Research on Joint Modeling of Intent Detection and Slot Filling

Dan Yang
School of Computer Science and Engineering
Xi’an Technological University
Xi’an, 710021, Shaanxi, China
E-mail: 1330311378@qq.com

Yi Li
School of Computer Science and Engineering
Xi’an Technological University
Xi’an, 710021, Shaanxi, China
E-mail: xatuliyi@163.com

Chaoyang Geng
School of Computer Science and Engineering
Xi’an Technological University
Xi’an, 710021, Shaanxi, China
E-mail: 541211200@qq.com

***Abstract*—In task-based dialogue system, the key of the natural language understanding module is intent detection and slot filling. At this stage, Joint modeling of intention detection and slot filling tasks has become the mainstream and achieved good results. In order to investigate the correlation between intention detection and slot filling tasks, Joint model of intention detection and slot filling based on attention mechanism in three dimensions: one-way modeling from intention to slot, Unidirectional modeling from slot to intention and bidirectional modeling from intention to slotSeparately.And experiments were conducted using the Chinese dataset CAIS, and the results showed three evaluation results for time slot F1.The intention accuracy and overall accuracy of joint models for intention detection and filling gaps are usually higher than those of unidirectional models.**

Keywords-Intent Detection; Slot Filling; Multi-head Attention Mechanism

I. INTRODUCTION

The Natural-language understanding module is the core part of task dialogue, which aims to transform user input into structured language. [1] The main task is to intentionally detect and supplement these two types of sub marriages.The former aims to roughly understand the intention of the target discourse. And determining the category they belong to is usually seen as sentence level text classification work.The latter translates the intention of the target discourse into specific instructions, namely. e. Identifying key semantic information contained in user discourse is considered as a character-level sequence annotation task [2].

Earlier, intent detection and filling tasks were independently modeled, commonly known as assembly line route [3]. The approach of the pipeline ignores the correlation between the two tasks, resulting in the problem that the intention detection results are difficult to match with the slot filling results and the problem of error propagation [4]. The joint modeling of the two is a method with better performance at present. A hot topic in recent years has been the linking of intent detection and socket execution tasks into a common model.Starting from these three types, this paper will analyze the correlation between Intent detection and slot filling tasks. And analyze the impact of the individual task on the performance of joint pattern discovery.

II. RELATED WORK

Intention detection can be seen as a sentence-level classification task. The traditional method is to obtain important physical information from text through n-grams[9], but this method is limited to simple sentences.Traditional machine learning

DOI: 10.2478/ijanmc-2023-0059

algorithms, such as SVM [10] and Adapost [11], train models by labeling certain data.Deep learning methods are also more effective in intention detection tasks such as CNN, RCNN, LSTM, and FastTextFor filling in spaces, it is usually considered as a character hierarchy marking task.The traditional method is based on the Conditional Combination Field (CRF) architecture and has strong sequence labeling ability, but only applicable to relatively small datasets. [13] In addition, in the "time slot filling" task, deep learning methods are also superior to traditional models, such as CNN based [14] and RNN based [15].

People have found that traditional methods often overlook the strong relationship between two tasks, leading to the problem of error propagation.Therefore, scholars began to study the way of joint modeling and became the mainstream. It is broadly classified into three types: intent to plus one way synthesis, plus to intent one way synthesis, double synthesis model. GanniTur et al. [4] proposed to use RNNs to learn the joint work. The models consider the correlation between the two tasks by sharing parameters. One joint model guides intent detection and slot filling tasks through intent or slot information display. For example, Goo et al. [4] combined the loss functions of the two tasks for optimization, and used the gating mechanism as a special gate function to model intent detection and slot, the relationship between fillings. Li et al. [16] proposed an intention enhancement gating mechanism to mine the semantic association between slots and intentions. Qin et al.[17] used the stack propagation framework to directly input the word level intention detection information into the slot filling. While the two-way joint model models the two tasks in two directions, Wang et al. [18] proposed Bi-Model to consider the cross influence between intent and slot.

Although the joint model of these two tasks, namely intention detection and filling gaps, has made significant progress. However, the relationship between the two in joint modeling and the extent to which each task affects the overall model validation performance still need to be studied and tested. Let's compare based on the joint model proposed by Qin et al.[8] for intention detection and filling in gaps.

III. MODEL

As shown in the overall structure in the figure 1.This model includes an encoder module, a bidirectional relationship module, an intention and time channel, and a decoding module, among which a bilateral relationship module, an intention and channel Time includes the attention layer of intention and spatial labels, the attention layer of intention and spatial interaction, and the feeding network layer.

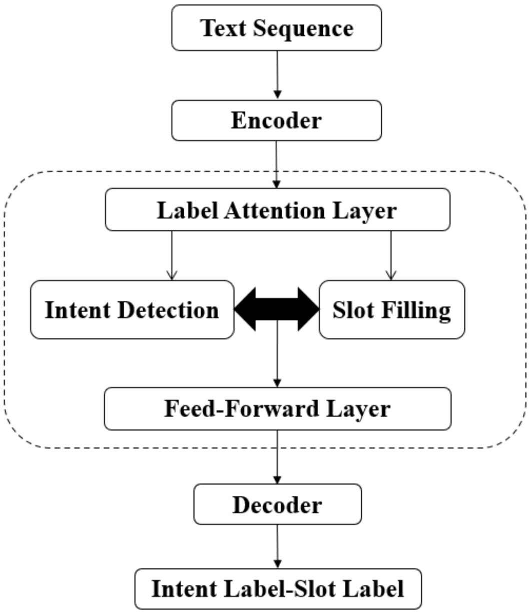

Figure 1. The Overall Structure of the Model.

A. Encoder Module

This module uses contextual semantic attributes as input to subsequent sub modules through Bi-LSTM, as shown in Figure 2.When encoding user text is transformed into an input sequence after a pre-training layer, and X is input to the Bi-LSTM layer to take advantage of temporal features in word sequences to obtain contextual semantic information. For the input sequence, each position i in the sequence has an LSTM to learn it from both positive and negative directions, respectively, if the hidden layer state of the LSTM output at position i positive is $\overrightarrow{h_i}$, the positive LSTM outputs the hidden layer state as $\overleftarrow{h_i}$. The forward and reverse results of LSTM are also combined to obtain the hidden layer state of each vector after encoding. BilSTM is as follows:

$$H = \{h_1, h_2, \ldots, h_n\} \tag{1}$$

$$h_i = \left[\overrightarrow{h_i}, \overleftarrow{h_i}\right] \tag{2}$$

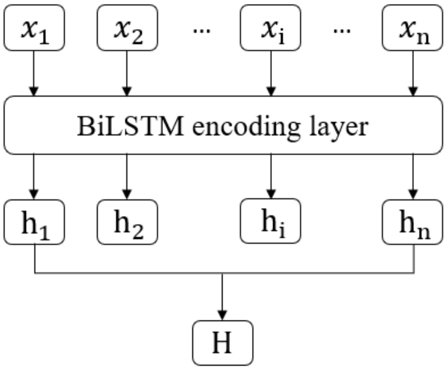

Figure 2. Encoder.

B. Intent and Slot Joint Module

1) Intent and slot labeling attention layer

The intent label and the slot label are given different levels of attention to obtain explicit intent representation and slot label representation, which are used for the subsequent direct interaction of the input co-interactive attention layer. In particular, the parameters of slot filling decoder and intent detection decoder layer are used as slot embedding matrix and intent embedding matrix (and the number of slots and intent tags are recorded). Use as query and askey and value to obtain intent and slot attention representation:

$$A = softmax\left(HW^v\right) \tag{3}$$

Enter the embedded state obtained from the partition coding module into the intent focus layer and the anecdote focus layer to obtain the semantic information of the intent or anecdote and to obtain the intent annotation focus or anecdote representation, as shown in Figure 3.

$$H_v = H + AW^v \tag{4}$$

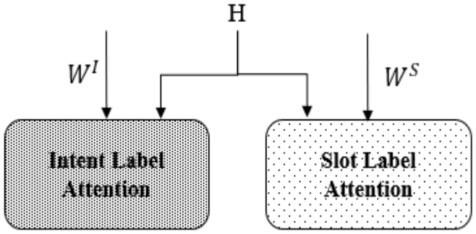

Figure 3. Label Attention Layer.

2) Intent and Slot Interaction Attention Layer

The intention representation and slot representation obtained by marking the attention layer capture the semantic information of the intention and slot respectively, and further explore in the collaborative interactive attention layer through and, as shown in Figure 4, to realize the two-way connection between the two tasks. Matrices and mappings, as well as matrices using different linear projections.

The intention of updating, displaying, and combining corresponding slot data is considered as a total weight query for keywords, values, and results, which is a normal layer with Dong intentionally obtains new expressions of intent from the attention layer of the tag, in order to avoid overfitting and reduce errors.Similarly, in order to improve the impression of time channels and integrate corresponding intention information, they are treated as queries, keywords, and values, resulting in a parity sum. Weight, using intention expression to obtain new intention expressions from the attention label layer to the normal layer.

$$C_I = softmax\left(\frac{Q_I K_S{}^T}{\sqrt{d_k}}\right) V_S \tag{5}$$

$$H_I{}' = LN\left(H_I + C_I\right) \tag{6}$$

$$C_S = softmax\left(\frac{Q_S K_I{}^T}{\sqrt{d_k}}\right) V_I \tag{7}$$

$$H_S{}' = LN\left(H_S + C_S\right) \tag{8}$$

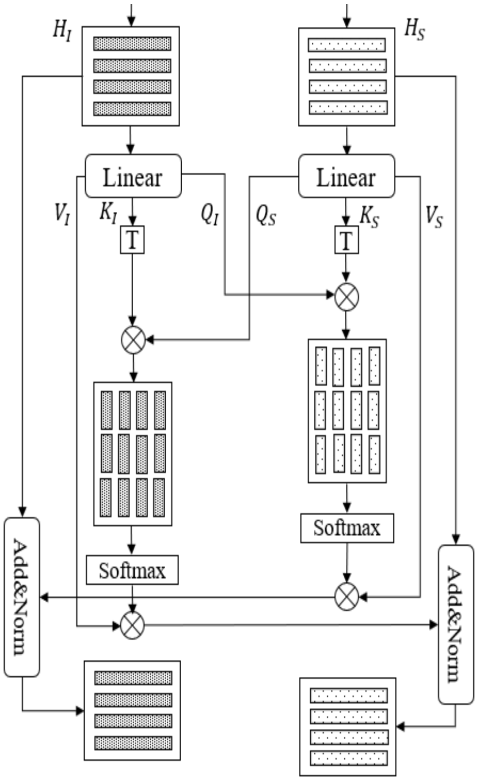

Figure 4. Intent and Slot Interaction Attention layer.

The above is the case when the intention is associated with time slots in two directions, and the unidirectional relationship is also similar.By treating the questionnaire as a keyword and representing the intention of the time period, one can obtain the intention to express a one-way relationship between the time period. The one-way relationship between time intervals is similar by treating them as questionnaires and keywords as values.

3) Feed-forward network layer

The intention and interval data are implicitly fused through the feed network layer, as shown in Figure 5.Firstly, the intention display and interval display of connection updates, including intention and spatial information, are connected through layers. Feed the network and ultimately obtain the latest intention display and interval display through the "layer normalization" function.

$$H_S{}' = LN\left(H_S + C_S\right) \tag{9}$$

$$H_{IS} = H_I{}' \oplus H_S{}' \tag{10}$$

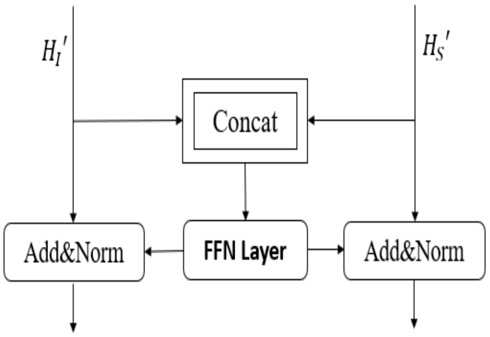

Figure 5. Feed-forward Network Layer.

C. Decoder

In order to have sufficient interaction between slot and intent detection tasks, a network with multi-layer stacked cooperative interactive attention is applied. After stacking the L layer, the final updated slot and intent representation are obtained, as shown in Figure 6. A maximum pooling operation is applied to obtain the representation C of the sentence as the input of intent detection:

$$\hat{H}_I^{(L)} = \left(\hat{H}_{(I,1)}^{(L)}, \hat{H}_{(I,2)}^{(L)}, \ldots, \hat{H}_{(I,n)}^{(L)}\right) \tag{11}$$

$$\mathrm{\hat{H}_S^{(L)}} = \left(\mathrm{\hat{H}_{(S,1)}^{(L)}}, \mathrm{\hat{H}_{(S,2)}^{(L)}}, \ldots, \mathrm{\hat{H}_{(S,n)}^{(L)}}\right) \tag{12}$$

$$\hat{y}^I = softmax\left(W^I c + bs\right) \tag{13}$$

$$o^I = argmax\left(y^I\right) \tag{14}$$

Here $\hat{y}^I$ denotes the output intent distribution, o^I denotes the intent label, and W^I denotes the trainable parameters of the model.

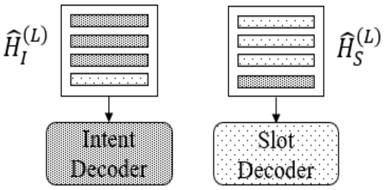

Figure 6. Decoder.

Here, conditional random fields are used to model the slot scale dependence between adjacent character, i.e:

$$O_S = W^S \hat{H}_S^{(L)} + bs \tag{15}$$

$$P\left(\hat{y}|O_S\right) = \frac{\sum_{i=1} exp\ f\left(y_{i-1}, y_i, O_S\right)}{\sum_{y'}\sum_{i=1} exp\ f\left(y_{i-1}^{'}, y_i^{'}, O_S\right)} \tag{16}$$

Here, the transition score is calculated from to represent the prediction label sequence.

IV. Experiment

A. Experimental setup and evaluation index

This experiment was conducted using the Chinese CAIS dataset, and Liu et al. [7] used the dataset to collect the voices of Chinese artificial intelligence (CAIS) speakers. And marked with slot door and intention label, exercise kit (7995 sentences), inspection kit (994 sentences) and Test suite. (Including 1024 sentences) Separate from intention allocation.

The following models are used for experiments: Slot-Gated [4], Stack-Propagation, SF-ID [5], CM-Net [6], and Co-Interactive Transformer [8]. Among them, slot gate control and stack communication are unidirectional common mode, SF-ID network, and CM network. And the interaction converter is a bidirectional joint format.At the same time, precision (acc) was used in the experiment to evaluate intention detection work, and F1 value was used to evaluate the work of filling gaps. And evaluate the overall performance of the model using sentence accuracy (sent_acc), defined as follows:

$$acc = \frac{\sum_{i=1}^{b}\begin{cases}1 & ID_i^* = ID_i \\ 0 & ID_i^* \neq ID_i\end{cases}}{b} \tag{17}$$

$$F_1 = \frac{2\times\sum_{i=1}^{b}\left|SF_i \cap SF_i^*\right|}{\sum_{i=1}^{b}\left|SF_i\right| + \sum_{i=1}^{b}\left|SF_i^*\right|} \tag{18}$$

B. Identify the Headings Experimental results and analysis

Table 1 shows the results of some mainstream models on the data set CAS. The models used in experiment 1 and experiment 2 are one-way combined models, and the models used in experiment 3-6 are two-way combined models. The data show that the performance of the two-way joint model is higher than that of the one-way joint model, but the performance of some one-way joint models is not excluded from the two-way joint model.

The one-way joint model slot gated used in experiment 1 learned the relationship between intent and slot attention vector by learning the relationship between intent and slot attention vector, but the information acquisition of intent is limited. The one-way joint model Stack Propagation used in Experiment 2 uses the word-level intention detection mechanism, uses the output of intention detection as the input of the slot filling task, and directly uses the information of intention to guide the slot filling task to improve the performance of the model, making the performance of the model higher than that of the two-way joint models SF-ID network and cm net used in experiments 3 and 4. However, the performance of Stack Propagation model is still limited by the limited filling capacity of Intent guide slot. The Co-Interactive Transformer model used in Experiment 5 is a relatively advanced two-way correlation model in recent years, and the two have achieved good results by establishing two-way connections in two related tasks to consider cross influence. The model in Experiment 6 was optimized based on Experiment

5. In order to establish bidirectional connections between intention and time, and improve model performance.

TABLE I. MODEL RESULTS OF CHINESE DATASET CAIS

Experiment	**Model**	**CAIS**		
		Slot F1	***Acc***	***Sent-acc***
1	Sloted Gated	81.8[a]	94.3	80.5
2	Stack-Propagation	87.8	94.7	84.7
3	SF-ID Network	84.9	94.5	82.4
4	CM-Net	86.2	94.6	84.6
5	Co- Interactive Transformer	88.6	95.2	86.2
6	Our Model	89.2	96.3	87.9

Validity check of each module of the model presented in this document. The experimental results of one-way joint modeling of slot, one-way joint modeling of slot to intention detection and two-way joint modeling of slot by intention detection are compared. The results are shown in Table 2. Although it is impossible to see which model performs better in the two-way joint model, the evaluation indexes of the two-way joint model are higher than those of the two-way joint model. This shows that the performance of bidirectional modeling is better than the performance of unidirectional collaborative modeling in joint modeling of intent detection and slot filling.

TABLE II. RESULTS OF MODEL IN CHINESE DATASET CAIS

Model	**CAIS**		
	Slot F1	***Acc***	***Sent-acc***
Intent➡Slot	88.4	95.8	85.5
Slot➡Intent	88.8	95.6	85.8
Our Model	89.2	96.3	87.9

C. Intent Detection and Slot Filling Analysis

To explore the extent to which Intent detection and slot filling tasks affect the overall performance of the model, using the results of CAIS data set on each model, using a discount plot to visualize the experimental detection results on each model, as shown in Figure 7. The horizontal coordinates of the folding diagram represent the individual model, and the vertical coordinates represent the correct rate of each category (sentence level, intention, and slot). Analysis shows that each model can accurately identify intent labels and slot labels corresponding to most examples.All three types of valid identification numbers are intendedslots Sentence level, therefore slot filling work affects the overall detection accuracy of the joint model.

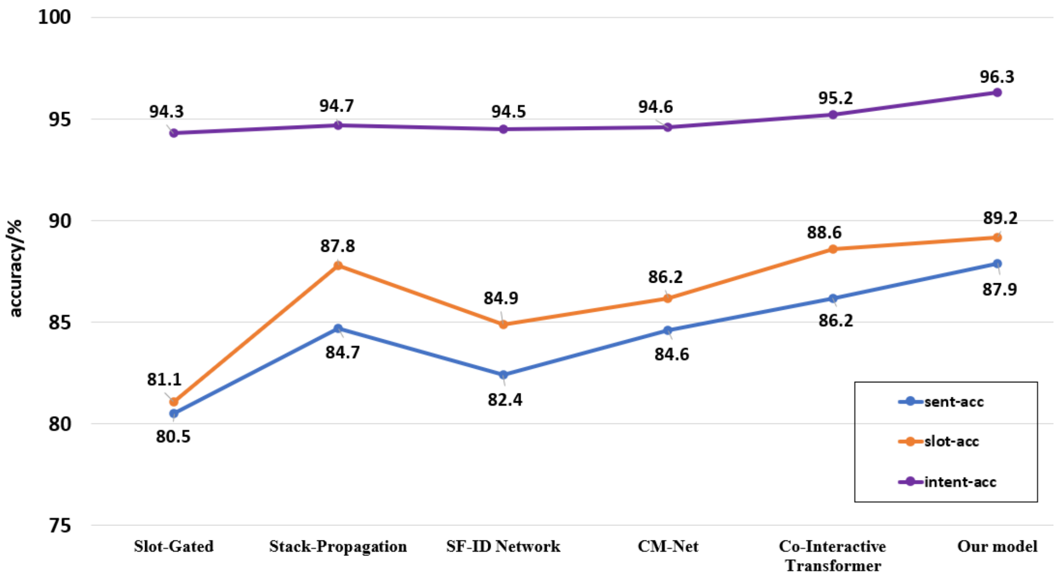

Figure 7. Sentence-level Analysis

V. CONCLUSIONS

In order to investigate the relationship between the two tasks of intention detection and space filling and the degree of influence on the joint model, a three-dimensional analysis of the joint model of intention detection and space filling based on attentional mechanisms was carried out, i.e. from intention detection to space extraction and then to filling.

Unidirectional modeling from slot filling to intention detection, and bidirectional modeling from intention detection to slot filling. The experimental results show that there are three evaluation results for F1 value. The intention accuracy and overall accuracy of bidirectional modeling for detecting and filling intentions are higher than those of joint modeling. Unidirectional.In addition, in both unidirectional and bidirectional common modes, slot filling has a greater impact on the overall detection performance of the model.

REFERENCES

[1] Research on how to understand natural language in task dialogues based on deep learning [D] Xi'an University of Electronic Science and Technology 2021, DOI: 10.27389/d. cnki. gxadu. 2021. 003266.

[2] Guo Suchao, Hao Xia, Yao Xiaobo, and Li Lin.Study on Quiz intention recognition and slot filling joint model of agricultural pest knowledge [J]. Journal of Agricultural Machinery, 2023, 54 (01): 205 215

[3] Zhang J, Bui T, Yoon S, et al. Few Shot Intent Detection via Contrastive Pre Training and Fine Tuning [J]. 2021.

[4] Goo CW, Gao G, Hsu YK, et al.Slot strobe modeling is used for joint time slot filling and intention prediction. In: Proc of Conf., North American Chapter of Computational Linguistics Association.: Human Language Technology, Vol. 2. 2018. 753 − 757.

[5] Haihong E, Beiqing Niu, Zhong Fuchen, and Meinasong, "a novel two-way interconnection model is used for joint intention detection and slotting," in ACL Proc., 2019.

[6] Liu Yijin, fan Dongmeng, Zhang Jinchao, Zhou Jie, Chen Yufeng, and Jinan Xu, "CM net: novel collaborative memory network for oral understanding," in Proc., 2019 of emnlp.

[7] Teng, Qin, Automobile, ecc.Improve your understanding of spoken Chinese through the [C]//International Conference on Voice and Signal ProcessingIEEE, 2020.

[8] Qin l, Liu T, cut W, etc.Interactive Codec for Joint Slot Filling and Intent Detection [C]//International Conference on Acoustics, Sound, and Signal ProcessingIEEE, 2021.

[9] Zhang, H. Wang. A Joint Model for Identifying Understanding Oral Intention and Filling Gaps. Extracted from: The Process of the International Federation of Artificial Intelligence 25th edition 2016.2993-2999

[10] Haffner P, Tur G, Wright JH. Optimize SVM for complex call classification. In: Process International Acoustics Conference and Signal Processing. IEEE (ICASSP 2003) Volume 1 IEEE, 2003.632-635

[11] Shapire RE, Singer Y. BoosTexter: A Text Sorting System Based on Boosting. Machine Learning, 2000,39 (2): 135-168.

[12] Wei P.F., Zeng B., Wang M.H., and Zeng A.Review of Speech Understanding Joint Modeling Algorithms Based on Deep Learning [J]Software Magazine, 2022, 33 (11): 41924216.DOI:10.13328/j.cnki.jos.006385.

[13] Yu Bengang, mladší bratr Fan ZhaoReview of Natural language processing conditions and airport model research [J]Journal of information resource management, 2020, 10 (05): 96111. doi:10.13365/j.jirm.2020.05.096.

[14] Xu P, Sarikaya R. Triangle CRF based on convolutional neural network is used for joint intention detection and slot filling [C] / / / automatic speech recognition and understanding (asru), IEEE workshop in 2013. IEEE, 2013.DOI:10.1109/ASRU.2013.6707709.Neter J R, Guzide O. Deep learning in natural language processing [J]. 2018(1).

[15] Ravuri S V, Stolcke A. Repetitive Neural Networks and LSTM Models for Word Pronunciation Classification [C]//2015DOI:10.21437/Interlingua2015 42.

[16] Ni j, Young T, pandelea V et coll.Research progress in dialogue systems based on deep learning [J]2021.

[17] Yu, Xie En, JinRNN semantic frame analysis model, using dual models for intention detection and time slot filling [J]2018.

[18] Zhang C, Li Yi, doon et al.Filling and intention testing combination based on neural capsule network [C]//annual paper collection 57 Society for Computational linguistics2019.