



计算机科学

COMPUTER SCIENCE

基于BERT和多特征门控机制的口语理解联合方法

王志明, 郑凯

引用本文

王志明, 郑凯. 基于BERT和多特征门控机制的口语理解联合方法[J]. 计算机科学, 2023, 50(11A): 230300002-6.

WANG Zhiming, ZHENG Kai. Joint Method for Spoken Language Understanding Based on BERT and Multiple Feature Gate Mechanism [J]. Computer Science, 2023, 50(11A): 230300002-6.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[接诉即办智能派单业务调度算法研究](#)

Study on Scheduling Algorithm of Intelligent Order Dispatching

计算机科学, 2023, 50(11A): 230300029-7. <https://doi.org/10.11896/jsjcx.230300029>

[基于局部上下文焦点机制和交谈注意力的方面级情感分析](#)

Aspect-based Sentiment Analysis Based on Local Context Focus Mechanism and Talking-Head Attention

计算机科学, 2023, 50(11A): 220900266-6. <https://doi.org/10.11896/jsjcx.220900266>

[基于MacBERT和对抗训练的审计文本命名实体识别](#)

Audit Text Named Entity Recognition Based on MacBERT and Adversarial Training

计算机科学, 2023, 50(11A): 230200083-6. <https://doi.org/10.11896/jsjcx.230200083>

[基于特征融合的无标复句关系识别](#)

Relation Recognition of Unmarked Complex Sentences Based on Feature Fusion

计算机科学, 2023, 50(11A): 221100065-6. <https://doi.org/10.11896/jsjcx.221100065>

[基于核心句的端到端事件共指消解](#)

End-to-End Event Coreference Resolution Based on Core Sentence

计算机科学, 2023, 50(11): 185-191. <https://doi.org/10.11896/jsjcx.221000078>

基于 BERT 和多特征门控机制的口语理解联合方法

王志明¹ 郑凯²

1 华南师范大学计算机学院 广州 510631

2 华南师范大学网络中心 广州 510631

(wangzhiming@m.scnu.edu.cn)

摘要 意图分类和槽位填充是口语理解任务的两个子任务,用于在对话系统中识别文本序列的意图以及从文本序列中提取出能进一步确定意图具体内容的槽位信息。近些年的研究表明,这两个任务具有相关性,且可以相互促进。然而,现有的大多数联合方法仅利用单一特征简单地通过共享参数建立两者之间的关联,而这往往会导致模型泛化能力差及特征利用程度低等问题。针对这些问题,提出了一种新的联合模型。模型在 BERT 的基础上引入了意图特征提取层和槽位特征提取层用于进一步提取文本特征,以增强文本向量表示能力,并通过多特征门控机制融合了多方特征,充分利用两个任务之间的语义关系预测标签。在公开数据集 ATIS 和 SNIPS 上的实验结果表明,提出的模型能有效提升意图分类和槽位填充的性能,相比已有方法取得了更优的结果。

关键词: 意图分类; 槽位填充; 联合学习; 门控机制; BERT

中图法分类号 TP391

Joint Method for Spoken Language Understanding Based on BERT and Multiple Feature Gate Mechanism

WANG Zhiming¹ and ZHENG Kai²

1 School of Computer Science, South China Normal University, Guangzhou 510631, China

2 Network Center, South China Normal University, Guangzhou 510631, China

Abstract Intent classification and slot filling are two subtasks of spoken language comprehension that are used to identify the intent of text sequences in a conversation system and to obtain slot information from the text sequences that may be used to further infer the exact substance of the intent. Recent research has revealed that these two tasks are connected and can reinforce one another. However, the majority of joint techniques now just use one feature to establish the relationship between the two by only exchanging parameters, which frequently results in issues like poor model generalization and low feature utilization. In order to solve these problems, a novel joint model is proposed that adds an intent feature extraction layer and a slot feature extraction layer for additional text feature extraction based on BERT to improve text vector representation. It also fuses features from different parties using gate mechanism to fully utilize the semantic relationship between the two tasks to predict labels. Experimental findings on the openly accessible datasets ATIS and SNIPS demonstrate the effectiveness of the proposed model in improving intent categorization and slot filling performance, outperforming current approaches.

Keywords Intent classification, Slot filling, Joint learning, Gate mechanism, BERT

1 引言

口语理解(Spoken Language Understanding, SLU)是人机对话系统中的一个重要组成部分,其目的是让计算机能够正确理解人类的自然语言。口语理解通常涉及意图分类和槽位填充两个子任务。意图分类侧重于预测用户的意图,可以将其视为一个文本分类问题。而槽位填充侧重于提取文本中的实体信息,为话语中的每个单词预测实体类型标签,该任务通常被认为是序列标注问题。获取用户问句中的意图和实体信息,可帮助计算机理解人类自然语言,完成问答任务。

早期的意图分类和槽位填充是作为两个任务分开进行

的,并没有考虑两个任务的依赖关系。但事实上,意图和槽位是相互关联的,这两个任务可以相互加强,因此近年来有许多研究把工作重点放在了探索这两个任务的交互上,如 Capsule-NLU^[1], SF-ID Network^[2], Attention BiRNN^[3], Slot-Gated^[4] 和 Stack-Propagation^[5]。虽然这些模型在联合任务上取得了一定的效果,但是它们仍存在一些问题。首先,大多数的联合方法往往只采用单一的模型来获取文本的意图和槽位特征,无法多维度地获取和利用文本特征,这在缺乏训练数据的情况下会导致模型泛化能力较差和模型性能弱等问题。其次,这些联合方法大多不利用意图分类和槽位填充之间的相关性以互相促进两个任务,又或者只是简单地通过其中一者

基金项目:国家自然科学基金(62237001)

This work was supported by the National Natural Science Foundation of China(62237001).

通信作者:郑凯(david@senu.edu.cn)

促进另一者,并不能真正做到互相促进。

针对上述提到的问题,本文提出一种基于 BERT 和多特征门机制的联合学习模型,用于意图分类和槽位填充联合任务。模型通过利用 BERT, CNN 和 BiLSTM 等模型获得带有丰富语义和时序特征的文本表示,再利用门机制自动学习捕捉来自意图和槽位的信息用于最终预测结果。实验结果表明了提出的模型的有效性。

2 相关工作

早期的意图分类方法主要采用 n-grams 作为特定实体的特征进行分类^[6],也会利用特征工程结合 SVM^[7]和 AdaBoost^[8]等传统的机器学习算法提升意图分类效果。随着深度神经网络的发展,卷积神经网络(CNN)^[9]和长短期记忆网络(LSTM)^[10]也被广泛应用在意图分类任务上,且都取得了不错的效果。随后,基于 CNN 和 LSTM 的各类变体如 GRU^[11],RCNN^[12],BiLSTM-Attention^[13]和 DCPNN^[14]也在意图分类任务上证明了其有效性。

槽位填充任务的传统方法是基于条件随机场(Conditional Random Field, CRF^[15]),该方法在序列标注任务上表现良好。近些年来,深度学习模型逐渐成为处理槽位填充任务的主要方式。Yao 等提出使用 RNN^[16],RNN-CRF^[17]和 LSTM^[18]模型去预测槽位标签,证明了基于 RNN 的模型表现优于传统的 CRF 模型。而随着注意力机制的提出,也有研究者尝试在 RNN 结构中引入注意力机制并用于槽位填充任务,更进一步增强了模型的预测能力^[19]。

由于意图分类和槽位填充两个任务有较强的关联性,因此越来越多的研究人员倾向于对这两个任务建立联合模型。最早的联合方式是通过三级 CRF 来对两个任务进行建模^[20],并证明了联合处理两个任务比单独处理两个任务更具优势。随后,利用 RNN^[21]和 CNN^[22]来学习文本的上下文表示的方式也逐渐应用到联合模型中。Liu 等^[3]则将注意力机制引入联合模型中,增强了 RNN 隐藏层的表示。

为了学习两个任务之间的关系,Goo 等^[4]在双向 LSTM 结构中加入一种门机制。而 Zhang 等^[1]则采用一种胶囊网络的结构来为两个任务建立联系。类似的还有 Haihong 等^[2]提出的 SF-ID 交互网络,以及 Chen 等^[23]基于 word attention 和门机制提出的联合模型。随着 Transformer 和 BERT 的兴起,Qin 等^[24]设计了一种交互 Transformer 结构用于共享意图和槽位信息,而 Chen 等^[25]则将 BERT 用于联合模型中,证明了预训练模型对于联合任务的有效性。

3 方法提出

本节详细介绍了多特征交互门联合模型。该模型建立在 Joint BERT^[25]的基础上,并额外增加了两个特征提取层和两个特征交互门,以解决前文提到的两个问题。如图 1 所示,模型共由 5 个部分组成:BERT 层用于编码输入的文本序列,获取基本的语义表示;意图和槽位特征提取层分别用于从 BERT 的输出中进一步提取意图和槽位信息,以丰富文本特征;交互门层则用于建立意图和槽位信息之间的双向联系;CRF 层用于学习槽位之间的依赖关系。

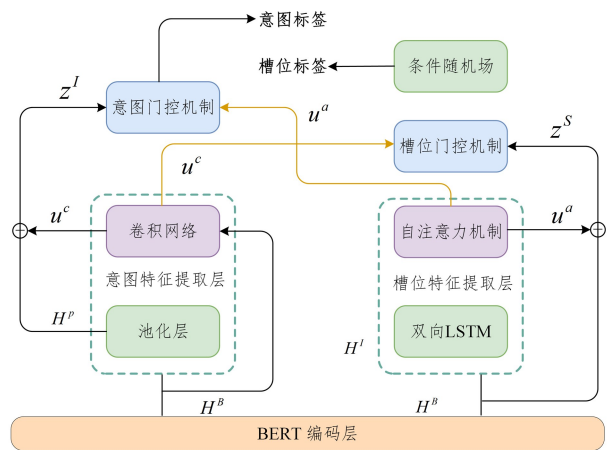


图 1 联合模型结构示意图

Fig. 1 Illustration of joint model structure

3.1 BERT 编码层

联合模型使用预训练模型 BERT 来对输入文本序列进行编码,以获取每个输入序列的上下文表示。BERT 由多个双向 Transformer^[26]编码层组成,经过在大规模无标记文本上训练,BERT 模型可以提供强大的上下文文本表示,并已被证明可以提升各种下游 NLP 任务的效果。使用 BERT 编码文本序列的步骤如下:对于一个问句文本序列 x_1, x_2, \dots, x_N ,在其前后加入特殊标记后得到 X ,其中 x_{CLS} 为文本序列开始标记, x_{SEP} 为文本序列之间的分隔标记。

$$X = x_{CLS}, x_1, x_2, \dots, x_N, x_{SEP} \quad (1)$$

接下来将 X 输入到 BERT 模型后,得到句子的上下文语义表示 $H_{CLS}^B \in \mathbb{R}^{1 \times D}, H_{CLS}^B \in \mathbb{R}^{N \times D}$,其中 N 表示句子长度, D 表示 BERT 的隐藏层维度, H_{CLS}^B 表示文本的句子级表示, H^B 表示文本的单词级表示。

$$\begin{cases} H_{CLS}^B, H^B = \text{BERT}(X) \\ H^B = (h_1^B, h_2^B, \dots, h_N^B) \end{cases} \quad (2)$$

3.2 意图特征提取层

意图特征提取的作用主要是提取出能够表示文本的句子级特征向量。过去的很多基于 BERT 的联合任务,往往将 H_{CLS}^B 作为输入文本的句子级表示。但 Reimers 等^[27]的研究表明,在 H^B 的基础上进行平均池化的方式,在句子级别表示上要比使用 BERT 的 H_{CLS}^B 效果好。同时,也有研究表明使用 CNN 模型对 H^B 进行特征提取的方式,在意图分类任务上具有优势^[28]。因此本文结合池化和卷积神经网络两种方式进行意图特征提取。

本文使用的卷积神经网络结构如图 2 所示,它由 5 种尺寸的卷积核构成,每一种尺寸有 80 个滤波器,共有 400 个滤波器。将 H^B 作为卷积网络的输入,并在卷积运算后,经过激活函数 Relu 和最大池化层得到输出 c ;接着将它们拼接起来得到 $u_c = \mathbb{R}^{1+d_c}$,其中 d 表示 CNN 的输出维度;最后,将 H^B 的平均池化结果 $H_p \in \mathbb{R}^{1 \times D}$ 和 c 再次拼接得到 $z^I \in \mathbb{R}^{1 \times (D+d_c)}$,将其作为文本的句子级表示。

$$\begin{cases} H^p = \text{Average Pooling}(H^B) \\ c_i = \text{CNN}_i(h_1^B, h_2^B, \dots, h_N^B) \\ u_c = \text{concat}(c_1, c_2, c_3, c_4, c_5) \\ z^I = \text{concat}(H^p, u_c) \end{cases} \quad (3)$$

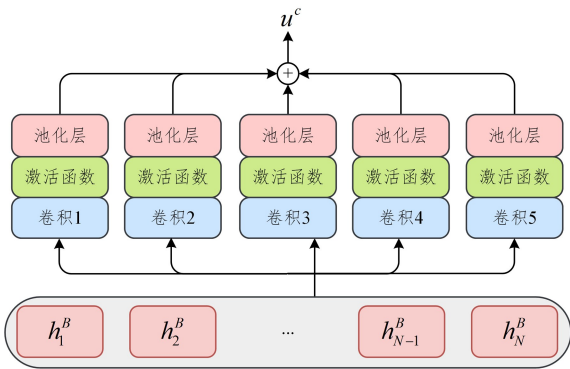


图2 卷积神经网络结构

Fig. 2 Structure of convolution neural network

3.3 槽位特征提取层

槽位特征提取的主要目标是获取能够表示文本的单词级特征向量。在过去的联合任务方法中,模型往往只采用单一的特征向量表示。考虑到双向长短期记忆网络(BiLSTM)在槽位填充任务上的有效应用^[2],本文采取将BERT和BiLSTM结合的方式来获取文本的单词级表示。

BiLSTM可以学习前向和后向的文本序列特征,并将两个LSTM网络连接到同一个输出层以得到最终表示。其具体计算过程如下:

$$\begin{cases} \vec{h}_t^{\rightarrow} = \text{LSTM}(x_t, \vec{h}_{t-1}^{\rightarrow}) \\ \vec{h}_t^{\leftarrow} = \text{LSTM}(x_t, \vec{h}_{t-1}^{\leftarrow}) \\ \vec{h}_t^{\rightarrow\leftarrow} = \text{concat}(\vec{h}_t^{\rightarrow}, \vec{h}_t^{\leftarrow}) \\ H^l = (\vec{h}_1^{\rightarrow\leftarrow}, \vec{h}_2^{\rightarrow\leftarrow}, \dots, \vec{h}_T^{\rightarrow\leftarrow}) \end{cases} \quad (4)$$

其中, \$\vec{h}_t^{\rightarrow}\$ 和 \$\vec{h}_t^{\leftarrow}\$ 分别表示前向和后向的LSTM在 \$t\$ 时刻的隐藏层状态, \$x_t\$ 和 \$\vec{h}_{t-1}^{\rightarrow}\$ 分别表示 \$t\$ 时刻BiLSTM的输入和输出, \$H^l \in \mathbb{R}^{N \times d_l}\$ 表示BiLSTM的最终输出, \$d_l\$ 表示BiLSTM的隐藏层大小。

为了能进一步从BiLSTM获取重要的信息表示,在BiLSTM后引入了自注意力机制^[26]。自注意力机制是注意力机制中的一种特例,它主要用于对一个序列中所有单词的关系进行建模,可以直接计算单词之间的依赖关系。其详细的计算过程如下:

$$\begin{cases} Q = W^Q \times H^l \\ K = W^K \times H^l \\ V = W^V \times H^l \\ \alpha = \text{Softmax}(Q^T K) \\ u^a = V\alpha \end{cases} \quad (5)$$

其中, \$W^Q \in \mathbb{R}^{N \times N}\$, \$W^K \in \mathbb{R}^{N \times N}\$, \$W^V \in \mathbb{R}^{N \times N}\$ 为可以训练的参数,分别用于计算Query, Key, Value。 \$\alpha \in \mathbb{R}^{d_l \times d_l}\$ 表示注意力权重, \$u^a \in \mathbb{R}^{N \times d_l}\$ 则表示自注意力机制的输出结果。最后,将 \$H^B\$ 和 \$u^a\$ 拼接起来,获得同时拥有丰富语义信息的文本单词级表示 \$z^s \in \mathbb{R}^{N \times (D+d_l)}\$:

$$z^s = \text{concat}(H^B, u^a) \quad (6)$$

3.4 多特征门控机制

为了能够增强意图和槽位信息之间的联系,以提高联合模型的性能,设计了特征交互门来获得意图和槽位的最终表示。过去的许多研究方法中,门机制主要是利用意图信息增强槽位信息,但是忽略了槽位信息对意图信息的作用。因此

我们设计了两个门机制,既将意图信息添加到槽位填充任务中,也将槽位信息添加到意图分类任务中。门机制可以自动学习分配意图和槽位信息的权重,使得意图分类和槽位填充能够相互促进性能。

在意图特征门中,结合来自意图特征提取层的 \$z^l\$ 和来自槽位特征提取层的 \$u^a\$,用于得到意图信息的最终表示。具体的计算步骤如下:

$$\begin{cases} u' = \max \text{pooling}(u^a) \\ I = \text{concat}(Z^l, u') \\ g^l = \text{sigmoid}(W^l \times I) \\ O^l = I \otimes g^l \end{cases} \quad (7)$$

其中, \$W^l\$ 是一个可训练的权重矩阵, \$g^l\$ 表示意图门权重, \$\text{sigmoid}\$ 是一种激活函数, \$\otimes\$ 表示哈达玛积。 \$O^l\$ 为意图门的最终输出,也作为意图信息的最终表示,利用它可以最终得到意图的预测结果 \$y^l\$。

$$y^l = \text{Softmax}(W_y^l \times O^l + b_y^l) \quad (8)$$

其中, \$W_y^l\$ 和 \$b_y^l\$ 是可训练的权重矩阵和偏置。同样的,在槽位特征门中结合了来自意图特征提取层的 \$u^c\$ 和来自槽位特征提取层的 \$z^s\$,用于表示槽位信息。具体的计算步骤如下:

$$\begin{cases} S_i = \text{concat}(z_i^s, u^c), i \in \{1, \dots, N\} \\ g_i^s = \text{sigmoid}(W^s \times S_i) \\ O_i^s = S_i \otimes g_i^s \\ O^s = [O_1^s, O_2^s, \dots, O_N^s] \end{cases} \quad (9)$$

其中, \$W^s\$ 是一个可训练的权重矩阵; \$g_i^s\$ 表示第 \$i\$ 个槽位对应的槽位门权重; \$O^s\$ 为槽位门的最终输出,也作为槽位信息的最终表示。

3.5 条件随机场

条件随机场能够学习相邻槽位标签之间的依赖关系,并预测出最优的槽位结果,因此在槽位门机制后加入了条件随机场用于提高槽位预测的准确率。条件随机场的计算方式如下:

$$\begin{cases} \text{score}(x, y^s) = \sum_{i=1}^N ([A]_{y_{i-1}^s, y_i^s} + [W^C \times O_i^s]_{y_i^s}) \\ P(y^s | x) = \frac{\exp(\text{score}(x, y^s))}{\sum_{y^s} \exp(\text{score}(x, y^s))} \end{cases} \quad (10)$$

其中, \$W^C\$ 是权重矩阵, \$A_{i,j}\$ 表示从 \$i\$ 标签转移至 \$j\$ 标签的转移分数, \$y^s\$ 表示槽位的预测结果。

3.6 损失函数

选用交叉熵损失函数(Cross Entropy Loss Function)作为联合模型训练的损失函数。对于意图分类和槽位填充任务,损失函数定义如下:

$$P(y^l, y^s | x) = P(y^l | x) \prod_{n=1}^N P(y_n^s | x) \quad (11)$$

其中, \$P(y^l, y^s | x)\$ 表示给定输入序列 \$x\$ 情况下联合模型预测结果的条件概率。

4 实验及评价

4.1 数据集

为了评估所提联合模型的预测效果,使用两个公开的对话系统数据集 ATIS^[31] 和 SNIPS^[32] 进行实验。

ATIS是航空公司旅行信息系统数据集,包含了预定航班的录音文本记录。训练集包括4478个语音文本,验证集包括500个语音文本,测试集包含893个语音文本。数据集中总共包含了120种槽位标签和21种意图类型。

SNIPS 数据集是 SNIPS 语音助手收集的对话数据集, 包含训练集 13084 个语音文本, 验证集 700 个语音文本, 测试集 700 个语音文本。数据集中总共包含了 72 种槽位标签和 7 种意图类型。

4.2 评价方法与实验设置

使用 3 种评估指标来评估模型性能。采用准确率(Intent Accuracy)来评估意图分类任务, 使用 F1 分数(F1 scores)来评估槽位填充任务, 使用句子级语义框架准确率(Sentence Accuracy)来评估两个任务的总体表现。

实验中使用的 BERT 编码器为 bert-base-cased, 编码器具有 12 个隐层, 输出 768 维张量, 12 个自注意力头, 共 110×10^6 参数量, 在不区分大小写的英文文本上进行训练而得到, 我们选用该模型并在其基础上对其参数进行更新。模型训练时的学习率为 5×10^{-5} , Dropout 设置为 0.1, 批次大小为 32, epoch 为 15。模型采用 Adam 优化算法来更新参数, 实验设备是 GeForce RTX 2080Ti。

4.3 基线模型

将提出的模型方法与多种先进的联合方法进行对比, 选择的基线模型如下。

(1) Joint Seq: Hakkani-Tür 等^[33]提出了一种基于 RNN-LSTM 的联合模型, 用于同时处理意图、槽位、领域 3 种任务。

(2) Attention BiRNN: Liu 等^[3]提出了 Attention BiRNN, 该模型提出了一种基于注意力的编码器-解码器模型, 利用注意力机制提高了联合模型的性能。

(3) Slot-Gated: Goo 等^[4]提出了 Slot-gated, 该模型引入

了一种槽位门机制, 用于将意图信息融入到槽位填充任务上。

(4) Capsule-NLU: Zhang 等^[1]采用基于胶囊的神经网络模型, 利用语义层次进行建模, 并通过动态路由协议实现意图分类和槽位填充。

(5) SF-ID Network: Haihong 等^[2]提出了 SF-ID 网络——一种双向交互网络, 该网络包含了意图和槽位两个子网, 用于共享意图和槽位之间的信息。

(6) SlotRefine: Wu 等^[34]采用了一种两步迭代机制, 将槽位填充任务分为两步进行, 增强了模型对槽位标签依赖性的学习。

(7) StackPropagation: Qin 等^[5]提出了一种基于栈传播机制的联合模型, 用于建立意图和槽位之间的联系。

(8) Joint BERT: Chen 等^[25]采用标准的 BERT 模型的分类和序列标注能力来构建联合模型

(9) Co-Interactive Transformer: Qin 等^[24]提出了一种交互式 Transformer 模型, 同时考虑意图分类和槽位填充两个任务之间的交互性。

4.4 实验结果与分析

从表 1 中能够看出, 本文模型在 ATIS 和 SNIPS 两个数据集上的各方面表现都超越了所有基线模型。基线模型中的联合模型只是利用了意图和槽位信息的单一特征, 并且没有充分利用两者之间的潜在关系进一步促进模型理解能力。而我们提出的模型则在利用多特征的基础上, 通过门控机制实现了意图和槽位信息的相互促进, 从而达到了比基线模型更好的效果。

表 1 各模型实验结果对比

Table 1 Comparison of experimental results

(单位: %)

模型	ATIS Dataset			SNIPS Dataset		
	Intent	Slot	Sent	Intent	Slot	Sent
Joint Seq	92.6	94.3	80.7	96.9	87.3	73.2
Attention BiRNN	91.1	94.2	78.9	96.7	87.8	74.1
Slot-Gated Full Attention	93.6	94.8	82.2	97.0	88.8	75.5
Capsule-NLU	95.0	95.2	83.4	97.3	91.8	80.9
SF-ID Network	96.6	95.6	86.0	97.0	90.5	78.4
SlotRefine	97.1	96.2	87.0	97.4	93.7	84.4
StackPropagation	96.9	95.9	86.5	98.0	94.2	86.9
Co-interactive Transformer	97.7	95.9	87.4	98.8	95.9	90.3
Joint BERT	97.8	95.7	88.2	98.4	96.6	92.6
SlotRefine+BERT	97.7	96.1	88.6	99.0	97.0	92.9
StackPropagation+BERT	97.5	96.1	88.6	99.0	97.0	92.9
本文模型	98.1	96.2	88.8	99.1	97.2	93.3

4.5 消融分析

为了研究模型中各个模块对实验结果的影响, 以累积的方式进行消融研究。我们将重点放在了特征门机制和意图提取层上, 通过逐步地减少模型中的模块进行实验, 最终得到表 2 所列的消融实验结果。

表 2 消融实验结果

Table 2 Results of ablation experiments

(单位: %)

消融模型	ATIS 数据集			SNIPS 数据集		
	Intent	Slot	Sent	Intent	Slot	Sent
完整模型	98.09	96.16	88.80	99.14	97.18	93.29
移除门控机制	97.98	96.08	88.60	98.29	97.05	92.86
移除槽位特征提取层	97.87	95.75	88.47	98.29	96.46	92.00
移除意图特征提取层	97.32	95.87	88.13	98.00	96.38	91.71

(1) 多特征门机制的作用: 将意图特征门和槽位特征门同时去除, 在 ATIS 数据集上, 各方面指标均有所下降。而在 SNIPS 数据集上, 各指标下降的幅度大更多。这说明特征门机制可以更好地共享意图和槽位的信息, 尤其是在 SNIPS 数据集上。从两个数据集上的表现情况来看, 也可以发现多特征门机制对意图分类任务的提升效果比较明显。

(2) 特征提取层的作用: 从表 2 可以发现, 在去除意图特征提取层后, 在两个数据集上意图分类的表现效果也有所下降, 而在去除槽位特征提取层后, 两个数据集中槽位填充的 F1 得分和句子语义得分下降得比较明显, 而意图分类准确率变化不大。这说明了使用多维度的特征文本表示有助于进一步提升两个任务的效果, 且特征提取层对联合任务的整体效果提升有一定的贡献。

结束语 本文设计了一种基于 BERT 的多特征门机制

模型用于意图分类和槽位填充联合任务。该模型在预训练模型 BERT 的基础上,引入了由 CNN 和 Pooling 组成的意图特征提取层,以及由 BiLSTM 和 Self-Attention 组成的槽位特征提取层,分别用于获取更加丰富的意图和槽位语义表示,解决了标准 BERT 模型特征单一的问题。另外,模型还引入了多特征门机制,用于将来自意图和槽位的不同特征进行融合,使得意图分类和槽位填充两个任务能够相互促进,相互提升。在 ATIS 和 SNIPS 数据集上的实验结果表明了所提模型的有效性,其有效提高了意图分类和槽位填充两个任务的精度。未来我们计划通过引入额外的领域知识来改进模型。

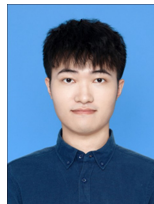
参 考 文 献

- [1] ZHANG C, LI Y, DUN, 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.
- [2] HAIHONG E, NIU P, CHEN Z, et al. A Novel Bi-directional Interrelated Model for Joint Intent Detection and Slot Filling [C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019;5467-5471.
- [3] LIU B, LANE I. Attention-based recurrent neural network models for joint intent detection and slot filling[J]. arXiv:1609.01454, 2016.
- [4] 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, Volume 2 (Short Papers). 2018;753-757.
- [5] QIN L, CHE W, LI Y, et al. A Stack-Propagation Framework with Token-Level Intent Detection for Spoken Language Understanding[C]// Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019;2078-2087.
- [6] ZHANG X, WANG H. A joint model of intent determination and slot filling for spoken language understanding[C]// IJCAI. 2016;2993-2999.
- [7] HAFFNER P, TUR G, WRIGHT J H. Optimizing SVMs for complex call classification[C]// 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'03). IEEE, 2003, 1: 1.
- [8] SCHAPIRE R E, SINGER Y. BoosTexter: A boosting-based system for text categorization [J]. Machine Learning, 2000, 39(2): 135-168.
- [9] KIM Y. Convolutional Neural Networks for Sentence Classification[C]// EMNLP. 2014.
- [10] RAVURI S, STOLCKE A. Recurrent neural network and LSTM models for lexical utterance classification[C]// Sixteenth Annual Conference of the International Speech Communication Association. 2015.
- [11] CHO K, VAN MERRIËNBOER B, GULCEHREC, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[J]. arXiv:1406.1078, 2014.
- [12] LAI S, XU L, LIU K, et al. Recurrent convolutional neural networks for text classification[C]// Twenty-ninth AAAI Conference on Artificial Intelligence. 2015.
- [13] 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 (volume 2: Short papers). 2016;207-212.
- [14] JOHNSON R, ZHANG T. Deep pyramid convolutional neural networks for text categorization[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2017;562-570.
- [15] RAYMOND C, RICCARDI G. Generative and discriminative algorithms for spoken language understanding[C]// 8th Annual Conference of the International Speech Communication Association (INTERSPEECH 2007). 2007.
- [16] YAO K, ZWEIG G, HWANG M Y, et al. Recurrent neural networks for language understanding[C]// Interspeech. 2013;2524-2528.
- [17] YAO K, PENG B, ZWEIG G, et al. Recurrent conditional random field for language understanding[C]// 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2014;4077-4081.
- [18] YAO K, PENG B, ZHANG Y, et al. Spoken language understanding using long short-term memory neural networks[C]// 2014 IEEE Spoken Language Technology Workshop (SLT). IEEE, 2014;189-194.
- [19] SIMONNET E, CAMELIN N, DELÉGLISE P, et al. Exploring the use of attention-based recurrent neural networks for spoken language understanding [C] // Machine Learning for Spoken Language Understanding and Interaction NIPS 2015 workshop (SLUNIPS 2015). 2015.
- [20] JEONG M, LEE G G. Triangular-chain conditional random fields[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2008, 16(7): 1287-1302.
- [21] GUO D, TUR G, YIH W, et al. Joint semantic utterance classification and slot filling with recursive neural networks[C]// 2014 IEEE Spoken Language Technology Workshop (SLT). IEEE, 2014;554-559.
- [22] XU P, SARIKAYA R. Convolutional neural network based triangular crf for joint intent detection and slot filling[C]// 2013 IEEE Workshop on Automatic Speech Recognition and Understanding. IEEE, 2013;78-83.
- [23] CHEN S, YU S. Wais: Word attention for joint intent detection and slot filling[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33(1): 9927-9928.
- [24] QIN L, LIU T, CHE W, et al. A co-interactive transformer for joint slot filling and intent detection[C]// 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2021). IEEE, 2021;8193-8197.
- [25] CHEN Q, ZHUO Z, WANG W. Bert for joint intent classification and slot filling[J]. arXiv:1902.10909, 2019.
- [26] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [J]. Advances in Neural Information Processing Systems, 2017, 30.
- [27] REIMERS N, GUREVYCH I, REIMERS N, et al. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks [C]// Proceedings of the 2019 Conference on Empirical Methods

in Natural Language Processing. Association for Computational Linguistics, 2019: 671-688.

- [28] CHOI H, KIM J, JOE S, et al. Evaluation of BERT and ALBERT sentence embedding performance on downstream NLP tasks[C]//2020 25th International conference on pattern recognition(ICPR). IEEE, 2021: 5482-5487.
- [29] HE C, CHEN S, HUANG S, et al. Using convolutional neural network with BERT for intent determination[C]//2019 International Conference on Asian Language Processing (IALP). IEEE, 2019: 65-70.
- [30] SAFAYA A, ABDULLATIF M, YURET D. Kuisail at semeval-2020 task 12: Bert-cnn for offensive speech identification in social media[C]//Proceedings of the Fourteenth Workshop on Semantic Evaluation. 2020: 2054-2059.
- [31] TUR G, HAKKANI-TÜR D, HECKL. What is left to be understood in ATIS? [C]//2010 IEEE Spoken Language Technology Workshop. IEEE, 2010: 19-24.
- [32] COUCKE A, SAADE A, BALL A, et al. Snips voice platform: an embedded spoken language understanding system for private-by-design voice interfaces[J]. arXiv:1805.10190, 2018.

- [33] 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.
- [34] WU D, DING L, LU F, et al. SlotRefine: A fast non-autoregressive model for joint intent detection and slot filling[J]. arXiv: 2010.02693, 2020.



WANG Zhiming, born in 1999, post-graduate, is a member of China Computer Federation. His main research interests include natural language processing and spoken language understanding



ZHENG Kai, born in 1978, Ph.D, senior engineer. His main research interests include education information technology and network security.