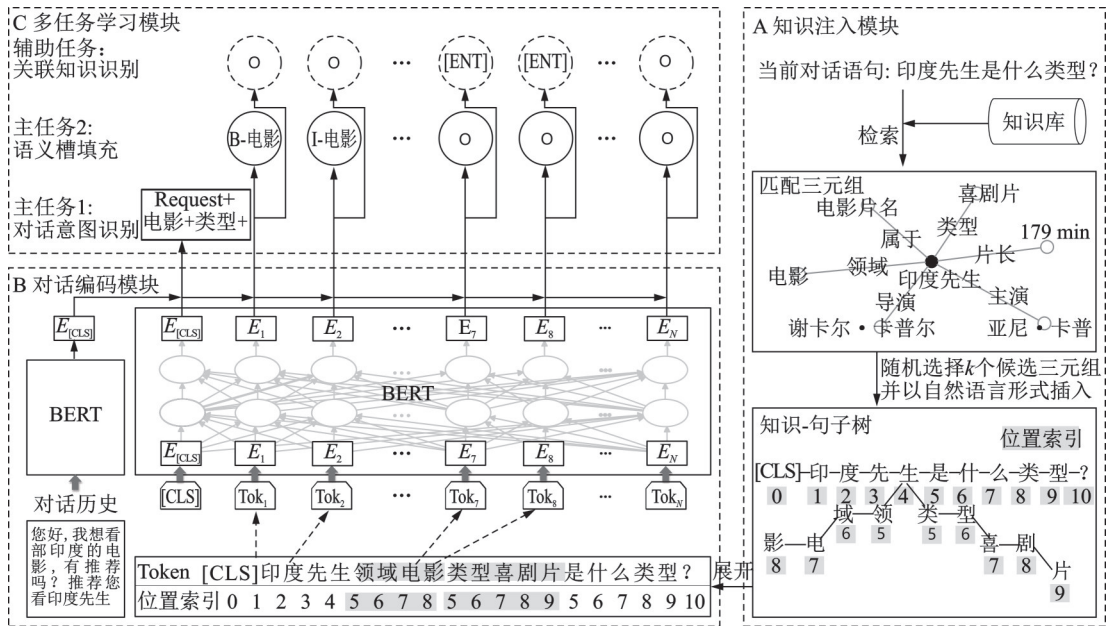


图2 K-CAM模型结构图

Fig.2 Framework of K-CAM model

当前对话的融合表示,通过意图识别和语义槽填充任务的联合训练,关联知识识别任务的辅助学习,最终实现对当前对话语义的精准捕获。

2.1 知识注入模块

知识注入模块用于给当前对话语句加入合适的知识,提升模型对于生僻或特有内容的理解能力。受知识增强预训练模型K-BERT^[12]的启发,本文将相关知识作为自然语言显式注入到当前对话文本中。不同的是,本文对匹配知识进一步筛选,以限制知识的数量,防止加入知识过多带来噪声。知识注入模块可以进一步分为随机知识筛选和知识-句子树构建两个环节,如图2 A模块所示。

2.1.1 随机知识筛选

对于当前对话语句 u_t , 在领域知识库 KB 中检索出 u_t 中的实体 w_i^t 匹配的所有知识三元组 $E_{all} = [(w_i^t, r_{i0}^t, w_{i0}^t), (w_i^t, r_{i1}^t, w_{i1}^t), \dots, (w_i^t, r_{ik}^t, w_{ik}^t)]$ 。由于 E_{all} 中不是每一个三元组都与该轮对话相关, 本文在 E_{all} 中随机选择 k 个知识作为候选三元组 E , 有

$$E = \text{random_select_k}(E_{all}) \tag{1}$$

2.1.2 知识-句子树构建

为将知识以自然语言形式显式注入到对话语句中, 本文将候选三元组与当前对话语句中的实体连接, 形成如图2 A模块所示的知识-句子树形式, 表达式为

$$\text{SentTree}_t = [w_0^t, \dots, w_i^t[(r_{i0}, w_{i0}), (r_{i1}, w_{i1}), \dots, (r_{ik}, w_{ik})], \dots, w_j^t] \tag{2}$$

同时, 为每一个词元添加特定的位置索引, 保存知识-句子树的结构信息, 记为 P 。

2.2 对话编码模块

对话编码模块用于获取融合知识和对话历史的当前对话表征, 为多任务学习模块提供基础特征向量。该模块选择预训练模型BERT^[15]作为基础编码器, 包括知识-句子树编码和对话历史编码两部分, 如图2 B模块所示。

2.2.1 知识-句子树编码

当前对话语句经过知识注入模块后,变为添加知识的句子树形式,因此需要将树结构转换为可供BERT输入的序列形式。知识-句子树展开形式如图2 B模块中的示例,其中起关键作用的是保留了知识-句子树结构信息的位置索引。然而,将知识作为分支直接加入到句子中作为自然语言的一部分,会分割原始语义,带来知识噪声。因此,需要构建合理的注意力矩阵,让分支上的知识只能影响关联它们的实体,而不会影响原句中的其他词语。本文首先构建可视矩阵 M ,表示关联知识和原句之间的影响关系。可视矩阵 M 的元素 M_{ij} 定义为

$$M_{ij} = \begin{cases} 0 & w_i \oplus w_j \\ -\infty & w_i \otimes w_j \end{cases} \quad (3)$$

式中: $w_i \oplus w_j$ 表示 w_i 和 w_j 在同一个分支上; $w_i \otimes w_j$ 表示 w_i 和 w_j 不在同一个分支上。

其次,为了充分利用 M 中的句子树结构可视关系,将 M 添加到BERT的自注意力机制中,其计算过程为

$$Q^{i+1}, K^{i+1}, V^{i+1} = h^i W_q, h^i W_k, h^i W_v \quad (4)$$

$$S^{i+1} = \text{softmax} \left(\frac{Q^{i+1} (K^{i+1})^T + M}{\sqrt{d_k}} \right) \quad (5)$$

$$h_{i+1} = S^{i+1} V^{i+1} \quad (6)$$

式中: Q, K, V 分别为查询矩阵、关键值矩阵和值矩阵; S 为影响力因子; W_q, W_k 和 W_v 为可训练的模型参数; h^i 是第 i 个掩码-自注意力块的隐藏状态; d_k 为比例因子。

对于知识注入模块得到的知识-句子树 SentTree_t ,结合定义的位置编码 P 和可视矩阵 M ,输入BERT中可以得到输出为

$$H_{\text{knowSent}} = \text{BERT}(\text{SentTree}_t, P, M) \quad (7)$$

式中: $H_{\text{knowSent}} = \{h_{s,0}, h_{s,1}, \dots, h_{s,l+k}\}$ 表示知识增强的当前对话表征,其中 $h_{s,i}$ 表示第 i 个词元对应的隐向量表示, l 表示原句的长度, k 表示插入知识的长度,因此 $l+k$ 为知识-句子树展开后的序列长度。

2.2.2 对话历史编码

将对话历史 $C = [u_1, u_2, \dots, u_{l-1}]$ 中的每一句话拼接起来,并在第1个词元前插入用于分类的特殊令牌[CLS],得到更新后的对话历史文本为 $C = [[\text{CLS}], w_{1,1}, w_{1,2}, \dots, w_{1,n}, \dots, w_{i,j}, \dots, w_{l-1,m}]$ 。然后,基于BERT编码,可以得到输出

$$H_{\text{context}} = \text{BERT}(C) \quad (8)$$

式中: $H_{\text{context}} = \{h_{c,0}, h_{c,1}, \dots, h_{c,cl}\}$ 表示对话历史的向量表征,其中 cl 表示对话历史的长度。本文取第1个特殊令牌[CLS]的隐藏状态 $h_{c,0}$ 作为对话历史的整体向量表示。

2.3 多任务学习模块

多任务学习模块基于融合知识和对话历史的当前对话表示,训练对话理解模型。意图识别和语义槽填充是对话理解的两个基本任务。由于这两个任务的相互依赖性,使用统一的模型进行联合训练可以学到更好、更丰富的表征,从而同时提升两个任务的效果。此外,为了防止关联的知识占据主要语义,更好地发挥出知识的辅助理解功能,本文增加关联知识识别任务,与对话意图识别和语义槽填充两个主任务一起进行联合训练,最终实现对当前对话语义的准确理解,如图2 C模块所示。该模块首先将对话编码模块的输出,即知识增强对话表示 $H_{\text{knowSent}} = \{h_{s,0}, h_{s,1}, \dots, h_{s,l+k}\}$ 和对话历史表示 $h_{c,0}$ 进行拼接,获取融合知识和对话历史的当前对话表示。其次,结合不同任务进行多任务学习,提高对话理解的准确性。

2.3.1 意图识别任务

意图识别任务旨在判断当前对话的目的,本文将其定义为一个多标签分类问题。将知识增强的对话表示 $H_{\text{knowSent}} = \{h_{s,0}, h_{s,1}, \dots, h_{s,l+k}\}$ 中特殊令牌[CLS]对应的编码 $h_{s,0}$ 和对话历史编码 $h_{c,0}$ 进行拼接,得到融合知识和对话历史的当前对话整体表示 $h_{\text{all},0}$ 。意图识别任务具体过程表示为

$$h_{\text{all},0} = \text{concat}(h_{s,0}, h_{c,0}) \tag{9}$$

$$\hat{y}^{\text{intent}} = \text{sigmoid}(W^{\text{intent}}h_{\text{all},0} + b^{\text{intent}}) \tag{10}$$

式中: \hat{y}^{intent} 表示当前对话语句在意图标签上的概率分布; W^{intent} 为可训练的模型参数。

2.3.2 语义槽填充任务

语义槽填充任务用于标记对话文本中有意义的内容,本文将其定义为1个序列标注问题。由于本文将知识显式注入到对话语句中,当前对话语句内容发生改变,因此需要在语义槽填充任务中给属于关联知识的词元增加标签“○”,区别于原句中的实体槽位,如图3中[主任务1:新语义槽标签]所示。首先,将对话历史编码 $h_{c,0}$ 和知识增强表示 $H_{\text{knowSent}} = \{h_{s,0}, h_{s,1}, \dots, h_{s,l+k}\}$ 中每个词元对应的编码 $h_{s,i}$ 进行拼接,得到每个词元融合知识和对话历史的表示 $h_{\text{all}} = \{h_{\text{all},0}, h_{\text{all},1}, \dots, h_{\text{all},l+k}\}$ 。其次,在语义槽标签上进行多分类,具体过程如下

$$h_{\text{all},i} = \text{concat}(h_{s,i}, h_{c,0}) \tag{11}$$

$$\hat{y}_i^{\text{slot}} = \text{softmax}(W^{\text{slot}}h_{\text{all},i} + b^{\text{slot}}) \tag{12}$$

式中: \hat{y}_i^{slot} 表示第*i*个词元在语义槽标签上的概率分布; W^{slot} 为可训练的模型参数; $i \in \{1, 2, \dots, l+k\}$ 。

2.3.3 关联知识识别任务

关联知识识别任务旨在区分原句和被插入的知识,本文将其定义为一个序列标注问题。对属于关联知识的词元,设定其标签为“[ENT]”,对于其他词元,设定其标签为“○”,如图3中[辅助任务:关联知识标签]所示。与语义槽填充任务类似,将每个词元对应的融合表示 $h_{\text{all},i}$ 在关联知识标签上进行分类,即

$$\hat{y}_i^{\text{entity}} = \text{softmax}(W^{\text{entity}}h_{\text{all},i} + b^{\text{entity}}) \tag{13}$$

式中: $\hat{y}_i^{\text{entity}}$ 表示第*i*个词元在标签“○”和“[ENT]”上的概率分布; W^{entity} 为可训练的模型参数。

知识句子树																				
句子树展开	印	度	先	生	领	域	电	影	类	型	喜	剧	片	是	什	么	类	型	?	
主任务1: 新语义槽标签	B+Inform+	I+Inform+	I+Inform+	I+Inform+	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
辅助任务: 关联知识标签	○	○	○	○	[ENT]	[ENT]	[ENT]	[ENT]	[ENT]	[ENT]	[ENT]	[ENT]	[ENT]	[ENT]	[ENT]	[ENT]	[ENT]	[ENT]	[ENT]	[ENT]
主任务2: 对话意图标签	Inform+电影+片名印度先生 Request+电影+类型+																			

图3 多任务学习标签示例
Fig.3 Example of multi-task learning labels

2.3.4 多任务学习

由于3个任务都属于分类问题,本文采用交叉熵损失函数作为3个任务的目标函数,即

$$L_{\text{intent}} = - \sum_{i=1}^{na} y_i^a \times \log(\hat{y}_i^a) \tag{14}$$

式中: y_i^a 为当前问句在意图识别任务上的标签; \hat{y}_i^a 为当前问句在意图标签上的概率分布; na 为意图识别任务中的问句数量。

$$L_{\text{slot}} = - \sum_{j=1}^{l+k} \sum_{i=1}^{nb} y_i^{bj} \times \log(\hat{y}_i^{bj}) \quad (15)$$

式中: y_i^{bj} 为当前问句中第 **bj** 个词元的语义槽标签; \hat{y}_i^{bj} 为当前问句第 **bj** 个词元在语义槽标签上的概率分布; nb 为问句实体识别任务中的问句数量。

$$L_{\text{entity}} = - \sum_{j=1}^{l+k} \sum_{i=1}^{ne} y_i^{ej} \times \log_2(\hat{y}_i^{ej}) \quad (16)$$

式中: y_i^{ej} 为当前问句中第 **ej** 个词元的关联知识标签; \hat{y}_i^{ej} 为当前问句中第 **ej** 个词元在关联知识标签上的概率分布; ne 为关联知识识别任务中的问句数量。

模型的损失 L_{all} 由这3部分损失共同组成,即

$$L_{\text{all}} = L_{\text{intent}} + L_{\text{slot}} + \lambda L_{\text{entity}} \quad (17)$$

式中 λ 为调和系数,用来调和意图识别和语义槽填充两个主任务和辅助任务关联知识识别之间的权重。

3 实验结果与分析

3.1 实验数据集

实验数据来自目前最大的中文任务型对话数据集 RiSAWOZ^[13],共包含 1.12 万组人-人 (Human to human, H2H)多轮语义注释对话,超过 15 万轮次,跨越 12 个领域。单领域对话和多领域对话分别占 65% 和 35%。本文基于该数据集的知识库,构建了 8 790 条知识三元组,用以辅助对话理解。RiSAWOZ 数据集说明如表 1 所示。

3.2 实验设置和评估指标

本文使用下列基线模型进行对比实验:

(1) Joint BERT^[7]:基于 BERT 的意图识别和语义槽填充联合训练对话理解模型。

(2) K-BERT^[12]:基于 BERT 的显式知识融合预训练模型。

(3) KnowBERT 模型:分别基于 BERT 获得问句表示和关联知识表示,并将两种表示进行拼接,用以完成问句解析的任务。

(4) ERNIE 模型^[17]:基于 BERT 模型进行优化,在预训练阶段增加了外部的知识,在各项中文 NLP 任务上取得了良好的结果。

(5) K-CAM:本文提出的融合知识和对话历史的多任务学习对话理解模型,显式注入知识获取增强表示后,通过多任务学习减少知识噪声。

具体实验中,本文采用 AdamW^[18]算法作为优化器,设置学习率为 $3e-5$,Dropout 比例为 0.1,批处理大小为 32,并对 BERT 的参数进行微调。本文所采用的评价指标包括宏平均精度 Macro-P (Macro averaged precision),宏平均召回率 Macro-R (Macro averaged recall)和宏平均 F_1 值 Macro- F_1 (Macro averaged F_1)。

3.3 实验结果与分析

在相同的实验数据集、实验设置和评估指标下,模型对比实验结果如表 2 所示。由实验结果可以看

表 1 RiSAWOZ 数据集说明

Table 1 Description of the RiSAWOZ dataset

类别	集合	对话 轮次	对话意图 标签种类/个	语义槽 标签种类/个	知识三元 组/条
单领域	训练集	79 062			8 490
	验证集	4 742	499	445	
	测试集	4 814			
多领域	训练集	55 518			
	验证集	3 374	377	276	
	测试集	4 472			

出,无论是在单领域数据上还是在多领域数据上,本文提出的K-CAM模型在对话意图识别和语义槽填充两个任务上的3项评价指标均高于基线模型,有效提升了任务导向型对话理解的精度。实验结果表明,本文提出的知识注入模块通过三元组的注入,将知识以自然语言的形式插入对话文本中,能够有效提升模型的对任务型对话意图的识别能力;同时知识可即插即用,满足知识源动态发展的需要。在意图识别和语义槽填充两个任务的宏平均 F_1 值在单领域数据上分别提升了4.87%和2.09%,在多领域数据上分别提升了3.90%和1.42%。相比于不添加知识的Joint BERT模型,K-CAM模型通过构建句子树显式注入知识,增强了模型对特有内容的理解能力,适应知识源的动态更新。与同样采用显式添加知识方法的预训练模型K-BERT相比,K-CAM模型更适合任务型对话理解场景,首先对关联知识进行筛选,限制三元组数量,从而减少不相关三元组注入的知识过多而破坏原始语义,并保证了添加知识的多样性;其次,针对对话理解场景,更新语义槽标签,并设置关联知识识别的辅助任务区分插入的知识,有效减少了知识噪声。

表2 不同模型对比结果

Table 2 Comparison results of different models

任务	模型	单领域			多领域			%
		Macro-P	Macro-R	Macro- F_1	Macro-P	Macro-R	Macro- F_1	
意图识别	Joint BERT	84.82	88.25	86.50	80.38	84.62	82.45	
	K-BERT	84.93	89.52	87.17	80.64	85.48	82.99	
	Know-BERT	84.40	87.00	85.68	81.44	82.60	82.20	
	ERNIE	84.74	90.52	87.54	80.60	85.82	83.13	
	K-CAM	92.45 (7.52↑)	91.63 (2.11↑)	92.04 (4.87↑)	87.66 (7.02↑)	86.14 (0.66↑)	86.89 (3.90↑)	
语义槽填充	Joint BERT	88.52	69.42	77.81	90.21	72.92	80.65	
	K-BERT	88.64	68.56	77.25	89.66	72.13	79.95	
	Know-BERT	87.65	66.48	75.61	86.92	67.51	76.00	
	ERNIE	88.83	69.96	78.27	89.41	72.18	79.88	
	K-CAM	94.58 (5.94↑)	70.86 (1.44↑)	79.90 (2.09↑)	92.21 (2.00↑)	73.95 (1.03↑)	82.07 (1.42↑)	

本文对K-CAM模型进行消融实验来进一步说明模型中各个模块的作用。消融实验通过单独去除知识注入模块、对话历史编码模块、关联知识识别模块来验证各个模块的有效性。具体地,为了验证注入知识的有效性,本文去除知识注入模块(w/o know),将当前对话语句直接作为对话编码模块的输入进行了实验;为了验证添加对话历史信息的有效性,本文去除对话历史编码部分(w/o context)进行了实验;为了验证关联知识识别辅助任务有效减少了知识噪声,本文构建了一个去除该任务的模型(w/o entity task),只进行意图识别和语义槽填充两个任务的联合训练。最后,如果以上3部分均不考虑(w/o all),即不注入知识,不融合对话历史,也不添加辅助学习任务,模型则与Joint BERT模型一致。

消融实验结果如表3所示,可以看出无论是意图识别还是语义槽填充任务上,单独去除知识注入、对话历史融合以及添加辅助任务的多任务学习策略实验宏平均精度、宏平均召回率、宏平均 F_1 值均有所下降,进一步证明本文所提模型的3个模块策略可以有效提高任务型对话理解的精度。值得注意的是,针对多领域数据,在意图识别任务上,去除知识注入模块(w/o know)后的模型的宏平均精度值略高于本文提出的模型K-CAM,但其宏平均召回率远低于K-CAM,能综合反映以上两个指标的宏平均 F_1 值也低于K-CAM。因此,不论是在单领域还是多领域,本文所提模型都能够很好地提升任务型对话理解的精度,同时也证明了K-CAM各个模块策略的有效性。

表3 消融实验结果

Table 3 Results of ablation experiments

%

类别	模型	单领域			多领域		
		Macro-P	Macro-R	Macro- F_1	Macro-P	Macro-R	Macro- F_1
意图识别	K-CAM	92.45	91.63	92.04	87.66	86.14	86.89
	w/o know	89.99	88.47	89.22	87.91	84.69	86.27
	w/o context	87.65	90.13	88.88	83.93	86.42	85.16
	w/o entity task	89.24	90.35	89.79	85.25	83.49	84.37
	Joint BERT(w/o all)	84.82	88.25	86.50	80.38	84.62	82.45
语义槽填充	K-CAM	94.58	70.86	79.90	92.21	73.95	82.07
	w/o know	91.14	70.11	79.25	91.88	73.66	81.77
	w/o context	89.66	69.43	78.25	90.50	73.09	80.87
	w/o entity task	90.41	71.02	79.55	91.38	73.33	81.36
	Joint BERT(w/o all)	88.52	69.42	77.81	90.21	72.92	80.65

最后,本文对K-CAM模型的性能进行评估,以探究模型的领域迁移能力。图4绘制了K-CAM模型和基线模型Joint BERT随训练轮次数的增加,模型预测对话动作的宏平均 F_1 值的变化情况。从图4可以看出,本文提出的显式知识注入的对话理解模型K-CAM比基线模型Joint BERT达到相同精度所需要学习的时间更短。这证明了尽管BERT作为K-CAM模型和Joint BERT模型的基础编码器,其本身不是基于对话文本进行的预训练,但当其应用到对话领域时,显式注入知识的K-CAM模型可以帮助模型更快地实现领域迁移。

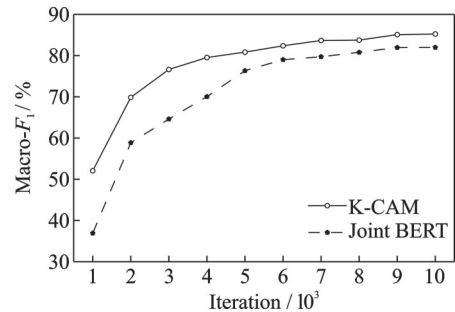


图4 实验性能评估

Fig.4 Experimental performance evaluation

4 结束语

本文提出一种显式知识注入的多任务学习对话理解模型。将知识以自然语言的形式显式插入到对话文本中,避免重复编码知识,增强模型对生僻内容的理解能力;此外,通过多任务学习策略,增加关联知识识别的辅助任务,与对话理解的主任务一同进行联合训练,减少知识噪声。实验结果表明,本文提出的模型K-CAM在任务导向型对话理解上具有更高的实验精度和更好的使用效果。未来工作将进一步研究针对加入知识的筛选方法,以最大可能地减少知识噪声,发挥知识融合的正向作用。此外,将研究如何将该模型迁移到开放型对话理解问题上,提升模型对常识知识的融合效果。

参考文献:

- [1] QIN L, XIE T, CHE W, et al. A survey on spoken language understanding: Recent advances and new frontiers[C]// Proceedings of International Joint Conferences on Artificial Intelligence Organization (IJCAI). Montreal: Morgan Kaufmann, 2021: 4577-4584.
- [2] 杨帆, 饶元, 丁毅, 等. 面向任务型的对话系统研究进展[J]. 中文信息学报, 2021, 35(10): 1-20.
YANG Fan, RAO Yuan, DING Yi, et al. Progress in task-oriented dialogue system[J]. Journal of Chinese Information Processing, 2021, 35(10): 1-20.
- [3] YAO K, ZWEIG G, HWANG M Y, et al. Recurrent neural networks for language understanding[C]// Proceedings of Interspeech. Lyon, France: International Speech Communication Association (ISCA), 2013: 2524-2528.
- [4] YAO K, PENG B, ZHANG Y, et al. Spoken language understanding using long short-term memory neural networks[C]//

- Proceedings of IEEE Spoken Language Technology Workshop. South Lake Tahoe, California and Nevada: IEEE, 2014: 189-194.
- [5] LIU B, LANE I. Attention-based recurrent neural network models for joint intent detection and slot filling[C]//Proceedings of Interspeech. San Francisco: International Speech Communication Association(ISCA), 2016.
- [6] GOO C W, GAO G, HSU Y K, et al. Slot-gated modeling for joint slot filling and intent prediction[C]//Proceedings of the Conference of the North American of the Association for Computational Linguistics. New Orleans: ACL, 2018: 753-757.
- [7] CHEN Q, ZHUO Z, WANG W. BERT for joint intent classification and slot filling[C]//Proceedings of the 17th Annual SIGDial Meeting. Los Angeles: ACL, 2016.
- [8] LIU B, LANE I. Joint online spoken language understanding and language modeling with recurrent neural networks[EB/OL]. (2016-09-06)[2024-05-19]. <https://doi.org/10.18653/v1/W16-3603>.
- [9] WANG Y, HE T, FAN R, et al. Effective utilization of external knowledge and history context in multi-turn spoken language understanding model[C]// Proceedings of 2019 IEEE International Conference on Big Data. Espoo, Finland: IEEE, 2019: 960-967.
- [10] HAO N, XIE P, CHEN B. Research on intent-slot recognition algorithm based on knowledge graph and user topic[C]// Proceedings of 2021 40th Chinese Control Conference(CCC). [S.l.]: IEEE, 2021: 7276-7281.
- [11] WU T W, JUANG B H. Knowledge augmented BERT mutual network in multi-turn spoken dialogues[C]//Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). [S.l.]: IEEE, 2022: 7487-7491.
- [12] LIU W, ZHOU P, ZHAO Z, et al. K-BERT: Enabling language representation with knowledge graph[C]//Proceedings of the 34th AAAI Conference on Artificial Intelligence. [S.l.]: AAAI, 2020, 34(3): 2901-2908.
- [13] QUAN J, ZHANG S, CAO Q, et al. RiSAWOZ: A large-scale multi-domain wizard-of-Oz dataset with rich semantic annotations for task-oriented dialogue modeling[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. [S.l.]: ACL, 2020.
- [14] YAO K, PENG B, ZWEIG G, et al. Recurrent conditional random field for language understanding[C]//Proceedings of 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). [S.l.]: IEEE, 2014: 4077-4081.
- [15] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [EB/OL]. (2019-05-24)[2023-3-05]. <https://arxiv.org/abs/1810.04805v2>.
- [16] BORDES A, USUNIER N, GARCIA-DURAN A, et al. Translating embeddings for modeling multi-relational data[C]// Proceedings of Advances in Neural Information Processing Systems. Lake Tahoe: MIT Press, 2013: 2787-2795.
- [17] SUN Y, WANG S, LI Y, et al. ERNIE: Enhanced representation through knowledge integration[EB/OL]. (2019-04-20) [2023-03-05]. <https://doi.org/10.48550/arXiv.1904.09223>.
- [18] LOSHCHILOV I, HUTTER F. Decoupled weight decay regularization[C]//Proceedings of the 7th International Conference on Learning Representations. New Orleans: OpenReview. net, 2019.

作者简介:



李帅鹏(1996-),男,博士研究生,研究方向:自然语言处理、知识图谱,E-mail:shuaipengli@stu.xjtu.edu.cn。



王平辉(1984-),通信作者,男,博士,教授,研究方向:机器学习与数据挖掘、自然语言处理,E-mail:phwang@mail.xjtu.edu.cn。



孙望淳(2000-),男,硕士研究生,研究方向:自然语言处理,E-mail:sun1123812254@stu.xjtu.edu.cn。



杨阳(1998-),女,硕士研究生,研究方向:自然语言处理,E-mail:jotrex@stu.xjtu.edu.cn。



杜友田(1980-),男,博士,教授,研究方向:互联网多媒体数据内容关联理解与跨媒体检索,E-mail:duyt@mail.xjtu.edu.cn。



马小科(1982-),男,博士,教授,研究方向:数据挖掘、机器学习,E-mail:xkma@xidian.edu.cn。



杜永杰(1984-),男,高级工程师,研究方向:自然语言处理,E-mail:tcldu@163.com。