



# 基于 S-LSTM 模型利用‘槽值门’机制的说话人意图识别

## 摘要

对说话人意图的识别极大地推进了自然语言理解任务的发展.之前的工作大多采用 Bi-LSTM 即双向 LSTM 模型进行词汇特征与词汇之间语义关系的提取,但这并不能很好地使句子整体和构成句子的词汇个体之间的信息进行交流.而 S-LSTM (Sentence-state LSTM) 模型,即句子状态 LSTM 模型可以很好地将自然语言中句子整体与词汇个体的信息相结合,以便于我们挖掘与利用意图检测与槽值填充之间的关系成立联合模型来更好地理解应答系统中蕴含的语义.因此,本文引入了‘槽值门’机制解决 S-LSTM 应用于意图检测与槽填充的联合任务时最新迭代时刻的句子状态信息没有得到充分利用的问题.最终的实验结果在 ATIS 数据集和 Snips 数据集上均取得了优于目前最先进算法的结果.

## 关键词

意图识别;槽填充;S-LSTM;槽值门

中图分类号 TP391

文献标志码 A

收稿日期 2019-10-15

资助项目 国家自然科学基金(61872199,61872424,61936005)

## 作者简介

王子岳,男,硕士生,研究方向为自然语言处理.846890074@qq.com

邵曦(通信作者),男,博士,副教授,研究方向为多媒体信息系统.shaoxi@njupt.edu.cn

<sup>1</sup> 南京邮电大学 通信与信息工程学院,南京,210003

## 0 引言

在自然语言的问答系统中,SLU (Spoken Language Understanding) 是必不可少的一环,它指引着后续对话的内容.SLU 的目的是形成一个语义框架,以获取用户话语或查询的语义,它通常包括两个任务:意图检测(intention detection)<sup>[1]</sup>以及槽填充(slot filling)<sup>[2]</sup>.

在问答系统,尤其是任务型问答系统中,当用户提出要求时,系统首先需要分辨出用户想要做什么,这便是意图检测的目的.意图检测任务可以被看作是一种分类任务,把一句话按照它包含的语义分到不同的意图类别中.之后,系统需要分辨出完成用户要求所必须的要素,例如当用户需要预定机票时,系统需要分辨出所要预定的航班的行程以及具体的时间,这就是槽填充任务所要解决的问题.槽填充任务可以被看作是一种序列标记任务<sup>[3]</sup>,它将单词的嵌入表示映射为与之对应的输出向量,据此为每一个单词打上不同的标签,这个标签取决于当前单词是否包含上述完成用户要求的要素以及是哪一种要素.

学者们常常将意图检测和槽填充单独考虑,而没有将两者间的联系进行探究.但随着研究的推进,人们发现同时实现意图检测和槽填充的联合模型往往能够取得更好的效果,这一现象预示着说话人的意图与槽值之间的相关性对意图识别结果的提升有着至关重要的效用.但这些工作都没有清晰地建模意图与槽值之间的关系,而 S-LSTM<sup>[4]</sup>模型为这种建模提供了可能,它很好地将句子中每个单词的信息与句子整体的信息进行融合.但将 S-LSTM 模型同时运用在槽填充和意图检测任务时,能够最详尽且最清晰地呈现句子所蕴含时序语义信息的句子状态信息并没有得到充分利用,这造成了严重的资源浪费也对结果的准确性造成了影响.为了解决这一问题,本文开创性地引入了‘槽值门’机制,在进行槽填充任务时,将最新迭代时刻的句子状态信息作为监督信息加以运用.论文的结构如下所述:第 1 节介绍了前人在意图识别领域所做的工作,第 2 及第 3 节分别介绍了 S-LSTM 模型以及‘槽值门’机制,第 4 节介绍了我们提出的模型,第 5 节讨论了我们具体的实验设置以及最终的实验结果,第 6 节是结论.

## 1 相关工作

在单独解决槽填充任务时,人们会选择条件随机场(CRF)<sup>[5]</sup>以及循环神经网络(RNN)<sup>[6]</sup>的方法.而在单独处理意图检测任务时,人们尝试过类似于支持向量机(SVM)<sup>[7]</sup>以及 Adaboost<sup>[8]</sup>的各种各样的分类器.

在面对意图检测和槽填充的联合任务时,学者们为了利用它们之间的关系做出了很多努力.从2014年 Guo 等<sup>[9]</sup>利用语义分析树构造了路径特征来对槽位与意图进行识别到2016年 Hakkani-Tür 等<sup>[10]</sup>构造多域联合的语义框架,人们试着通过对两者之间的关系进行增强从而得到较好的结果.同样是在2016年,注意力机制<sup>[11]</sup>也被用来构造联合模型<sup>[12]</sup>.这一模型通过意图检测任务和槽填充任务之间文本向量的共享实现了信息交互的意图从而取得了很好的效果.2018年,‘槽值门’(slot-gate)机制作为控制两种任务之间信息结合的方式被引入了模型<sup>[13]</sup>.

## 2 S-LSTM

### 2.1 结构介绍

S-LSTM 模型全称是 Sentence-state LSTM,也就是句子状态 LSTM,它在对每一句话进行嵌入的时候,不仅仅聚焦于单词层面的嵌入,还会训练出一种句子嵌入来对句子的整体状态进行表示,其具体结

构如图所 1 示.

每个单词的隐藏状态  $h$  由前一时刻句子中一定步长范围内的单词的隐藏状态与上一个时刻句子整体的状态  $g^{t-1}$  所决定.句子的整体状态数量是可以调整的超参数,但一般仅选用一个节点来表示句子的整体状态.步长也是可以调整的超参数,例如步长  $s$  为 1 代表此时刻单词  $i$  的隐藏状态  $h_i^t$  受前一时刻隐藏状态  $[h_{i-1}^{t-1}, h_i^{t-1}, h_{i+1}^{t-1}]$  的影响,此时每一个隐藏状态和前一时刻的句子状态都对应着一个‘门’机制,来控制其对当前时刻隐藏状态的影响:

$$\xi_i^t = [h_{i-1}^{t-1}, h_i^{t-1}, h_{i+1}^{t-1}], \quad (1)$$

$$\hat{i}_i^t = \sigma(W_i \xi_i^t + U_i x_i + V_i g^{t-1} + b_i), \quad (2)$$

$$\hat{l}_i^t = \sigma(W_l \xi_i^t + U_l x_i + V_l g^{t-1} + b_l), \quad (3)$$

$$\hat{r}_i^t = \sigma(W_r \xi_i^t + U_r x_i + V_r g^{t-1} + b_r), \quad (4)$$

$$\hat{f}_i^t = \sigma(W_f \xi_i^t + U_f x_i + V_f g^{t-1} + b_f), \quad (5)$$

$$\hat{s}_i^t = \sigma(W_s \xi_i^t + U_s x_i + V_s g^{t-1} + b_s), \quad (6)$$

$$o_i^t = \sigma(W_o \xi_i^t + U_o x_i + V_o g^{t-1} + b_o), \quad (7)$$

$$u_i^t = \tanh(W_u \xi_i^t + U_u x_i + V_u g^{t-1} + b_u), \quad (8)$$

$$i_i^t, l_i^t, r_i^t, f_i^t, s_i^t = \text{softmax}(\hat{i}_i^t, \hat{l}_i^t, \hat{r}_i^t, \hat{f}_i^t, \hat{s}_i^t), \quad (9)$$

$$c_i^t = l_i^t \cdot c_{i-1}^{t-1} + f_i^t \cdot c_i^{t-1} + r_i^t \cdot c_{i+1}^{t-1} + s_i^t \cdot c_g^{t-1} + i_i^t \cdot u_i^t, \quad (10)$$

$$h_i^t = o_i^t \cdot \tanh(c_i^t), \quad (11)$$

其中,  $\xi_i^t$  代表一个文本窗所覆盖单词的前一时刻隐藏状态,  $l_i^t, r_i^t, f_i^t, s_i^t$  以及  $i_i^t$  都是控制对  $c_i^t$  产生影响的信息流的‘门’.具体来说,  $i_i^t$  是控制此时刻对应单词

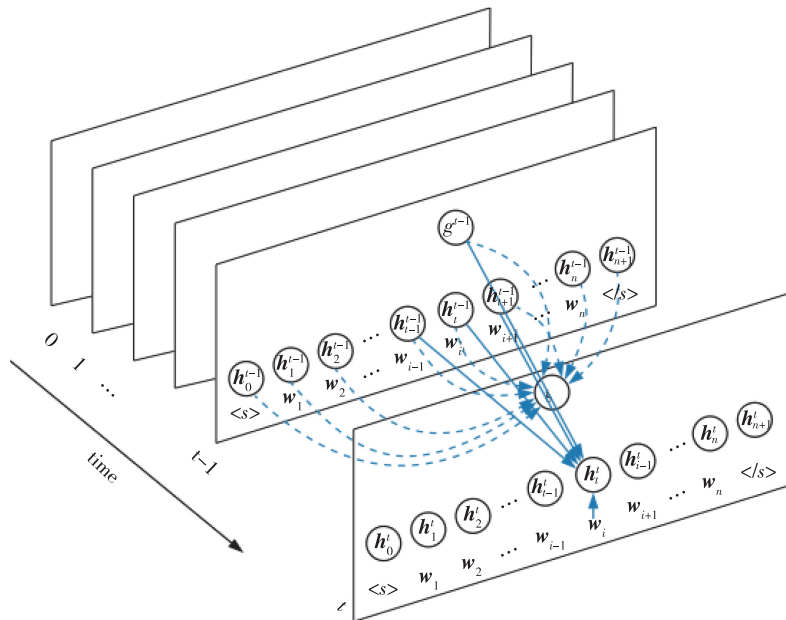


图 1 S-LSTM 结构图

Fig. 1 Structure of S-LSTM

输入的信息流,  $r'_i$  与  $r'_i$  分别控制前一个时刻中上一个单词的隐藏状态和下一个单词的隐藏状态对当前单词隐藏状态的影响, 而  $f'_i$  则是控制前一时刻当前单词对当前时刻当前单词的影响, 因此 S-LSTM 可以通过不断地迭代将上下文信息结合到每一个单词的隐藏状态. 同时, 采用  $s'_i$  来控制上一时刻句子整体状态对当前时刻当前单词的影响. 上述这些参数, 经过归一化后使其总和为 1.  $o'_i$  将初步得出的隐藏状态做出进一步的变换之后成为最终输出的此时刻当前单词的隐藏状态.  $W, U, V, b$  均为可以训练的模型参数.  $\sigma$  代表激活函数 Sigmoid 函数.

此时刻句子状态  $g^t$  是基于前一时刻所有单词隐藏状态以及前一时刻的句子状态计算得来的:

$$\bar{h} = \text{avg}(h_0^{t-1}, h_1^{t-1}, \dots, h_{n+1}^{t-1}), \quad (12)$$

$$\hat{f}_g = \sigma(W_g g^{t-1} + U_g \bar{h} + b_g), \quad (13)$$

$$\hat{f}_i = \sigma(W_f g^{t-1} + U_f h_i^{t-1} + b_f), \quad (14)$$

$$o^t = \sigma(W_o g^{t-1} + U_o \bar{h} + b_o), \quad (15)$$

$$f_0^t, \dots, f_{n+1}^t, f_g^t = \text{softmax}(\hat{f}_0^t, \dots, \hat{f}_{n+1}^t, \hat{f}_g^t), \quad (16)$$

$$c_g^t = f_g^t \cdot c_g^{t-1} + \sum_i f_i^t \cdot c_i^{t-1}, \quad (17)$$

$$g^t = o^t \cdot \tanh(c_g^t), \quad (18)$$

$f_0^t, \dots, f_{n+1}^t$  和  $f_g^t$  是控制来自前一时刻所有单词隐藏状态和前一时刻句子状态的信息流的‘门’.  $o^t$  是输出门,  $W, U, b$  为可以训练的模型参数.

由于 S-LSTM 将句子作为一个整体进行循环, 使每一个单词与句子整体状态不断得到更新, 而不是类似于 Bi-LSTM<sup>[14]</sup> 仅仅在单词序列上按照固定的顺序进行循环, S-LSTM 可以比 Bi-LSTM 更好地探究整体(句子)与部分(单词)的关系, 并且在句子较长时训练迭代的次数也会相对较少. 参考文献中的实验结果也证明, S-LSTM 在多种任务上对信息的提取效果好于 Bi-LSTM. 因此我们考虑将 S-LSTM 作为一种特征提取器运用于意图检测与槽填充的联合任务中.

## 2.2 存在的问题

在将 S-LSTM 运用于意图检测与槽填充的联合任务中时, 最直接的想法是将 S-LSTM 输出的每一个单词状态作为输入, 输入到分类器当中从而得到最后的结果. 但由于每一个时刻的单词隐藏状态都仅受上一个时刻句子状态的影响, 与当前句子状态没有直接的监督关系, 而最新时刻的句子隐藏状态蕴含着最为准确以及详尽的句子语义信息, 直接进行分类就无法利用最新迭代时刻的句子信息, 这样会

造成资源的浪费, 且通过我们的实验表明这样做对最终的结果造成了一定的影响. 为了充分利用最新迭代时刻的句子状态, 我们想到运用‘槽值门’机制使最新迭代时刻的句子状态对单词状态产生监督.

## 3 ‘槽值门’机制

‘槽值门’(slot-gate)机制<sup>[13]</sup>在2018年被提出, 作者将由 Bi-LSTM<sup>[14]</sup>提取出的每个单词的隐藏状态和 Bi-LSTM 中最后一个单元的隐藏状态分别经过槽值自注意力(slot self-attention)机制和意图自注意力(intent self-attention)机制之后得到的结果输入到‘槽值门’中, 用意图分类的结果监督槽值的填充. 而在我们的模型中, ‘槽值门’的输入为单词的状态表示和句子整体的状态表示. 本文中所使用的‘槽值门’结构如图2所示.

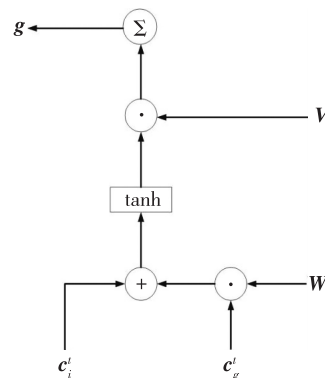


图2 “槽值门”结构图

Fig. 2 Structure of slot-gate

用公式表达即为

$$g = \sum V \cdot \tanh(c_i^t + W \cdot c_g^t), \quad (19)$$

其中,  $c_i^t$  代表  $t$  时刻第  $i$  个单词的隐藏状态,  $c_g^t$  代表  $t$  时刻句子的整体状态,  $W$  和  $V$  为可以训练的模型参数, 输出  $g$  为对应的权值. 在实际运用的过程中, 不论是单词还是句子的状态都选用固定长度的向量来表示, 因此两者之间可以进行直接相加, 且不需要做矩阵乘法来达到改变维度的目的. 通过  $W$  矩阵将句子状态变换到与单词状态相同的向量空间, 之后通过  $\tanh$  函数作为激活函数, 将较小的差距放大使其更加明显. 然后通过  $V$  进一步改变向量空间之后求和得出最终的输出权值.

## 4 模型介绍

在提出‘槽值门’的参考文献<sup>[13]</sup>中, 作者采用了 Bi-LSTM 的方式提取句子中的信息, 包括句子整

体状态和单词状态.而 S-LSTM 可以在速度不变慢的前提下更好地提取上述两种信息,且门机制可以很好地将最新时刻的句子状态加以利用.因此本文将 S-LSTM 与‘槽值门’机制结合起来,用当前时刻的句子状态(可以在某种程度上被理解为句子的意图)监督槽填充的过程,并将槽填充与意图检测任务在同一系统中接受训练.

对每一句输入的语句首先经过 S-LSTM 的特征提取,得到此时刻每个单词的表示  $c_i^t$  与此时刻句子整体状态表示  $c_g^t$ , 在此基础上,完成槽填充任务和意图检测任务.

#### 4.1 槽填充

在得到  $c_i^t$  与  $c_g^t$  之后,将其输入到‘槽值门’机制中,计算出此单词在此句子状态下的权重.最后的输出如下式所示:

$$o_i^t = n \cdot c_i^t + c_g^t \cdot g, \quad (20)$$

权值  $g$  代表此时刻句子状态对此单词状态的监督.为了保证原本单词状态中来自单词之间和之前时刻句子状态的信息不会被过度修改,我们在接受监督的单词信息之外加上了原本的单词信息. $n$  为可以调整的超参数,当迭代次数较多时,因为单词状态已经较好地掌握了句子的整体状态,所以可以将  $n$  设置得比较大,赋予原单词状态较大的权重,同理当迭代次数较少时,可以将  $n$  设置得较小.

我们将得到的  $y_i^s$  输入到条件随机场<sup>[15]</sup>(CRF)当中来计算每一个单词对应槽值的条件概率:

$$P(Y_1^n | \mathbf{o}, \mathbf{W}_s, \mathbf{b}_s) = \frac{\prod_{i=1}^n \psi_i(y_{i-1}, y_i, \mathbf{o})}{\sum_{Y'} \prod_{i=1}^n \psi_i(y'_{i-1}, y'_i, \mathbf{o})}, \quad (21)$$

$$\psi_i(y_{i-1}, y_i, \mathbf{o}) = \exp(\mathbf{W}_s^{y_{i-1}, y_i} o_i + b_s^{y_{i-1}, y_i}), \quad (22)$$

其中,  $\mathbf{W}_s^{y_{i-1}, y_i}$  和  $b_s^{y_{i-1}, y_i}$  对于每一对连续的标签来说都是不同的.将最终 CRF 的输出结果输入到 softmax 当中得到最终的槽值填充结果.

#### 4.2 意图检测

过去的许多研究表明仅仅使用最终的句子隐藏状态所得到的效果比不上通过注意力机制(attention)将每一个单词的隐藏状态进行加权求和之后分类的结果,因此我们在意图检测中引入注意力机制.我们采用加性注意力机制<sup>[16]</sup>,  $t$  时刻进行分类判别的句子整体状态  $g^t$  为

$$g^t = \sum_i \alpha_i^t h_i^t, \quad (23)$$

其中:

$$\alpha_i^t = \frac{\exp(\mathbf{u}^T \theta_i^t)}{\sum_n \exp(\mathbf{u}^T \theta_n^t)}, \quad (24)$$

$$\theta_i^t = \tanh(\mathbf{W}_\alpha h_i^t + \mathbf{b}_\alpha), \quad (25)$$

$\mathbf{W}_\alpha, \mathbf{b}_\alpha$  与  $\mathbf{u}$  为可以训练的模型参数.

对  $t$  时刻句子整体状态的意图分类判别通过 softmax 进行:

$$\mathbf{y} = \text{softmax}(\mathbf{W}_c \mathbf{g}^t + \mathbf{b}_c), \quad (26)$$

$\mathbf{y}$  为预测的分类标签结果,  $\mathbf{W}_c$  与  $\mathbf{b}_c$  为可训练的模型参数.

## 5 实验

### 5.1 数据集

ATIS<sup>[17]</sup>(Airline Travel Information Systems)数据集,是一个英文数据集,它被广泛应用于 SLU 的研究中.数据集中记录了人们在预定机票时所作的表述,其中训练集中包含了 4 478 条语句,测试集中包含了 893 条语句,验证集中包含了 500 条语句.在训练集中包含了 120 种不同种类的槽标签和 21 种不同的意图种类.

为了更全面地验证模型的能力,本文还引入另外一个由 Snips2017 年收集的消费意图引擎<sup>[18]</sup>(custom-intent-engines)的 NLU 数据集.其中训练集包括了 13 084 条语句,测试集中包括了 700 条语句,验证集包含 700 条语句.此数据集中每一种意图的语句数量基本相同,且包括了 72 种不同的槽标签以及 7 种不同的意图种类.两种数据集的具体情况如表 1 所示.

表 1 ATIS 数据集以及 Snips 数据集的统计信息

Table 1 Statistics of ATIS and Snips datasets

	ATIS	Snips
词汇数量/个	722	11 241
槽值种类数/种	120	72
意图种类数/种	21	7
训练集大小/条	4 478	13 084
验证集大小/条	500	700
测试集大小/条	893	700

相较于单独领域的 ATIS 数据集, Snips 数据集涉及了更多的领域,且意图更为多样.例如: Snips 数据集中的意图从“播放音乐”到“查询天气”,每种意图之间差别很大,而 ATIS 数据集中,所有的意图都始终围绕着机票信息.这也就导致 ATIS 数据集由于

始终关注机票这一话题,所以涉及的单词重复率很高.由于所涉及话题较为多样,Snips 数据集所涉及的单词量较多.

同时,ATIS 数据集的训练样本极不均衡,训练样本中询问航班信息及预定机票的样本超过了 50%,而询问最低票价的样本仅出现了一次.而 Snips 数据集的训练样本较为均衡,不同类别样本数量差异较小.

## 5.2 设置

经过多次实验,我们将 S-LSTM 的迭代次数设定为 6 次,单词窗口长度设定为 2,句子状态数为 1 个,由于迭代次数较多, $n$  设置为 2,每一批数据的大小(batch-size)为 30 条,为了防止过拟合,将‘dropout’<sup>[19]</sup>值设置为 0.5.同时,为了防止梯度爆炸,需要引入梯度裁剪<sup>[20]</sup>(gradient client),将每一批数据裁剪为 3 批各 10 条数据.在实际实验的过程中引入预训练词向量数据集 GloVe<sup>[21]</sup>中 300 维的嵌入,将所有内部隐藏状态的维度设定为 300 维.所有的梯度下降均采用 Adam 调整器来实现.

## 5.3 实验结果及分析

对于意图检测的结果,本实验选择使用准确率作为衡量,而对于槽值填充的结果,本实验选择使用 F1 值进行衡量,它将精确率和召回率视为同等重要.

本实验选取基于注意力机制的循环神经网络模型和‘槽值门’模型作为此次实验的参照对象.并且,本实验也将没有经过任何处理的 S-LSTM 输出直接进行槽填充任务的结果与叠加槽值门之后的模型(没有使用 CRF)进行了比较,具体数据如表 2 所示.

表 2 实验结果与对比

模型	ATIS 数据集		Snips 数据集	
	槽值 (F1 值)	意图 (准确率)	槽值 (F1 值)	意图 (准确率)
文献[12]	94.2	91.1	87.8	96.7
文献[13]	95.2	94.1	88.8	97.0
直接输出	92.5		90.3	
本文 不使用 CRF	92.6	94.4	90.5	97.8
使用 CRF	95.4		93.4	

由此可见,在结合‘槽值门’之后,模型的效果得到了提升.经过 CRF 处理之后,本实验的模型在 Snips 数据集上不论是槽值填充还是意图检测均取得了极大的进步,而在 ATIS 数据集上也达到并超越了目前最先进算法的结果.

之所以在 ATIS 数据集上取得的提升没有 Snips 数据集上大,我们认为可能是因为以下两点原因:

1) ATIS 数据集中的话题集中性较高,基本都是围绕着机票这一主题展开,这方便了模型对其整体信息的学习与把握,同样也难以体现出 S-LSTM 对整体和个体信息交流的优势.

2) ATIS 数据集中由于训练样本的不均衡,导致模型比较容易学会样本较多的分类,而对于出现次数较少的样本模型难以学习透彻,无法具有很好的泛化能力,这就导致对 ATIS 数据集的学习可以较为容易地达到一个较高水平,但想要继续突破较为困难.

## 6 结论

本文将 S-LSTM 模型应用到意图检测和槽填充的联合任务上,用以替代 Bi-LSTM 作为信息的提取工具,并引入‘槽值门’机制,将最新时刻句子整体状态信息加以利用,从而进一步提升模型表现.最终得到了优于目前最先进算法的表现,这既证明了 S-LSTM 可以更好地提取句子中的信息也预示着对意图和槽值之间关系的建模与进一步挖掘会为意图检测和槽填充的联合任务的进步提供动力.在未来的工作中,可以考虑使用图神经网络(Graph Neural Network)<sup>[22]</sup>以便更为明确地对句子中信息进行建模与应用.

## 参考文献

### References

- [1] Brenes D J, Gayo-Avello D, Pérez-González K. Survey and evaluation of query intent detection methods [C] // Proceedings of the 2009 Workshop on Web Search Click Data, 2009:1-7
- [2] Tur G, De Mori R. Spoken language understanding: systems for extracting semantic information from speech [M]. Chichester, UK: John Wiley & Sons, 2011
- [3] Gupta A, Hewitt J, Kirchoff K. Simple, fast, accurate intent classification and slot labeling [J]. arXiv Preprint, 2019, arXiv:1903.08268
- [4] Zhang Y, Liu Q, Song L F. Sentence-state LSTM for text representation [C] // Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2018:317-327
- [5] Gorin A L, Parker B A, Sachs R M, et al. How may I help you? [C] // Proceedings of Workshop on Interactive Voice Technology for Telecommunications Applications, DOI:10.1109/ivtta.1996.552741
- [6] Yao K S, Peng B L, Zhang Y, et al. Spoken language understanding using long short-term memory neural

- networks [ C ] // IEEE Spoken Language Technology Workshop ( SLT ), 2014:189-194
- [ 7 ] Sarikaya R, Hinton G E, Ramabhadran B. Deep belief nets for natural language call-routing [ C ] // 2011 IEEE International Conference on Acoustics, Speech and Signal Processing ( ICASSP ), 2011, DOI: 10.1109/ICASSP.2011.5947649
- [ 8 ] Schapire R E, Singer Y. BoosTexter: a boosting-based system for text categorization [ J ]. Machine learning, 2000, 39 ( 2/3 ): 135-168
- [ 9 ] Guo D, Tur G, Yih W T, et al. Joint semantic utterance classification and slot filling with recursive neural networks [ C ] // 2014 IEEE Spoken Language Technology Workshop ( SLT ), 2014, DOI: 10.1109/SLT.2014.7078634
- [ 10 ] Hakkani-Tür D, Tur G, Celikyilmaz A, et al. Multi-domain joint semantic frame parsing using bi-directional RNN-LSTM [ C ] // Interspeech 2016, ISCA, 2016:715-719
- [ 11 ] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate [ J ]. arXiv Preprint, 2014, arXiv:1409.0473
- [ 12 ] Liu B, Lane I. Attention-based recurrent neural network models for joint intent detection and slot filling [ J ]. arXiv, 2016, arXiv:1609.01454
- [ 13 ] 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, 2:753-757
- [ 14 ] Schuster M, Paliwal K K. Bidirectional recurrent neural networks [ J ]. IEEE Transactions on Signal Processing, 1997, 45( 11 ):2673-2681
- [ 15 ] Ma X Z, Hovy E. End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF [ J ]. arXiv Preprint, 2016, arXiv:1603.01354
- [ 16 ] Zhang X D, Wang H F. A joint model of intent determination and slot filling for spoken language understanding [ C ] // Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, 2016:2993-2999
- [ 17 ] Tur G, Hakkani-Tur D, Heck L. What is left to be understood in ATIS? [ C ] // 2010 IEEE Spoken Language Technology Workshop, 2010, DOI: 10.1109/SLT.2010.5700816
- [ 18 ] 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 Preprint, 2018, arXiv:1805.10190
- [ 19 ] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting [ J ]. The Journal of Machine Learning Research, 2014, 15 ( 1 ):1929-1958
- [ 20 ] Pascanu R, Mikolov T, Bengio Y. On the difficulty of training recurrent neural networks [ C ] // International Conference on Machine Learning, 2013:1310-1318
- [ 21 ] Pennington J, Socher R, Manning C. Glove: global vectors for word representation [ C ] // Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing ( EMNLP ), 2014:1532-1543
- [ 22 ] Zhou J, Cui G Q, Zhang Z Y, et al. Graph neural networks: a review of methods and applications [ J ]. arXiv Preprint, 2018, arXiv:1812.08434

## Speaker intention recognition based on S-LSTM model and slot-gate

WANG Ziyue<sup>1</sup> SHAO Xi<sup>1</sup>

<sup>1</sup> College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003

**Abstract** The recognition of the speaker's intention has greatly promoted the development of natural language understanding. In previous studies, the bidirectional long short-term memory (Bi-LSTM) model has been mostly employed in natural language processing to extract the features of words and the relationships between them. However, Bi-LSTM cannot establish a well-enough relation between the information contained in a sentence and its individual vocabulary. Another previously proposed model, i.e., the S-LSTM (Sentence-state LSTM) model, can establish a relation between sentence information and its individual words. This, in turn, facilitates the establishment of the relationship between intention detection and slot filling, for the purpose of proposing a joint model to better understand the semantics contained in the question-answer system. Therefore, in this paper, slot-gate mechanism is introduced to solve the waste of the latest iteration sentence state when S-LSTM is applied to the joint task of intention detection and slot filling. The experimental results based on ATIS and Snips datasets confirm that the proposed mechanism is superior to other state-of-the-art approaches.

**Key words** intention detection; slot filling; S-LSTM; slot-gate