A Collaborative-Transformer for Joint Spoken Language Understanding

Anonymous EMNLP submission

Abstract

A spoken language understanding (SLU) system mainly includes two tasks: slot filling task and intent detection task. These two tasks are closely related and highly dependent on each other. In this paper, we propose a collaborative-transformer for joint SLU to explicitly model the information interaction between intent and slot, which can mutually promote the performance of two tasks, thus establishing the correlation between intent and slot deeply. In addition, our model can also obtain the explicit representation of intent and slot, which is the basis of explicit interaction between intent and slot. The experimental results on the two data sets are explicit, and our model performs better than other models, achieving the state of the art. The experimental results on the two data sets show that our model performs better than other models, and achieves the state-of-the-art.

1 Introduction

Spoken language understanding is an important part of the task-based dialogue system, which can identify the intent and extract the corresponding slots form user's utterances. Therefore, SLU consists of two subtasks: intent detection and slot filling (Tur and De Mori, 2011). For example, as shown in Table 1, the intent of the sentence "from seattle to salt lake city" is "atis_flight", and each word corresponds to a slot label.

from	seattle	to	salt	lake	city
O	B-fromloc.city_name	O	B-toloc.city_name	I-toloc.city_name	I-toloc.city_name
atis_flight					

Table 1: An example with intent and slot annotation(BIO format).

The traditional method uses pipeline to solve these two tasks. Specifically, intent detection is a separate classification task, and slot extraction is a separate sequence labeling task, but there is a strong correlation between these two tasks (Goo et al., 2018).

For instance, the intent of the sentence "from seattle to salt lake city" is "flight", so the slot label of the word "seattle" is more inclined to "fromloc.city_name" rather than "city_name". Conversely, when the slot labels for the word "seattle" and the phrase "salt lake city" are "fromloc.city_name" and "toloc.city_name", the intent of the discourse is most likely "flight". Due to the correlation between these two tasks, some work uses joint model to explore the connection between them.(Zhang and Wang, 2016; Hakkanitur et al., 2016; Liu and Lane, 2016)

Benefiting from the advantages of parameter sharing brought by multi-task learning, these models are superior to the pipeline-based method. However, these works did not further explore the relationship between intent detection and slot filling. Goo et al. (2018) and Li and Li (2018) use the gating mechanism to selectively add intent information to the slot filling task.

Qin et al. (2019) proposed a joint model with Stack-Propagation for SLU tasks, which will identify the intent of each token and input the intent representation of each token into the slot filling task. If the intent of some tokens is predicted incorrectly, the intent of other tokens will also be useful for the slot filling of each token. Wang et al. (2018) established two task networks for the intent detection task and the slot filling task. The two task networks are iterative during the training process, and the prediction results of the other task will be used when training one task.

Goo et al. (2018), Li and Li (2018) and Qin et al. (2019) are only to explore how to better integrate the intent information into the slot filling task. Although Wang et al. (2018) considers inputting the results of slot filling into the intent detection task, iterative training cannot model the interaction directly between them. Zhang et al. (2019) propose a hierarchical capsule neural network to model the the hierarchical relationship among word, slot, and intent in an utterance. E et al. (2019) introduced an SF-ID network, which includes SF sub-network and ID sub-network. The two sub-networks iteratively achieve the flow of information between intent and slot. Although these works have achieved good results on SLU tasks, their models still suffer from two issues:

1)in these models, information is only transferred in one direction at the same time, such as from intent to slot, and then from slot to intent. The information interaction between intent and slot is not completed at the same time, which limits their performance and make these networks unable to adequately model the complex interactive information between intent and slot. 2)None of these models obtain an explicit representation of intent and slot, which is very important for the information interaction between intent and slot.

In this paper, we propose a collaborative-transformer to solve the above two challenges.

- First, inspired by the LAN (Cui and Zhang, 2019)(Hierarchically-Refined Label Attention Network for Sequence Labeling) model, we introduced intent label embedding and slot label embedding and designed the intent label attention layer and slot label attention layer. In this way, we can explicitly get the intent representation and slot representation.

- Second, we designed collaborative interactive attention mechanism, which can complete the two-way information interaction between the intent and the slot inside the encoder. And the token level and semantic level association between these two task will also be captured.

- Finally, the stacking of multi-layer collaborative transformer blocks allows intent representation and slot representation to deeply integrate the information of each other to improve the performance of the two tasks.

We conduct experiments on two benchmark datasets of ATIS (Coucke et al., 2018) and SNIPS (Goo et al., 2018), Experimental results show that the performance of our model exceeds the current state-of-the-art in multiple evaluation indicators.

The contributions of Collaborative-Transformer Network are as follows:

1)We propose an collaborative-interactive attention mechanism, which can not only use intent knowledge to improve the performance of slot filling tasks, but also use slot knowledge to guide the detection of intent, and the information interaction between intent and slot is explicit and simultaneous . 2)We constructed intent label attention and slot label attention to explicitly obtain the representation of intent and slot, which established a foundation for the interaction of information between intent and slot. 3)The experimental results on two public data sets show the superiority of our model in performance, and our model has achieved state-of-the-art on multiple indicators.

2 Proposed model

In this section, we will introduce each sub-network of the Collaborative-Transformer model in detail. First, we use BiLSTM as a shared encoder to obtain the context sequence information of the input sentence, which can capture the shared features between the intent detection task and the slot filling task. Next we designed stacked collaborative-transformer blocks to model the interaction between intent and slot. Finally, the intent representation and slot representation of the collaborative-transformer block output of the last layer are input to the softmax layer and the CRF layer respectively for intent detection and slot filling. We use joint learning to train the model in figure 1.

2.1 Shared encoder

LSTM is widely used in classification tasks and sequence labeling tasks, and as a baseline model for these tasks. We choose BiLSTM as the shared encoder layer of our model. The output of each utterance word embedding layer is $E = (e_1, e_2, ..., e_t)$. After input to BiLSTM, we can get the hidden state of this utterance $H = (h_1, h_2, ..., h_t)$. As a shared representation of intent and slot, H can provide context information and time series information for collaborative-transformer block, which are useful for slot filling tasks.

2.2 Collaborative-Transformer block

Collaborative-transformer block is the main innovation of this paper. Each collaborative-transformer block includes three networks: intent and slot label attention layer; collaborative interactive attention layer for intent representation and slot representation, and a feed-forward network layer for informa-

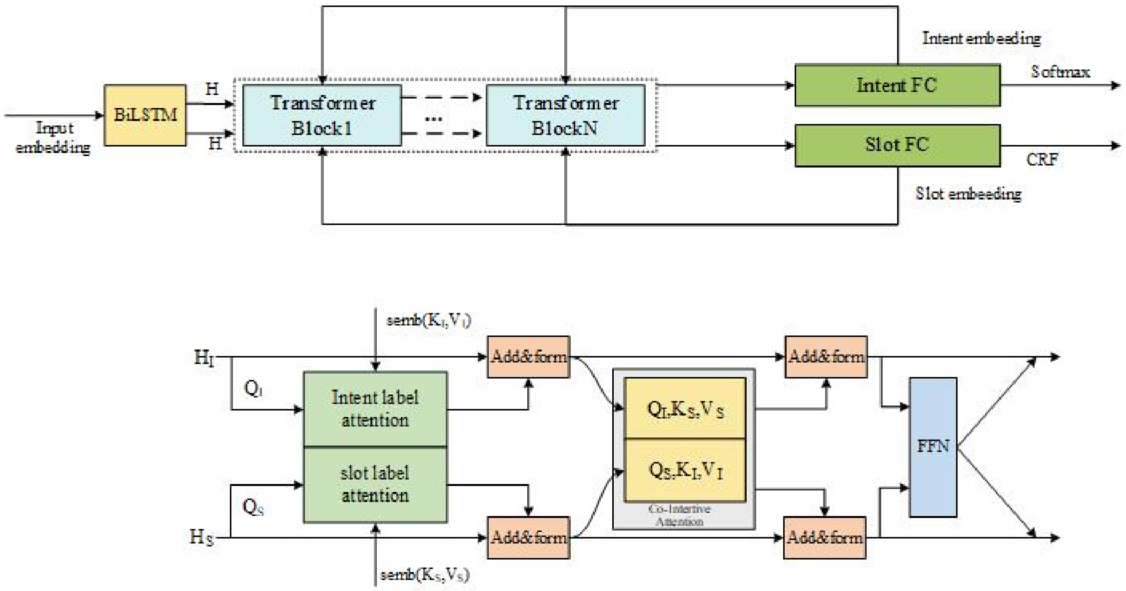

Figure 1: Illustration Collaborative-Transformer for joint SLU. It contains a shared BiLSTM encoder and stacked Collaborative-Transformer Block.

tion fusion.

2.3 Intent and slot label attention layer

In order to get the intent representation and slot representation explicitly, before collaborative interactive attention, we built the intent label attention layer and slot label attention layer. Specifically, we use the parameters of the Slot-FC layer and Intent-FC layer as slot embedding S_{emb} and intent embedding I_{emb} . The parameter matrix W of the output layer can be regarded as the distribution of labels in a certain sense.

For the intent filling task, we use the hidden state H as the query, S_{emb} as the key and value, and then obtain the intent representation through the intent label attention layer. The attention calculation process is as follows

$$A = softmax(HS_{emb}^T) \tag{1}$$

$$H_I = H + AI_{emb}^T \tag{2}$$

There is no additional parameter in the calculation process. And the slot attention layer is the same as intent attention layer. A represents the attention matrix, which can also be regarded as a fuzzy classification result of intent in token level. Therefore, we can get the intent representation H_I and the slot representation H_S. The role of the label attention layer in the first layer of collaborative transformer block is to explicitly obtain H_I and H_S, and the label attention layer in the subsequent $N-1$ layer is mainly for the interaction between label and representation.

2.4 Collaborative interactive attention layer

In our collaborative-transformer block, the role of the collaborative interactive attention layer is information interaction between intent and slot. Intent knowledge and slot knowledge are integrated into intent representation and slot representation respectively, so we can get a slot-aware intent representation and a slot-aware intent representation. The input of the collaborative interactive attention layer is the intent representation $H_I = (h_1, h_2, .., h_I)$ and slot representation $H_S = (h_1, h_2, ...h_S)$ generated by the label attention layer. If the head is set to 1, the calculation process of collaborative interactive attention is as follows:

$$Q_I, K_I, V_I = H_I W_I^Q, H_S W_I^K, H_S W_I^V \tag{3}$$

$$Q_S, K_S, V_S = K_I^T, Q_I^T, H_S, W_S^V \tag{4}$$

$$A_I = softmax\frac{(Q_I K_I^T)}{\sqrt{d_k}} \tag{5}$$

$$A_S = softmax\frac{(Q_S K_S^T)}{\sqrt{d_k}} \tag{6}$$

$$A_I^T = softmax\frac{(Q_I K_I^T)^T}{\sqrt{d_k}} \tag{7}$$

$$H_I^{new} = H_I + A_I V_I \tag{8}$$

$$H_S^{new} = H_S + A_S V_S \tag{9}$$

Where Q_I, K_I, V_I are intent query, key and value, Q_S, K_S, V_S represent query, key and value of slot, W_I^Q,W_I^K, W_I^V, W_S^V are weight matrix corresponding to Q_I, K_I, V_I, V_S; Q_I andK_I are the transpose of K_S and Q_S, and A_I is the transpose of A_S.

In fact, the attention matrix A_I is generated by the inner product of intent query $Q_I(K_S)$ and slot query $Q_S(K_S)$, and the collaborative attention is symmetrical about intent and slot. In this way, the intent detection task and the slot filling task complete the knowledge interaction and establish a token-level connection.

2.5 Feed-forward network layer for information fusion

In the previous section, we explicitly modeled the interaction between intent and slot through collaborative interactive attention, and got the new intent representation H_I^{new} and slot representation H_S^{new}. In this section, we use feed-forward network layer to implicitly fuse intent knowledge and slot knowledge.

First, we concatenate H_I and H_S as r_{is}, and then we concatenate each r_{is}^t of r_{is} with r_{is}^{t-1} and r_{is}^{t+1} as r^t :

$$r_{is} = H_I^{new} \oplus H_S^{new} \tag{10}$$
$$r^t = r_{is}^{t-1} \oplus r_{is}^t \oplus r_{is}^{t+1} \tag{11}$$

where $\oplus$ is concatenation operation.

Next, the FFN layer fuses the window features, intent features and slot features contained in r.

$$FFN(r) = max(0, rW_1 + b_1)W_2 + b_2 \tag{12}$$
$$O_I = H_I^{new} + FFN(r) \tag{13}$$
$$O_S = H_S^{new} + FFN(r) \tag{14}$$

The fusion of intent and slot in the FFN layer is equivalent to the implicit interaction between intent and slot. FFN layer automatically integrates intent knowledge and slot knowledge, and window features are very useful for slot filling task (Zhang and Wang, 2016).

2.6 Joint Training

The output of the last layer of the collaborative transformer block, $O_I = (O_1, O_2, ..., O_t), O_S = (O_1, O_2, ..., O_t)$, serves as the final representation of intent and slot. For the intent detection task, O_I is used for intent prediction after the output layer is mapped:

$$v_I = maxpool(O_I^{final}) \tag{15}$$
$$y_{I\_pred} = softmax(W_I v_I + b_I) \tag{16}$$

Where y is the predicted distribution of intent and the loss function is formulated as:

$$L_{inte} = \sum y_{I\_label}^i \log(y_{I\_label}^i) \tag{17}$$

For the slot filling task, to take advantage of the global dependencies between labels we use CRF to search for the global optimal prediction sequence. O_S is first mapped to the label distribution Y_S by the output layer:

$$Y_S = WO_S + b_S \tag{18}$$

Then we input Y_S into the CRF, and use the Viterbi algorithm to decode the prediction sequence y_{I_pred}. Given a sequence $x = [x_1, x_2, ..., x_t]$, the corresponding golden label sequence is $y = [y_1, y_2, ..., y_t]$, the loss function is formulated as:

$$L_{slot} = \log(p(y_{S\_label} \mid x)) \tag{19}$$
$$= -(s(x, y_{S\_label}) - \log(\sum_{\widetilde{y} \in Y_x} e^{s(x,\widetilde{y})}) \tag{20}$$

Where $S(\cdot)$ represents the score of a path, Y_x is all possible label sequences.

We use joint training to optimize these two tasks simultaneously. Compared with the pipeline model, the joint training method can effectively reduce the error propagation (Zhang and Wang, 2016). The final joint objective is formulated as :

$$L = \lambda L_{intent} + (1 - \lambda)L_{slot} \tag{21}$$

Where λ is the hyperparameter combining the two losses.

3 Experiments

3.1 Experimental Settings

To fully evaluate our model, we conduct experiments on two public benchmark datasets, ATIS (Hemphill et al., 1990) and Snips (Coucke et al., 2018). In order to ensure the fairness of the comparative experiment, our data processing method

	ATIS	SNIPS
Num of slots label	120	72
Num of intents label	21	7
Vocabulary size	722	11241
Traing set	4478	13084
Dev set	500	700
Test set	893	700
Tagging scheme	BIO	BIO

Table 2: Details of dataset and tagging scheme.

is consistent with Qin et al. (2019). The details of these two dataset are shown in the Table 2. For both ATIS and Snips, $300d$ GloVe vector (Pennington et al., 2014)is used as word embedding. The BiLSTM and transformer-block hidden size are set as 128. $L2$ regularization is used on our model is 1×10^{-6} and dropout ratio is adopted is 0.1 for reducing overfit. We use Radam (Liu et al., 2019) to optimize the parameters in our model. The layer of collaborative-transformer block is set to 3 as hyper-parameters of the model. The weight coefficient A of the loss function is set to 0.2. For all the experiments, we select the model which works the best on the dev set, and then evaluate it on the test set.

3.2 Baselines

We compare our model with the existing baselines including:

- Joint Seq. Hakkanitur et al. (2016) proposed a multi-task modeling approach for jointly modeling domain detection, intent detection, and slot filling in a single recurrent neural network (RNN) architecture.
- Attention BiRNN. Liu and Lane (2016) leveraged the attention mechanism to allow the network to learn the relationship between slot and intent.
- Slot-Gated Atten. Goo et al. (2018) proposed the slot-gated joint model to explore the correlation of slot filling and intent detection better.
- Self-Attentive Model. Li and Li (2018) proposed a novel self-attentive model with the intent augmented gate mechanism to utilize the semantic correlation between slot and intent.
- Bi-Model. Wang et al. (2018) proposed the Bi-model to consider the intent and slot filling cross-impact to each other.
- CAPSULE-NLU. Zhang et al. (2019) proposed a capsule-based neural network model with a dynamic routing-by-agreement schema to accomplish slot filling and intent detection.
- SF-ID Network. E et al. (2019) introduced an SF-ID network to establish direct connections for the slot filling and intent detection to help them promote each other mutually.
- Stack-Propagation. Qin et al. (2019) proposed a joint model with Stack-Propagation which can explicitly use the intent information as input for slot filling, thus to capture the intent semantic knowledge.

For all the models above, we adopt the reported results from Qin et al. (2019).

3.3 Overall Results

Following Qin et al. (2019), we treated slot filling and intent prediction as sequence labeling problem and classification problem respectively, thus evaluate the proposed model performance by F1 score and accuracy. The sentence-level semantic frame parsing using overall accuracy. Table 3 gives the experiment results of the proposed models on SNIPS and ATIS.

As shown in Table 3, our model significantly outperforms the baseline models on both datasets and achieves the state-of-the-art results. Compared with the model of Qin et al. (2019) on Snips, our model achieves intent classification accuracy of 98.9% (from 98.0%), slot filling F1 of 96.0% (from 94.2%), and sentence-level semantic frame accuracy of 90.3% (from 86.9%). On ATIS, our model achieves intent classification accuracy of 95.95% (from 95.9%), slot filling F1 of 97.5% (from 96.9%), and sentence-level semantic frame accuracy of 87.4% (from 86.5%).

Compared to ATIS, Snips includes multiple domains and has a larger vocabulary, For the more complex Snips dataset, our model achieves a large gain in the sentence-level semantic frame accuracy. The reason is that our model explicitly models the interaction.

Compared with the model that only passes the intent information to the slot filling task, our model has achieved the expected improvement in the intent detection task. The improvement of the intent detection task will also pass more accurate intent knowledge to the slot filling task to achieve the effect of mutual promotion.

3.4 Ablation experiment

In Section 4.3, we describe the improvement of the model on two data sets. In this section, we will explore the reasons why our model works by ablation experiments. First, we explored the effect of the label attention layer on model performance. Next, we studied the role of collaborative interactive attention in the co-transformer block. Finally, we studied the effect of the FNN layer on our model.

3.4.1 Effect of label attention layer

In this section, we set up the following ablation experiments to study the impact of the label attention layer.

1)We remove the intent attention layer and replace Hi with Hidden State H of BiLSTM. This means that we only get the slot representation explicitly.

2)We remove the slot attention layer and replace Hi with Hidden State H of BiLSTM. This means that we only get the intent representation explicitly.

As shown in Table 4, compared with the complete model, when we only use the explicit representation of the slot (with out intent label attention), the performance of the model is degraded on both tasks. And when we only use the explicit representation of intent (with out slot label attention), we also get a similar result. We believe that this is because only the explicit representation of the slot or only the explicit representation of the intent will affect the interaction of the collaborative interactive attention model information, which makes the slot representation only interact with the shared representation. This also proves the importance of obtaining explicit representation of intent. The lack of explicit representation of intent or slot will make the transfer of knowledge between intent and slot incomplete.

3.4.2 Effect of collaborative interactive attention layer

In order to confirm the importance of our proposed collaborative interactive attention and verify the impact of collaborative interactive attention on model performance, we use self-attention instead of collaborative interactive attention in the ablation experiment, and the input is changed to the concatenation of intent representation and slot representation. This replacement means the interaction is adapted from explicit querying each other's information to implicit fusion. And we name it as self-interactive attention.

As shown in Table 4, after using self-attention instead of collaborative interactive attention, the performance of the model on the three indicators in the two data sets has a relatively large decrease. The experimental results show that compared to implicit interactions, using collaborative interactive attention to explicitly model the interaction between intent representation and slot representation can better capture the complex correlation of two tasks.

3.4.3 Effect of FFN layer

In order to explore the effect of the FFN layer on the final effect of the model, we also removed the FFN layer in the ablation experiment. The experimental results show that slot F1 has different degrees of decline on the two data sets, intent accuracy drops slightly. The reason is that the slot filling task is essentially a sequence labeling task, and the window information captured by the FFN layer is useful for the sequence labeling task. The sample implicit fusion of the intent representation and slot representation in FFN layer will further enhance the two representations, by the stacking of multiple layers of collaborative transformer blocks.

4 Related Work

The intent detection task is regarded as a sentence classification task. At present, many text classification methods are used to solve this task, such as support vector machine (SVM), recurrent neural network (RNN), convolutional neural network (CNN) (Haffner et al., 2003; Sarikaya et al., 2011; Hashemi et al., 2016). The slot filling task is solved as a sequence labeling task. Common methods are conditional random field (CRF), recurrent neural network (RNN), and LSTM-CRF (Huang et al., 2015). Cui and Zhang (2019) Proposed LAN instead of CRF to capture the dependencies between labels.

In recent years, some methods based on joint training have avoided the error propagation of the pipeline method. Zhang and Wang (2016) first uses the GRU network for the slu task, and the intent detection and slot filling share the same GRU encoder. Liu and Lane (2016) proposed an attention-based encoder-decoder joint model for the two task. Benefiting from parameter sharing and joint training

Model	SNIPS			ATIS		
	Slot(F1)	intent(Acc)	Overall(Acc)	Slot(F1)	Intent(Acc)	Overall(Acc)
Join Seq (Hakkanitur et al., 2016)	87.3	96.9	73.2	94.3	92.6	80.7
Attention BiRNN (Liu and Lane, 2016)	87.8	96.7	74.1	94.2	91.1	78.9
Slot-Gated Full Atten (Goo et al., 2018)	88.8	97	75.5	94.8	93.6	82.2
Slot-Gated Intent Atten (Goo et al., 2018)	88.3	96.8	74.6	95.2	94.1	82.6
Self-Attentive Model (Li and Li, 2018)	90	97.5	81	95.1	96.8	82.2
Bi-Model (Wang et al., 2018)	93.5	97.2	83.8	95.5	96.4	85.7
CAPSULE-NLU (Zhang et al., 2019)	91.8	97.3	80.9	95.2	95	83.4
SF-ID Network (E et al., 2019)	90.5	97	78.4	95.6	96.6	86
Stack-Propagation(Qin et al., 2019)	94.2	98.0	86.9	95.9	96.9	86.5
Our model	**96.2**	**99.0**	**90.58**	**95.95**	**97.5**	**87.4**

Table 3: Details of dataset and tagging scheme.

Model	SNIPS			ATIS		
	Slot(F1)	intent(Acc)	Overall(Acc)	Slot(F1)	Intent(Acc)	Overall(Acc)
With out intent label attention	95.8	98.5	89.7	95.5	97.3	86.7
With out slot label attention	95.77	98.7	89.9	95.3	97.3	86.2
Self-interactive attention	95.1	98.0	87.5	95.6	96.6	86
With out FFN layer	95.7	98.8	90.2	95.6	96.6	86
Our model	**96.2**	**99.0**	**90.58**	**95.95**	**97.5**	**87.4**

Table 4: Ablation experiments on the Snips and ATIS datasets

between the two tasks, these models perform better than the pipeline model. However, these methods only rely on parameter sharing to establish the connection between the two tasks, and do not specifically model the representation of the slot or the representation of intent.

Recently, some work has begun to explore the enhancement of slot filling tasks by intent information. Goo et al. (2018) designed slot-gate to control the flow of information between intent filling and slot filling tasks. Li and Li (2018) proposed an intention enhancement gate mechanism to capture the dependency between intent and slot. Qin et al. (2019) proposed the Stack-Propagation framework to predict the intent of each token and pass it to the slot decoder, and in this way, the model can avoid slot filling errors caused by incorrect prediction of the overall intent of utterance. These models only model the flow of intent information to slot filling tasks, and do not pay attention to the impact of slot information on intent detection. Wang et al. (2018) proposed Bi-model to iteratively enhance the other task with the result of one task during the training process. Zhang et al. (2019) propose a hierarchical capsule neural network to model the the hierarchical relationship among word, slot, and intent in an utterance. E et al. (2019) introduced an SF-ID network, which includes SF sub-network and ID sub-network. The two sub-networks iteratively achieve the flow of information between intent and slot. This limits their performance and makes these networks unable to adequately model the complex interactive information between intents and slots. However, in these networks, information is only transferred from intent to slot, and then from slot to intent. The information interaction between intent and slot is not completed at the same time, which limits their performance and make these networks unable to adequately model the complex interactive information between intent and slot. Compared with these models, in our proposed collaborative transformer network, intent representation and slot representation can be obtained explicitly and interact at the same time. In this way, our model can better establish the correlation between intent and slot at the word level and semantic level.

Acknowledgments

The acknowledgments should go immediately before the references. Do not number the acknowledgments section. Do not include this section when submitting your paper for review.

References

Alice Coucke, Alaa Saade, Adrien Ball, Thodore Bluche, Alexandre Caulier, David Leroy, Clment Doumouro, Thibault Gisselbrecht, Francesco Caltagirone, and Thibaut Lavril. 2018. Snips voice platform: an embedded spoken language understanding system for private-by-design voice interfaces.

Leyang Cui and Yue Zhang. 2019. Hierarchically-refined label attention network for sequence labeling. pages 4113–4126.

Haihong E, Niu Peiqing, Chen Zhongfu, and Song Meina. 2019. A novel bi-directional interrelated model for joint intent detection and slot filling. pages 5467–5471.

Chihwen Goo, Guang Gao, Yunkai Hsu, Chihli Huo, Tsungchieh Chen, Kengwei Hsu, and Yunnung Chen. 2018. Slot-gated modeling for joint slot filling and intent prediction. 2:753–757.

Patrick Haffner, Gokhan Tur, and J H Wright. 2003. Optimizing svms for complex call classification. 1:632–635.

Dilek Hakkanitur, Gokhan Tur, Asli Celikyilmaz, Yunnung Chen, Jianfeng Gao, Li Deng, and Yeyi Wang. 2016. Multi-domain joint semantic frame parsing using bi-directional rnn-lstm. pages 715–719.

Homa B Hashemi, Amir Asiaee, and Reiner Kraft. 2016. Query intent detection using convolutional neural networks.

Charles T Hemphill, John J Godfrey, and George R Doddington. 1990. The atis spoken language systems pilot corpus. pages 96–101.

Zhiheng Huang, Wei Xu, and Kai Yu. 2015. Bidirectional lstm-crf models for sequence tagging. *arXiv: Computation and Language*.

Changliang Li and Liang Li. 2018. A self-attentive model with gate mechanism for spoken language understanding. pages 3824–3833.

Bing Liu and Ian Lane. 2016. Attention-based recurrent neural network models for joint intent detection and slot filling. pages 685–689.

Liyuan Liu, Haoming Jiang, Pengcheng He, Weizhu Chen, Xiaodong Liu, Jianfeng Gao, and Jiawei Han. 2019. On the variance of the adaptive learning rate and beyond. *arXiv: Learning*.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Libo Qin, Wanxiang Che, Yangming Li, Haoyang Wen, and Ting Liu. 2019. A stack-propagation framework with token-level intent detection for spoken language understanding. *arXiv: Computation and Language*.

Ruhi Sarikaya, Geoffrey E. Hinton, and Bhuvana Ramabhadran. 2011. Deep belief nets for natural language call-routing. In *IEEE International Conference on Acoustics*.

Gokhan Tur and Renato De Mori. 2011. *Spoken Language Understanding: Systems for Extracting Semantic Information from Speech*. Ph.D. thesis.

Yu Wang, Yilin Shen, and Hongxia Jin. 2018. A bimodel based rnn semantic frame parsing model for intent detection and slot filling. 2:309–314.

Kai Zhang, Hefu Zhang, Qi Liu, Hongke Zhao, and Enhong Chen. 2019. Interactive attention transfer network for cross-domain sentiment classification.

Xiaodong Zhang and Houfeng Wang. 2016. A joint model of intent determination and slot filling for spoken language understanding. pages 2993–2999.