A Novel Bi-directional Interrelated Model for Joint Intent Detection and Slot Filling

Haihong E[*], Peiqing Niu[*], Zhongfu Chen[*], Meina Song
Beijing University of Posts and Telecommunications, Beijing, China
{ehaihong,niupeiqing,chenzhongfu,mnsong}@bupt.edu.cn

Abstract

A spoken language understanding (SLU) system includes two main tasks, slot filling (SF) and intent detection (ID). The joint model for the two tasks is becoming a tendency in SLU. But the bi-directional interrelated connections between the intent and slots are not established in the existing joint models. In this paper, we propose a novel bi-directional interrelated model for joint intent detection and slot filling. We introduce an SF-ID network to establish direct connections for the two tasks to help them promote each other mutually. Besides, we design an entirely new iteration mechanism inside the SF-ID network to enhance the bi-directional interrelated connections. The experimental results show that the relative improvement in the sentence-level semantic frame accuracy of our model is 3.79% and 5.42% on ATIS and Snips datasets, respectively, compared to the state-of-the-art model.

1 Introduction

Spoken language understanding plays an important role in spoken dialogue system. SLU aims at extracting the semantics from user utterances. Concretely, it identifies the intent and captures semantic constituents. These two tasks are known as intent detection and slot filling (Tur and De Mori, 2011), respectively. For instance, the sentence *'what flights leave from phoenix'* sampled from the ATIS corpus is shown in Table 1. It can be seen that each word in the sentence corresponds to one slot label, and a specific intent is assigned for the whole sentence.

Sentence	what	flights	leave	from	phoenix
Slots	O	O	O	O	B-fromloc
Intent	atis_flight				

Table 1: An example sentence from the ATIS corpus

[*] Authors contributed equally.

Traditional pipeline approaches manage the two mentioned tasks separately. Intent detection is seen as a semantic classification problem to predict the intent label. General approaches such as support vector machine (SVM) (Haffner et al., 2003) and recurrent neural network (RNN) (Lai et al., 2015) can be applied. Slot filling is regarded as a sequence labeling task. Popular approaches include conditional random field (CRF) (Raymond and Riccardi, 2007), long short-term memory (LSTM) networks (Yao et al., 2014).

Considering the unsatisfactory performance of pipeline approaches caused by error propagation, the tendency is to develop a joint model (Chen et al., 2016a; Zhang and Wang, 2016) for intent detection and slot filling tasks. Liu and Lane (2016) proposed an attention-based RNN model. However, it just applied a joint loss function to link the two tasks implicitly. Hakkani-Tür et al. (2016) introduced a RNN-LSTM model where the explicit relationships between the slots and intent are not established. Goo et al. (2018) proposed a slot-gated model which applies the intent information to slot filling task and achieved superior performance. But the slot information is not used in intent detection task. The bi-directional direct connections are still not established. In fact, the slots and intent are correlative, and the two tasks can mutually reinforce each other. This paper proposes an SF-ID network which consists of an SF subnet and an ID subnet. The SF subnet applies intent information to slot filling task while the ID subnet uses slot information in intent detection task. In this case, the bi-directional interrelated connections for the two tasks can be established. Our contributions are summarized as follows: 1) We propose an SF-ID network to establish the interrelated mechanism for slot filling and intent detection tasks. Specially, a novel ID subnet is proposed to apply the slot information to intent detec-

Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 5467–5471
Florence, Italy, July 28 - August 2, 2019.

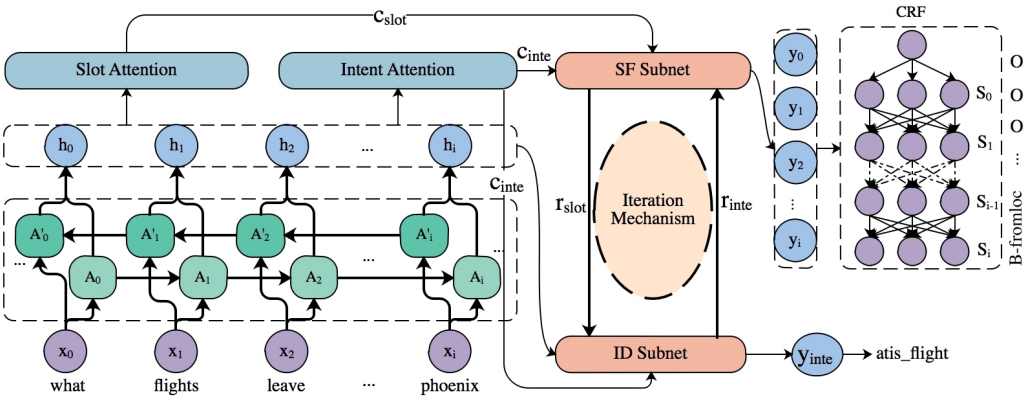

Figure 1: The structure of the proposed model based on SF-ID network

tion task. 2) We establish a novel iteration mechanism inside the SF-ID network in order to enhance the connections between the intent and slots. 3) The experiments on two benchmark datasets show the effectiveness and superiority of the proposed model.

2 Proposed Approaches

This section first introduces how we acquire the integration of context of slots and intent by attention mechanism. And then it presents an SF-ID network which establishes the direct connections between intent and slots. The model architecture based on bi-directional LSTM (BLSTM) is shown in Figure 2.[1]

2.1 Integration of Context

In SLU, word tags are determined not only by the corresponding terms, but also the context (Chen et al., 2016b). The intent label is also relevant with every element in the utterance. To capture such dependencies, attention mechanism is introduced.

Slot filling: The i^{th} slot context vector c^i_{slot} is computed as the weighted sum of BLSTM's hidden states $(h_1, ..., h_t)$:

$$c^i_{slot} = \sum_{j=1}^{T} \alpha^S_{i,j} h_j \quad (1)$$

where the attention weight α is acquired the same way as in (Liu and Lane, 2016).

Intent detection: The intent context vector c_{inte} is calculated as the same way as c_{slot}, in particular, it just generates one intent label for the whole sentence.

[1]The code is available at https://github.com/ZephyrChenzf/SF-ID-Network-For-NLU

2.2 SF-ID Network

The SF-ID network consists of an SF subnet and an ID subnet. The order of the SF and ID subnets can be customized. Depending on the order of the two subnets, the model have two modes: SF-First and ID-First. The former subnet can produce active effects to the latter one by a medium vector.

2.2.1 SF-First Mode

In the SF-First mode, the SF subnet is executed first. We apply the intent context vector c_{inte} and slot context vector c_{slot} in the SF subnet and generate the slot reinforce vector r_{slot}. Then, the newly-formed vector r_{slot} is fed to the ID subnet to bring the slot information.

SF subnet: The SF subnet applies the intent and slot information (i.e. c_{inte} and c_{slot}) in the calculation of a correlation factor f which can indicate the relationship of the intent and slots. This correlation factor f is defined by:

$$f = \sum V * tanh(c^i_{slot} + W * c_{inte}) \quad (2)$$

In addition, we introduce a slot reinforce vector r_{slot} defined by (3), and it is fed to the ID subnet to bring slot information.

$$r^i_{slot} = f \cdot c^i_{slot} \quad (3)$$

ID subnet: We introduce a novel ID subnet which applies the slot information to the intent detection task. We believe that the slots represent the word-level information while the intent stands for the sentence-level. The hybrid information can benefit the intent detection task. The slot reinforce vector r_{slot} is fed to the ID subnet to generate the reinforce vector r, which is defined by:

$$r = \sum_{i=1}^{T} \alpha_i \cdot r^i_{slot} \quad (4)$$

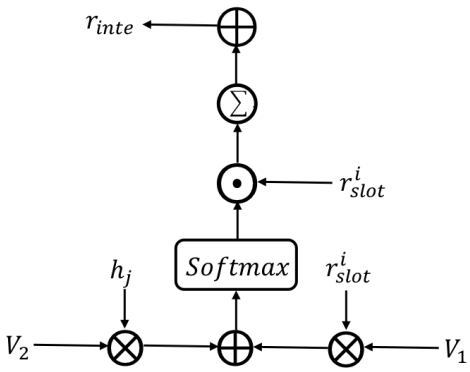

Figure 2: Illustration of the ID subnet

where the weight α_i of r^i_{slot} is computed as:

$$\alpha_i = \frac{exp(e_{i,i})}{\sum_{j=1}^{T} exp(e_{i,j})} \quad (5)$$

$$e_{i,j} = W * tanh(V_1 * r^i_{slot} + V_2 * h_j + b) \quad (6)$$

We also introduce an intent reinforce vector r_{inte} which is computed as the sum of the reinforce vector r and intent context vector r_{inte}.

$$r_{inte} = r + c_{inte} \quad (7)$$

Iteration Mechanism: The intent reinforce vector r_{inte} can also be fed into the SF subnet. In fact, this intent reinforce vector r_{inte} can improve the effect of relation factor f because it contains the hybrid information of intent and slots, and (2) can be replaced by:

$$f = \sum V * tanh(c^i_{slot} + W * r_{inte}) \quad (8)$$

With the change in the relation factor f, a new slot reinforce vector r_{slot} is acquired. Thus, the ID subnet can takes a new r_{slot} and exports a new r_{inte}. In this case, both SF subnet and ID subnet are updated, one iteration is completed.

In theory, the interaction between the SF subnet and ID subnet can repeat endlessly, which is denoted as the iteration mechanism in our model. The intent and slot reinforce vectors act as the links between the SF subnet and the ID subnet and their values continuously change during the iteration process.

After the iteration mechanism, the r_{inte} and r_{slot} participate in the final prediction of intent and slots, respectively. For the intent detection task, the intent reinforce vector r_{inte} and the last hidden state h_T of BLSTM are utilized in the final intent prediction:

$$y_{inte} = softmax(W^{hy}_{inte} concat(h_T, r_{inte})) \quad (9)$$

For the slot filling task, the hidden state h_i combined with its corresponding slot reinforce vector r^i_{slot} are used in the i^{th} slot label prediction. The final expression without CRF layer is:

$$y^i_{slot} = softmax(W^{hy}_{slot} concat(h_i, r^i_{slot})) \quad (10)$$

2.2.2 ID-First Mode

In the ID-First mode, the ID subnet is performed before the SF subnet. In this case, there are some differences in the calculation of ID subnet in the first iteration.

ID subnet: Unlike the Slot-First mode, the reinforce vector r is acquired by the hidden states and the context vectors of BLSTM. Thus, (4) (5) (6) can be replaced by:

$$r = \sum_{i=1}^{T} \alpha_i \cdot h_i \quad (11)$$

$$\alpha_i = \frac{exp(e_{i,i})}{\sum_{j=1}^{T} exp(e_{i,j})} \quad (12)$$

$$e_{i,j} = W * \sigma(V_1 * h_i + V_2 * c^j_{slot} + b) \quad (13)$$

The intent reinforce vector r_{inte} is still defined by (7), and it is fed to the SF subnet.

SF subnet: The intent reinforce vector r_{inte} is fed to the SF subnet and the relation factor f is calculated the same way as (8). Other algorithm details are the same as in SF-First mode.

Iteration Mechanism: Iteration mechanism in ID-First mode is almost the same as that in SF-First mode except for the order of the two subnets.

2.3 CRF layer

Slot filling is essentially a sequence labeling problem. For the sequence labeling task, it is beneficial to consider the correlations between the labels in neighborhoods. Therefore, we add the CRF layer above the SF subnet outputs to jointly decode the best chain of labels of the utterance.

3 Experiment

Dataset: We conducted experiments using two public datasets, the widely-used ATIS dataset (Hemphill et al., 1990) and custom-intent-engine dataset called the Snips (Coucke et al., 2018), which is collected by Snips personal voice assistant. Compared with the ATIS dataset, the Snips dataset is more complex due to its large vocabulary and cross-domain intents.

Evaluation Metrics: We use three evaluation

Model		ATIS Dataset			Snips Dataset		
		Slot (F1)	Intent (Acc)	Sen. (Acc)	Slot (F1)	Intent (Acc)	Sen. (Acc)
Joint Seq (Hakkani-Tür et al., 2016)		94.30	92.60	80.70	87.30	96.90	73.20
Atten.-Based (Liu and Lane, 2016)		94.20	91.10	78.90	87.80	96.70	74.10
Sloted-Gated (Goo et al., 2018)		95.42	95.41	83.73	89.27	96.86	76.43
SF-ID Network	SF-First (with CRF)	95.75	**97.76**	86.79	91.43	**97.43**	**80.57**
	SF-First (without CRF)	95.55	97.40	85.95	90.34	97.34	78.43
	ID-First (with CRF)	**95.80**	97.09	**86.90**	**92.23**	97.29	80.43
	ID-First (without CRF)	95.58	96.58	86.00	90.46	97.00	78.37

Table 2: Performance comparison on ATIS and Snips datasets. The improved cases are written in bold.

Model	ATIS		Snips	
	Slot	Intent	Slot	Intent
Without SF-ID	95.05	95.34	88.9	96.23
ID subnet Only	95.43	95.74	89.57	97.42
SF subnet Only	95.14	95.75	90.7	96.71
SF-ID (no interaction)	95.56	95.75	90.97	97.01
SF-ID (SF-First)	95.75	**97.76**	91.43	**97.43**
SF-ID (ID-First)	**95.80**	97.09	**92.23**	97.29

Table 3: Analysis of seperate subnets and their interaction effects

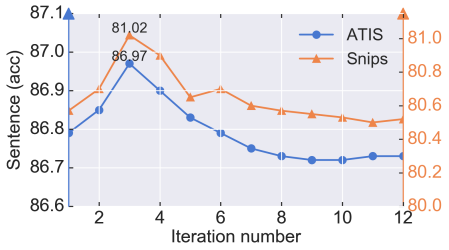

Figure 3: Effect of iteration number on the model performance in SF-First mode

metrics in the experiments. For the slot filling task, the F1-score is applied. For the intent detection task, the accuracy is utilized. Besides, the sentence-level semantic frame accuracy (sentence accuracy) is used to indicate the general performance of both tasks, which refers to proportion of the sentence whose slots and intent are both correctly-predicted in the whole corpus.

Training Details: In our experiments, the layer size for the BLSTM networks is set to 64. During training, the adam optimization (Kingma and Ba, 2014) is applied. Besides, the learning rate is updated by $\eta_t = \eta_0/(1+pt)$ with a decay rate of $p = 0.05$ and an initial learning rate of $\eta_0 = 0.01$, and t denotes the number of completed steps.

Model Performance: The performance of the models are given in Table 2, wherein it can be seen that our model outperforms the baselines in all three aspects: slot filling (F1), intent detection (Acc) and sentence accuracy (Acc). Specially, on the sentence-level semantic frame results, the relative improvement is around 3.79% and 5.42% for ATIS and Snips respectively, indicating that SF-ID network can benefit the SLU performance significantly by introducing the bi-directional interrelated mechanism between the slots and intent.

Analysis of Seperate Subnets: We analyze the effect of seperate subnets, and the obtained results are given in Table 3. The experiments are conducted when the CRF layer is added. As we can see, both models including only the SF subnet or the ID subnet have acheived better results than the BLSTM model. Therefore, we believe that both SF subnet and ID subnet have significance in performance improvement.

Beside, we also analyse the condition with independent SF and ID subnet, in other words, when there is no interaction in SF and ID subnet. We can see it also obtains good results. However, the SF-ID network which allows the two subnets interact with each other achieve better results. This is because the bi-directional interrelated mechanism help the two subnets promote each other mutually, which improves the performance in both tasks.

Analysis of Model Mode: In Table 2, it can be seen that the ID-First mode achieves better performance in the slot filling task. This is because the ID-First mode treats the slot filling task as a more important task, because the SF subnet can utilize the intent information output from the ID subnet. Similarly, the SF-First mode performs better in the intent detection task. In general, the difference between the two modes is minor.

Iteration Mechanism: The effect of iteration mechanism is shown in Figure 3. The experiments are conducted in SF-First mode. Sentence accuracy is applied as the performance measure because it can reflect the overall model performance. It increases gradually and reaches the maximum value when the iteration number is three on both ATIS and Snips dataset, indicating the effective-

ness of iteration mechanism. It may credit to the iteration mechanism which can enhance the connections between intent and slots. After that, the sentence accuracy gradually gets stabilized with minor drop. On balance, the iteration mechanism with proper iteration number can benefit the SLU performance.

CRF Layer: From Table 2 it can be seen that the CRF layer has a positive effect on the general model performance. This is because the CRF layer can obtain the maximum possible label sequence on the sentence level. However, CRF layer mainly focuses on sequence labeling problems. So the improvement of the slot filling task obviously exceeds that of the intent detection task. In general, the performance is improved by the CRF layer.

4 Conclusion

In this paper, we propose a novel SF-ID network which provides a bi-directional interrelated mechanism for intent detection and slot filling tasks. And an iteration mechanism is proposed to enhance the interrelated connections between the intent and slots. The bi-directional interrelated model helps the two tasks promote each other mutually. Our model outperforms the baselines on two public datasets greatly. This bi-directional interrelated mechanism between slots and intent provides guidance for the future SLU work.

Acknowledgments

The authors would like to thank the reviewers for their valuable comments. This work was supported in part by the National Key R&D Program of China under Grant SQ2018YFB140079 and 2018YFB1403003.

References

Yun-Nung Chen, Dilek Hakanni-Tür, Gokhan Tur, Asli Celikyilmaz, Jianfeng Guo, and Li Deng. 2016a. Syntax or semantics? knowledge-guided joint semantic frame parsing. In *Spoken Language Technology Workshop (SLT), 2016 IEEE*, pages 348–355. IEEE.

Yun-Nung Chen, Dilek Hakkani-Tür, Gökhan Tür, Jianfeng Gao, and Li Deng. 2016b. End-to-end memory networks with knowledge carryover for multi-turn spoken language understanding. In *INTERSPEECH*, pages 3245–3249.

Alice Coucke, Alaa Saade, Adrien Ball, Théodore Bluche, Alexandre Caulier, David Leroy, Clément Doumouro, Thibault Gisselbrecht, Francesco Caltagirone, Thibaut Lavril, et al. 2018. Snips voice platform: an embedded spoken language understanding system for private-by-design voice interfaces. *arXiv preprint arXiv:1805.10190*.

Chih-Wen Goo, Guang Gao, Yun-Kai Hsu, Chih-Li Huo, Tsung-Chieh Chen, Keng-Wei Hsu, and Yun-Nung Chen. 2018. Slot-gated modeling for joint slot filling and intent prediction. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, volume 2, pages 753–757.

Patrick Haffner, Gokhan Tur, and Jerry H Wright. 2003. Optimizing svms for complex call classification. In *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, volume 1, pages I–I. IEEE.

Dilek Hakkani-Tür, Gökhan Tür, Asli Celikyilmaz, Yun-Nung Chen, Jianfeng Gao, Li Deng, and Ye-Yi Wang. 2016. Multi-domain joint semantic frame parsing using bi-directional rnn-lstm. In *Interspeech*, pages 715–719.

Charles T Hemphill, John J Godfrey, and George R Doddington. 1990. The atis spoken language systems pilot corpus. In *Speech and Natural Language: Proceedings of a Workshop Held at Hidden Valley, Pennsylvania, June 24-27, 1990*.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Siwei Lai, Liheng Xu, Kang Liu, and Jun Zhao. 2015. Recurrent convolutional neural networks for text classification. In *AAAI*, volume 333, pages 2267–2273.

Bing Liu and Ian Lane. 2016. Attention-based recurrent neural network models for joint intent detection and slot filling. *arXiv preprint arXiv:1609.01454*.

Christian Raymond and Giuseppe Riccardi. 2007. Generative and discriminative algorithms for spoken language understanding. In *Eighth Annual Conference of the International Speech Communication Association*.

Gokhan Tur and Renato De Mori. 2011. *Spoken language understanding: Systems for extracting semantic information from speech*. John Wiley & Sons.

Kaisheng Yao, Baolin Peng, Yu Zhang, Dong Yu, Geoffrey Zweig, and Yangyang Shi. 2014. Spoken language understanding using long short-term memory neural networks. In *Spoken Language Technology Workshop (SLT), 2014 IEEE*, pages 189–194. IEEE.

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