Big Data Mining Analysis Technology For Natural Language Processing Robot Design

Yongqiang Wang[1*], Li Yang[1], and Zhixin Lun[2]

[1] College of Artificial Intelligence, Tangshan University, Tangshan, 063000, China
[2] Computing Center, Tangshan University, Tangshan, 063000, China
[*] Corresponding author. E-mail: wang.yqyq@yandex.com

Received: June 20, 2023; Accepted: Nov. 02, 2023

In the context of information big data, breakthroughs have been made in artificial intelligence, natural machine language and other technologies, and robot interactive dialogue has become a reality. Seq2seq is a common natural machine language processing technology, but the traditional Seq2seq dialogue model faces the problem of lack of semantic information and long-distance dependence. Therefore, the traditional Seq2seq technology is studied, and BiLSTM and attention mechanism are used to optimize the Seq2seq dialogue model. The simulation experiment test shows that in the iterative loss performance test of the dialogue model with an increased attention mechanism, the overall curve of the BiLSTM model has a gentle trend, and the loss is lower than that of the LSTM model. At 100 iterations, the loss value of the LSTM model is 0.36 and the loss value of the BiLSTM model is 0.17 . In the music scene dialogue test, the LSTM model could not accurately understand the meaning of the dialogue, and the satisfaction rate was 45%. The BiLSTM model accurately recognized the meaning of the dialogue and responded correctly, with a satisfaction rating of 69%. The innovation of research content adopts BiLSTM and attention mechanism to optimize the traditional Seq2seq dialogue model, improve the language analysis ability of robots, and provide important reference significance for the development of robot intelligence.

Keywords: Data mining; Machine language; BiLSTM model; Seq2seq model
© The Author('s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are cited.
http://dx.doi.org/10.6180/jase.202412_27(12).0008

1. Introduction

Natural language science is an important research direction in the field of modern artificial intelligence. Relying on big data, information science, probability, and other disciplines, it has promoted the transformation of intelligent robot language processing technology. At present, major technology companies are investing heavily in artificial intelligence technology, the core of which lies in the voice interaction between robots and humans. Major Internet companies have launched their own robot voice systems, such as Microsoft's Xiaobing, Apple's Siri, Baidu's Xiaodu, etc., all of which are intelligent dialogue robots. However, the current level of intelligence of dialogue technology is not high, mainly in the problems of low language recognition rate, lack of semantic information, and low dialogue intelligence. In the future, chatbots will still be an important battlefield for the development of major technology companies. There are four main types of mainstream intelligent chatbot technologies, including retrieval robots, artificial template robots, translation robots, and deep learning robots. The first three are less intelligent, mainly rely on the set dialogue template or language database, and lack effective dialogue scenes. Deep learning technology simulates human dialogue scenes in the form of bionics and could analyze language and characters, learn abilities, and solve problems. The advantage of deep learning robots is that they do not limit dialogue scenarios and have strong language scalability. The disadvantage is that they cannot be separated from supervised learning and rely on many corpus templates in the early stage. In the future, deep learning needs to be improved on corpus databases and deep machine learning. The upgrade of software and hardware will improve the limitations of current deep language technology. Therefore, based on deep learning technology, the problems faced by chatbots are analyzed, and the improved Seq2seq model is used for experimental tests to improve the dialogue performance of current chatbots. Adopting intelligent machine learning technology to improve dialogue models provides important reference opinions for the intelligent development of dialogue robots.

2. Related work

Natural language technology is the key direction of artificial intelligence innovation and development and has been applied in many industries. Scholars at home and abroad have done a lot of data research on this. Russo A et al. studied the characteristics of the robot dialogue system to achieve the effect of communication and interaction between the machine and the patient. It starts with a discussion of potential techniques, methods, and related issues, and provides a spoken dialogue model. The analysis of systems and methods is concentrated. According to qualitative experiments, 12 items were determined to meet the inclusion criteria. This study assessed the mental status of patients with dementia [1]. Hedges and Sadrzadeh [2] analyzed the distributional semantic model of categorical composition and presented abstract setting instances as the representations of sets and spatial mappings. The equivalence of relation instantiation and truth-theoretic semantics of generalized quantifiers is proved by experiments. A survey by Parr and Gobbel [3] found that much medical data are recorded in free text, not obtained through structured data. On the other hand, unstructured free-text information can be used based on natural language processing. It is confirmed by experimental design that natural language has a prominent effect in dealing with nephrology. Li and Mao [4] found that natural language faces a huge problem in the process of causality, and the process of using the rule method to deal with this problem is huge. Considering the complexity of its causal relationship expression, the convolutional network is used to complete the collection of data relationships, and the pre-knowledge is fused to obtain the causal relationship. The results confirmed that the scheme has a reliable causal extraction effect. Scholars such as Plab found that the non-parametric attention machine strategy used in natural language reasoning lacks effective supervision. Therefore, a language model with supervised attention is adopted to deal with the current problem. Relevant syntactic labeling of attention modules will be carried out within this framework while ensuring the unification of the training part of internal supervision. And use multiple datasets to verify the method, the results show that the adopted supervised attention scheme is efficient and worthy of promotion [5]. Sorin discovered that natural language technology can represent free text as structured data. And the latest deep technology can improve its learning effect. Therefore, the relevant radiology content over the years was studied, and the effect of the deep learning model in the natural language processing process was verified through the relevant database. The verification results show that the natural language technology based on deep learning technology has important application value in the field of radiology [6]. Ötleş et al. found that high-quality feedback can improve the learning effect. It is not clear whether large-scale feedback quality can be achieved using natural language technology. Therefore, a variety of natural language technologies were evaluated, and 5 postgraduate-related surgical feedback was collected. Training and evaluating the natural language model, the results show that the average maximum accuracy of SVM technology is 0.64 , and the maximum average accuracy is 0.83 when modified to distinguish high-quality feedback from low-quality feedback, indicating the feasibility of natural language technology [7].

The innovation of natural language technology has promoted the development of related technologies of intelligent dialogue robots. Zhu et al. found that the current computer has a limited ability to generate humor, and the corresponding humor language processing is still at the primary level. Humor computing ability is widely used in the field of machine dialogue and has important value in the treatment of patients with mental diseases. Therefore, a humor mechanism is proposed, and relevant humor resources are obtained with the help of big data to deal with the lack of humorous scenes in robot dialogue. The application results show that the scheme is practical and can make up for the defects of the existing machine dialogue [8]. Sono and other teams found that intelligent robots have been used in many scenarios and can communicate with people. However, the overall communication effect is not ideal, so the Google Cloud Vision library is used to obtain the word list and generate the corresponding topic vector, thus effectively making up for the lack of robot dialogue. The experimental results show that this scheme has a better dialogue effect than the machine dialogue which randomly selects utterances [9]. Experts such as Heath found that robots can accomplish simple human interaction effects.

However, due to technical factors, there are differences in the effect of interaction with people. Therefore, the social robot is optimized and has a dialogue window to realize the survey of satisfaction. Interact through volunteer games to answer the robot's questions. Experiments show that the participants can effectively reply to the relevant questions, but the robot cannot deal with the relevant questions [10]. Chiou et al. found that existing robots have inherent advantages in some special fields, especially in high-risk rescue and character search. However, existing robots still have major problems in communication and interaction with humans and cannot effectively respond to some complex scenarios. Therefore, a virtual robot task experiment was carried out to optimize the human interaction effect of the robot through experiments. The results show that the effective interpretation of the information pushed by the robot will greatly help the application of the robot, reduce unnecessary time consumption in the task process, and improve the task effect [11].

It can be seen from the above research data that natural language technology has a wide range of applications and effectively promotes the development of artificial intelligence technology. Chatbots that combine big data and natural language technologies have made breakthroughs in recent years. Therefore, combining the big data technology and the current mainstream Seq2seq technology to conduct research, improve the recognition performance of the existing dialogue robots, and improve the practicability of dialogue robots.

3. Building a robot dialogue model for natural language processing

3.1. Construction of Robot Dialogue Generation Model

Robot dialogue is an extension of modern artificial intelligence technology. With the application of big data mining in machine language processing, more intelligent robot voice interaction has entered people's daily life. The robot dialogue system based on big data technology includes a speech recognition section, natural language processing section, dialogue management section, etc. Fig. 1 is the dialogue robot language processing framework [12].

In the robot dialogue system, the language information data is collected through the speech recognition section, and the language expression information is identified through natural language processing (NLP) technology and converted into specific semantic data, and the dialogue management section is used to process and generate natural language. The section realizes the function of dialogue interaction [13]. Among them, the natural language generation part is the key to the robot dialogue system, and its main function is to generate the reply language. There are two types of existing dialogue generation, which are based on retrieval to generate preset responses, which mainly rely on a huge language database and avoid the problem of grammatical errors [14]. The other is based on big data deep learning technology, which is more intelligent. The latter is used here, and a robot dialogue system is built based on deep dialogue (Sequence to sequence, seq2seq) technology. seq2seq dialogue generation adopts Long Short-Term Memory (LSTM) network, which has an excellent ability to deal with time series. The schematic diagram of the structure of the LSTM network is shown in Fig. 2 [15].

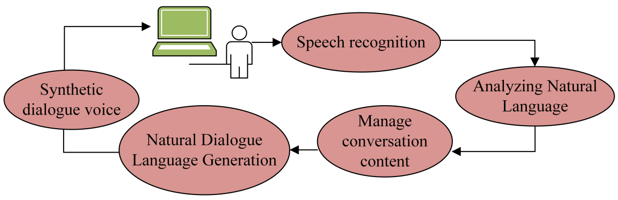


Fig. 1. A Framework for Language Processing of Dialogue Robots

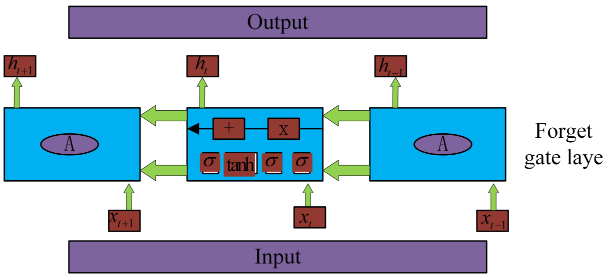


Fig. 2. The schematic diagram of the structure of the LSTM network

In LSTM, the LSTM network is divided into three parts: output, input and forget gate, and the choice of information is realized through three parts [16]. The forgetting gate is mainly responsible for the control of the forgetting degree of the unit information. The output information of the previous unit is defined as h_{t-1}, the input of this unit is, x_{t-1} and all elements in the output matrix are mapped to the interval $[0, 1]$ by the activation function processing c_{t-1}, then the output expression of the forgetting gate is as Eq. (1) seen.

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{1}$$

In Eq. (1), σ represents the activation function, b_f represents the forgetting gate bias term, and W_f represents the forgetting gate weight matrix. In the output gate, tanh the filtering of the added information will be realized through the function

[17]. Among them, the information data tanh will generate a new vector after passing through the function $\tilde{c}_t$, and the output gate will be mainly responsible for mapping the vector information to the $[0, 1]$ interval. After such a screening operation, the in and out of the information can be controlled. The fixed input gate input is i_t, its mathematical expression is as shown in Eq. (2).

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \tag{2}$$

In Eq. (2), it b_i represents the input gate bias term and W_i represents the input gate weight matrix. The expression of the vector generated by the function is $\tilde{c}_t$ as shown in Eq. (3).

$$\tilde{c}_t = \tanh\left(W_c \cdot [h_{t-1}, x_t] + b_c\right) \tag{3}$$

In Eq. (3), it b_c represents the bias term of W_i this process, which represents the weight matrix of this process, W_c represents input and unit layer weights. Then the expression of the input gate is as shown in Eq. (4).

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \tag{4}$$

The output gate is mainly responsible for screening the current neural unit information, and its structure is shown in Fig. 3.

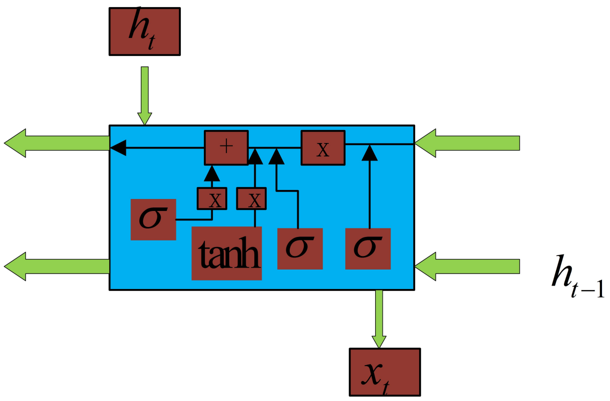


Fig. 3. Schematic diagram of output door

The output gate maps the information of each element of the matrix to the $[0, 1]_{\text{interval}}$, and finally outputs the final language information required by filtering. The output expression of the output gate is shown in Eq. (5).

$$o_t = \sigma\left(W_o \cdot [h_{t-1}, x_t] + b_o\right) \tag{5}$$

In Eq. (5), it b_c represents the output gate bias term and W_i represents the output gate weight matrix, ^{W_o} represents the weight of the output layer. Then the new vector expression of the output gate is as shown in Eq. (6).

$$h_t = o_t^* \tan\left(c_t\right) \tag{6}$$

In the LSTM model, if the robot wants to form a sentence expression, it usually adopts a simple addition method, that is, adding multiple words. It can also take the form of the average equal method, but the latter cannot take into account the sequential expression of sentences. For example, if the sentence is expressed as "I don't like him", the emotion expressed is derogatory. The use of LSTM can well describe the relationship between long distances. Dependency, taking LSTM can learn and memorize the corresponding information. However, simply using LSTM still faces problems in sentence construction, and it is impossible to encode the information from front to back. Especially in more delicate classification scenes, some words and sentences expressing praise, derogation or neutrality cannot effectively express the emotional intensity of sentences [18]. Then, the forward LSTM and the backward LSTM can be combined to build a long and short-term memory model (Bi-directional Long Short-Term Memory, BiLSTM). For example, the former LSTM inputs three vectors as (h_{r1}, h_{r2}, h_{r3}), and the latter LSTM inputs as (h_{l1}, h_{l2}, h_{l3}). Combining the front LSTM with the post LSTM, and splicing the front and back hidden vectors, we get $[(h_{r1}, h_{l3}), (h_{r2}, h_{l2}), (h_{r3}, h_{l3})]$, ie (h_1, h_2, h_3). BiLSTM is used to better capture semantic dependencies and improve the accuracy of robot language expression. The model diagram of the BiLSTM model is shown in Fig. 4.

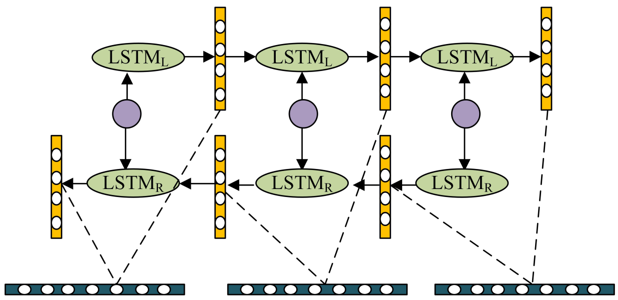


Fig. 4. Model diagram of BiLSTM model

3.2. Construction of Robot Dialogue Generation Model

The seq2seq technology is used to build a conversational chatbot, and the translation model is based on seq2seq to realize the translation of language sequences. Map one sequence to another output sequence through BiSTLM to capture bidirectional semantic dependencies. The schematic diagram of the seq2seq model is shown in Fig. 5 [19].

The seq2seq model is mainly composed of two parts: Decoder and Encoder. It can be seen from the model structure in Figure 5 that the seq2seq model has a large degree

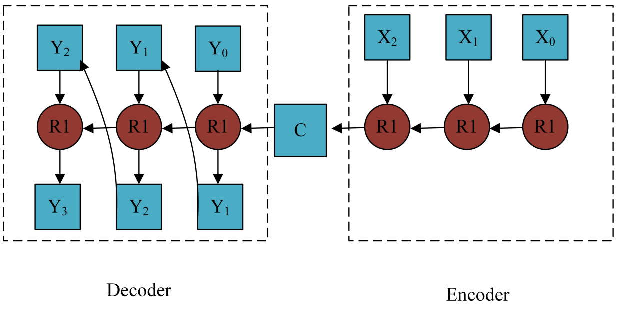


Fig. 5. Schematic diagram of seq2seq model principle

of freedom and can be combined with a variety of neural networks. BiSTLM is mainly used here. The current state input information and output information X are calculated to calculate, and the expression is as shown in Eq. (7) [20].

$$h_t = f(h_{t-1}, x_t) \tag{7}$$

In Eq. (7), it h_{t-1} represents the output information of the previous unit, and x_t represents the input information of the unit, which is the input $f(\cdot)$ of the forget gate. The encoding process is mainly to convert the sequence into a fixed-length vector through the Encoder, and obtain the semantic vector through nonlinear transformation as shown in Eq. (8) [21].

$$C = q(h_1, h_2, h_3, \ldots, h_t) \tag{8}$$

In Eq. (8), h_t is the output of the hidden layer, ^{t} the number of words in the sequence, and the q parameter value. During the decoding process, the Decoder gets the output series $y_1, \ldots, y_{t-1}$ and predicts the next output word, as shown in Eq. (9).

$$y_t = g(\{y_1, y_2, \ldots, y_{t-1}\}, C) \tag{9}$$

In Eq. (9), it $g(\cdot)$ represents the nonlinear activation function. During the decoding process, the end character and length of the text will be specified in the Decoder, and the encoding will be terminated when both requirements are met. However, in the traditional seq2seq model, fixed-length vectors C cannot effectively represent all sequence information, and text front-end information is easily covered by back-end information. In order to solve this problem, the attention mechanism Attention is introduced, and the input data is carefully allocated to improve the training effect of seq2seq [22]. The schematic diagram of the attention mechanism based on seq2seq is shown in Figure 6. The decoder X obtains the word prediction probability according to the input as shown in Eq. (10).

$$p(yi \mid y_1, .., y_{t-1}, x) = g(y_{i-1}, s_i, c_i) \tag{10}$$

In Eq. (10), p^p is the prediction probability, s_i which refers to the network ^{i} moment state, and c_i is a i moment vector, then the s_i expression is as shown in Eq. (11) [23].

$$s_i = f(s_{i-1}, y_{i-1}, c_i) \tag{11}$$

To obtain better results, each output h_i will introduce different weights, as shown in Eq. (12).

$$c_i = \sum_{j=1}^{T_x} a_{ij} h_j \tag{12}$$

In Eq. (12), T_x is the matrix parameter, which h_i represents the hidden layer, and a_{ij} the calculation is shown in Eq. (13) [24].

$$a_{ij} = \frac{\exp\left(e_{ij}\right)}{\sum_{k=1}^{T_x} \exp(e_{ik})} \tag{13}$$

In Eq. (13), the e_{ij} expression is as seen in Eq. (14).

$$e_{ij} = V^T \tanh\left(W_{s s_{i-1}} + W_h h_t\right) \tag{14}$$

In Eq. (14), it V represents the network weight, W_s represents the offset vector, and W_h represents the parameters of the embedding layer.

$$e_{ij} = V^T \tanh\left(W_{s s_{i-1}} + W_h h_t\right) \tag{15}$$

The traditional seq2seq selects the maximum probability as the output value, but some words in the corpus have a high probability and are often defined as safe responses. To avoid the problem of safe reply output, use beam search, use pruning operation, and multi-layer space search to improve model performance. The pruning process will filter low-probability nodes in each search, reduce the number of searches to enhance the model search efficiency, and then build a dialogue robot model. The overall process of the robot dialogue model is shown in Fig. 6 [25].

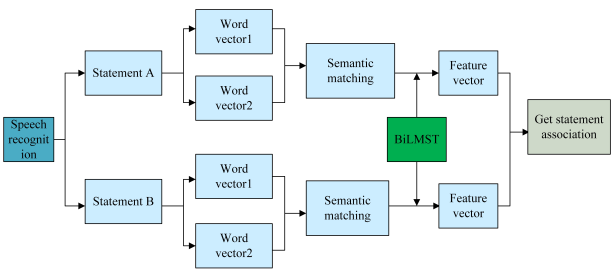


Fig. 6. Overall process of the robot dialogue model

4. Chatbot dialogue simulation test

In the experimental test of the dialogue robot, TensorFlow developed by Google will be selected as the model training and learning platform. The test system is Win7 platform, the running memory is 16G, the processor is 16 -core 3.5Ghz, and the graphics card used in the test is NVIDIA GTX2070. To test the performance of the constructed chatbot model, a large amount of corpus test data is needed in the experimental test. Considering that the robot needs to be more colloquial to simulate human dialogue scenes, choosing movie dialogue scenes as the experimental test data meets the requirements of robot learning and training. Considering that there is a large amount of dialogue data from movies and TV series abroad, the open subtitles corpus is selected as the robot text. Converted into Chinese dialogue corpus through translation tools. The corpus is divided into topics, including music, games, war, news, and more. The size of the test corpus is shown in Table 1.

In the experimental test, it is necessary to verify the language generation effect of LSTM and BiLSTM in the dialogue robot, and the performance effect of the two models based on seq2seq technology will be compared. The number of hidden layer nodes is set to 512 , the word vector dimension is set to 64 , and the batch size is set to 128 . The iterative loss results of the robot dialogue model are shown in Fig. 7.

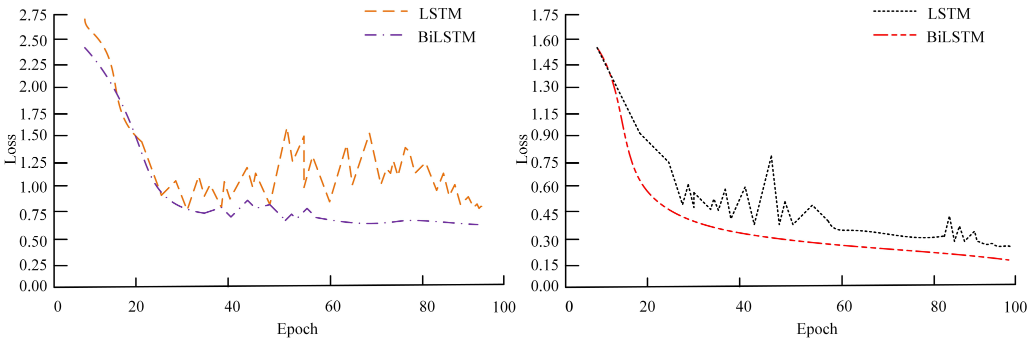
Fig. 7 is the iterative loss result of the robot dialogue model, and Fig. 7(a) is the test result of the dialogue robot model without adding the attention mechanism. As can be seen from the data in the figure, without adding the attention mechanism, as the number of Epoch training iterations increases, the value of the Loss function of the two models shows a downward trend. The yellow curve represents the trend of the LSTM model, and the purple curve represents the trend of the BiLSTM curve. Overall, both models have fluctuations, but the BiLSTM fluctuation is small, and the loss is smaller. Fig. 7(b) is the test result of the dialogue robot model with the added attention mechanism. Compared with Fig. 7(a) without the attention mechanism, the loss of both models is smaller. The black curve represents LSTM, and the red curve represents BiLSTM. It can be seen from the image data that after adding the attention mechanism, the overall curve of the BiLSTM model changes smoothly, and the loss is lower than that of the LSTM model. At 100 iterations, the loss value of the LSTM model is 0.36 and the loss value of the BiLSTM model is 0.17 . Adding an attention mechanism to the model can effectively improve its performance of the model and reduce the loss of model performance. At the same time, in corpus training, BiLSTM has better performance and lower loss than the LSTM model. To further verify the performance of the BiLSTM model in the robot dialogue, the accuracy of the model will be tested in two scenarios of music and games. The test results are shown in Fig. 8.

Figure 8 shows the results of the dialogue test accuracy of the robot dialogue model. Taking the music dialogue scene and the game dialogue scene as the test background, the corpus accuracy of the two models was tested under 100 loop iterations. Fig. 8(a) is the test result of the music corpus scene. It can be seen from the fluctuation of the curve in the figure that with the gradual increase in the number of model test iterations, the accuracy of the two dialogue models gradually increases. The LSTM model tends to converge at 52 iterations, and the model semantic test accuracy at this moment is 0.81 ., while the BiLSTM model tends to converge at 44 iterations, and the model semantic test accuracy at this moment is 0.93 . The BiLSTM model has better convergence performance and higher test accuracy. Fig. 8(b) is the test result of the game corpus scene. Compared with the music corpus scene, the game scene is more complex and more professional. Compared with the music corpus scene, the dialogue accuracy of the two models in the game scene has decreased, but the BiLSTM model has less impact, and tends to converge after 44 iterations. At this moment, the corpus test accuracy rate is 0.89 . In the game scene, it has a great influence on the LSTM model, which tends to converge after 61 iterations, and the accuracy rate of the corpus test at this moment is 0.58. In the conventional dialogue environment, both dialogue models have higher dialogue accuracy, but in complex and professional dialogue scenarios, the BiLSTM model has a higher overall dialogue accuracy. Compared with the LSTM model, the BiLSTM model has a higher dialogue accuracy. The dialogue accuracy rate has increased by 46.65%. As shown in Table 2, the experimental results of the two dialogue models are shown.

Table 2 shows the actual dialogue test results of the two models and evaluates the dialogue effects of the two dialogue models through satisfaction. In dialogue sample 1, the music dialogue scene is selected, the LSTM model cannot accurately understand the meaning of the dialogue, and the satisfaction rate is 45%, which is unqualified. In the same scene, the BiLSTM model accurately identified the meaning of the dialogue and replied, "I especially like listening to music", with a satisfaction rating of 69%. In a simple dialogue scenario, such as dialogue sample 6 asking "Is today the weekend?", both models can reply, but the BiLSTM model's dialogue reply is more accurate, and specific to a certain day, and the satisfaction rate is 92%, 69% higher than the LSTM model. In actual dialogue scenarios,

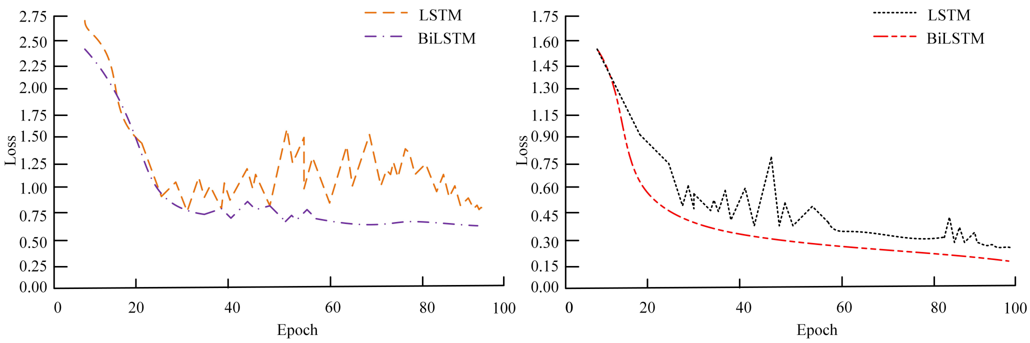
Table 1. Scale of test corpus

Corpus type	Open subtitles corpus data quantity
Developing corpus	2100
Test corpus	2100
Training ji corpus	44063055
Dataset corpus	44063054



(a) Attention mechanism sample data test is not added (b) Sample data test under attention set

Fig. 7. Iterative loss results of robot conversation model



(a) Attention mechanism sample data test is not added (b) Sample data test under attention set

Fig. 8. Results of dialog test accuracy of robot dialog model

there are differences in the responses of dialogue models in different scenarios, and both models can respond simply by replying. However, in complex scenarios, the BiLSTM model has more accurate and standard answers than the LSTM model. To effectively improve the dialogue ability of the model and avoid the appearance of model security response, the cluster search parameters are set to 9 in the experimental experiment, and the experimental dialogue experiment test is carried out, as shown in Table 3.

Table 3 is the test results of the experimental dialogue part under the BiLSTM model. During the experiment, to avoid the generation of safe replies, the cluster search parameters are set to 9, so that the robot will search for more information when answering questions and avoid safe replies from appearing. From the reply to information in Table 3, in the case of not setting the cluster search, among the 9 replies given by the BiLSTM model, the security reply set by the system appears at most 4 times, which is "I'm sorry", and the security reply "I don't know". appeared twice. After the cluster search is set, the data sources that can be obtained by the model are more abundant, the repetition rate of safe responses is significantly reduced, and the language responses generated are more accurate and richer. Finally, the emotion recognition rates of different models are tested, and the results are shown in Table 4.

Table 4 shows the recognition rate of the model in different emotional scenarios. From the data in Table 4, under the separate LSTM and BiLSTM models, the accuracy of emotional attitude recognition in the dialogue of the two models is not high, and the recognition rate of the LSTM model in multiple emotional scenes is less than 60%, which is an unqualified realization. Compared with the LSTM model, the emotion recognition rate of the BiLSTMM model is bet-

Table 2. Shows the experimental results of two dialogue models

Interactive object	LSTM model reply result	Reply result of BiLSTM model	LSTM model satisfaction/%	BiLSTM model satisfaction/%
1. What kind of music do you like best?	Thank you very much.	I especially like listening to music.	45	69
2. What game are you playing recently?	I don't know.	Sorry, I don't like games very much.	35	75
3. Which city do you think you like best?	I like it very much.	I prefer Asian cities.	36	75
4. What did you have for breakfast today?	I only like sweet food.	I had pizza for breakfast.	65	90
5. Do you like going to the beach to bask in the sun	Sorry, I understand.	Yes, I like the sunshine on the beach very much	56	91
6. Is today a weekend?	yes.	Yes, today is Sunday	69	92
7.Let's go for an outing Are you going?	I don't know.	I'd love to join you.	60	92

Table 3. Test Results of Experimental Conversation under the BiLSTM Model

Interactive object	BiLSTM model	BiLSTM model+beam search
1.What color do you like?	I like green clothes.	I like new clothes very much.
2.What color do you like?	I like new clothes.	I'm very interested in green.
3.What color do you like?	I'm sorry.	I haven't thought about it yet.
4.What color do you like?	I'm sorry.	I don't know.
5.What color do you like?	I don't know.	I'm sorry.
6.What color do you like?	I'm sorry.	Tell me again.
7.What color do you like?	I'm sorry.	thank you very much.
8.What color do you like?	Please say it again.	Let me think again.
9.What color do you like?	I don't know.	

Table 4. Test results of experimental dialogue part under the BiLSTM model (unit/%)

Conversation model type	Sadness	Fear	Uneasiness	Happy	Excitement	Angry	Melancholy	Impatience
LSTM	63.54	56.65	54.52	45.54	59.45	53.55	57.45	45.45
BiLSTM	66.45	64.54	67.45	63.45	64.45	67.45	69.45	60.45
LSTM + Attention	63.45	67.43	69.46	69.46	67.46	69.46	68.46	64.46
BiLSTM + Attention	78.54	77.45	73.45	76.45	76.48	76.45	74.45	73.45
LSTM + Attention + beam search	70.45	72.43	69.45	67.45	70.45	71.45	69.45	67.45
BiLSTM + Attention + beam search	89.45	81.55	84.45	80.46	82.45	91.45	86.45	84.56

ter, but the recognition accuracy cannot meet the requirements. After adding the attention mechanism, the emotion recognition rate of the two models has been improved to a certain extent, but the overall effect is not outstanding. Therefore, by adding a cluster search and attention mechanism to the model, compared to the LSTM model in complex emotional dialogue scenarios, the BiLSTM model that integrates attention and cluster search has better overall emotional recognition performance. The recognition rates for the four emotional scenarios of sadness, fear, unease, and happiness are 89.45%, 81.55%, 84.45%, and 80.46%, respectively. Compared to LSTM and BiLSTM, the recognition ability has increased by 35.66% and 21.36%, respectively. From this, it can be seen that the BiLSTM dialogue model has better dialogue performance in intelligent dialogue robots compared to traditional LSTM models, meeting more stringent dialogue scenario requirements.

5. Conclusion

In the context of big data, artificial intelligence dialogue robots will be an important exploration direction of modern science and technology. Research on natural language-oriented robot dialogue technology shows that the traditional seq2seq dialogue model cannot handle long dialogue problems and cannot fully express semantic information. Therefore, the BiLSTM model is selected to fully represent the semantic information and construct the Decoder and Encoder spaces. Considering that the traditional seq2seq model cannot represent all sequence information, an attention mechanism is introduced for optimization, and beam search is used to overcome the problem of safe reply in dialogue systems. The performance test results show that in the dialogue accuracy test of the dialogue model, in the music corpus scene, the LSTM model tends to converge at 52 iterations, and the model semantic test accuracy at this moment is 0.81 , while the BiLSTM model tends to converge at 44 iterations. Convergence, the model semantic test accuracy at this moment is 0.93. In the different emotional scene recognition rate tests (Attention plus beam search), compared with the LSTM model, the proposed BiLSTM has better emotional recognition performance. The recognition rates of the four emotional scenes of sadness, fear, anxiety, and happiness are 89.45%, 81.55%, 84.45%, and 80.46%, while the LSTM model recognition rates are 70.45%, 72.43%, 69.45%, and 67.45%, respectively. At the same time, the BiLSTM model has an excellent performance in the dialogue recognition effect and dialogue response satisfaction test. However, there are still shortcomings in the research. The collected dialogue material information is mainly based on Chinese, with a small number and relatively simple scenes. In the later stage, a foreign language corpus database should be added, and the corpus resources should be enriched.

References

[1] A. Russo, G. D'Onofrio, A. Gangemi, F. Giuliani, M. Mongiovi, F. Ricciardi, F. Greco, F. Cavallo, P. Dario, D. Sancarlo, et al., (2019) *"Dialogue systems and conversational agents for patients with dementia: The human–robot interaction"* **Rejuvenation research** **22**(2): 109–120. DOI: https://doi.org/10.1089/rej.2018.2075.

[2] J. Hedges and M. Sadrzadeh, (2019) *"A generalised quantifier theory of natural language in categorical compositional distributional semantics with bialgebras"* **Mathematical Structures in Computer Science** **29**(6): 783–809. DOI: https://doi.org/10.1017/S0960129518000324.

[3] S. K. Parr and G. T. Gobbel, (2020) *"Considerations for advancing nephrology research and practice through natural language processing"* **Kidney International** **97**(2): 263–265. DOI: https://doi.org/10.1016/j.kint.2019.12.001.

[4] P. Li and K. Mao, (2019) *"Knowledge-oriented convolutional neural network for causal relation extraction from natural language texts"* **Expert Systems with Applications** ***115***: 512–523. DOI: https://doi.org/10.1016/j.eswa.2018.08.009.

[5] P. Li, H. Yu, W. Zhang, G. Xu, and X. Sun, (2020) *"Sa-nli: A supervised attention based framework for natural language inference"* **Neurocomputing** ***407***: 72–82. DOI: https://doi.org/10.1016/j.neucom.2020.03.092.

[6] V. Sorin, Y. Barash, E. Konen, and E. Klang, (2020) *"Deep learning for natural language processing in radiology—fundamentals and a systematic review"* **Journal of the American College of Radiology** ***17***(5): 639–648. DOI: https://doi.org/10.1016/j.jacr.2019.12.026.

[7] E. Ötleş, D. E. Kendrick, Q. P. Solano, M. Schuller, S. L. Ahle, M. H. Eskender, E. Carnes, and B. C. George, (2021) *"Using natural language processing to automatically assess feedback quality: Findings from 3 surgical residencies"* **Academic Medicine** ***96***(10): 1457–1460. DOI: https://doi.org/10.1097/ACM.0000000000004153.

[8] D. Zhu, (2019) *"Humor robot and humor generation method based on big data search through IOT"* **Cluster Computing** ***22***: 9169–9175. DOI: https://doi.org/10.1007/s10586-018-2097-z.

[9] T. Sono, S. Satake, T. Kanda, and M. Imai, (2019) *"Walking partner robot chatting about scenery"* **Advanced Robotics** ***33***(15-16): 742–755. DOI: https://doi.org/10.1080/01691864.2019.1610062.

[10] S. Heath, J. Liddle, and J. Wiles, (2020) *"The challenges of designing a robot for a satisfaction survey: Surveying humans using a social robot"* **International Journal of Social Robotics** ***12***(2): 519–533. DOI: https://doi.org/10.1007/s12369-019-00604-0.

[11] E. K. Chiou, M. Demir, V. Buchanan, C. C. Corral, M. R. Endsley, G. J. Lematta, N. J. Cooke, and N. J. McNeese, (2021) *"Towards human–robot teaming: Tradeoffs of explanation-based communication strategies in a virtual search and rescue task"* **International Journal of Social Robotics**: 1–20. DOI: https://doi.org/10.1007/s12369-021-00834-1.

[12] M. Shiomi, A. Nakata, M. Kanbara, and N. Hagita, (2021) *"Robot reciprocation of hugs increases both interacting times and self-disclosures"* **International Journal of Social Robotics** ***13***: 353–361. DOI: https://doi.org/10.1007/s12369-020-00644-x.

[13] G. Chen, W. Sheng, Y. Li, Y. Ou, and Y. Gu, (2022) *"Humanoid Robot Portrait Drawing Based on Deep Learning Techniques and Efficient Path Planning"* **Arabian Journal for Science and Engineering *47***(8): 9459–9470. DOI: https://doi.org/10.1007/s13369-021-06245-8.

[14] K. E. Gunther, A. L. Hild, and S. L. Bieber, (2018) *"Natural Resource Experience Affects Engagement with Emotionally Primed Presentations of Science"* **Rangeland Ecology & Management *71***(2): 163–170. DOI: https://doi.org/10.1016/j.rama.2017.12.004.

[15] A. V. Karhade, M. E. Bongers, O. Q. Groot, E. R. Kazarian, T. D. Cha, H. A. Fogel, S. H. Hershman, D. G. Tobert, A. J. Schoenfeld, C. M. Bono, et al., (2020) *"Natural language processing for automated detection of incidental durotomy"* **The Spine Journal *20***(5): 695–700. DOI: https://doi.org/10.1016/j.spinee.2019.12.006.

[16] J. Vermassen, K. Colpaert, L. De Bus, P. Depuydt, and J. Decruyenaere, (2020) *"Automated screening of natural language in electronic health records for the diagnosis septic shock is feasible and outperforms an approach based on explicit administrative codes"* **Journal of Critical Care *56***: 203–207. DOI: https://doi.org/10.1016/j.jcrc.2020.01.007.

[17] N. Maganti, H. Tan, L. M. Niziol, S. Amin, A. Hou, K. Singh, D. Ballouz, and M. A. Woodward, (2019) *"Natural language processing to quantify microbial keratitis measurements"* **Ophthalmology *126***(12): 1722–1724. DOI: https://doi.org/0.1016/j.ophtha.2019.06.003.

[18] S. K. Kang, K. Garry, R. Chung, W. H. Moore, E. Iturrate, J. L. Swartz, D. C. Kim, L. I. Horwitz, and S. Blecker, (2019) *"Natural language processing for identification of incidental pulmonary nodules in radiology reports"* **Journal of the American College of Radiology *16***(11): 1587–1594. DOI: https://doi.org/10.1016/j.jacr.2019.04.026.

[19] S. Mor and G. Gupta, (2021) *"Artificial intelligence and technical efficiency: The case of Indian commercial banks"* **Strategic Change *30***(3): 235–245. DOI: https://doi.org/10.1002/jsc.2406.

[20] N. Haristiani, A. Danuwijaya, M. Rifai, and H. Sarila, (2019) *"Gengobot: A chatbot-based grammar application on mobile instant messaging as language learning medium"* **Journal of Engineering Science and Technology *14***(6): 3158–3173.

[21] M. S. Baig, A. Imran, A. U. Yasin, A. H. Butt, and M. I. Khan, (2022) *"Natural language to sql queries: A review"* **International Journal of Innovations in Science Technology *4***: 147–162.

[22] R. L. Boyd and H. A. Schwartz, (2021) *"Natural language analysis and the psychology of verbal behavior: The past, present, and future states of the field"* **Journal of Language and Social Psychology *40***(1): 21–41. DOI: https://doi.org/10.1177/0261927X20967028.

[23] J. Roh, S. Park, B.-K. Kim, S.-H. Oh, and S.-Y. Lee, (2021) *"Unsupervised multi-sense language models for natural language processing tasks"* **Neural Networks *142***: 397–409. DOI: https://doi.org/10.1016/j.neunet.2021.05.023.

[24] V. D. Badal, S. A. Graham, C. A. Depp, K. Shinkawa, Y. Yamada, L. A. Palinkas, H.-C. Kim, D. V. Jeste, and E. E. Lee, (2021) *"Prediction of loneliness in older adults using natural language processing: exploring sex differences in speech"* **The American Journal of Geriatric Psychiatry *29***(8): 853–866. DOI: https://doi.org/10.1016/j.jagp.2020.09.009.

[25] V. D. Badal, S. A. Graham, C. A. Depp, K. Shinkawa, Y. Yamada, L. A. Palinkas, H.-C. Kim, D. V. Jeste, and E. E. Lee, (2021) *"Prediction of Loneliness in Older Adults Using Natural Language Processing: Exploring Sex Differences in Speech"* **The American Journal of Geriatric Psychiatry *29***(8): 853–866. DOI: https://doi.org/10.1016/j.jagp.2020.09.009.