

Retraction

Retracted: Research on Oral English Dialogue Understanding Based on Deep Learning

Scientific Programming

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

Copyright © 2023 Scientific Programming. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] C. Li and R. Tian, "Research on Oral English Dialogue Understanding Based on Deep Learning," *Scientific Programming*, vol. 2022, Article ID 8005321, 8 pages, 2022.

Research Article

Research on Oral English Dialogue Understanding Based on Deep Learning

Chan Li and Rui Tian 

English Teaching and Research Section of Shijiazhuang Medical College, Shijiazhuang, He Bei 050000, China

Correspondence should be addressed to Rui Tian; rui@st.btbu.edu.cn

Received 14 December 2021; Revised 7 January 2022; Accepted 22 January 2022; Published 14 March 2022

Academic Editor: Sikandar Ali

Copyright © 2022 Chan Li and Rui Tian. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Oral English dialogue is a crucial part of a dialogue system that enables a computer to “understand” the input language as a human does, so the performance of a dialogue system is closely related to the performance of oral English dialogue understanding. In task-based human-machine dialogue systems, external knowledge bases can provide the machine with valid information beyond the training data, helping the model to better perform the oral English dialogue comprehension task. In this paper, we propose a deep recurrent neural network based on feature fusion, which directly stacks multiple nodes at a single time node to deepen the complexity of nonlinear transformations. The feature fusion network structure is applied to the ATIS dataset for oral English dialogue comprehension experiments, and the experimental results demonstrate that the feature fusion RNN network can further improve the effectiveness of the RNN network and the GRU network structure unit can obtain better results among different RNN node units.

1. Introduction

With the rapid development of deep learning, the field of artificial intelligence has undergone a cross-generational change [1–3]. As one of the core technologies in the field of natural language processing, human-computer dialogue technology has now become a hot spot for research in this field. Human-computer dialogue systems are used in a wide range of applications, such as smart home appliances, navigation systems, human-machine customer service, and mobile phone assistants, to provide convenient and personalized services for staff and users [4, 5]. Therefore, human-computer dialogue technology is highly valued in both academia and industry, and the research on the human-computer dialogue has important academic significance and application value.

The study of human-computer dialogue technology first originated in the 1960s, when the famous computer scientist Alan Turing proposed the Turing test [6] to test whether machines really have human intelligence. Since then, researchers have been working on human-computer dialogue.

The authors of [7] simulate a psychotherapist to psychologically de-escalate and treat psychiatric patients. On the basis of the chat system, the authors of [8] developed the Parry [9] system.

In recent years, with the spread of social networks and new media, the emergence of massive amounts of data has acted as an invisible hand to drive the development of human-computer dialogue technology. Unlike template- and rule-based dialogue systems, deep learning-based human-computer dialogue technologies, driven by large amounts of data, are able to use neural networks to derive effective features from large-scale training data to achieve a level of understanding of user conversations and learning language expressions. The rapid growth of deep learning techniques today is largely dependent on massive amounts of chart data and rising hardware computing performance [10]. Currently, the most representative human-computer dialogue systems include Apple’s Siri, Microsoft’s Cortana, Taobao’s Aliyun Xiaomi, Baidu’s in-car system Apollo, and Alibaba’s Tmall Genie [11–13]. These products not only use human-machine dialogue to efficiently help users complete

specific tasks and save them time but also allow them to have an excellent service experience. The core systems of these products are called task-based dialogue systems, which can help users complete specific tasks, with the main application scenario being personal assistants.

Users generally interact with the system via voice and text to enable various queries and agent functions such as car navigation, weather enquiries, e-mail sending and receiving, intelligent search, and schedule reminders. The system enables users to complete various daily tasks quickly and accurately [14].

We focus on spoken language comprehension tasks under a modular task human-computer dialogue system. The modular human-computer dialogue system is a modulation task-based human-computer dialogue system, which is divided into the following modules as shown in Figure 1, namely, speech recognition and speech synthesis module, oral English dialogue comprehension module, dialogue management module, and dialogue generation module. In the task-based dialogue system, the oral English dialogue understanding module, as the first part of the system, is the basis of the whole dialogue system, and its function is to understand the user's questions, convert the user's dialogue information into structured data that can be recognized by the computer, and serve as the data input for the dialogue management module. The oral English dialogue comprehension module, therefore, plays a key role in the overall task-based human-computer dialogue system [15]. Currently, the oral English dialogue understanding module is also being developed by many companies as a cloud service open to the public, for example, Microsoft's LUIS platform [16].

In summary, oral English dialogue comprehension as an important module in task-based dialogue systems is a fundamental part of current human-computer interaction technology and is a challenging and innovative research task in both academia and industry. With the rapid development of deep learning, the use of deep learning techniques can effectively improve the effectiveness of spoken language comprehension tasks, promote the development of HCI research, and better meet the needs of practical application scenarios.

2. Related Work

As one of the core technologies in the field of artificial intelligence, human-computer dialogue systems have always been a hot topic of interest to researchers [17]. As human-computer dialogue technology continues to mature and the range of applications in human-computer dialogue systems continues to expand, people's requirements for dialogue systems are becoming higher and higher. In task-based human-computer dialogue systems, spoken language comprehension plays an important and indispensable role in the whole system, and although the research work on spoken language comprehension tasks is relatively mature, there are still many difficulties and challenges in this field. In [18], a graph-and-rule-based approach was proposed to obtain intentional templates for the consumer intention

recognition task, which has yielded good results in the consumer domain. The researchers of [19] showed that the difference in natural language expressions under a single domain can lead to a dramatic increase in the number of templates and rules, which requires a lot of human and material resources. Thus, although rule- and template-based approaches to oral English dialogue understanding do not require the support of huge amounts of training data and can guarantee the accuracy, the human and material resources consumed rise dramatically as the complexity of the task increases and cannot address the high costs associated with changing domains, intention templates, and rules [20].

Reference [21] used DBN to solve the task of intention recognition in oral English dialogue understanding. RNNs, an important model in the field of NLP, were used by [22] to build and perform the task of intention recognition. The authors of [23] compared and investigated traditional recurrent neural networks with the above two variants of the structure on sentence classification tasks and demonstrated that long and short-term memory networks and recurrent networks of gate units were significantly better than ordinary recurrent neural networks. The authors of [10] used recurrent neural networks for the first time for the slot filling task and obtained good results. CRF, a paradigmatic model for solving slot filling tasks in machine learning, was used by [24] in combination with CNN for the slot filling task. Bidirectional recurrent neural networks (BiRNN) were used in [25] to solve the slot filling problem. LSTM with conditional random fields was proposed in [2] to solve the slot filling problem with good results. A joint model using recurrent neural networks and the Viterbi algorithm was first proposed in [12] to solve both the intention recognition and slot filling tasks. Their work, despite the improvement in model performance, still suffers from information loss caused by recurrent neural networks. The authors of [3] adopted the idea of a joint model for intent recognition and slot filling using GRU to perform a sequence annotation task based on word representation learned by GRU to predict slot labels and a pooling layer to capture global information about the user's utterance for the intent recognition task.

Deep learning, supported by a large amount of data and reliable hardware, has gradually become an important technical support for human-computer dialogue research. Deep learning techniques have also been used because of the good effect as an important module in human-computer dialogue systems.

3. Joint Model of Oral Language Comprehension

3.1. Description of the Problem. The knowledge-driven multi-round dialogue oral English dialogue comprehension task makes use of an external knowledge base to supplement the multi-round dialogue oral English dialogue comprehension model with external information, thus improving the accuracy of the intention recognition and slot filling tasks. The task can be defined as given multiple rounds of user dialogue $U = \{u_1, u_2, u_3, \dots, u_t\}$, where u_t is taken as the current round of dialogue and $\{u_1, u_2, u_3, \dots, u_{t-1}\}$ as the

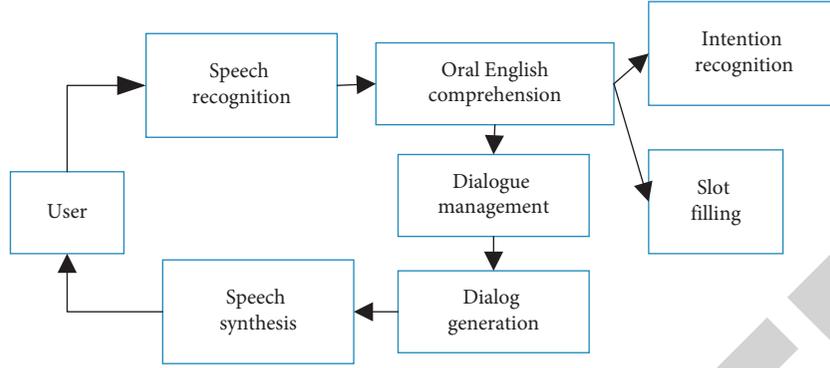


FIGURE 1: The modular task-based human-computer oral English dialogue system.

historical information for each round of dialogue u_i , it is divided into $u_i = \{x_1^i, x_2^i, \dots, x_k^i\}$; and the external knowledge base is taken as $K = \{k_1, k_2, \dots, k_M\}$. The goal of the joint model of oral English dialogue comprehension, etc., is to learn the current round of dialogue based on the historical information provided by H to be able to identify the intent I and slot of the current u_t dialogue label S . The formula is

$$\{I, S\} = F(u_t, H, K). \quad (1)$$

3.2. Model Design. The addition of an external knowledge base enables the oral English dialogue comprehension model to obtain better inferences, and how to obtain good quality candidate knowledge is the main problem that needs to be solved.

Firstly, we design a candidate knowledge recall rule for roughly extracting the candidate knowledge with higher relevance from the external knowledge base; secondly, we propose a candidate knowledge attention module to filter the candidate knowledge based on the implicit information of words so as to obtain a higher quality candidate knowledge vector; and finally, we use the joint model of oral English dialogue comprehension to complete the task of simultaneously completing intention recognition and semantic slot filling [19]. In addition, we build on the model by using the BERT model as a subcoding layer for candidate knowledge, historical information, and current sentences, resulting in a better-quality representation of word vectors.

Based on this, this paper proposes a joint model for a multiround oral English dialogue comprehension task based on an external knowledge base and historical information. We selected ConceptNet [22] as a supplement to the external knowledge base and used its semantic network properties to recall the candidate knowledge related to each word in the user's conversation from the knowledge base; meanwhile, we used the slot attention layer to weigh the average of the historical information and slot candidate knowledge, respectively; finally, we used the joint model to complete the task with the addition of external knowledge and historical information to the oral English dialogue comprehension task.

The oral English dialogue comprehension task contains not only an intention recognition task for sentence-level classification but also a sequence annotation task for word-level classification. In the candidate knowledge selection phase, this paper focuses on the sequence annotation task, which looks for relevant candidate knowledge through the information of each word in the user's dialogue.

As shown in Table 1, for the sample utterance "Please lead me to the MIT," the word MIT refers to "Massachusetts Institute of Technology," which stands for a university acronym and also the geographical location of the university "Cambridge, Massachusetts, USA." This information is inaccessible to the human-computer dialogue system, and the introduction of an external knowledge base can solve this problem. This paper chooses to use "word" as the "query information" of the query knowledge base. For example, the ten most relevant information of "MIT" are queried, and the results contain most of the related knowledge of the word "MIT," which can help the computer better understand the user's dialogue [13].

For each dialogue text $M = \{m_1, m_2, \dots, m_n\}$, the ConceptNet knowledge base was used to conduct an intrabase search for m and recall the 10 most relevant candidate knowledge (or all of them if less than 10), denoted as $K_i = \{k_1, k_2, \dots, k_p\}$, where $0 < p < 11$.

At present, encoding user dialogue utterances into a word vector that the computer can recognize is the first step. We use the BERT pretraining model proposed by Gogel in 2018 as the word encoding layer and define the input data as follows: for each dialogue text $M = \{m_1, m_2, \dots, m_n\}$, BERT is used to transform it into an $X = \{x_1, x_2, \dots, x_n\}$ text vector. For the current word m_i ($0 < i < n + 1$), the recall knowledge $K_i = \{k_1, k_2, \dots, k_p\}$ is encoded using BERT, and the vector of CLS bits in BERT is taken as the sentence representation of each knowledge K to obtain the candidate knowledge encoding $L = \{l_1, l_2, \dots, l_p\}$ corresponding to the current word m_i , where the value of p does not exceed 10.

3.3. Joint Model. The federated model has two main layers of BiLSTM as the base model, including a historical information attention module and a candidate knowledge attention module attached to the first layer of BiLSTM

TABLE 1: Sample candidate knowledge query.

Sample statement:	Please lead me to the MIT
Query words:	MIT
Query results	
MIT is a type of a university	
MIT is in Massachusetts	
MIT is an institute of higher learning	

(current sentence encoding layer) and an auxiliary gate structure attached to the second layer of BiLSTM (sequence annotator), as shown in Figure 2.

The same encoding layer as designed in Chapter 3 is used in the encoding layer of the current utterance to encode the current $u_t = \{x_1, x_2, \dots, x_n\}$ using BiLSTM, which is calculated as shown in equation (2) to obtain the output hidden layer o^1 and the final representation of the utterance c^1 , respectively. Here, $o^1 = \{o_1^1, o_2^1, \dots, o_n^1\}$ represents the hidden state of each word of the user's utterance after $BiLSTM_{\text{encoder}}$ encoding; c^1 is the final state representation of $BiLSTM_{\text{encoder}}$, which is used to represent the information of the whole user's current utterance, as follows:

$$\{O^1, C^1\} = BiLSTM_{\text{encoder}}(u_t). \quad (2)$$

Our model encapsulates the attention module in the same way for both historical information and candidate knowledge filtering. The focus of the attention module is to calculate the attention weights, α_k representing the attention weight between the k th word of the current statement after the encoding layer and the corresponding hidden layer o_k^1 and the historical information H or candidate knowledge L (uniformly denoted by h in the formula), $v_k = \{v_1^k, v_2^k, \dots, v_s^k\}$ representing the set of information between the corresponding hidden layer o_k^1 and the historical information or candidate knowledge after the attention weight layer, and the attention weights are calculated as follows:

$$\alpha_k = \frac{\exp(\text{score}(o_k^1, h_i))}{\sum_{i=1}^n \exp(\text{score}(o_k^1, h_i))}, \quad (3)$$

$$v_k = \sum_{i=1}^n \alpha_{k,i} h_i.$$

After the attention module, both historical information and candidate knowledge were weighted to obtain the most relevant and important information about the word, where the final representation of the current word x_i was v_i^1 and the final representation of the candidate knowledge was v_i^1 . A layer of a forward propagation neural network was used to fuse V_h, V_l , and o_k , and a ReLU activation function was added for the nonlinear transformation. $D = \{d_1^1, d_2^1, \dots, d_n^1\}$ represents the set of vectors over which the historical information representation, the candidate knowledge representation, and the current hidden layer are fused, and d_k^1 is computed as follows:

$$d_k^1 = \text{relu}(FF(V_h, V_l, o_k)). \quad (4)$$

The input of the decoder is D and c^1 of the output of the attention module, and the output of the decoder is O^2 and I , which are used for the semantic slot filling and intention recognition tasks, respectively.

$$\{O^2, I\} = BiLSTM_{\text{tag}}(D, C^1). \quad (5)$$

Using the correlation between intention recognition and semantic slot filling in oral English dialogue comprehension, adding intention information to slot recognition can effectively improve the performance of the semantic slot filling model. The combination of the context vector O^2 and the intention vector I of the slot is used to obtain the weight vector g through the gate structure and use the g congregation as the weight for predicting the slot y_i^s , where w_1 and w_y^s are trainable parameters and are calculated as follows:

$$g = \sum w_1 \cdot \tan h(o_i^2 + w_2 \cdot I),$$

$$y_i^s = \text{softmax}(w_y^s (o_i^2 \cdot g + o_i^2)). \quad (6)$$

The intention recognition task uses the final cell state/prediction from the sequence annotation, where w is the trainable parameter and y is calculated as follows:

$$y_i^l = \text{softmax}(w_y^l \cdot I). \quad (7)$$

4. Experiment

4.1. Experimental Environment. The experimental environment is a 64 bit Linux environment using Ubuntu 16.04.3. Development tools used are Python 2.7.14 in the anaconda environment, and the python deep learning libraries scene and eras are used. Numpy and Scipy packages can be installed directly from the Linux command line using the pipe command. To better handle deep neural network algorithms, think of a deep learning library that can perform fast operations on both CPUs and GPUs and is designed to handle large neural network algorithms. Eras is an application programming interface for high-level neural network applications using thumb, tomorrow, and CNTK as backends. It is written in python, is highly modular in nature, and runs seamlessly on both CPUs and GPUs. This paper experiments with deep neural networks are based on Python2.7 using thumb and keras deep learning libraries.

4.2. Data Preparation. This paper applies data information from the Airline Travel Information System (ATIS) domain. This database was developed in the early 1990s as a project of the US Defense Advanced Research Projects Agency (DARPA) in the area of airline travel information. The task consists of voice queries for flight-related information, where the input speech is converted into corresponding textual information by speech recognition, and the understanding of the text is reduced to the extraction of task-specific parameter problems such as destination and departure. The ATIS data include 4978 sentences in the training set from the ATIS-2 and ATIS-3 datasets of Class A data and 893 sentences in the test set from the Nov93 and

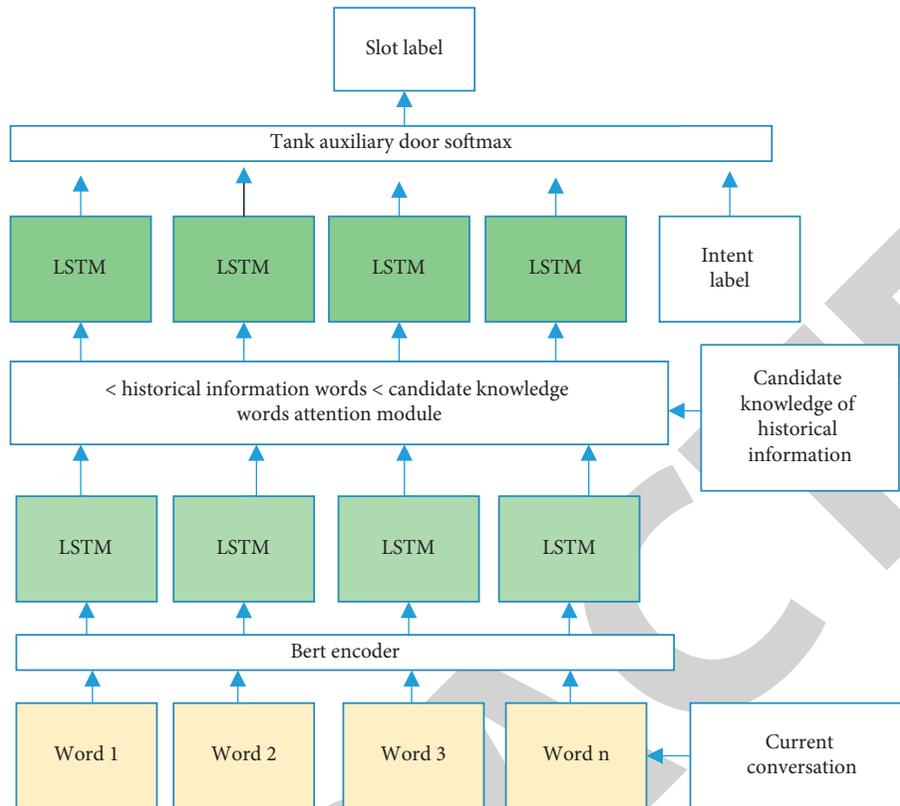


FIGURE 2: Structure of the oral comprehension model.

Dec94 datasets of ATIS-3, where the words in each sentence can be found in a table of named corpora, including domain-specific entities such as city, airline, airport name, and date [23].

The ATIS database uses a semantic framework to represent that each sentence has an intention and each word in the sentence has a corresponding semantic slot. The slot values shown in Table 2 have not been normalized. There are 128 tags in the ATIS database.

4.3. Evaluation Indicators. A more common evaluation metric for sequence annotation units in spoken language comprehension is the F value, and the F value is the reconciled balance of recall and precision.

4.4. Experimental Results and Analysis. CRF is the most commonly used method to deal with sequence problems in oral English dialogue comprehension, and it has achieved good results in sequence annotation. The results of the experiments are shown in Table 3.

From Table 3, we can see that the F1 values of Elman and Jordan are 0.19% and 1.3% higher in the normal RNN model compared to the CRF model. Therefore, ordinary RNN is more advantageous in solving the problem of sequence annotation for oral English dialogue understanding. The LSTM and GRU variants of the RNN improved by 1.41% and

1.68%, respectively, compared to the Jordan of the normal RNN, and by 0.3% and 0.51% compared to the Elman, with GRU the LSTM and GRU variants. The main reason for this is the inclusion of gate control units in the recurrent network nodes, which makes the network more convenient for controlling history and input information and solves the gradient disappearance problem of ordinary recurrent neural networks.

Several variants of RNNs are proposed, and we also perform oral English dialogue comprehension experiments on the ATIS database. The purpose of this series of experiments is to verify which of the different LSTM variants achieves better results in this experiment and to analyze the results of this experiment.

From the experimental results in Table 4, it can be seen that the experimental results of different long and short-term memory networks on the ATIS dataset are different. From the experimental results, it seems that the LSTM model achieves a 0.48% improvement over the PeepholeLSTM experimental results and a 0.13% improvement relative to the CoupledLSTM model, and it can be seen through these experiments that, for many variants of the LSTM, the authors of [4] compared eight different structural variants of the LSTM model and concluded that none of the variants achieved results that exceeded those of ordinary long and short-term memory networks. In this paper, the two variants of PeepholeLSTM and CoupledLSTM were not introduced

TABLE 2: An example of an ATIS dataset.

Sentence	How much is the cheapest flight from Boston to New York tomorrow morning?
Objectives	Airfare
Cost relative	Cheapest
Depart city	Boston
Arrival city	New York
Depart date. Relative	Tomorrow
Depart time period	Morning

TABLE 3: Experimental results of CRF and RNN.

Model	Precision (%)	Recall (%)	F1 (%)
CRF	93.15	92.73	92.24
Jordan-RNN	93.25	93.02	93.13
Elman-RNN	94.36	94.13	94.24
LSTM	94.69	94.4	94.54
GRU	95.03	94.6	94.81

TABLE 4: LSTM, PeepholeLSTM, and CoupledLSTM structural model experiments.

LSTM variant	Precision (%)	Recall (%)	F1 (%)
LSTM	94.69	94.4	94.54
PeepholeLSTM	94.14	93.97	94.06
CoupledLSTM	94.18	94.04	94.41

into the experimental comparison of the structure of the hidden layer units of deep neural networks for feature fusion. The main three types of structural units, namely, Basic RNN, LSTM, and GRU, are used for training and learning in the deep neural network structure of feature fusion.

In the deep recurrent neural network structure based on feature fusion, the data are processed by changing the recurrent neurons in the network structure, and the experimental results obtained are shown in Table 5.

In comparing the experimental results of the deep recurrent neural network based on feature fusion with the experimental results of the basic RNN and variants of the LSTM and GRU in Table 3, the performance of the Elman with feature fusion improved by 0.47% over the Elman neural network structure for the same recurrent neurons, the LSTM with feature fusion improved by 0.51% compared to the LSTM neural network structure, and the GRU with feature fusion improved by 0.38% compared to the GRU. The experimental performance of the GRU with feature fusion was improved by 0.38% compared to the GRU. These experiments show that the deep recurrent neural network model with feature fusion can achieve better performance metrics in oral English dialogue understanding than the normal RNN and its variants. In contrast, among the different recurrent neurons in the feature recurrent neural network structure, the recurrent neurons of the normal

TABLE 5: Deep feature fusion network with different recurrent neurons.

Circulating neuron	Precision (%)	Recall (%)	F1 (%)
Basic	95.1	94.34	94.71
LSTM	95.48	94.57	95.02
GRU	95.5	94.89	95.19

RNN were compared with the LSTM and GRU, the recurrent neurons of the variants obtained better performance, and we can compare the performance plots of the normal RNN with those of the feature fusion recurrent neural network by means of a bar chart.

The experiments in this paper are built using tensorflow and keras deep learning libraries in Python, and the number of training epochs for the neural network set in this paper is 50.

Table 6 shows that the neural network model is a normal RNN model with LSTM and GRU. The results obtained from the first training are higher for the variant than for the two models of the basic RNN. However, when compared with the variants, the LSTM and the gated recurrent network structures yielded better results mainly because these two structures mainly incorporate gating and cellular memory states on the neuron nodes, which can control the reception or rejection of historical information and input information, thus solving the gradient disappearance problem of ordinary RNNs. Due to the large number of parameters contained in the neurons of the LSTM and GRU, the training time is 2-3 times longer than that of a normal RNN model when training and learning, but the performance of the network is obtained. In order to visualize the training and learning process, the performance of the network is described by depicting the change in $F1$ value, and a line graph is drawn as shown in Figure 3.

TABLE 6: F1 values during training of different network structures.

Epoch	0	9	19	29	39	49
Elman	78.18	91.35	92.67	93.5	93.29	94.24
Jordan	81.44	91.25	92.33	92.78	93.04	93.13
LSTM	85.95	90.78	93.09	94.05	94.2	94.54
GRU	88.75	91.82	93.78	94.4	94.72	94.81
Feature Elman	80.95	91.63	92.7	93.78	94.5	94.71
Feature LSTM	80.69	90.2	92.56	94.13	94.69	95.02
Feature GRU	90.3	92.5	93.69	94.54	95.03	95.19

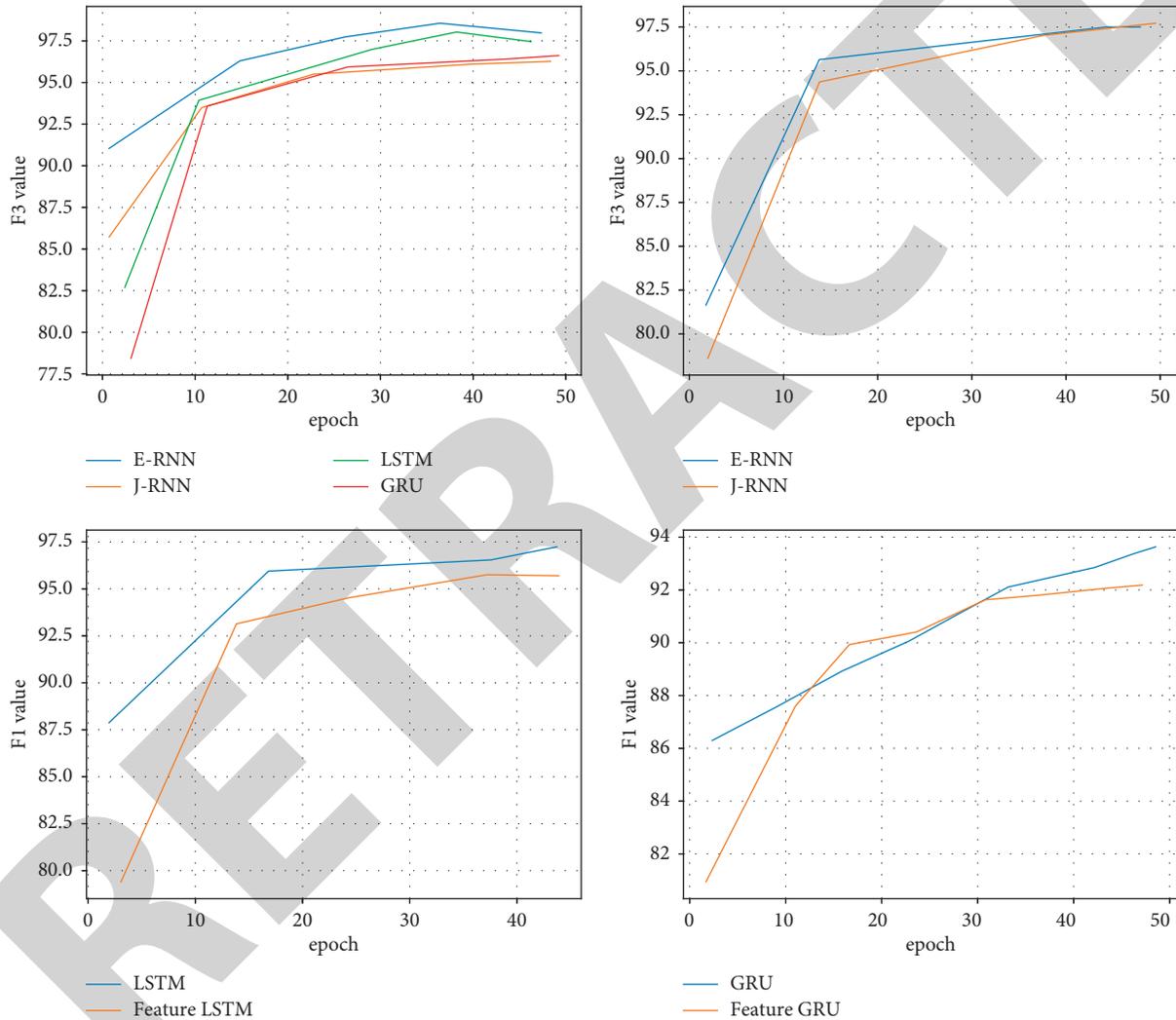


FIGURE 3: Training process of different neural network structures.

5. Conclusions

The current technique of combining deep learning models with external knowledge bases is widely admired by researchers in the field of NLP. In task-based human-machine dialogue systems, external knowledge bases can provide the machine with valid information beyond the training data to help the model better perform the oral English dialogue comprehension task. Firstly, we propose a candidate

knowledge recall rule to extract highly relevant knowledge sets from external knowledge bases; secondly, we propose a “word” related knowledge attention module to assign relevance weights to the knowledge in candidate knowledge sets to obtain the most effective knowledge vectors; and finally, we introduce a pretrained language model based on the idea of transfer learning. Finally, a pretrained language model BERT is introduced to replace Word2Vec as the word vector support for the model based on the idea of transfer

learning. Experiments show that the model proposed in this paper has better results on the restructured ATIS dataset.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

References

- [1] J. Li and E. Hovy, "A model of coherence based on distributed sentence representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 2039–2048, Doha, Qatar, 2014, October.
- [2] W. Y. Wang, R. Artstein, A. Leuski, and D. Traum, "Improving spoken dialogue understanding using phonetic mixture models," in *Proceedings of the Twenty-Fourth International FLAIRS Conference*, Palm Beach, FL, USA, 2011, March.
- [3] F. Morbini, K. Audhkhasi, R. Artstein et al., "A reranking approach for recognition and classification of speech input in conversational dialogue systems," in *Proceedings of the 2012 IEEE oral English dialogue Technology Workshop (SLT)*, pp. 49–54, IEEE, Miami, FL, USA, 2012, December.
- [4] E. Nouri, R. Artstein, A. Leuski, and D. Traum, "Augmenting conversational characters with generated question-answer pairs," in *Proceedings of the 2011 AAAI Fall Symposium Series*, Arlington, Virginia, 2011, November.
- [5] X. Ning, W. Li, B. Tang, and H. He, "BULDP: biomimetic uncorrelated locality discriminant projection for feature extraction in face recognition," *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2575–2586, 2018.
- [6] A. Tammewar, M. Pamecha, C. Jain, A. Nagvenkar, and K. Modi, "Production ready chatbots: generate if not retrieve," in *Proceedings of the Workshops at the Thirty-Second AAAI Conference on Artificial Intelligence*, New Orleans, LA, USA, 2018, June.
- [7] M. Zembrzowski, H. Jeon, J. Marhula et al., "Automatic speech recognition adaptation to the IoT domain dialogue system," in *Proceedings of the International Symposium on Methodologies for Intelligent Systems*, pp. 215–226, Springer, Warsaw, Poland, 2017, June.
- [8] L. Wang, C. Zhang, Q. Chen et al., "A communication strategy of proactive nodes based on loop theorem in wireless sensor networks," in *Proceedings of the 2018 Ninth International Conference on Intelligent Control and Information Processing (ICICIP)*, pp. 160–167, IEEE, Wanzhou, China, 2018, November.
- [9] T. Kawahara, "Spoken dialogue system for a human-like conversational robot ERICA," in *Proceedings of the 9th International Workshop on Spoken Dialogue System Technology*, pp. 65–75, Springer, Singapore, 2019.
- [10] C. Li, L. Li, and J. Qi, "A self-attentive model with gate mechanism for oral English dialogue understanding," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 3824–3833, Brussels, Belgium, 2018.
- [11] R. Masumura, T. Tanaka, A. Ando et al., "Improving conversation-context language models with multiple oral English dialogue understanding models," in *INTERSPEECH*, pp. 834–838, Springer, Singapore, 2019.
- [12] Y. N. Chen, A. Celikyilmaz, and D. Hakkani-Tur, "Deep learning for dialogue systems," in *Proceedings of the 27th International Conference on Computational Linguistics: Tutorial Abstracts*, pp. 25–31, Vancouver, Canada, 2018, August.
- [13] X. Ning, K. Gong, W. Li, and L. Zhang, "JWSAA: joint weak saliency and attention aware for person re-identification," *Neurocomputing*, vol. 453, pp. 801–811, 2021.
- [14] Z. Zhao, S. Zhu, and K. Yu, "A hierarchical decoding model for oral English dialogue understanding from unaligned data," in *Proceedings of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7305–7309, IEEE, Brighton, UK, 2019, May.
- [15] H. Li, D. Zeng, L. Chen, Q. Chen, M. Wang, and C. Zhang, "Immune multipath reliable transmission with fault tolerance in wireless sensor networks," in *Proceedings of the International Conference on Bio-Inspired Computing: Theories and Applications*, pp. 513–517, Springer, Xi'an, China, 2016, October.
- [16] D. Wu, C. Zhang, L. Ji, R. Ran, H. Wu, and Y. Xu, "Forest fire recognition based on feature extraction from multi-view images," *Traitement du Signal*, vol. 38, no. 3, pp. 775–783, 2021.
- [17] C. H. Cao, Y. N. Tang, D. Y. Huang, G. WeiMin, and Z. Chunjiang, "IIBE: an improved identity-based encryption algorithm for wsn security," *Security and Communication Networks*, vol. 2021, Article ID 8527068, 8 pages, 2021.
- [18] H. Che and J. Wang, "A nonnegative matrix factorization algorithm based on a discrete-time projection neural network," *Neural Networks*, vol. 103, pp. 63–71, 2018.
- [19] Y. Wang, J. Wang, and H. Che, "Two-timescale neurodynamic approaches to supervised feature selection based on alternative problem formulations," *Neural Networks*, vol. 142, pp. 180–191, 2021.
- [20] X. Ning, K. Gong, W. Li, L. Zhang, X. Bai, and S. Tian, "Feature refinement and filter network for person re-identification," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 9, pp. 3391–3402, 2021.
- [21] Y. J. Zhao, Y. L. Li, and M. Lin, "A review of the research on dialogue management of task-oriented systems," *Journal of Physics: Conference Series*, vol. 1267, no. 1, Article ID 012025, 2019.
- [22] W. Shang, S. Zhu, and D. Xiao, "Research on human-computer dialogue based on improved Seq2seq model," in *Proceedings of the 2021 IEEE/ACIS 20th International Fall Conference on Computer and Information Science (ICIS Fall)*, pp. 204–209, IEEE, Xi'an, China, 2021, October.
- [23] K.-C. Li, M. Chang, and K.-H. Wu, "Developing a task-based dialogue system for English language learning," *Education Sciences*, vol. 10, no. 11, p. 306, 2020.
- [24] P. Biyani, C. Akkaya, and K. Tsioutsoulouklis, "Identifying domain independent update intents in task based dialogs," in *Proceedings of the 19th Annual SIGDial Meeting on Discourse and Dialogue*, pp. 410–419, Melbourne, Australia, 2018, July.
- [25] L. Matějů, D. Griol, Z. Callejas, J. M. Molina, and A. Sanchis, "An empirical assessment of deep learning approaches to task-oriented dialog management," *Neurocomputing*, vol. 439, pp. 327–339, 2021.