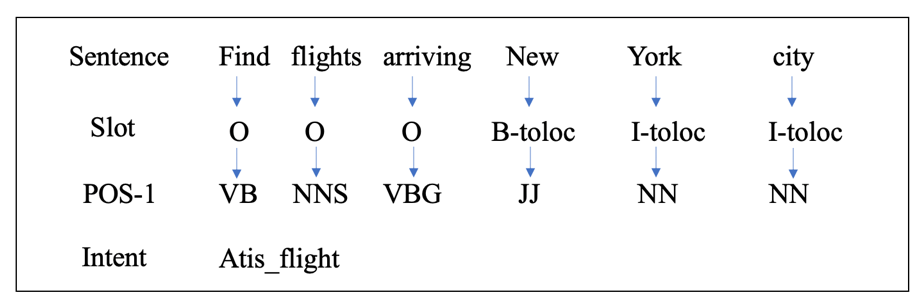
POS Scaling Attention Modeling for Joint Slot Filling and Intent Classification

**Abstract**

With the development of natural language processing technology, the joint intent recognition with attention mechanism is also actively explored. However, in the actual application process, the slot filling accuracy, domain perturbation and model convergence speed in dialog system have major problems. In order to further improve the accuracy of intent recognition and the speed of convergence, this paper will introduce a part-of-speech scaling attention mechanism to exploit the semantic features of short texts and improve the detection performance of slot filling. The experiments show that our proposed model significantly improves sentence-level semantic frame accuracy with 3.2% and 2.2% on the benchmark ATIS and Snips datasets respectively. Furthermore, our proposed model’s convergence speed is also improved.

**Introduction**

Natural language understanding technology (NLU) is an important component of the dialogue system. NLU is dedicated to improving the extraction of deep semantic information. It is mainly divided into two tasks: intent classification and slot filling. These two tasks are mainly focus on predicting user intent and identify semantic components. Take a book flight-related utterance as an example, "find flights arriving New York city" as shown in Figure 1.



Both intent classification and slot filling can be considered as classification tasks. The only difference between the two is that the granularity of the classification is inconsistent. The intent classification task is to predict the intent of the entire sentence, and the slot filling task is to predict the semantic component corresponding to each word. Usually, slot filling can be treated as a sequence labeling task, popular algorithm contains conditional random fields (CRF) [1], HMM [2], intent classification task is a classification task in sentence level, such as SVM [3], RNN [4] [6] [7].

In NLU, attention mechanism was introduced to the model in order to provide the precise focus, which allows the network to learn where to pay attention in the in- put sequence for each output label. Liu et al. [8] use attention on the hidden layer of two-way recurrent neural network to capture the important semantics of a sentence. Chen et al. [9] proposed a word-level attention mechanism.

Considering the mutual influence of intent classification and slot filling and the error propagation of different tasks, joint training has gradually occupied the field of intent recognition in recent years [11] [12][13]. Goo et al. [10] proposed a Slot-Gate structure to build establishment of intent classification and slot filling; Zhang et al. [14] introduced the capsule network into the field of intent recognition and modeled the semantic hierarchy to further improve performance; E et al. [15] explore the way of Interaction of slot filling task and intent classification.

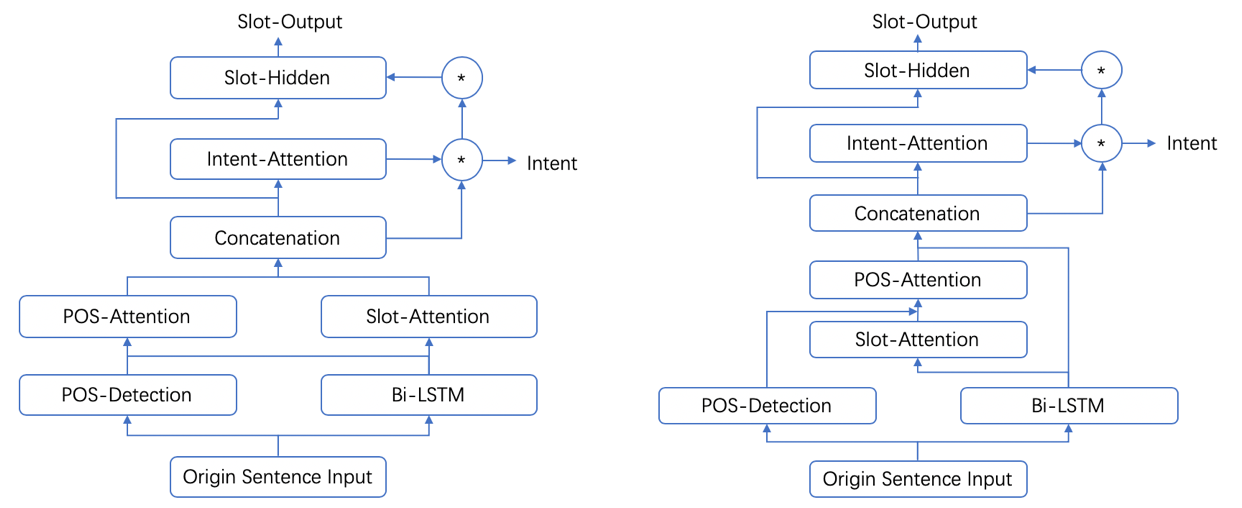
In addition, BERT and memory networks have also been continuously explored and applied in intent detection. However, the prior work did not utilize part-of-speech information in attention mechanism. The contributions of this paper include three-folds:

1. Our proposed architecture can improve the accuracy of slot filling and intent classification.
2. Our proposed architecture can improve the convergence speed.
3. The experiments on SLU datasets show the generalization and the effectiveness of the proposed POS scaling attention.

**Architecture**

This section mainly introduces our POS-Scaling attention model. With the excellent performance of the attention mechanism in NLP, the attention mechanism is roughly divided into self-attention and attention. But self-attention mechanism will greatly increase the calculation in practical application scenarios. Compared with the self-attention mechanism, providing traditional external information to attention mechanism can reduce the amount of calculation.

The model architecture is illustrated in Figure 2, where there are two different model: (a) is one with part-of-speech encoding and slot filling parallelly and (b) is one with part-of-speech encoding after slot filling module.

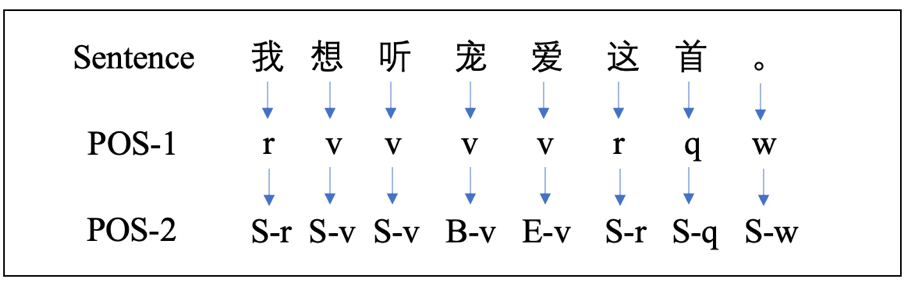


**Pre-Encoding module**

The bidirectional long short-term memory (BLSTM) model takes a word sequence as input, and then generates the forward hidden state and the backward hidden state . The final hidden state at time step i is a concatenation of and , .

**POS detection module**

The part-of-speech of text sequence contains nouns, verbs, etc., then this information apply one-hot encoding and obtain part-of-speech vector .



In the Chinese scenario, due to Chinese part-of-speech corresponds to multiple words, there are two parts of speech tagging mechanisms, as shown in Figure 3: one is to apply parts of speech to each word equally, and the other uses BIE labels.

**POS scaling attention module**

The part-of-speech scaling attention module takes part-of-speech vector and the hidden vector as inputs, and calculates the weight of each word.

where the POS scaling weights α are calculated as below:

where is the activation function, is the weight matrix of a feed-forward neural network, is the cumulative sum of specific dimension.

**Parallel POS encoding**

**Slot-Filling**

Slot filling module aims to map the hidden vector to . For each hidden state , we compute the slot context vector as the weighted sum of LSTM’s hidden states by the slot attention weights：

where the slot attention weights α are calculated as below:

In parallel POS encoding mode, the part-of-speech scaling attention output and the slot-filling attention output can combine a new slot state hidden vector:

Then the hidden state vector is utilized for slot filling:

**Intent Classification**

The intent context vector cI can also be computed in the same manner as cS, but the intent detection part takes the last hidden state of BLSTM, POS scaling attention and slot filling attention:

Then the hidden state vector is utilized for slot filling

**Joint optimization**

To obtain both slot filling and intent prediction jointly, the objective is formulated as:

It is assumed that the slot filling task and the intent classification task are relatively independent, and jointly optimize the corresponding probability loss function.

**Slot-POS encoding mode**

Unlike parallel POS encoding mode, Slot-POS encoding mode changes the input of the part-of-speech scaling attention mechanism to the output of the slot-filled attention module:

Then the hidden state vector is utilized for slot filling:

**Experimental**

To evaluate our proposed model, we conduct experiments on the benchmark datasets, ATIS and Snips datasets. The statistics are shown in Table 4~5:

Dataset

In the domain of natural language understanding, the Airline Travel Information System Dataset (ATIS) is widely used in research. The dataset contains audio recordings of people making flight reservations, the training set contains 4,978 utterances, the test set contains 893 utterances. There are 120 slot tags and 21 intent types in the training set. In addition, in order to ensure the general performance of the model, the same experiment was performed on the Snips dataset. The Snips dataset is a dialog collected by the Snips voice assistant. The training set contains 13,784 utterances, the test set contains 700 utterances. There are 72 slot labels and 7 intent types. Table 1 shows the intents and associated utterance examples.

The ATIS and Snips datasets are both English datasets. Snips is more complex than ATIS. As shown in Table 1, Snips has more types of intent and has richer data, including weather and information, book restaurant, and play music, while the ATIS is all about flight information with similar vocabularies across them .

|  |  |
| --- | --- |
| Intent | Utterance Example |
| GetWeather | will it snow in mt on June 13 2038 |
| BookRestaurant | can you get me reservations for a highly rated restaurant in seychelles |
| SearchCreativeWork | find a movie called living in america |
| RateBook | rate the current novel four of 6 stars |
| PlayMusic | play music from clark kent in the year 1987 |
| SearchScreeningEvent | is unbeatable harold at century theatres |

Considering the differences between different languages, this paper conduct experiment on the Chinese dataset (CISD); the Chinese data set (CISD) is the dialog data from the Chinese dialogue platform. The training set contains 3567 utterances and the test set contains 891 utterances. There are 8 intent types and 60 slot tags. Table 3 show the intents and associated utterance examples.

|  |  |
| --- | --- |
| Intent | Utterance Example |
| get\_news\_info | 看娱乐新闻 |
| play\_music | 火箭少女101的卡路里 |
| system\_control | 激活密码设置 |
| schedule | 请把菜单上的是什么从周二改为晚上5点这周五 |
| alarm | 请删除6:45的闹铃 |
| weather\_info | 丽水的天气怎么样 |

|  |  |
| --- | --- |
| **Intent** | **Size** |
| System\_control | **314** |
| Play\_music | **2598** |
| Schedule | **249** |
| Alarm | **385** |
| Ask\_knowledge | **87** |
| Get\_place\_info | **378** |
| Weather\_info | **302** |
| Get\_news\_info | **145** |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **ATIS** | | | **Snips** | | |
| **Slot**  **(F1)** | **Intent**  **(Acc)** | **Sentence**  **(Acc)** | **Slot**  **(F1)** | **Intent**  **(Acc)** | **Sentence**  **(Acc)** |
| **Attention-Based** | 94.2 | 91.1 | 78.9 | 87.8 | 96.7 | 74.1 |
| **Slot-Gate (Full)** | 94.8 | 93.6 | 82.2 | 88.8 | 97.0 | 75.5 |
| **Slot-Gate (Intent)** | 95.2 | 94.1 | 82.6 | 88.3 | 96.8 | 74.6 |
| **Parallel POS-Scaling** | **95.64** | **95.7** | **85.8** | **89.18** | 96.86 | **76.85** |
| **Slot-POS** | 95.34 | 93.7 | 83.42 | **89.18** | **97.71** | 76.57 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **CSID** | | |
| **Slot**  **(F1)** | **Intent**  **(Acc)** | **Sentence**  **(Acc)** |
| **Slot gate (Full)** | 85.72 | 97.53 | 76.76 |
| **Parallel POS-Scaling (BIES)** | 85.91 | 97.64 | 78.11 |
| **Slot-POS (BIES)** | 83.19 | 94.05 | 72.27 |
| **Parallel POS-Scaling (average)** | **86.57** | **97.97** | **78.45** |
| **Slot-POS (average)** | 85.75 | **97.97** | 77.77 |

**Experimental parameters**

Given the size of the dataset, we set the size of the hidden vector to 64, the learning rate to 0.001, the optimizer is Adam, the maximum epoch is set to 25 with an early-stop strategy.

**Result analysis**

To evaluate the performance of our proposed model, we use the F1 score to evaluate the performance of slot filling, the accuracy to evaluate the performance of intent classification, the sentence-level frame accuracy to evaluate the overall performance of the model.

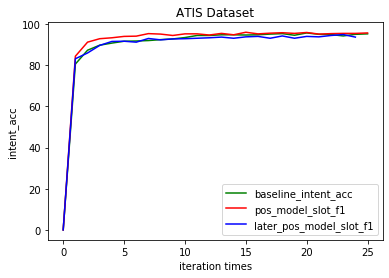
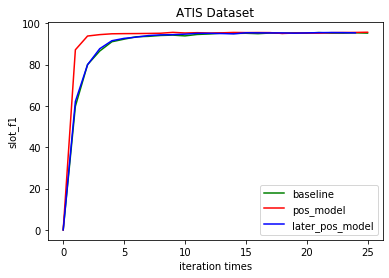


图5 ATIS Slot-F1 图6 ATIS Intent-acc

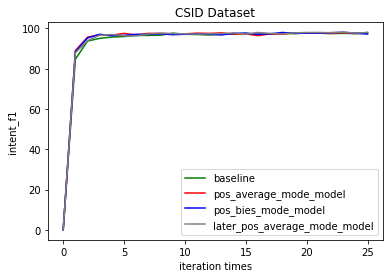
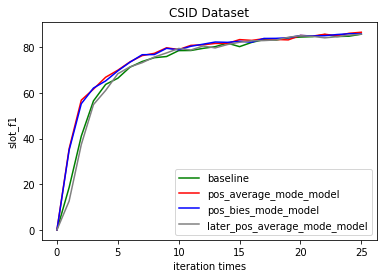


图7 CAIS Slot-F1 图8 CAIS Intent-acc

Table 4 shows that our proposed model outperforms the baselines on both datasets. On the ATIS dataset, we achieve 95.64% on slot filling, 95.7% on intent detection and 85.8% on semantic frame. For the Snips dataset, we achieve 89.18% on slot filling, 97.71% on intent detection and 76.85% on semantic frame. Parallel POS scaling mode outperforms Slot-POS mode on ATIS dataset， while Slot-POS mode outperforms Parallel POS scaling mode. The probable reason is that a simpler SLU task, such as ATIS, contains the enough distribution of part-of-speech.

Table 4 shows that our proposed model outperforms the baselines on Chinese datasets. We achieve 86.57% on slot filling, 97.97% on intent detection and 78.45% on semantic frame. Parallel POS scaling mode outperforms Slot-POS mode on ATIS dataset while Slot-POS mode outperforms Parallel POS scaling mode. Besides, applying parts of speech to each word equally performs better than BIES labels mode.

According to Figs. 5-8, it can be easily found that the convergence speed of our proposed model has been greatly improved.

It is obvious that our proposed model performs better than baseline, and it has also been verified that POS scaling attention module works in both Chinese and English datasets.

**Conclusion**

This paper focuses on promoting model performance and convergence speed. The experiments show that part-of-speech scaling attention can improve the performance and convergence speed, and generalization.

In addition, since part-of-speech information is more interpretable, part-of-speech scaling attention is more valuable for NLU, especially short text understanding tasks.

**Future Work**

This paper introduces part-of-speech scaling attention modules to optimize the model, and explore the position of part-of-speech module. But we need to further explore interaction the part-of-speech and intent classification.

**References**

1. Lafferty, J., McCallum, A., & Pereira, F. C. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data.
2. Rabiner, L., & Juang, B. (1986). An introduction to hidden Markov models. ieee assp magazine, 3(1), 4-16.
3. Wang, L. (Ed.). (2005). Support vector machines: theory and applications (Vol. 177). Springer Science & Business Media.
4. Mikolov, T., Karafiát, M., Burget, L., Černocký, J., & Khudanpur, S. (2010). Recurrent neural network based language model. In Eleventh annual conference of the international speech communication association.
5. Liu, B., & Lane, I. (2016). Attention-based recurrent neural network models for joint intent detection and slot filling. *arXiv preprint arXiv:1609.01454*.
6. Ravuri, S., & Stolcke, A. (2015). Recurrent neural network and LSTM models for lexical utterance classification. In *Sixteenth Annual Conference of the International Speech Communication Association*.
7. Ravuri, S., & Stolcke, A. (2016, March). A comparative study of recurrent neural network models for lexical domain classification. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 6075-6079). IEEE.
8. Liu, B., & Lane, I. (2016). Attention-based recurrent neural network models for joint intent detection and slot filling. *arXiv preprint arXiv:1609.01454*.
9. Chen, S., & Yu, S. (2019, July). WAIS: Word Attention for Joint Intent Detection and Slot Filling. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 33, pp. 9927-9928).
10. Goo, C. W., Gao, G., Hsu, Y. K., Huo, C. L., Chen, T. C., Hsu, K. W., & Chen, Y. N. (2018, June). 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)* (pp. 753-757).
11. Li, C., Li, L., & Qi, J. (2018). A self-attentive model with gate mechanism for spoken language understanding. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 3824-3833).
12. Wang, Y., Shen, Y., & Jin, H. (2018). A bi-model based rnn semantic frame parsing model for intent detection and slot filling. *arXiv preprint arXiv:1812.10235*.
13. Gupta, S., Shah, R., Mohit, M., Kumar, A., & Lewis, M. (2018). Semantic parsing for task oriented dialog using hierarchical representations. *arXiv preprint arXiv:1810.07942*.
14. Zhang, C., Li, Y., Du, N., Fan, W., & Yu, P. S. (2018). Joint slot filling and intent detection via capsule neural networks. *arXiv preprint arXiv:1812.09471*.
15. Haihong, E., Niu, P., Chen, Z., & Song, M. (2019, July). A novel bi-directional interrelated model for joint intent detection and slot filling. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 5467-5471).
16. Schmid, H. (1994, August). Part-of-speech tagging with neural networks. In *Proceedings of the 15th conference on Computational linguistics-Volume 1* (pp. 172-176). Association for Computational Linguistics.
17. Chen, Q., Zhuo, Z., & Wang, W. (2019). Bert for joint intent classification and slot filling. *arXiv preprint arXiv:1902.10909*.
18. Shridhar, K., Dash, A., Sahu, A., Pihlgren, G. G., Alonso, P., Pondenkandath, V., ... & Liwicki, M. (2019, July). Subword Semantic Hashing for Intent Classification on Small Datasets. In *2019 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-6). IEEE.
19. Liu, Y., Meng, F., Zhang, J., Zhou, J., Chen, Y., & Xu, J. (2019). CM-Net: A Novel Collaborative Memory Network for Spoken Language Understanding. *arXiv preprint arXiv:1909.06937*.
20. 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 preprint arXiv:1909.02188*.