Joint Intention Detection and Semantic Slot Filling Based on BLSTM and Attention

Chen Tingting
Inner Mongolia Normal University
School of Computer Science and Technology,
Hohhot, China
e-mail: 18347389831@163.com

Lin Min
Inner Mongolia Normal University
School of Computer Science and Technology,
Hohhot, China
e-mail: linmin@imnu.edu.cn

Li Yanling
Inner Mongolia Normal University
School of Computer Science and Technology,
Hohhot, China
e-mail: liyanling7871397@163.com

***Abstract*—Spoken language understanding (SLU) of the dialogue system usually involves two tasks: intent detection and semantic slot filling. The current Joint intention detection and semantic slot filling has become the mainstream method of SLU research. A Bidirectional long short-term memory (BLSTM)model based on the attention mechanism is used to jointly identify the intent and semantic slot filling of the Hohhot bus query. The experimental results show that the model achieves a good performance in the intent detection and semantic slot filling, and the result based on the character mark is better than the one based on word mark. The F1 score is better than the others based on the LSTM model.**

Keywords-long short-term memory; joint intention detection; attention mechanism

I. Preface

With the development of speech detection, speech synthesis and natural language processing technology, the spoken dialogue system [1] has gradually become a reality. So far, many oral dialogue systems in different languages have been developed internationally, and some have already been put into commercial use. The spoken dialogue system is widely used, such as information inquiry, online reservation service, and onboard system. It can provide convenient consultation services for people at any time and place.

The SLU system is a key component of the spoken dialogue system. SLU systems often need to identify the speaker's intent and extract semantic components from the user's utterance. These two tasks are often referred to as intent detection and semantic slot filling [2].

At present, in order to improve performance, the joint intention detection and semantic slot filling has become the mainstream method of SLU research. This paper mainly focuses on joint modeling of intent detection and semantic slot filling, and uses the BLSTM method based on attention mechanism to identify the intention and semantic slot filling for Hohhot bus inquiry sentences, so as to recognize the origin, destination and travel intention of Hohhot residents. It can be used to construct a question answering system of public transport, serving the construction of information technology in Western cities, aiming at providing a convenient urban travel service for tourists.

II. Related Technology

Semantic slot filling is similar to named entity detection tasks and is a richer representation of named entities. There have been extensive researches on named entity detection. The research methods mainly include rule-based methods [3] and statistical-based methods. The rule-based method is costly to implement, the system performance is heavily dependent on the artificially designed rule method, and the portability is poor, so the method is gradually eliminated. In the past decade, large-scale corpus-based statistical methods have become mainstream research methods. Multiple machine learning methods have been successfully applied to named entity detection tasks, including supervised learning methods, semi-supervised learning methods, unsupervised learning methods, and hybrid methods. Among which the main models used in supervised learning are hidden Markov models [4] (HMM), maximum entropy models (ME) [5], support vector machines (SVM) [6], conditional random fields (CRF) [7] et al. Semi-supervised learning [8] using the labeled small data set (seed data) bootstrap learning. Unsupervised learning using lexical resources (such as WordNet) [9] for context clustering. The hybrid approach combines several models or uses statistical methods and manually summarized knowledge bases[10]. These statistical-based learning methods use the feature-based corpus for feature extraction. Their disadvantages are that a large amount of manually labeled training data is needed on the training corpus, and manual construction of features is required.

The deep learning method has the advantage of automatically extracting features. For example, the recurrent neural network (RNN) can give the neural network the ability to explicitly model time by adding a self-joining hidden layer across time points. In the RNN, the output of the hidden layer will not only enter the output but also enter the hidden layer of the next moment. Because of the problem

978-1-7281-1410-1/19/$31.00 ©2019 IEEE

of gradient explosion and gradient disappearance in RNN, in order to overcome this defect, a long short term memory network (LSTM) has emerged. LSTM is a special RNN that can learn long-term dependencies. LSTM was originally proposed by Hochreiter and Schmidhuber, and many researchers have carried out a series of optimization and improvement work for LSTM. LSTM has been rapidly developed and is now widely used in various aspects of natural language processing. In order to solve the problem that the unidirectional LSTM cannot capture the relevant semantic information of the text, the BLSTM appears. The BLSTM can not only consider the bidirectional word sequence input but also solve the problem of slow learning of the feedforward neural network.

The attention mechanism breaks the structure of traditional encoder and decoder, which are dependent on the internal limitation of a fixed length vector. By retaining the intermediate output of the input sequence of the LSTM encoder, a model is trained to selectively learn the input, and the output sequence is associated with the model output. It is widely used in machine translation, speech detection, image annotation, and many other fields. D. Bahdanau [12] et al first applied attention to neural network machine translation, and then Luong [13] et al proposed two attention mechanisms, global mechanism and local mechanism, which were extended in RNN.

Intent detection can be regarded as a classification problem, which can be solved using classification methods, such as SVM model [14], Adaboost method [15] and so on. The general steps of these traditional classification methods are to extract the features of the input utterances, then train the classifier, and finally test with the trained classification model. The disadvantage is that the features need to be extracted manually. When the data set changes, features need to be redesigned. This often evolves into problems such as feature design and feature selection, which consumes a lot of manpower.

The above research methods have achieved good performance in the general field for semantic slot filling and intent detection respectively. The joint detection model can combine two tasks and solve them at the same time. Jeong et al used a joint approach, the triangular-chain CRF model, to solve the semantic slot filling and intent detection tasks, and jointly capture the intrinsic relationship between them [16]. Making a certain contribution to joint detection, but it takes time and effort, and there must be enough training corpus. In 2013, Xu from Microsoft used the convolutional neural network and triangular-chain conditional random field (CNN-TriCRF) model for joint intent detection and semantic slot filling [17]. Compared to Jeong, the proposed Triangular-chain CRF model increases the intent detection and semantic slot filling tasks by 1.02% and 1%, respectively, but the parameters of the model training are more and more complex.

BLSTM method based on attention mechanism is adopted in this paper. BLSTM can accumulatively memorize different input words to obtain vectors containing semantic and grammatical information. After adding the attention mechanism, it can pay attention to the influence of different input sequences on output, identify context information clearly and grasp keywords accurately, which can better improve the performance of the model.

III. BLSTM Model Based on Attention Mechanism

In this paper, intention detection and semantic slot filling are jointly modeled, and the method of BLSTM based on the attention mechanism is adapted to realize the joint intention detection and semantic slot filling for the bus information query in Hohhot.

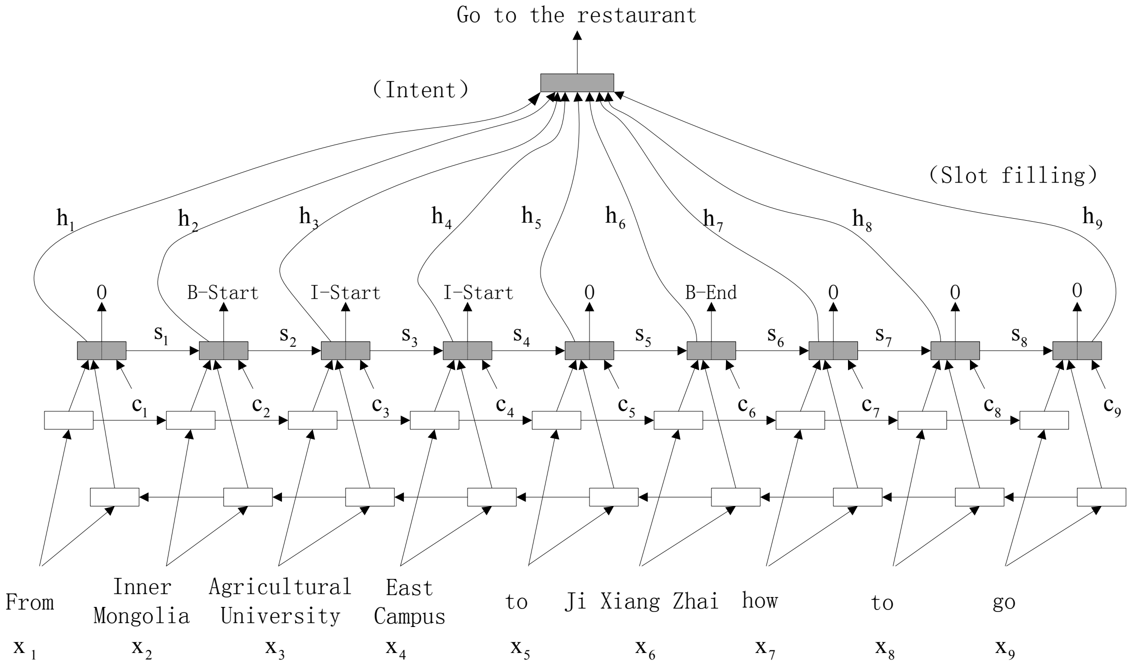

Figure 1. Joint model of BLSTM based on attention mechanism

The attention-based BLSTM model for joint intent detection and semantic slot filling is shown in Figure 1. In semantic slot filling, the goal is to map the sequence $x=(x_1,...,x_T)$ to its corresponding annotation sequence $y=(y_1,...,y_T)$. The BLSTM model reads the word sequence forward and backward. The forward LSTM reads the word sequence in its original order and generates a hidden state fh_i at each time step. Similarly, the backward LSTM reads the word sequence in reverse order and generates a hidden state sequence $(bh_T,...,bh_1)$. The final encoder hidden state h_i at each time step *i* is a concatenation of the forward state fh_i and the backward state bh_i, ie $h_i=[fh_i,bh_i]$.

The last state of the forward and backward LSTM carries information for the entire source sequence. The final state of the LSTM is used to calculate the initial decoder hidden state. The decoder is a unidirectional LSTM. At each decoding step *i*, the decoder state s_i is related to the previous decoder state s_{i-1}, the previously issued tag y_{i-1}, the aligned encoder hidden state h_i, and the context vector c_i:

$$s_i=f(s_{i-1},y_{i-1},h_i,c_i) \qquad (1)$$

wherein the context vector c_i is calculated as the weighted sum of the encoder states $h=(h_1,...,h_T)$:

$$c_i=\sum\nolimits_{j=1}^{T}\alpha_{i,j}h_j \qquad (2)$$

$$\alpha_{i,j}=\frac{\exp(e_{i,j})}{\sum\nolimits_{k=1}^{T}\exp(e_{i,k})} \qquad (3)$$

$$e_{i,k}=g(s_{i-1},h_k) \qquad (4)$$

where g is the feedforward neural network. At each decoding step, the encoder state h_i is taken as input and the context vector c_i provides additional information to the decoder.

For joint modeling of intent detection and semantic slot filling, this experiment adds an additional decoder to the intent detection part that shares the same encoder as the semantic slot fill decoder. The intent decoder only generates one output, so no alignment is required. For joint intention detection and semantic slot filling, the intent probability distribution is generated using the hidden state h pre-calculated by the bidirectional LSTM, and the weighted average of the hidden state h is changed.

IV. Experiment

A. Dataset

This paper collects data for the public transportation sector in Hohhot and selects five locations that people often visit, including hospitals, schools, views, restaurants, and shopping malls, as the intent category of the statement. The departure and destination are populated as two semantic slots. The data set description is shown in Table I.

TABLE I. DATA SET DESCRIPTION

Field	Number of questions	Number of Intentions	Intent example	Number of semantic slots	Semantic slot example
Bus	695,880	5	Go to school...	2	Departure place, destination

The data used in this experiment was collected by the laboratory itself, with a total of 695,880 statements. There are roughly four types of statements, as shown in Table II. Among them, there are 601,110 sentences in training sets, 63,360 sentences in test sets, and 31,410 sentences in development sets. The departure, destination, and irrelevant items are labeled S, E, and O, respectively. Among them, the training sets, the test sets, and the development sets are devided the data according to a certain proportion are shown in Table III.

TABLE II. FOUR STATEMENT TYPES

Statement type	Example
E-O	Inner Mongolia Normal University (E)- specific location (O)
O-E	How to get to (O)- Inner Mongolia Normal University (E)
O-E-O	Go to (O)-Inner Mongolia Normal University (E)- route (O)
O-S-O-E-O	From (O)-Zhao Jun Museum (S)- to (O)-Inner Mongolia Normal University (E)- route (O)

TABLE III. NUMBER OF SETS OF STATEMENTS

Statement type	Number of statements		
	Training set	*Test set*	*Development set*
E-O	12,600	3,600	1,800
O-E	3,150	900	450
O-E-O	18,900	5,400	2,700
O-S-O-E-O	566,460	53,460	26,460

Next, the sentences are preprocessed, word vectors are generated, and the data are labeled based on word and character: Jieba segmentation is used for word-based segmentation, and some places' name is added into the custom dictionary to ensure the accuracy of word segmentation. The two groups of data are marked in the same way. The labeling instructions are shown in Table IV and Table V, and the examples of word-based labeling are shown in Table VI.

TABLE IV. INTENT TAG DESCRIPTION

Intention mark	Intention category
hospital	Go to the hospital
school	Go to school
shopping	Go to the mall
view	Go to attractions
restaurant	Go to the restaurant

TABLE V. SEMANTIC SLOT TAG DESCRIPTION

Semantic slot tag	**Semantic slot category**
B-Start	Departure
I-Start	
B-End	Destination
I-End	
O	Other

TABLE VI. JOINT IDENTIFICATION

Statement	**From**	**Inner Mongolia**	**Agricultural University**	**East Campus**	**to**	**Ji Xiang Zhai**	**how**	**to**	**go**
Semantic slot tag	O	B-Start	I-Start	I-Start	O	B-End	O	O	O
Intention mark	restaurant								

B. Evaluation Standard

In this paper, accuracy is used to evaluate intention detection. In data sets, some statements may have more than one intention. In this paper, according to the type of destination they want to reach in the data, they are categorized as intention categories.

The experiment uses the F1 score to measure the filling result of the semantic slot. A semantic slot is extracted correctly, indicating that both the origin and destination are correct. The evaluation index of the F1 score is defined as follows:

$$F\mathbf{1} = \frac{\mathbf{2}\times P\times R}{(P+R)} \tag{5}$$

P is the accuracy rate and R is the recall rate.

C. Experimental Results and Analysis

In this paper, the LSTM model was used as the baseline of the experiment, and the LSTM+attention, BLSTM, and BLSTM+attention models were compared. The character-based and word-based results are shown in Table VII and Table VIII.

TABLE VII. CHARACTER-BASED RESULTS

Model	**Semantic slot filling F1 score (%)**	**Intent detection accuracy (%)**
LSTM	84.15	84.05
LSTM+attention	99.02	84.20
BLSTM	99.03	85.95
BLSTM+attention	**99.84**	**88.79**

TABLE VIII. WORD-BASED RESULTS

Model	**Semantic slot filling F1 score (%)**	**Intent detection accuracy (%)**
LSTM	98.62	76.11
LSTM+attention	95.03	77.50
BLSTM	98.78	**80.71**
BLSTM+attention	**99.17**	80.2

The results of character segmentation show that the F1 score of the LSTM model in semantic slot filling is 84.15%, and the intent detection accuracy rate is 84.05%. When the attention mechanism is added, the model can focus on the keyword better, and the performance of semantic slot filling is improved 14.87%, the performance of intent detection increased by 0.15%. When changed to the BLSTM model, not only the forward information of semantics but also the reverse information of semantics can be obtained. The performance of semantic slot filling is improved by 14.88%, and the performance of intention detection is improved by 1.9%. By introducing the BLSTM+attention model, the model can focus on the keywords well and the semantic reverse information is added. The performance of the semantic slot filling is improved by 15.69%. It has increased by 4.74% and achieved an optimal result.

In the results based on word segmentation, the F1 score of the LSTM model in semantic slot filling is 98.62%, and the intent detection accuracy rate is 76.11%. When the attention mechanism is added, the performance of semantic slot filling decreases by 3.59%, which is a slight decrease, the performance of intent detection increased by 1.39%. The reason for the slight decrease may be the uncertainty of word segmentation. When changed to the BLSTM model, the performance of semantic slot filling increased by 0.16%, the performance of intent detection increased by 4.6%. And BLSTM+attention was introduced. For the model, the performance of semantic slot filling is improved by 0.55%, and the intent detection performance is improved by 4.09%. Overall, the BLSTM+attention model results are still optimal.

Compared with the two methods, the performance of word segmentation is better than that of word segmentation. The reason is that in the word segmentation, the accuracy of the segmentation and the size of the custom dictionary will affect the results of the model training.

V. CONCLUSION

In this paper, the model based on BLSTM and attention mechanism is used for joint intention detection and semantic slot filling. And compared with LSTM, LSTM+attention, BLSTM and BLSTM+attention methods on word-based and character-based data sets. The experimental results show that the BLSTM+attention model achieves optimal results in terms of the accuracy of intent detection and semantic slot filling, and verifies the effectiveness of the proposed method.

There are still some shortcomings in this paper: Firstly, the types of entities involved in this paper are relatively few, and the semantic slot types in the extended routing statements will be further refined in the following so that the semantic slot types of the query statements are more abundant. Secondly, the advantage of the BLSTM+attention model is that adding attention mechanism can pay attention to the influence of different input sequences on output. However, since the global normalization problem of the output sequence is not taken into account, the output of the semantic slot is prone to bias problems. Further research is needed.

Acknowledgment

Thanks to the National Natural Science Foundation of China (61562068, 11704229, 61640204, 61806103). Inner Mongolia Natural Science Foundation (2017MS0607). Inner Mongolia Civil Affairs Commission Mongolian Informationization Special Support Subproject (MW-2014-MGYWXXH-01). Inner Mongolia Autonomous Region "Engineering Youth Innovation and Entrepreneurship Talent Project. Inner Mongolia Science and Technology Plan Project (Mongolian MOOC Online Interactive Learning Platform R&D Operation and National Education Big Data Analysis Research)support for this article.

References

[1] Y.Wang and A.Acero.Rapid development of spoken language understanding grammars. Speech Communication, 48(3-4):390–416, 2006.

[2] Liu B, Lane I. Attention-Based Recurrent Neural Network Models for Joint Intent Detection and Slot Filling[J]. 2016.

[3] J.Dowding, J. M.Gawron, D.Appelt, J. Bear, L. Cherny, R. Moore, and D. Moran, “Gemini: A natural language system for spoken-language understanding,” in Proc. Association for Computational Linguistics, pp. 54–61,1993.

[4] Zhao Linying. Research on Chinese Named Entity Recognition Based on Hidden Markov Model [D]. Xidian University, 2008.

[5] Wang Jiangwei. Chinese Named Entity Recognition Based on Maximum Entropy Model [D]. Nanjing University of Science and Technology, 2005.

[6] Chen Xiao. Chinese name recognition based organization support vector machine [D]. Shanghai Jiaotong University, 2007.

[7] Guo Jiaqing. Research on Named Entity Recognition Based on Conditional Random Fields[D]. Shenyang Institute of Aeronautical Engineering, 2007.

[8] Sui Chen. Research on Chinese Named Entity Recognition Based on Deep Learning [D]. 2017.

[9] Zhu Huifeng, Zuo Wanli, He Fengling, et al. An ontology-based text clustering method[J].Journal of Jilin University: Sci Ed, 2010, 48(2): 277-283.

[10] Huang Jizhou. Research and implementation of automatic extraction algorithm for chat robot knowledge base [D]. Chongqing University, 2006.

[11] Jin Liuke. Biomedical named entity recognition based on recurrent neural network [D]. Dalian University of Technology, 2016.

[12] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” arXiv preprint arXiv:1409.0473, 2014.

[13] Luong M T, Pham H, Manning C D. Effective Approaches to Attention-based Neural Machine Translation[J]. Computer Science, 2015.

[14] Haffner P, Tur G, Wright J H.Optimizing SVMs for complex call classification[C]//IEEE International Conference on Acoustics. CiteSeer, 2003: I-632-I-635vol.1.

[15] Schapire R E, Singer Y. BoosTexter: a boosting-based system for text categorization[J]. Machine Learning,2000,39(2-3):135-168.

[16] Jeong M , Lee G G . Triangular-chain conditional random fields[J]. IEEE Transactions on Audio Speech & Language Processing, 2008, 16(7):1287-1302.

[17] Xu P, Sarikaya R. Convolutional neural network based triangular crf for joint intent detection and slot filling[C]//Automatic Speech Recognition and Understanding (ASRU), 2013 IEEE Workshop on. IEEE, 2013: 78-83.

[18] Xu Yuxiang, Che Wanxiang, Liu Ting. Semantic Slot Recognition Based on Bi-LSTM-CRF Network[J]. Intelligent computers and applications, 2017, 7(6): 91-94.

[19] Sun Xin, Wang Houfeng. Analysis of Intent Recognition and Constraint of Question in Question and Answer[J]. Journal of Chinese Information Processing, 2017, 31(6).

[20] Vu N T, Gupta P, Adel H, et al. Bi-directional recurrent neural network with ranking loss for spoken language understanding[C]// IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2016.

[21] Hori C, Hori T, Watanabe S, et al. Context-Sensitive and Role-Dependent Spoken Language Understanding Using Bidirectional and Attention LSTMs[C]// INTERSPEECH. 2016:3236-3240.

[22] Barahona L M R, Gasic M, Mrkšić N, et al. Exploiting Sentence and Context Representations in Deep Neural Models for Spoken Language Understanding[J]. 2016.

[23] Vukotic V, Raymond C, Gravier G. Is it time to switch to Word Embedding and Recurrent Neural Networks for Spoken Language Understanding?[J].2015.