PAPER • OPEN ACCESS

Review of Intent Detection Methods in the Human-Machine Dialogue System

View the article online for updates and enhancements.

Review of Intent Detection Methods in the Human-Machine Dialogue System

Jiao Liu, Yanling Li[*] and Min Lin

College of computer science and technology, Inner Mongolia Normal University

[*]liyanling7871397@163.com

Abstract. Spoken language understanding is an important part of the human-machine dialogue system, intent detection is a sub-task of spoken language understanding, and it is very important. The accuracy of intent detection is directly related to the performance of semantic slot filling, and it is helpful to the following research of the dialogue system. Considering the difficulty of intent detection in human-machine dialogue system, the traditional machine learning method cannot understand the deep semantic information of user's discourse. This paper mainly analyzes, compares and summarizes the deep learning methods applied in the research of intent detection in recent years, and further considers how to apply deep learning model to multi-intent detection task, so as to promote the research of multi-intent detection methods based on deep neural network.

1. Introduction

With the advent of the era of artificial intelligence, more and more intelligent products have been widely used in our daily life, such as emotional care robot, personal phone assistant Siri, voice assistants Google Now and intelligent chat robot Xiao Bing from Microsoft Research Asia, etc. These intelligent dialogue systems bring a lot of convenience to users' lives. The dialogue system is mainly composed of five parts: Automatic Speech Recognition (ASR), Spoken Language Understanding (SLU), Dialogue Management (DM), Dialogue Generation (DG) and Text to Speech (TTS) [1], as shown in figure 1. In order to better understand the user's expression, and then feedback the correct information for the user, spoken language understanding plays an extremely important role. Intent Detection (ID), as a sub-module of spoken language understanding, is also the key to human-machine dialogue system. Traditional spoken language understanding is mainly divided into two sub-tasks -- intent detection and semantic slot filling. Because early research was constrained by application scenarios, data, and computing power, most spoken language understanding was limited to the specific domain.However, with technological innovation and the emergence of the multi-domain dialogue system, the current spoken language understanding is often divided into three tasks -- domain recognition, intent detection and semantic slot filling [2].

In a dialogue system, intent detection is crucial. The intent is the will of the user, that is, what the user wants to do. Intents are sometimes referred to as "Dialog Act" [3], which is the action of information that users share in the dialogue and are constantly updated. The intent is generally named "verb + noun", such as query weather, book a hotel, etc. Intent detection, also known as intent classification, classifies user utterances into previously defined intent categories according to the domains and intents involved in user utterances [4].
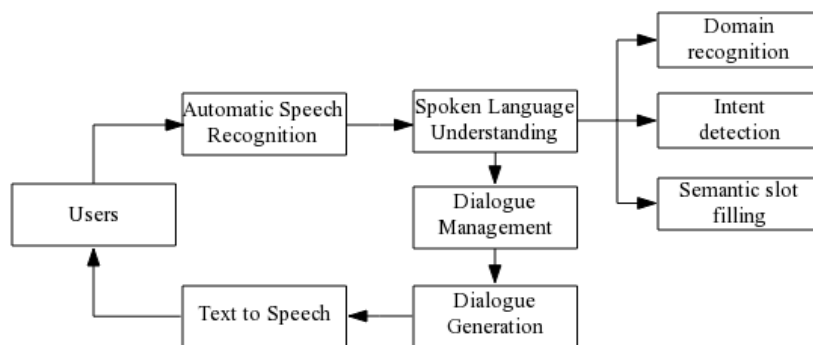
Figure 1. Schematic diagram of the dialogue system.

Nowadays, in the application process of the human-machine dialogue system, users may have multiple intents in different occasions, which will trigger multiple domains in the human-machine dialogue system, including task-oriented vertical domains and chats, etc. In the dialogue system, only when the user's topic domain is clearly defined, can the specific needs of the user be correctly analyzed, otherwise it will lead to wrong intent detection behind. Figure 2 is an instance diagram of the application of three tasks in spoken language understanding. When a user enters a query, it first needs to clear the user's input text belonging to a topic domain to "train" or "flight", due to the intent category is finer-grained than the topic domain, so we need to determine the user's intent, which is booking the ticket or refunding the ticket or querying time, according to the user's specific semantic information. And the semantic slot filling can also help the user intent judgment.
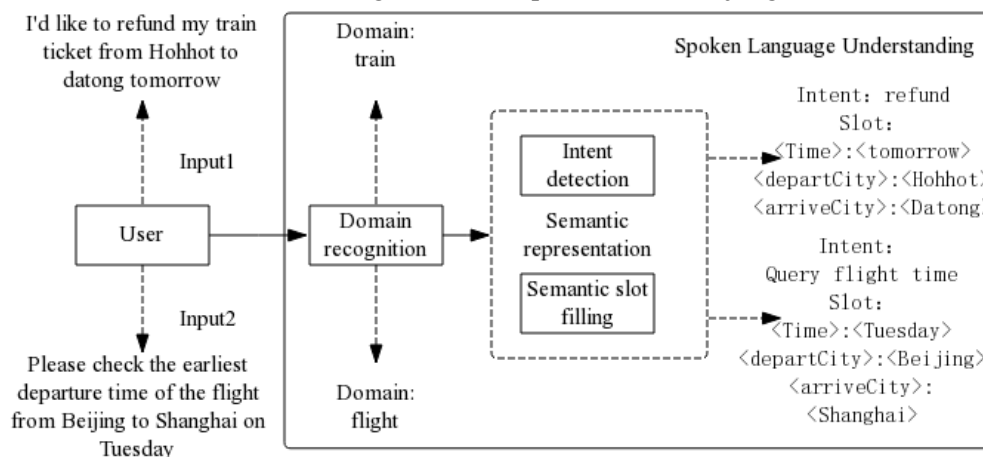

Figure 2. An instance diagram of intent detection.

2. The difficulties ofintent detection

2.1. *Lack of data sources*   annotated data
With the development of artificial intelligence technology, large Internet companies have launched chat robots. Due to the less user experience, it is difficult for most researchers to obtain the chat text between users and robots, which leads to the limited amount of dialogue text to be studied, which has become a major problem faced by intent detection tasks [5]. In the actual process of intent detection, there are very few intent texts with annotations and they are very difficult to obtain, which also brings challenges to the research and development of intent detection [6].

2.2. *The irregularity of user expression*   challenges of NLP : ambiguities, etc.
In the chat system, the user's intent text is generally characterized by colloquial expression, short sentences and broad content, which makes it difficult to identify the user's intent. For example, "I want to look for a dinner place", the corresponding intent of this colloquial daily expression is

"looking for a restaurant", so the colloquial intent text makes the domain topic is not clear, which is not conducive to the identification of user intent. For "Hanting", this semantically poor expression of the intent text, although "Hanting" often appears with "Hotel", it is very difficult for the machine to identify the user's topic domain as "Hotel". For "I want to book a ticket", which may be to book air tickets, train tickets, bus tickets and so on. Due to the user's expression is too broad, the machine cannot give feedback to the user in time.

2.3. Implicit intent detection
With the continuous expansion of the application scope of the human-machine dialogue system, there are more and more ways to express intent. According to the types of expression, intents can be divided into explicit intents and implicit intents [7]. Explicit intents refer to that the user clearly points out his or her intent requirements in text content, including topic domain, intent category and so on. Implicit intent refers to the fact that the user does not have clear intent requirements and it is necessary to infer the user's real intent by analyzing the user's potential intent [8]. Such as explicit intent text "Book a hotel near the People's Park for one night" and implicit intention text "I'm going to Shenzhen for two days next week". Although their intent is booking a hotel, the latter needs to judge the user's potential intent and speculate on the user's true intent. Therefore, implicit intent detection without explicit topic domain and category information is very difficult in the intent detection task.

2.4. Multiple intents detection
Multi-intent detection is similar to multi-label classification, but it is different from multi-label classification. Multi-label classification usually deals with long text, while multi-intent detection mainly deals with short text. How to detect multiple intents of users in short text is another difficulty of intent detection. In the process of multi-intent detection, we need to pay attention to three problems. Firstly, how to find that the user's intent text contains multiple intents; Secondly, how to determine the number of intents contained in the user's intent text; Finally, it is worth thinking about how to accurately identify users' various intents.

3. Main methods of intent detection

3.1. Traditional intent detection methods
In recent years, most scholars regard intent detection as a Semantic Utterance Classification (SUC) problem [9]. Traditional intent detection mainly includes rule-based template semantic recognition method [10] (1993) and classification algorithm based on statistical features [11,12] (2002-2014). Although the rule-based template matching method does not require a lot of training data, it can guarantee the accuracy of detection, but it cannot solve the problem of high cost of template reconstruction caused by changing the intent category. However, the method based on statistical feature classification needs to extract the key features of corpus text. However, the method of manually extracting features is not only costly, but also the accuracy of features cannot be guaranteed, which also leads to data sparse problems. Common methods include Naive Bayes [13] (1998), Adaboost [14] (2000), Support Vector Machine(SVM) [15] (2003) and logistic regression [16] (2007). Considering the non-standard and ambiguous information of user text, the traditional intent detection method cannot accurately understand the deep semantics of user text. How to accurately identify the user's real intent, this is still a very challenging task.

3.2. The current mainstream methods
With the development of deep learning, more and more scholars apply word embedding, Convolution Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) Network, Gated Recurrent Unit (GRU), Attention Mechanism and Capsule Networks to intent detection task. Compared with the traditional machine learning methods, the deep learning model has a great improvement in detection performance.

3.2.1. Intent detection based on word embedding. In recent years, word embedding is gradually used in semantic analysis tasks, due to the use of original lexical features will lead to data sparse problems, and continuous representation learning can solve data sparse problems [17] (2003). Kim et al. [18] (2015) took word embedding as lexical features for intent classification. Compared with the traditional word bag model, the intent classification method based on word embedding has better representational ability and domain extensibility for different classification contents. Considering the insufficiency of semantic information of word embedding, Kim et al. [19] (2016) used semantic vocabulary dictionary information to rich word embedding and improved the semantic representation of intent text. This model achieved good results, and show that rich word embedding will be helpful to improve the performance of intent detection.

3.2.2. Intent detection based on convolution neural network. CNN was originally used for image processing [20]. With the emergence of word embedding technology, CNN has been widely used in the field of natural language processing and achieved good research results. Kim et al. [21] (2014) attempted to use CNN in text classification tasks and achieved very ideal results. Hashemi et al. [22] (2016) used CNN to extract text vector representation as query classification feature to identify the user's intents each query. Compared with traditional artificial feature extraction method, this method not only reduces a lot of feature engineering, but also obtains deeper feature representation. However, CNN has representational limitations.

3.2.3. Intent detection based on recurrent neural networks and their variants. RNN is different from CNN, it represents a word sequence and can learn semantic information of word order according to the context. Bhargava A[23] (2013) reduced the error rate of intent detection by incorporating context information into intent detection tasks, indicating that context information is beneficial to intent detection. A simple RNN has problems such as gradient explosion or gradient disappearance, which cannot well simulate long-term dependence.

LSTM [24] (1997) can solve this problem by introducing a memory unit into the RNN structure, which can control the information to be retained and forgotten. This model is often also used to solve the problem of intent detection. Ravuri et al. [25] (2015) proposed using RNN and LSTM to solve the problem of intent classification. Experiments on Air Travel Information System (ATIS) dataset show that the error rate of intent detection of LSTM is 1.48% lower than that of RNN. The main reason is that LSTM has a good ability to modeling the temporal relationship of text, and has a good memory function for the input of long text.

GRU is an improvement of the LSTM model [26], which has the ability to retain information on long sequences and can learn contextual semantic information. For intent detection task, Ravuri et al. [27] (2016) used GRU and LSTM to comprehensively compare the performance of intent detection on ATIS and Cortana datasets. Experiments show that GRU and LSTM have almost the same performance in the intent detection task, but GRU has fewer parameters and the model is simpler.

3.2.4. Intent detection based on the combination of deep learning models. Considering the advantages and disadvantages of various deep learning models, most researchers combine the deep learning models with different advantages to classify users' intents. Qian et al. [8] (2017) proposed a travel consumption intent detection model based on Convolutional-LSTM, which took advantage of CNN can extract intent text features at a deeper level and LSTM can build the temporal relationship of text, and achieved good performance. Yu et al. [28] (2018) aimed at the problem of data sparse caused by the short text, and proposed a multi-turn dialogue intent detection model based on Biterm Topic Model (BTM) and Bidirectional Gated Recurrent Unit (BGRU). This combined model has achieved good results in the users' medical intent detection, and is superior to the performance of literature [8]. Huang [29] (2018) proposed the Character-CNN-BGRU deep learning combined model. The combined model not only makes use of the character-based method to make the list of words smaller, but also can solve the problem of unknown words, coupled with CNN can extract local features of the

intent text and BGRU can guarantee the temporal relationship of text, highlighting the advantages of the combined model in the intent detection task. However, the structure of the combined model is complex and the training time is long, how to simplify the combined model is a problem worth considering.

3.2.5. Intent detection based on Bidirectional Long short-term Memory (BLSTM) self-attention model. With the development of deep learning models, the expression of various sentence level vectors has emerged, such as: using the maximum pooling or average pooling of CNN to obtain sentence vectors, using the hidden state or the final hidden state of RNN to create sentence representation, etc. Lin et al. [30] (2017) proposed an improved method by introducing a self-attention mechanism to extract sentence representation. Sentence vectors were represented by two-dimensional matrices and different semantic information of sentences was represented by multiple vectors. The model is implemented on BLSTM, and obtains the sentence vector representation by weighted summation of the hidden state of LSTM, and realizes intent classification. This model can obtain various semantic information of sentence through the self-attention mechanism, which is helpful to the research of multi-intent detection.

Figure3 shows the self-attention model, as shown in figure3(a), assume that the input sequence is $S = (w_1, w_2, ..., w_n) \in R^{n \times d}$, input S into BLSTM, and the hidden state of the front and rear terms of the t word is calculated as follows:

$$\overrightarrow{h_t} = \overrightarrow{LSTM}(w_t, \overrightarrow{h_{t-1}}) \tag{1}$$

$$\overleftarrow{h_t} = \overleftarrow{LSTM}(w_t, \overleftarrow{h_{t+1}}) \tag{2}$$

If the number of LSTM hidden cells is k, and the hidden state of the preceding and trailing items is connected to get h_t, and $h_t \in R^{2k}$, $H \in R^{n \times 2k}$ represents the set of all hidden states h_t, and $H = (h_1, h_2, ..., h_n)$.

The self-attention weight matrix is expressed as:

$$a = soft\max(W_{s2} \tanh(W_{s1} H^T)) \tag{3}$$

As shown in figure3(b), $W_{s1} \in R^{m \times 2k}$ and $W_{s2} \in R^{1 \times m}$ are the self-attention weight matrix. m is the number of hidden cells in self-attention model, which is a super parameter and can be set arbitrarily. $H^T \in R^{2k \times n}$ is the transpose matrix of H. $a \in R^{1 \times n}$ can be obtained by calculation, and finally normalization by softmax function. Each dimension in a represents the attention of the corresponding words in the sentence. So the intent text vector can be expressed as: $d = a \cdot H$, then $d \in R^{1 \times 2k}$.

If we want to extract r semantic features from the intent text, we need r self-attention headers to extract semantic features. Then $W_{s2} \in R^{r \times m}$, gets self-attention weight matrix $A \in R^{r \times n}$, and the final intent text vector can be expressed as : $D = A \cdot H$, then $D \in R^{r \times 2k}$.

3.2.6. Intent detection based on the capsule network model. The concept of "capsule" was first proposed by Hinton et al. [31] (2011) to solve the representation limitations of CNN. A capsule contains the vector representation of a group of neurons, the direction of the vector represents the entity attributes, and the length of the vector represents the probability of the existence of the entity. Sabour et al. [32] (2017) proposed the capsule network, replacing CNN scalar output feature detector with vector output capsule, and replacing max-pooling with protocol routing. Compared with the original CNN, the capsule network will maintain the accurate location information of the entity in the

region. Therefore, Zhao et al. [33] (2018) used capsule network for text classification tasks for the first time, and proposed three dynamic routing strategies to improve the performance of dynamic routing process, and reduced the interference of noise (stop words and words unrelated to categories) capsules. Experiments on six standard datasets show that the capsule network performs well in text classification tasks, and the capsule network also has good performance in multi-label text classification tasks.

In the intent detection task, Xia et al. [34] (2018) proposed an intent capsule model based on capsule network, which utilizes the advantages of capsule model in the text modeling to process text hierarchically. As shown in figure 4, first of all, using the method of literature [30] in the intent text extracted semantic features with the self-attention mechanism, named semantic vector. Since different users express the same intent in different ways, but they contribute more to one intent than others, the appropriate contribution of each semantic is dynamically allocated by using dynamic routing mechanism to aggregate them into a higher level prediction vector, that is, the semantic expression of intent, and classify intents. The model achieves good results in the intent detection task.
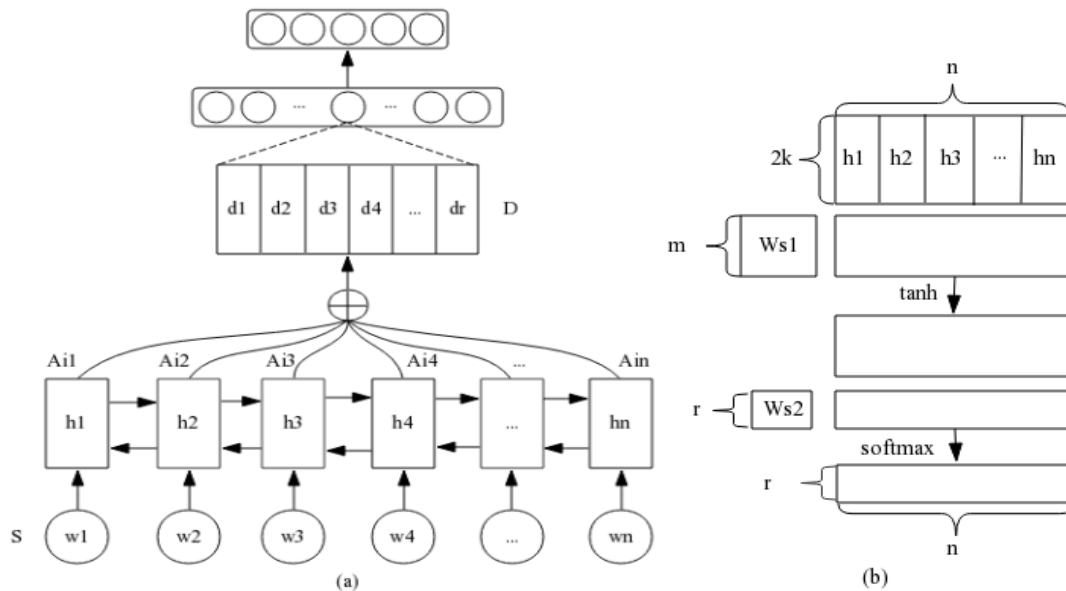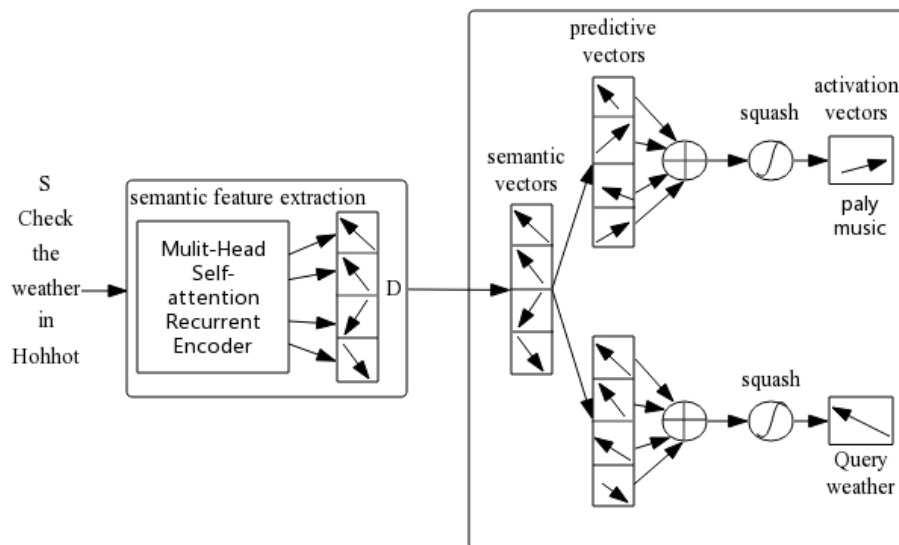

Figure 3. Self-attention model diagram [30].

Figure 4. Intent detection process diagram based on intent capsule.

3.2.7. Intent detection based on the method of joint recognition. With the continuous research and improvement of intent detection methods, considering that single-task research is prone to error propagation due to its independent model, some scholars have proposed a joint model of semantic slot filling and intent detection. Li et al. [35] (2017) conducted a joint model of intent detection and semantic slot filling through triangular chain conditional random fields. Compared with the cascade model that takes semantic slot filling result as intent detection feature, the joint model performs well in the intent detection task, highlighting the relevance between them. Liu et al. [36] (2016) captured important semantic components of sentences by adding attention mechanism on the hidden layer of bidirectional recurrent neural network (BRNN), and improved the accuracy of intent detection. In ATIS dataset, the error rate of intent detection of BRNN based on attention mechanism is 2.35%. Through the joint experiment of semantic slot filling and intent detection, the error rate of intent detection in ATIS dataset is reduced to 1.79%. It can be seen that semantic slot filling is helpful for intent detection.

3.2.8. Multiple intent detection methods. For multi-intent detection tasks, Yang et al. [37](2018) relied on syntax analysis to determine whether the user's intent text contains multiple intents, then determined the number of intent by using the word frequency-inverse document frequency (TF-IDF) and trained word embedding to calculate the matrix distance. Finally, classified the intents by combining syntactic features and CNN, and determined the user's multiple intents. However, there are few studies on multi-intent detection based on deep learning model, so this is a direction worth studying.

4. The evaluation methods of intent detection

4.1. The evaluation method of single intent detection
At present, intent detection is generally regarded as a semantic discourse classification problem, so the performance of intent detection method is evaluated by the one which used in text classifier [28]. That is, accuracy, recall rate, F1-score and error rate, classification speed, etc.

$$(1)\ \text{Accuracy:}\ Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

$$(2)\ \text{Precision:}\ Precision = \frac{TP}{TP+FP} \tag{5}$$

$$(3)\ \text{Recall:}\ Recall = \frac{TP}{TP+FN} \tag{6}$$

$$(4)\ \text{F1-score:}\ F1-score = \frac{2\times Precision\times Recall}{Precision+Recall} \tag{7}$$

Among them, TP indicates True Positive, that is, class A samples were correctly classified into class A. TN indicates True Negative, which do not belong to the class A samples were classified correctly into classes other than class A. FP indicates False positives, which does not belong to the class A samples were classified erroneously into class A. FN indicates False negatives, which belongs to the class A samples were erroneously into the classes other than class A.

4.2. The evaluation method of multiple intent detection
The evaluation method of multi-intent detection is different from that of single intent detection, it needs to judge the types of intent prediction in sentences. This paper adopts evaluation index of literature [37]. Suppose that there are N multiple intent samples recorded as (x_i, A_i), $0 \le i \le N$. A_i denotes the set of correct intent results of the sample set. $B_i = S_{(x_i)}$ denotes the set of predicted intent results of the sample set to be tested. The specific calculation formulas are as follows:

(1) Multi-Intent Accuracy(MIA): $MIA = \dfrac{1}{N} \sum\limits_{i=1}^{N} \dfrac{|A_i \cap B_i|}{|A_i \cup B_i|}$ (8)

(2) Multi-Intent Precision(MIP): $MIP = \dfrac{1}{N} \sum\limits_{i=1}^{N} \dfrac{|A_i \cap B_i|}{|B_i|}$ (9)

(3) Multi-Intent Recall(MIR): $MIR = \dfrac{1}{N} \sum\limits_{i=1}^{N} \dfrac{|A_i \cap B_i|}{|A_i|}$ (10)

5. Comparison of experimental performance results

Literature [34] has experimented with various deep learning models mentioned above on SNIPS English dataset and Commercial Voice Assistant (CVA) Chinese dataset, and the performance results of intent detection were shown in table 1:

Table 1. Performance comparison results of various deep learning models on intent detection tasks.

Model	SNIPS (Five kinds of intents)		CVA (80 kinds of intents)	
	Accuracy	F1	Accuracy	F1
CNN	0.9595	0.9595	0.8223	0.8210
RNN	0.9516	0.9518	0.8286	0.8275
GRU	0.9535	0.9534	0.8239	0.8216
LSTM	0.9569	0.9569	0.8319	0.8306
BLSTM	0.9501	0.9502	0.8428	0.8419
Self-attention BLSTM	0.9524	0.9522	0.8521	0.8513
Intent Capsnet	0.9621	0.9620	0.9088	0.9023

According to the results in the table, BLSTM can make full use of context to represent sentence features, which is better than the local feature representation of CNN, and the self-attention model can more fully capture the deep semantic information of sentences. Due to intent capsule can not only capture the deep semantics of intent text, but also ensure the exact location of semantic information of intent text, its intent detection performance is better than other deep neural network models.

6. Summary and prospect

This paper mainly introduces the difficulty and method of intent detection in the human-machine dialogue system. It summarizes and compares the intent detection methods of the deep learning model. Traditional intent detection methods can't understand user's intent in depth, while the deep learning model shows its advantages. The capsule network model achieves good performance in the intent detection task, and also has a good effect on multi-label classification. The self-attention model can extract various semantic features of sentences in the process of intent detection, thus contributing to the research of multi-intent detection.

At present, intent detection is not only applied in various fields such as e-commerce, travel consumption, medical treatment and chat, but also applied to network intrusion, network fraud and air target combat fields, to provide the guarantee for network security problems. Traditional dialogue systems are mainly oriented at single intent detection in specific fields. With the increasingly frequent interaction between humans and machines, the users' discourse expression is not limited to only one intent. How to accurately identify multiple intents of users will be our next research work.

References

[1]   Hongshen Chen, Xiaorui Liu, Dawei Yin, et al. A survey on dialogue systems: recent advances and new frontiers[J]. *Acm Sigkdd Explorations Newsletter*, 2017, 19(2):25-35.

[2]   Tur G. Spoken Language Understanding: Systems for extracting semantic information from speech[D]. NewYork, NY: John Wiley and Sons, 2011.

[3]   Austin J A. *How to do things with words*[M]// Harvard University Press. Cambridge: 1962.

[4]   Celikyilmaz A, Hakkani-Tur D, Tur G, et al. Exploiting distance based similarity in topic models for user intent detection[C]// *Automatic Speech Recognition & Understanding*. IEEE, 2011:425-430.

[5]   Bingfeng Luo, Yansong Feng, Zheng Wang, et al. 2018.Marrying up regular expressions with neural networks:A case study for spoken language understanding. arXiv preprint arXiv:1805.05588.

[6]   Yanling Li, Yonghong Yan. Weakly-supervised training method about Chinese spoken language understanging[J]. *Application of Computers*, 2015, 35(7):1965-1968.

[7]   Zhiyuan Chen, Bing Liu, Hsu Meichun, et al. Identifying intent posts in discussion forums[C]// *Proc. of Conference of the North American Chapter of the Association for ComputationalLinguistics-HumanLanguageTechnologies.*Atlanta,2013:1041–1050.

[8]   Yue Qian. Research on the identification method of users' travel consumption intent in chat robot[D]. Harbin: Harbin Institute of Technology, 2017.

[9]   Dauphin Y N, Tur G, Hakkani-Tur D, et al. "Zero-shot learning for semantic utterance classification," arXiv preprint arXiv:1401.0509, 2013.

[10]   Appelt D, Bear J, Cherny L, et al. GEMINI: A natural language system for spoken-language understanding[C]// *Proc. Meeting of the Association for Computational Linguistics*. 1993:54-61.

[11]   Yan Pengju. Research on natural language understanding in conversational systems[D]. Beijing: Tsinghua University, 2002.

[12]   Ahmad A S, Hassan M Y, Abdullah M P, et al. A review on applications of ANN and SVM for building electrical energy consumption forecasting[J]. *Renewable & Sustainable Energy Reviews,* 2014, 33(2):102 -109.

[13]   Andrew McCallum, Kamal Nigam, et al.A comparison of event models for naive bayes text classification[C]// *In AAAI-98 workshop on learning for text categorization*. 1998:41–48.

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

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

[16]   Genkin A, Lewis D D, Madigan D. Large-Scale Bayesian Logistic Regression for Text Categorization[J]. *Technometrics*, 2007, 49(3):291-304.

[17]   Bengio Y, Ducharme R, Vincent P, et al. A neural probabilistic language model[J].*Journal of Machine Learning Research*, 2003, (3): 1137-1155.

[18]   KimD, LeeY, ZhangJ, et al. Lexical feature embedding for classifying dialogue acts on Korean conversations[C]//*Proc. of 42th Winter Conference on Korean Institute of Information Scientists and Engineers,* 2015: 575–577.

[19]   Kim JK, Tur G, Celikyilmaz A, et al. Intent detection using semantically enriched word embeddings[C]// *Spoken Language Technology Workshop*. IEEE, 2016:414-419.

[20]   Lecun Y L, Bottou L, Bengio Y, et al. Gradient-Based Learning Applied to Document Recognition[J]. *Proceedings of the IEEE*, 1998, 86(11):2278-2324.

[21]   KimY.Convolutional Neural Networks for Sentence Classification[C]// *Proc. of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014:1746–1751.

[22]   HashemiHB., AsiaeeA, KraftR, Query intent detection using convolutional neural networks[C]// *International Conference on Web Search and Data Mining, Workshop on Query Understanding* , 2016.

[23]   Bhargava A, Celikyilmaz A, Hakkanitur D, et al. Easy Contextual Intent Prediction and Slot Detection[C]// *IEEE International Conference on Acoustics*. IEEE, 2013: 8337-8341.

[24]   Hochreiter S, Schmidhuber J. Long Short-Term Memory[J]. *Neural Computation*, 1997, 9(8):1735-1780.

[25]   Ravuri S V, Stolcke A. Recurrent neural network and LSTM models for lexical utterance classification[C]// *16th Annual Conference of the International Speech Communication Association*. 2015:135-139.

[26]   Dey R, Salemt F M. Gate-Variants of Gated Recurrent Unit (GRU) Neural Networks[C]//*IEEE 60th International Midwest Symposium on Circuits and Systems*.IEEE, 2017:1597-1600.

[27]   Ravuri S, Stolcke A. A comparative study of recurrent neural network models for lexical domain classification[C]// *Proc. of the 41th IEEE International Conference on Acoustics, Speech, and Signal Processing*. IEEE, 2016: 6075-6079.

[28]   Hui Yu, Xupeng Feng, Lijun Liu, et al. Identification method of user's medical intent in chatting robot[J].*Journal of Computer Applications*, 2018, 38(8): 2170-2174.

[29]   Jiawei Huang. Research on the classification method of user intent in the human-machine dialogue system[D]. Wuhan: Central China Normal University, 2018.

[30]   Zhouhan Lin, Minwei Feng, Santos CN D, et al. A structured self-attentive sentence embedding. arXiv preprint arXiv:1703.03130, 2017.

[31]   Hinton G E, Krizhevsky A, Wang S D. Transforming Auto-Encoders[C]//*In International Conference on Artificial Neural Networks*, 2011:44–51.

[32]   Sabour S, Frosst N, Hinton G E. Dynamic routing between capsules[J]. *In Advances in Neural Information Processing Systems*, 2017:3859–3869.

[33]   Wei Zhao, Jianbo Ye, Min Yang, et al. Investigating capsule networks with dynamic routing for text classification.Computing Research Repository, arXiv:1804.00857. Version 1, 2018.

[34]   Congying Xia, Chenwei Zhang, Chenwei Yan, et al. Zero-shot User Intent Detection via Capsule Neural Networks[C]//*Proc. of the 2018 Conference on Empirical Methods in Natural Language Processing*. 2018:3090-3099.

[35]   Yanling Li, Yonghong Yan. Research on execution strategy about statistical spoken language understanding[J]. *Computer Science and Technology*, 2017, 11(6):980-987.

[36]   Liu B, Lane I. Attention-Based Recurrent Neural Network Models for Joint Intent Detection and Slot Filling[C]//*17th Annual Conference of the International Speech Communication Association*. 2016:685-689.

[37]   Chunni Yang, Chaosheng Feng. Multi-intent recognition model with combination of syntactic feature and convolution neural network[J]. *Journal of Computer Applications*, 2018, 38(7): 1839-1845.