



A joint model based on interactive gate mechanism for spoken language understanding

Chengai Sun¹ · Liangyu Lv¹ · Tailu Liu¹ · Tangjun Li^{1,2}

Accepted: 19 May 2021 / Published online: 30 August 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Slot filling and intent detection are two important tasks in a spoken language understanding (SLU) system, it is becoming a tendency that two tasks are jointing learn in SLU. However, many existing model only conduct joint model by share parameters on the surface level rather than bi-directional interaction for slot filling and intent detection tasks. In this paper, we designed a dual interaction model based on the gate mechanism. First, We utilize a Dilated Convolutional Neural Networks (DCNN) block with self-attention to better capture the semantic of utterance. Besides, for the two tasks we adopt gate mechanism to get the interaction information of intent and slot, which can control the passing rate and make fully use of semantic relevance between slot filling and intent detection. Finally, the experiments results show that our model has significantly improved in the slot filling F1, intent detection accuracy on the ATIS and SNIPS datasets and overmatch other prior methods.

Keywords Bidirectional interaction · Gate mechanism · Dilated convolutional neural networks · Self-attention

1 Introduction

Spoken language understanding (SLU) is a critical component in human/machine spoken dialogue system, it is aimed to identify users' domain and intent in the natural language utterances and extract relevant parameters or slots to achieve the goal. SLU typically consists of two tasks, which intent detection and slot filling (Tur and De Mori) [1]. Intent detection focused on predicting the intent of users', which slot filling extracts semantic concepts. For instance, as shown in Table 1, give a about airticket utterance, which query "first class fares from Boston to Denver". Each token in the sentence has a slot, and there are a special intent for the whole sentence.

Previously, traditional pipeline methods that slot filling and intent detection were implemented separately, without

taking into account the dependent relationships for two tasks. Qin et al. [2] proposed that slot and intent are closely linked. In fact, the slot and intent are correlative, and the two tasks can mutually reinforce each other. For example, if the intent of a sentence is an air ticket query, its corresponding slot may include the departure and arrival cities rather than the name of the movie. Hence, it is necessity to combine slot filling and intent detection.

Considering the pipeline methods can lead to error propagation, and the two tasks has strong connection, the joint model for intent detection and slot filling has become a tendency and made greatly process, many joint model based on the multi-tasking learning framework has proposed (Zhang and Wang; Liu and Lane; Hakkani-Tur et al.) [3–5]. However, the prior model only conduct joint model by share parameters, they use a joint loss function to implicitly model the between the two tasks. Recently, Goo et al. [6] introduced slot-gated mechanism and Li et al. [7] introduced self-attention with gated mechanism which apply intent information to slot task, but the slot task not apply intent task, the bi-directional interaction for slot filling and intent detection tasks has not created. In fact, the joint model should treat as parallel tasks that take advantage of the relationship between slot and intent.

To solve the above-mentioned problems, we propose a bidirectional interaction model based on gate mechanism.

✉ Tangjun Li
litangjun126@126.com

¹ College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, China

² Present address: No. 579, Qianwan'gang Road, Huangdao District, Qingdao City, Shandong Province, 266590, People's Republic of China

Table 1 An example about air ticket enquiry from the ATIS datasets , with a slot BIO format

| Sentence | first | class | fares | from | Boston | to | Denver |
|----------------|--------------|--------------|-------|------|-----------|----|---------|
| Cold Slots | B-Class_type | I-Class_type | O | O | B-fromLoc | O | B-toLoc |
| Gold Intention | airfare | | | | | | |

First, we use DCNN as the sentence encoder to capture information at a larger distance by increasing the expansion width in the convolution kernel without increasing the number of parameters. A major advantage of self-attention can built directly connections between any two tokens in a sentence. Due to this strong connection relationship between the two tasks, we proposed a bidirectional interaction model based on a gating mechanism to improve the existing joint model. By digging into the connection between intention detection and slot filling, a bidirectional interaction mechanism was established to make up for the shortcomings of the existing joint model.

The rest of the paper is as follows: Section 2 discusses related work, Section 3 gives a detailed description of our model, Section 4 presents experiments results and analysis, and Section 5 summarizes this work and the future direction.

2 Related work

Spoken language understanding has a long research history, which originated from some call classification systems (Gorin et al.) [8] and the ATIS project in the 1990s. SLU typically consists of two tasks, which include intent detection and slot filling. The intent detection can be looked upon classification tasks to predict speakers intent general methods like support vector network (SVM) (Haffner et al.) [9] and recurrent neural network (RNN) [10] (Lai et al.) can be used. The slot filling can be looked upon sequence labeling task to extract semantic components, popular methods like conditional random field (CRF) [11] (Raymond and Ricciardi), long short-term memory (LSTM) networks (Yao et al. [12], (Sun et al.) [13] can be used.

Due to appear error propagation problem in traditional pipeline methods, recently, some model overcome the problem. Xu et al. [14] proposed using based on convolutional neural network(CNN) for joint slot filling and intent detection. Guo et al. [15] proposed using RNN for joint training of slot filling and intent detection. Zhang and Wang et al. [3] proposed using RNNs for joint slot filling and intent detection. Hakkani-Tur et al. [5] proposed a single recurrent neural network for joint slot filling and intent detection. Liu and Lane [4] proposed an attention-based encoder-decoder neural network model for joint slot

filling and intent detection, which an encoder for input and a decoder for output. However, those model only conduct joint model by share parameters, they use a joint loss function to implicitly model the between the two tasks.

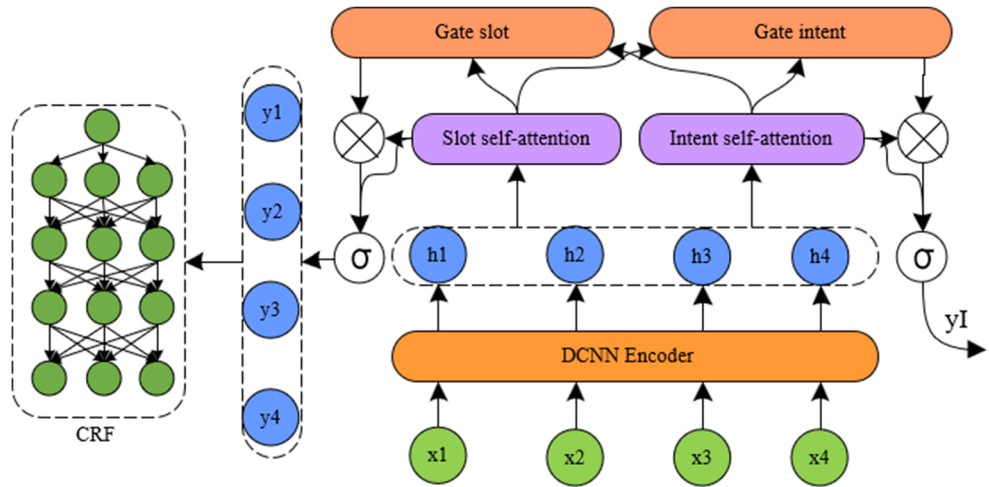
Recently, some joint model starts to explicitly for slot filling and intent detection, Wang et al. [16] proposed the Bi-model RNN structures, which take cross-impact into account for slot filling and intent detection jointly. Goo et al. [6] proposed using slot-gated mechanism, which apply intent detection to slot filling. Zhang et al. [17] proposed a hierarchical capsule neural network to model the the hierarchical relationship among word, slot, and intent in an utterance. E et al. [18] proposed an SF-ID network to build the direct connection for slot filling and intent detection, and introduce a interrelated mechanism to increase bi-directional interrelated. Qin et al. [2] proposed a token-level intent detection with Stack-Propagation framework, which use directly the intent detection as input for slot filling. Liu et al. [19] proposed a collaborative memory block to implicitly consider the mutual interaction between slot filling and intent detection.

Although much progress has been made in these approaches, intent detection and slot filling a challenging and open task. Therefore, a bidirectional interaction mechanism was established to make up for the shortcomings of the existing joint model to improve the performance of the SLU system.

3 Proposed approaches

This section describes the details of our framework, it consists of a encoder, a self-attention layer, bidirectional interaction gate layer that establishes connection between slot filling and intent detection and decoder for the two tasks. As shown in Fig. 1, we use DCNN encoder to extract the shared features of intent detection and slot filling, instead of using the RNN encoder. Then, we enter shared representations into two self-attention structures to get intent and slot representations. Finally, the intent representation and slot representation are input to the interaction gate structure to obtain interactive information. The purpose is to integrate the slot information into the intent detection task and the intent information into the slot filling task.

Fig. 1 Illustration of dual interaction model based on the gate mechanism for joint intent detection and slot filling



3.1 DCNN encoder

Compared with ordinary CNN, DCNN has great advantages in processing long-distance sequence information. As shown in Fig. 2, it is a three-layer convolutional neural networks, in the third layer, ordinary convolutional neural network can only capture 3 inputs before and after, but 2-dilated CNN can capture 7 inputs before and after. The number and speed of parameters remain unchanged. Therefore, the same post-stacked void convolution can better capture the global information of the input. Therefore, the DCNN is selected as the encoder of the input sentence in the model. Since the encoder stacks multi-layer void convolution, residual network are added between each void convolution block to prevent model degradation and make model training more stable, the specific structure is as follows:

The DCNN block is composed of 5 convolutional layers. The relationship information between two words at a distance can be obtained through layer superposition in Fig. 3, its input is the word embedding of the user’s utterance $X = \{X_1, X_2, \dots, X_n\}$ (n is the number of tokens in the input utterance). Then the specific calculation process of this DCNN block is as follows:

$$X_L = g_L \otimes DConv1D_1(X_{L-1}) + (1 - g_L) \otimes X_{L-1} \quad (1)$$

$$g_L = \sigma(DConv2D_2(X_{L-1})) \quad (2)$$

Where X_L is the output of DCNN block, X_{L-1} is the output of the previous convolutional layers, \otimes is sigmoid activation function, $DConv1D_1$ and $DConv2D_2$ are 1-dimensional dilated convolution which has a wider receptive field, g_L is the gating function of the second layer, which can control the proportion of X_L output from this layer

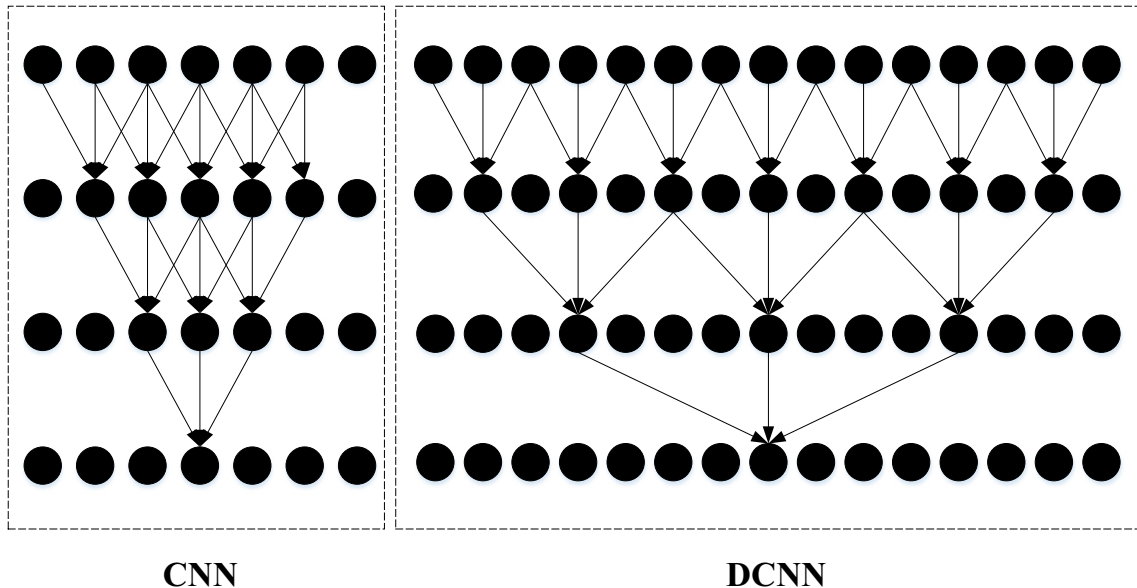


Fig. 2 Compare of CNN and DCNN

and the input of this layer, so as to control the transfer of information between each layer to the next layer, and the information can be transmitted across multiple channels [20]. In addition, this method can also alleviate the problem of possible gradient disappearance when the gradient is back propagated in a single path, making the gradient propagation more smoothly.

The encoder is composed of 5 CNN blocks, the size of the convolution window of each convolution block is 3, and the dilated rate is [1, 2, 4, 1, 1]. The dilated rates of the first three layers were 1, 2, and 4, the number of layers from low to high could capture window information, fragment information, and global information respectively. The last two layers were 1. Therefore, the 5-layer dilated CNN block can not only extract the semantic features required for intent recognition, but also capture the window features and context semantic features required for the sequence of labeling tasks of slot filling. Finally, the convolution result is input to the self-attention layer to better capture the semantic relationship between each token.

3.2 Self-attention

Self-attention is a special case of the attention mechanism and has been successfully applied to a variety of tasks. It refines the representation by matching a single sequence against itself, and then capture the internal structure of the sentence, in other words, give a sentence, each word should be conducted attention calculate to all the words in the sentence. In our model, we employ DCNN with self-attention mechanism to the advantages of temporal features and contextual information, which are useful for intent detection and sequence labeling tasks (Zhong et al.; Yin et al. [21, 22]).

Self-attention module is a very effective method for natural language processing tasks, it can be understood as a mapping. The self-attention mechanism learns the dependency relationship between any two tokens in a sentence and capture the internal structure information of the sentence. When the self-attention mechanism establishes a connection between each token and other tokens in the sentence, it will

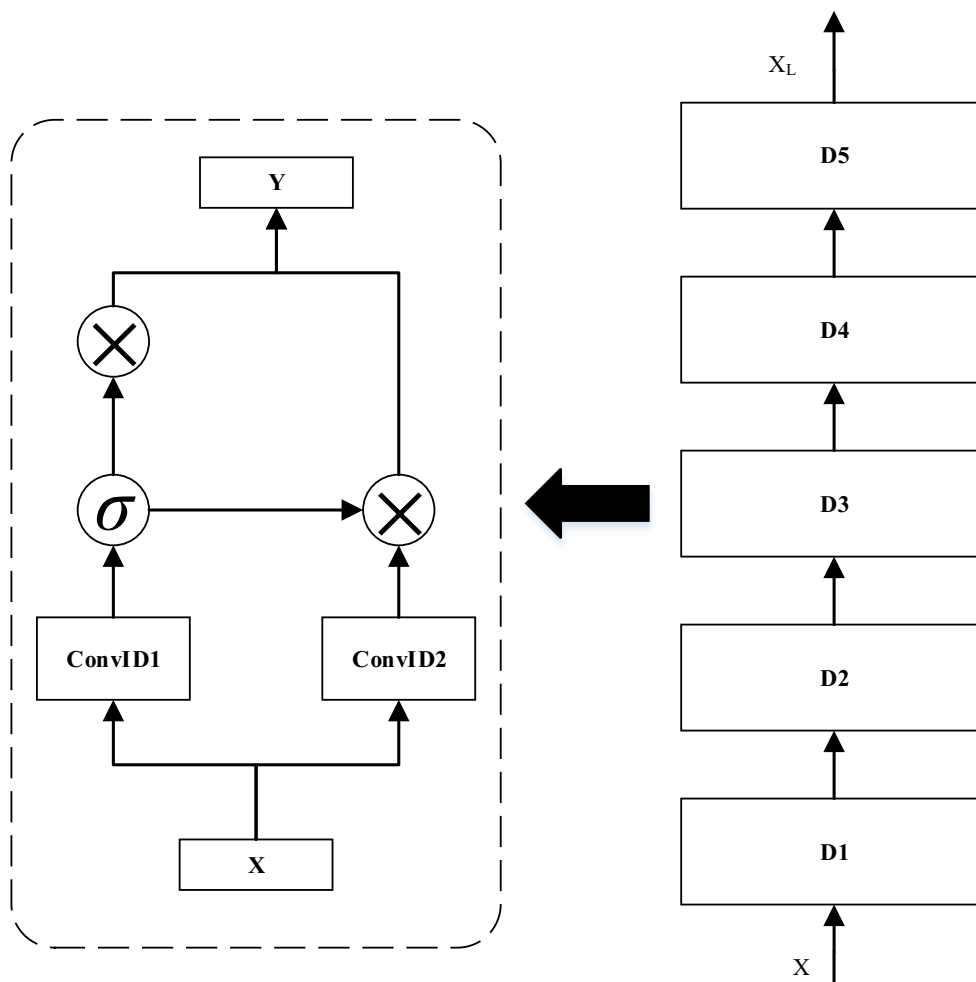


Fig. 3 The structure of DCNN

give greater weight to the important tokens in the sentence. We use scaled dot-product attention, following Vaswani et al. [23], where we map the input vectors $X \in \mathbb{R}^{n \times d}$ (d represents the mapped dimension) to queries(Q), keys(K), value(V), the calculation process is as follows:

$$Q = W_Q X_5 + b_Q \tag{3}$$

$$K = W_K X_5 + b_K \tag{4}$$

$$V = X_5 + b_V \tag{5}$$

$$C = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) \otimes V \tag{6}$$

Where $W_Q, W_K, W_V, b_Q, b_K, b_V$ are the weight matrix and bias of Q, K, and V respectively. d_k is usually set to a large value (e.g., 100), so the scaling factor could reduce the extremely small gradients effect. A masking operation (which masks the diagonal of the affinity matrix) is applied before the softmax to avoid high matching scores between identical vectors of query and key.

The model introduces two self-attention layers, the intent self-attention layer and the slot self-attention layer, which are independent of each other and do not share weights. The input of the intent self-attention layer and the slot self-attention layer is the output X_5 of the dilated convolutional layer, and the output is C^I and C^{slot} respectively.

3.3 Bidirectional interaction gating mechanism

In order to increase the interaction and connection between intent and slot, we designed an interaction gate structure to obtain the final representation of slot and intent. Therefore, our model can add intent information to the slot filling task, and also add slot information to the intent detection task. This way can make intent and slot promote the performance of each other.

In gated-intent module, the slot context vector C^{slot} and intent context vector C^I and combine to calculate a intent gate g^I , and the calculation process is as follows:

$$g^I = \tanh(W^I C^I + C^{slot}) \tag{7}$$

$$r^I = \text{maxpooling}(g^I C^I) \tag{8}$$

$$\hat{y}^I = \text{softmax}(W_y^I r^I + b_y^S) \tag{9}$$

where \tanh is the activation function, and W_y^I, b_y^S, W^I is trainable weight respectively. The element of gate g^I is mapped between $(-1, 1)$, which can be seen as a weight feature of the intent context vector C^I to control the information passing rate, we get the final representation r^I and the prediction y^I of the intent:

Unlike other methods that only use slot gates, we think that slot and intent information can also improve the performance of each other's tasks with the premise of strong correlation between two tasks. Therefore, similar to the intent gate, we

also designed a slot gate to control the passing rate of slot information.

In slot gate module, as described above, intent information is useful for slot filling task. The slot gate g^S can also be computed in the same method as g^I , but the C^{ave} is the average of C^I . The slot gate and y^S are modeled similarly, the calculation process is as follows:

$$C^{ave} = \text{meanpooling}(C^I) \tag{10}$$

$$g^S = \tanh(W^S C^{slot} + C^{ave}) \tag{11}$$

$$r^S = C^{slot} g^S \tag{12}$$

$$O^S = W_o^S r^S + b_o^S \tag{13}$$

$$P(\hat{y}^S | O^S) = \frac{\sum_{i=1} \text{expf}(y_{i-1}, y_i, O^S)}{\sum_{y'} \sum_{i=1} \text{expf}(y'_{i-1}, y'_i, O^S)} \tag{14}$$

Where W^S is the weight matrix, r^S is the slot final represent, $f(y_{i-1}, y_i, O^S)$ computes the transition score from y_{i-1} to y_i and y^S represents the predicted label sequence.

3.4 Loss function

The loss function of intention detection is cross entropy loss function, and it is calculated as follows:

$$L_1 = - \sum_{i=1}^{c_{y_i^I Label}} \log p(y_i^I) \tag{15}$$

And, the slot filling objection is formulated as:

$$L_2 = - \log P(\hat{y}^S | O^S) \tag{16}$$

Where $y_i^{I Label}$ is intent label and $y_i^{S Label}$ is slot label, To obtain both slot filling and intent detection jointly, the final objective is formulated as:

$$L = \alpha L_1 + (1 - \alpha) L_2 \tag{17}$$

Where α is hyper-parameter.

4 Experiments

4.1 Data set

In the section, in order to evaluate the proposed model, we used two public datasets, one is the public Airline Travel Information System (ATIS) (Hemphill et al.) dataset, which contains audio recordings of people making flight reservations, it is widely used in SLU research. The training set uses 4478 utterances and the test set uses 893 utterances, the another 500 utterances as development set, there are in total 120 slot labels and 21 intention types. Other is the custom-intent-engines collected by SNIPS (SNIPS data set) (Coucke et al.), which collected from the SNIPS personal voice assistant, The training set uses 13084 utterances and the test set uses 700 utterances, the another 700 utterances as

Table 2 Joint model results on two datasets

| Model | SNIPS | | ATIS | |
|---------------------------------|-------------|-------------|-------------|-------------|
| | Slot(F1) | intent(Acc) | Slot(F1) | Intent(Acc) |
| Joint Seq [5] | 87.3 | 96.9 | 94.3 | 92.6 |
| Attention BiRNN [4] | 87.8 | 96.7 | 94.2 | 91.1 |
| Slot-Gated Full Atten [6] | 88.8 | 97.0 | 94.8 | 93.6 |
| Slot-Gated Intent Atten [6] | 88.3 | 96.8 | 95.2 | 94.1 |
| Self-Attentive Model [7] | 90.0 | 97.5 | 95.1 | 96.8 |
| Bi-Model [16] | 93.5 | 97.2 | 95.5 | 96.4 |
| SF-ID Network [18] | 90.5 | 97.0 | 95.6 | 96.6 |
| Stack-Propagation Framework [2] | 94.2 | 98.0 | 95.9 | 96.9 |
| Our model | 94.7 | 98.3 | 96.3 | 97.5 |

Bold entries are the proposed models which are compared with previous models

development set, there are in total 72 slot labels and 7 intent types. Different from the ATIS data set, SNIPS is more complex because of its intent diversity and large vocabulary.

4.2 Training setup and evaluation criteria

In our model, the use of data set is following previous related work (Qin et al. [2]), which include format and partition. The data set dimensionalities of the word embedding is 300, the hidden units of the DCNN encoder are set as 128. The CNN is 5 layers, the void ratio is (1, 2, 4, 1, 1). L_2 regularization is used on our model is 1×10^{-5} . We use Adam (Kingma and Ba, 2014) to optimize the parameters in our model. For all the experiments, we select the model which works the best on the *dev* set, and then evaluate it on the test set.

Intent detection and slot filling tasks can be defined as classification tasks and sequence labeling tasks, respectively. The common indicators for classification tasks and sequence labeling tasks are also applicable to intent detection and slot filling tasks. We evaluate the SLU performance about slot filling using F1 score, intent prediction using

accuracy and it is calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (18)$$

$$P = \frac{TP}{TP + FP} \quad (19)$$

$$R = \frac{TP}{TP + FN} \quad (20)$$

$$F1 = \frac{2 \times P \times R}{P + R} \quad (21)$$

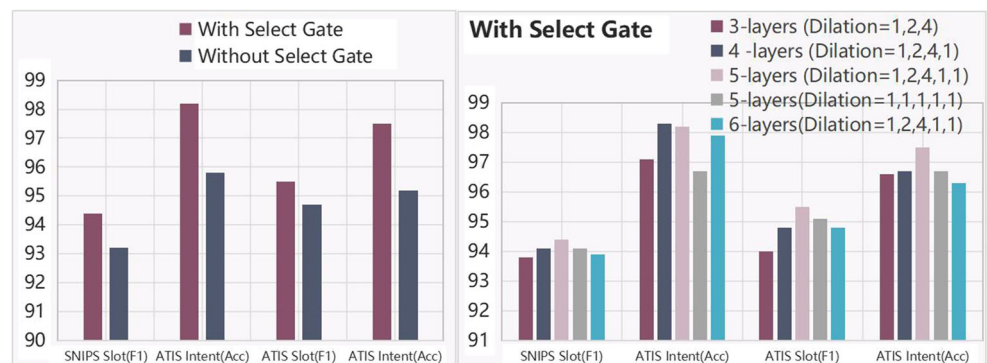
Where P and R are the accuracy rate and recall rate of the slot filling task.

4.3 Overall results

Following Qin et al., we evaluate the SLU performance about slot filling using F1 score, intent prediction using accuracy. The experiment results of the proposed models on SNIPS and ATIS datasets are shown in Table 2.

From the Table, we can see that our model significantly improvement than other existing model. In the ATIS datasets, we compared with the recent joint model with

Fig. 4 The analysis of independent gate and interaction result



stack-propagation framework, the improvement is around 0.4% and 0.6% in Slot F1 and Intent Acc. In the SNIPS datasets, the improvement is around 0.5% and 0.3% in Slot F1 and Intent Acc. This indicates the effectiveness of dual interaction model based on the gate mechanism, it can significantly benefit the SLU performance.

4.4 Analysis of gated mechanism

Table 2 shows the superior performance of our model. In order to better guide our follow-up work, we need to know the role of each module. So we performed ablation experiments to evaluate the impact of the various modules of the model on the final effect. First we tested the encoding ability of the DCNN Encoder. Then, we investigated the effect of the self-attention module on the results. Finally, we focused on the effectiveness of the interactive gate structure.

4.4.1 Effect of DCNN encoder

In order to verify the effect of DCNN layers, hole rate, and selection gate on the results, we set up the following ablation experiments.

We set up a comparison test, which includes different numbers of DCNN layers and different void ratios, and whether to use selection gates to control the flow of information between each layer of CNNs.

As can be seen from Fig. 4, the model has the best effect when the DCNN is 5 layers, the void ratio is (1, 2, 4, 1, 1) and the selection gate is used. However, with the same number of DCNN layers and void ratio, the model without selection gate performs very poorly. This is a common problem in DCNN networks—model degradation. When the CNN network becomes deeper and deeper, the model will become difficult to train and the performance will deteriorate. After adding the selection gate, the final output of each layer of DCNN block comes from its weighted sum of its inputs and inputs (such as (1) and (2)). It can alleviate model degradation and control information flow during training.

When the DCNN block hole rate of each layer is (1, 1, 1, 1, 1) (equivalent to the general 1-dimensional CNN), the performance of the model on the two data sets has significantly decreased. The accuracy of slot filling F1 and intent detection on the SNIPS dataset decreased by 0.6% and 1.5%, respectively. Similarly, the accuracy of slot filling F1 and intent detection on the ATIS data set decreased by 0.7% and 0.8%, respectively. This is because the receptive field of the hole convolution is larger and it is possible to obtain the global information of the short sentence after superimposing multiple DCNN blocks. Experimental results also verify the effectiveness of multi-layer dilated convolution.

Finally, the experimental results show the importance of the DCNN block with a void ratio of 1 in the last two layers in the slot filling task. However, the performance of intent detection is 3-layer, 4-layer, and 5-layer is hardly affected. The most likely reason is that intent detection requires more global information, and slot filling requires more detailed token information.

4.4.2 Effect of self-attention and interactive gate structure

In order to test the effect of self-attention and specially designed interaction gate mechanism, we designed the following ablation experiments.

First, we deleted slot self-attention and intent self-attention respectively, to observe the effect of self attention on intent detection and slot filling performance. It can be seen from Table 3 that after removing the intent self-attention and directly using the output of the DCNN encoder instead of the intent representation, the intent accuracy of the SNIPS and ATIS datasets decreased by 1% and 1.1%, respectively, but slot F1 remained almost constant. However, when we delete slot-attention in the same way, the intent accuracy rate remains the same without decreasing, and the slot F1 value drops slightly. We analyze this because the self-attention mechanism does not encode the relative position information into the final slot representation [24], and the relative position information is important for slot

Table 3 The analysis of separate training and interaction result

| Model | SNIPS | | ATIS | |
|-----------------------------------|-------------|-------------|-------------|-------------|
| | Slot(F1) | intent(Acc) | Slot(F1) | Intent(Acc) |
| Without intent self-attention | 94.5 | 97.3 | 96.2 | 97.0 |
| Without slot self-attention | 94.3 | 97.2 | 95.9 | 97.5 |
| Without intent gate | 93.5 | 97.3 | 95.8 | 96.4 |
| Without slot gate | 94.3 | 97.6 | 94.9 | 96.8 |
| Without intent gate and slot gate | 93.2 | 96.3 | 94.7 | 96.0 |
| our model | 94.7 | 98.3 | 96.3 | 97.5 |

Bold entries are the proposed models which are compared with previous models

filling, so self-attention has less impact on the slot filling task than intent recognition.

Then, we also explored the effect of the interaction gate mechanism on the model effect. As shown in Table 3, when only the intent gate is deleted, the intent accuracy rate decreases by about 1% and slot F1 decreases by about 0.5%; in contrast, after only the slot gate of the model is deleted, the slot F1 value and The intent accuracy has also dropped significantly. The experimental results show that the interaction of intent information and slot information by using an interactive gate structure can promote both intent detection and slot filling. When we remove the intent gate and slot gate at the same time, the model effect drops significantly. These also proves the effectiveness of our intent-slot interactive gate structure on two tasks.

5 Conclusions

In this paper, we designed a dual interaction model based on the gate mechanism for spoken language understanding, we further proposed DCNN for joint learning on both intent detection and slot filling tasks. In our model, DCNN block with self-attention to better capture the semantic of utterance, and gate mechanism is introduced to control the passing rate and make fully use of semantic relevance between slot filling and intent detection. Experiment results on ATIS data set have shown efficiency of our model and outperforms the state-of-the-art approach on both tasks. Besides, our model also shows consistent performance gain over the independent training models. In future works, we plan to improve our model by introducing extra knowledge.

Acknowledgments This work was supported by the National Natural Science Foundation of China under grant nos. 61702305, 11971270, and 61903089, the China Postdoctoral Science Foundation under grant no. 2017M622234 and Science and Technology Support Plan of Youth Innovation Team of Shandong higher School under grant no. 2019KJN2014.

References

1. Tur G (2011) Spoken language understanding: Systems for extracting semantic information from speech. Ph.D. Thesis
2. Qin L, Che W, Li Y, Wen H, Liu T (2019) A stack-propagation framework with token-level intent detection for spoken language understanding. arXiv: Computation and Language

3. Zhang X, Wang H (2016) A joint model of intent determination and slot filling for spoken language understanding, pp 2993–2999
4. Liu B, Lane I (2016) Attention-based recurrent neural network models for joint intent detection and slot filling, pp 685–689
5. Hakkaniur D, Tur G, Celikyilmaz A, Chen Y, Gao J, Deng L, Wang Y (2016) Multi-domain joint semantic frame parsing using bi-directional rnn-lstm, pp 715–719
6. Goo C, Gao G, Hsu Y, Huo C, Chen T, Hsu K, Chen Y (2018) Slot-gated modeling for joint slot filling and intent prediction 2:753–757
7. Li C, Li L (2018) A self-attentive model with gate mechanism for spoken language understanding, pp 3824–3833
8. GORIN A (1997) How may i help you? Speech Comm 23
9. Haffner P, Tur G, Wright JH (2003) Optimizing svms for complex call classification. 1:632–635
10. Lai S, Xu L, Liu K, Zhao J (2015) Recurrent convolutional neural networks for text classification, pp 2267–2273
11. Raymond C, Riccardi G (2007) Generative and discriminative algorithms for spoken language understanding, pp 1605–1608
12. Yao K, Peng B, Zhang Y, Yu D, Zweig G, Shi Y (2014) Spoken language understanding using long short-term memory neural networks, pp 189–194
13. Sun C, Lv L, Tian G, Liu T (2020) Deep interactive memory network for aspect-level sentiment analysis. 20(1)
14. Xu P, Sarikaya R (2013) Convolutional neural network based triangular crf for joint intent detection and slot filling, pp 78–83
15. Guo D, Tur G, Yih W, Zweig G (2014) Joint semantic utterance classification and slot filling with recursive neural networks, pp 554–559
16. Wang Y, Shen Y, Jin H (2018) A bi-model based rnn semantic frame parsing model for iintent detection and slot filling 2:309–314
17. Zhang C, Li Y, Du N, Fan W, Yu PS (2018) Joint slot filling and intent detection via capsule neural networks. arXiv: Computation and Language
18. Haihong E, Niu P, Chen Z, Song M (2019) A novel bi-directional interrelated model for joint intent detection and slot filling, pp 5467–5471
19. Liu Y, Meng F, Zhang J, Zhou J, Chen Y, Xu J (2019) Cmnet: A novel collaborative memory network for spoken language understanding
20. Gehring J, Auli M, Grangier D, Yarats D, Dauphin YN (2017) Convolutional sequence to sequence learning. arXiv: Computation and Language
21. Zhong V, Xiong C, Socher R (2018) Global-locally self-attentive encoder for dialogue state tracking. 1:1458–1467
22. Yin Q, Zhang Y, Zhang W, Liu T, Wang WY (2018) Deep reinforcement learning for chinese zero pronoun resolution. 1:569–578
23. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017) Attention is all you need. arXiv: Computation and Language
24. Yan H, Deng B, Li X, Qiu X TENER: Adapting Transformer Encoder for Named Entity Recognition

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.