IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 7 June 2024, accepted 12 July 2024, date of publication 24 July 2024, date of current version 2 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3432860

RESEARCH ARTICLE

Intelligent Human–Computer Interaction Dialog Model Based on End to End Neural Networks and Emotion Information

JINHUANG CHEN[1], PEIQI TU[2], ZEMIN QIU[1], ZHIJUN ZHENG[1], AND ZHAOQI CHEN[1]
[1]School of Information and Intelligent Engineering, Guangzhou Xinhua University, Guangzhou 510520, China
[2]School of Biomedical Engineering, Guangzhou Xinhua University, Guangzhou 510520, China

Corresponding author: Peiqi Tu (tpq@xhsysu.edu.cn)

This work was supported in part by the Key Disciplines at the School Level, Computer Application Technology Guangdong Province Key Construction Discipline Scientific Research Capacity under Grant 2020XZD02; in part by the Improvement Project, Video Object Detection, Tracking and Application Research under Grant 2022ZDJS151; and in part by the School-Level Teaching Quality Engineering Project, Virtual Department of Intelligent Medical Imaging Technologies Project under Grant 2023JYS003.

ABSTRACT The current dialog model suffers from the lack of grammatical accuracy and content relevance. To provide interactive content with high quality, the study constructs an intelligent human-computer interaction dialog model using end to end neural network and emotion information. The study utilizes the transformer-based bidirectional coding model and the end to end neural network model to realize the interactive dialogue, and embeds emotional information in the model to realize the intelligent human-computer interaction dialog. The outcomes demonstrated that the classification accuracy of the research-designed sentiment classification model on the three datasets was 89.97%, 90.15%, and 91.44%, respectively. The research-designed model consistently outperformed the other models in terms of emotion accuracy as well as Distinct-1 and Distinct-2 score values, which were 0.068, 0.204, and 80.425%, respectively. Meanwhile, the average accuracy of the research-designed model was higher than 90% on all three datasets. The emotional quality and content quality of the model during the dialog process were higher than several more popular emotional interaction dialog models. The comprehensive analysis outcomes reveal that the research-designed model is able to generate high-quality emotion-based responses in the conversation, providing users with a more intelligent human-computer interaction experience.

INDEX TERMS End to end neural network, emotional information, human-computer interaction dialogue, intelligence, emotion classification.

I. INTRODUCTION

Deep learning has led to a rapid development in the theory of machine learning, machine vision, reinforcement learning, natural language processing, and other related fields [1]. Furthermore, robotics represents a significant domain of application for artificial intelligence. As the field of artificial intelligence continues to evolve, the potential for human-robot interaction is expanding. This encompasses a range of functionalities, including medical assistance, intelligent customer service, knowledge education, and other services. Joukhadar et al. carried a experiment on the task of conversation behavior classification in order to correctly identify the intention behind the speaker during human-computer conversations. Meanwhile, it proposed a model of Arabic conversational behavior using bidirectional encoder representations from transformer (BERT). The superiority of the model was demonstrated through comparative experiments with a classification accuracy of over 90% [2]. McMillan et al. proposed a natural language processing algorithm-based investor sentiment score method for further research on bitcoin investment, using social data from human-computer dialogue systems. The method analyzed the relationship between investor sentiment and bitcoin investment situation by comparing the sentiment score with the change in bitcoin price [3]. Kobayashi and service,

The associate editor coordinating the review of this manuscript and approving it for publication was Giuseppe Desolda.

knowledge education proposed eye contact for enriching human-computer interface modes in order to enrich human-computer communication. Their neural network based on target detection detected the eyes of interacting users, and judged whether the service was on or not by the line of sight of interacting users. Experimental results demonstrated that the method had high usability [4]. Li et al. proposed a human-computer dialog model on the basis on voiceprint recognition and image recognition in order to protect user privacy in human-computer dialog. A sound noise reduction method based on Kalman filter algorithm was proposed in this model to improve the accuracy of voiceprint recognition. Experimental results indicated that the recognition accuracy of this method reached more than 90% [5]. Although many good results have been achieved for chatbot systems, the rule and template-based techniques used have major limitations. This new climax has been ushered in by the study effort on deep learning related technologies, which has been made possible by the rapid advancement of computer science and technology [6]. The present deep learning-based large-scale medical data centralized model for disease recognition, event prediction, and other purposes was the subject of a literature analysis by Somani et al. To help researchers grasp the principles of deep learning and make use of the newest technology in ECG analysis, it examined the implementation of deep learning in medical services and conducted a systematic assessment of the field [7]. Saleem et al. employed statistical analysis and a review of the literature to illustrate the utilization of machine learning and deep learning algorithms in agricultural robots over the past decade. This study aimed to examine the most recent research on deep learning in agricultural automation. It studied the effectiveness of deep learning in certain agricultural operations due to traditional machine learning models by plotting the performance graphs [8].

In general, there are two main types of dialog systems based on deep learning techniques, namely retrieval-based models and generative-based models. Large amounts of training data are typically needed for these models, and the resulting responses may include issues such syntactic mistakes [9]. From the above studies, the current dialog bots are generally designed to help users complete simple business processing and customer support and other specific task applications. However, as time passes and people's standards of spirituality evolve, there is an increasing necessity for human-computer interaction dialogue (HCID) that incorporates emotional considerations [10]. In recent years, tasks related to emotional dialog generation have also received the attention of many domestic and foreign scholars. Raamkumar and Yang to identify the current key gaps and future challenges in incorporating emotion into dialogue systems, studied the development of this field in five dimensions. The five review dimensions were conceptual empathy model and framework, adopted empathy concepts, developed data and algorithms, evaluated strategies, and advanced methods. The results showed that currently text-based emotion perception played a dominant role [11]. To bridge the gap between language and vision in multi-modal dialogue interaction systems, Firdaus et al. proposed a method to control response in the generation aspect of task-oriented multi-modal dialogue systems. This method used multi-modal hierarchical memory network to generate the response of text and image information, and then realized multi-modal dialogue. It also created a multi-domain, multi-modal dialogue dataset that included text and images for hotels, restaurants, electronics, and furniture. The results showed that the proposed method had significant advantages over the baseline model in controlling response generation tasks [12]. Qamar et al. found that with the increasing use of social networks, relationships in text data can be inferred from interaction data. Based on this finding, a new method of interpersonal relationship recognition was proposed. The approach used a psychological model to label the conversation corpus and applies machine learning techniques to determine the emotion-to-relationship mapping. The experimental results showed that the method had a correct rate of 85% in detecting interpersonal relationships, and compared with the traditional dialogue system based on language model, its ability to perceive emotions has been greatly improved [13]. Zhang et al. proposed a multi-modal and multi-task interactive graph attention network to solve the problem that there is no unified framework to study session context dependence, multi-modal interaction and multi-task association. The network can be used to establish local and global conversation context connections while enabling cross-task connections between two tasks. Finally, comprehensive experiments were conducted on MELD, MEISD and MSED and compared with the single-task learning framework. The experimental results showed that the marginal rate of the method in sentiment analysis and recognition was 1.88% and 1.99%, which was significantly higher than that of the single-task learning framework [14]. Synthesizing the current state of research on emotion dialog generation, it is known that the current methods often lead to lack of grammatical correctness and content relevance of responses in order to incorporate emotion information (EI) into the model [15]. The study used the end to end neural network (Seq2Seq) and EI to build a new intelligent HCID model to address the issues raised by the related research. The model sought to balance the link between the sentiment and content of the responses and enhance the quality of the responses produced by the intelligent HCID model. The innovation of the study is that a hierarchical sentiment model is built to prevent different levels of EI from interfering with each other. Compared with other research results, the contribution of this research is to provide more detailed and accurate emotion processing ability for intelligent HCID system by constructing hierarchical emotion model. With the application of Seq2Seq, it can be ensured that the generated responses were grammatically correct, while incorporating EI to keep the responses highly relevant in content to the topic of the conversation.

The research combines Seq2Seq and hierarchical affective model, mainly because it can simultaneously solve two key problems in intelligent HCID system: syntactic correctness of reply and relevance of content. When dealing with these problems, the traditional methods tend to pay attention to one thing and lose the other, either sacrificing the grammatical correctness to pursue the richness of emotional expression, or ignoring the transmission of emotion when pursuing the relevance of content. This approach has not been widely studied, possibly because building hierarchical affective models requires detailed segmentation and modeling of emotions, while ensuring that these different levels of affective information can be efficiently processed and fused in neural networks.

The study contains four parts. The first part is the introduction section, which identifies the direction of the study by analyzing the background of the study as well as the current state of research. The second part is methods and materials, which introduces the techniques used in the model designed for the study and the application of the techniques. The performance of the constructed model is analyzed through a series of experiments in the third part. The fourth part is discussion and conclusion, which further summarizes and analyzes the results of the study. At the same time, the study summarizes the content and significance, as well as the outlook of future research directions.

II. METHODS AND MATERIALS

To realize the intelligence of HCID, the study firstly constructs an interactive dialogue model using the Seq2Seq model, and introduces BERT into the dialogue model to improve it. On the basis of this model, the study introduces the emotion classification (ECF) model as well as hierarchical emotion coding (HEC) and emotional cost function (EC) into the dialog generation model. As a result, an intelligent HCID model based on Seq2Seq model and EI is constructed.

A. HUMAN-COMPUTER INTERACTION DIALOGUE GENERATION BASED ON SEQ2SEQ

Emotions play an indispensable role in cognition and social behavior, and integrating emotional factors into HCID can strengthen the emotional connection between the dialogue system and the human user and increase the user's involvement in the dialogue process [16], [17], [18]. Traditional human-computer dialogue technology has many limitations, such as the lack of coherence in dialogue generation, the inability to accurately understand the user's intention and the lack of emotional color. To solve these problems, this study introduces an HCID generation model based on Seq2Seq. The Seq2Seq model is a deep learning architecture capable of mapping one sequence to another, and is well suited for tasks such as machine translation, text summarization, and conversation generation in natural language processing [19]. However, conventional Seq2Seq is not ideal for capturing contextual information of sequences. For this purpose, BERT and gated recurrent unit (GRU) are used to improve the Seq2Seq model, to better capture contextual information about the input sequence. BERT is a bidirectional encoder based on Transformer, which learns the deep representation of language through pre-training and achieves significant performance improvement in various natural language processing tasks [20]. Bidirectional GRU is a recurrent neural network with gating mechanism, which can capture long-term dependencies in the sequence while maintaining the bidirectional flow of information, so as to better capture the context information of the input sequence [21]. The study combines BERT and bidirectional GRU as the encoder part of Seq2Seq. In addition, when the traditional Seq2Seq model deals with complex text sequences, due to its limited encoder feature extraction ability, it is often easy to lose the long-term dependency relationship within the text, which makes the input information obtained by the decoder insufficient, and then affects the quality of the generated dialogue content. The bidirectional coding model based on Transformer contains a lot of syntax information, and its attention mechanism can better solve long-term dependency problems. Therefore, on the basis of the above improvements, the attention mechanism is also introduced to improve the encoder. The structure of the improved dialog generation model is shown in Figure 1.

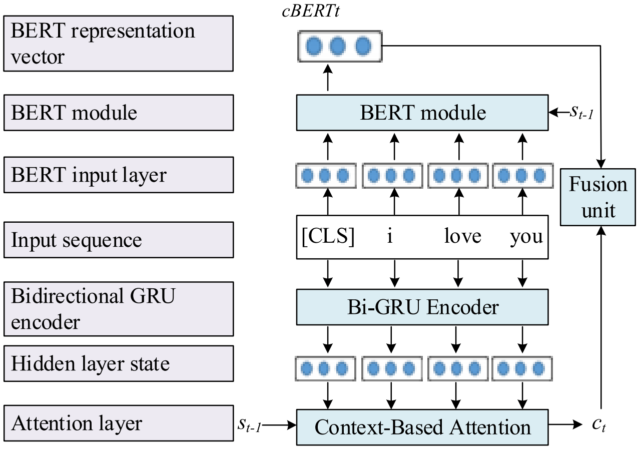


FIGURE 1. The improved dialog generation model structure with BERT.

In Figure 1, the model structure contains structural components such as attention layer (AL), hidden layer (HL), bidirectional GRU coding, BERT input layer, etc., in which the study uses bidirectional GRU to improve the model's comprehension of textual contextual information in the data pre-processing stage. Bidirectional GRU contains two units, forward and reverse, and the output vector of the HL of the forward GRU is calculated as shown in Equation (1).

$$\vec{h}_i = GRU\left(x_i, \vec{h}_{i-1}\right) \tag{1}$$

In Equation (1), $\vec{h}_{i-1}$ and $\vec{h}_1$ are the HL states outputted by the GRU unit at the previous and current moments, respectively, and x_i is the input word vector (WV) at the i-th moment. Equation (1) allows for a further deduction of the HL state in the opposite direction at that particular moment. Equation (2) illustrates how the HL states in both directions

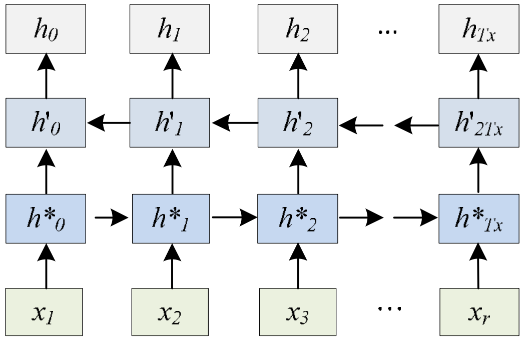


FIGURE 2. Bidirectional GRU model structure.

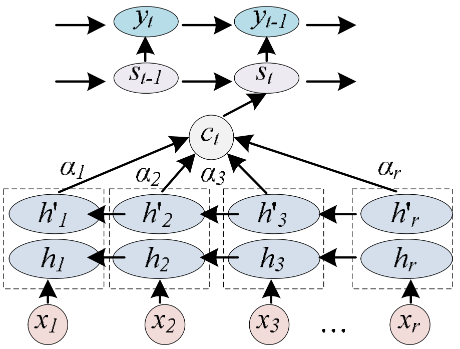


FIGURE 3. Structure of the Seq2Seq model with an attention mechanism.

are joined to create the input of the finished bidirectional GRU coding model.

$$h_i = \left[\vec{h}_i, h'_i\right] \tag{2}$$

In Equation (2), h'_i is the HL state at the i-th moment in the opposite direction. Figure 2 depicts the bidirectional GRU model's construction.

As shown in Figure 2, a bidirectional GRU is composed of two unidirectional GRUs that handle the forward and reverse of the sequence, respectively. Such a design can make the output of the model at any time contain both past and future information, thus improving the ability of the model to capture contextual information. The BERT model can capture rich feature information of text, including text surface features, syntactic features and semantic features, and it has significant advantages in dealing with long-distance dependent information. However, it is prone to lose some of the sequential information in the field of single round dialog generation [22], [23], [24]. For this reason, the study combines BERT and Seq2Seq modeling. The study combines BERT and bi-directional GRU to form the encoder part of the model, and the study designs the BERT module in the model to first convert all the words in the input sequence into BERT input representation. Next, it inputs them into the BERT model to acquire the collection of outputs. Equation (3) particularly displays the output set.

$$\begin{cases} H = BERT\left(X_{new}\right) = \left(T^0, T^1, T^2, \ldots, T^{T_x}\right) \\ X_{new} = \left[\left[CLS\right], X\right] \end{cases} \tag{3}$$

In Equation (3), $[CLS]$ is the input data flag in the BERT module, X is the original sequence, X_{new} is the original sequence with $[CLS]$ flag, T^i is the i-th data in the sequence, and T_x is the sequence data. The output vector through the BERT model is shown in Equation (4).

$$cBERT_t = W_{cB} \cdot \left[T^0, s_{t-1}\right] \tag{4}$$

In Equation (4), s_{t-1} is the hidden state of the decoder in the last moment, $cBERT_t$ is the final output vector, and W_{cB} is the parameter matrix of the fully connected layer. Aiming at the problem that Seq2Seq model tends to lose information when dealing with long sentences, the study introduces attention mechanism to improve it. The structure with attention mechanism-Seq2Seq (AM-Seq2Seq) is shown in Figure 3.

As shown in Figure 3, the attention vector is weighted by measuring the matching degree between the input and output positions at the decoding time. The specific calculation method is shown in Equation (5).

$$\begin{cases} c_t = \sum_{j=1}^{T_x} \alpha_{tj} h_j \\ \alpha_{tj} = \dfrac{\exp\left(e_{tj}\right)}{\sum_k \exp\left(e_{tk}\right)} \\ e_{tj} = a\left(s_{t-1}, h_j\right) \end{cases} \tag{5}$$

In Equation (5), e_{tj} and α_{tj} are variable parameters, and k is the data ordinal number. h_j is the hidden state, a is the dot product operation, and c_t is the attention vector. Thereafter, the study updates the current state using the last hidden state, the last output, and the attention vector as shown in Equation (6).

$$s_t = f\left(s_{t-1}, y_{t-1}, c_t\right) \tag{6}$$

In Equation (6), $f(\cdot)$ is the state function and y_{t-1} is the previous output. The current output content is obtained from the current state as well as the previous output and the attention vector as well. After improving the traditional Seq2Seq model using BERT and bi-directional GRU as well as AM, the features extracted by the improved model module cover multiple dimensional information of the text, although they are comparatively limited in their capacity for temporal capture [25], [26], [27]. For this reason, research inspired by the idea of gate control proposes a fusion unit (Fu) to combine BERT output vectors with contextual information. The structure of the Fu is shown in Figure 4.

As shown in Figure 4, the Fu first calculates a gated signal using the attention vector of the basic Seq2Seq model and the senttion-level feature vector output of the BERT module. Then, according to these two data vectors, the integrated information of the current moment is calculated under the control of the gate signal. The integrated information is

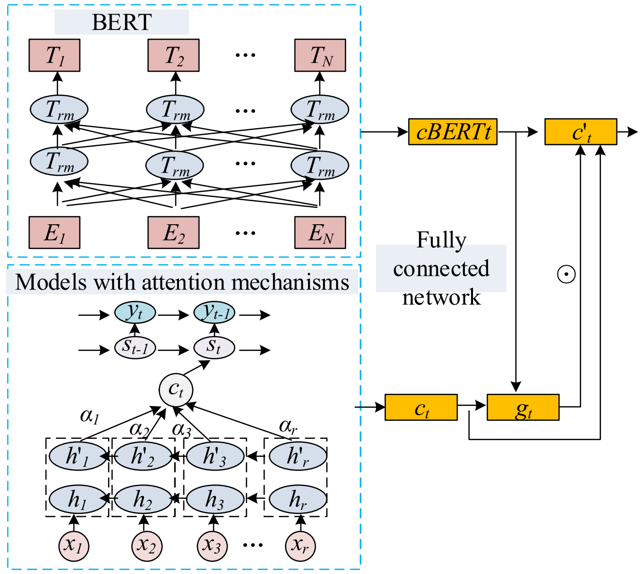


FIGURE 4. The proposed fusion unit structure.

calculated as shown in Equation (7).

$$\begin{cases} c'_t = g_t \odot c_t + (1 - g_t) \odot cBERT_t \\ g_t = \sigma\left(W_{FC} \cdot [c_t; cBERT_t]\right) \end{cases} \tag{7}$$

In Equation (7), g_t is the gating signal, σ is the activation function, and $\odot$ is the dot-multiplication operation symbol. W_{FC} is the fully connected network, and c'_t is the integrated output information. Combining the above, the study addresses the shortcomings of the traditional Seq2Seq dialogue generation model in capturing longer textual content with LTD, and proposes a method to improve and optimize the Seq2Seq model by using the AM and BERT models. Through this a new HCID model is established. In order to enhance the dialog model's capacity for emotion awareness and give consumers interaction services of a better dialog quality, the study employs this model as the foundation for including EIs in further research.

B. CONSTRUCTION OF A GENERATIVE MODEL FOR HUMAN-COMPUTER INTERACTION DIALOGUE WITH EMBEDDED EMOTIONAL INFORMATION

The study implements HCID generation using AM as well as BERT to improve the Seq2Seq model. To further improve the intelligence of HCID, it embed EI on the basis of the constructed dialogue generation method. However, in the traditional research work on emotional dialogue generation, the model often loses the grammatical correctness and content relevance of the response to a certain extent because of the inclusion of emotional factors [28], [29], [30]. To solve this problem, an ECF model based on Bidirectional long and short-term memory network (BiLSTM) and attention mechanism is proposed. Only classifying emotions to achieve dialogue can not give better feedback to emotions in dialogue. In view of this, this study introduces ECF model and uses EI at word level and sentence level to enhance the model's attention to EI. The attention-BiLSTM (AT-BiLSTM) ECF model designed by the research is shown in Figure 5.

As shown in Figure 5, the BiLSTM layer contains two LSTMs passed in opposite directions, and the model contains an input layer, a WV layer, a BiLSTM layer, an AL, and an output layer. The WV layer maps each word into the form of a WV, while the BiLSTM layer extracts semantic features from the text. The AL calculates the weights of the word-level features at each time step and obtains the sentence-level feature representation by weighted summation. The output layer uses the sentence-level features for sentiment classification [31], [32], [33]. Equation (8) illustrates the procedure by which every word in the WV layer is converted into a WV.

$$\vec{c}_i = W^k v^i \tag{8}$$

In Equation (8), v^i is a one-hot vector of size $|V|$, $\vec{c}_i$ is the WV after transformation, and W^k is a one-parameter matrix obtained through training and learning. The study introduces a gate mechanism in the BiLSTM layer to control the extent to which each LSTM unit retains historical information and memorizes the current input to extract important features. Weights are assigned to the corresponding WVs in the AL. The final sentence representation used for classification is shown in Equation (9).

$$h^* = tanh\left(H\left[softmax\left(w^T tanh\,(H)\right)^T\right]\right) \tag{9}$$

In Equation (9), H is the set of vectors input from the previous layer, w^T is the transpose of the parameter vectors obtained from training and learning, and h^* is the sentence output from the AL for classification. In addition, to make the emotion reflection of the model coherent, the study adds an emotion label of the input utterance to the input of the encoder. Meanwhile, the study splices the target vector in the model context vector for controlling the output emotion of the decoder. After introducing the ECF model as well as emotion labeling, the study proposes HEC to further enhance the model's focus on EI. The HCID model based on Seq2Seq and EI proposed by the study is shown in Figure 6.

As shown in Figure 6, the study divides the model into two different modules, which are hierarchical sentiment-based encoding module and EC-based decoding module. In the encoding module, the study introduces both word-level and sentence-level EI into the encoding module to further improve the dialog quality. In the decoding module, the study combines the loss function of traditional Seq2Seq, the specified EC, and the emotional polarization term (EPT) to construct the EC. Compared to the traditional dialog generation, all the studies on emotional dialog generation need to provide additional textual EIs. Based on the granularity of the EIs, they can be categorized into the word-level EIs and the sentence-level EIs. The current most of the studies are only based on one of them as the textual EI, this is due to the fact that too much EI will affect the model's understanding of the content of the text itself, which can lead to errors in the content and syntax of

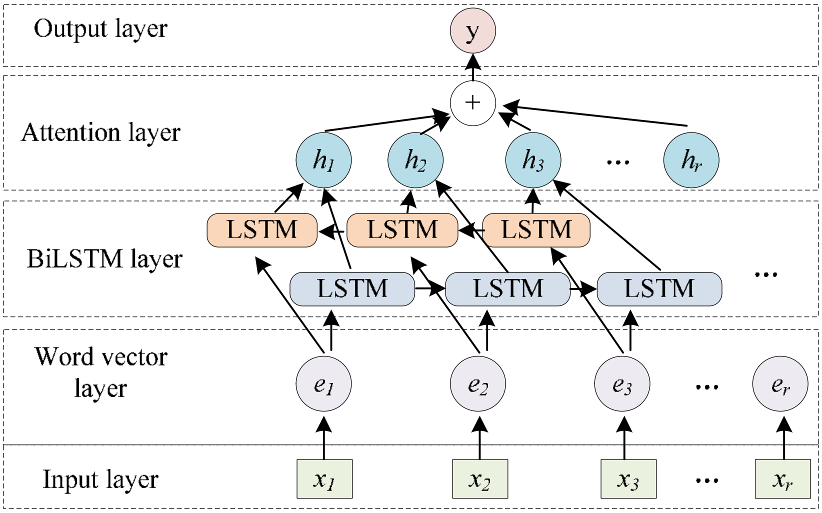


FIGURE 5. ECF model based on BiLSTM and attention mechanism.

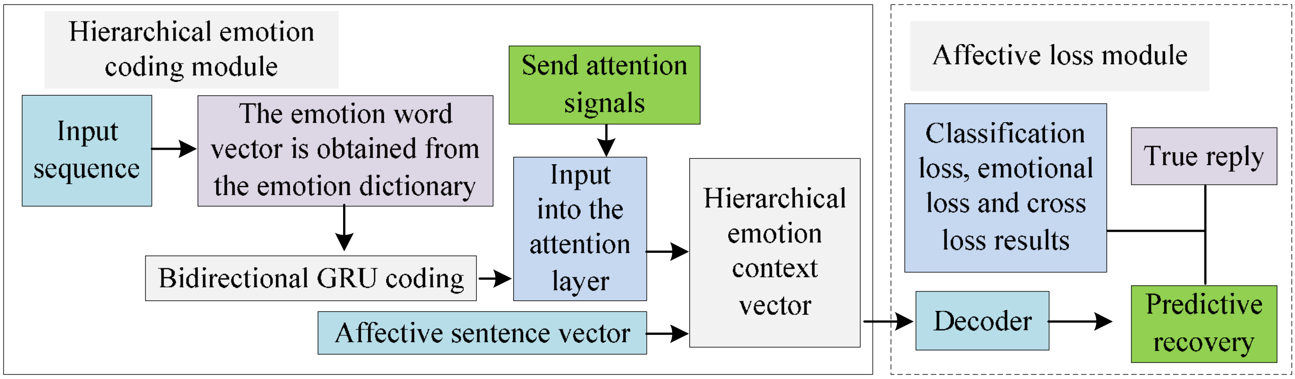

Hierarchical emotion coding module
Input sequence
The emotion word vector is obtained from the emotion dictionary
Send attention signals
Input into the attention layer
Bidirectional GRU coding
Hierarchical emotion context vector
Affective sentence vector
Affective loss module
Classification loss, emotional loss and cross loss results
True reply
Decoder
Predictive recovery

FIGURE 6. Human-computer interactive dialogue model based on Seq2Seq and emotional information.

the replies [34], [35], [36]. For this reason, the study encodes each of the two EIs independently and hierarchically at the encoding stage, resulting in hierarchical sentiment context vectors. Subsequently, the decoder is employed to generate the corresponding dialog reply content, with each word being generated word by word. The sentiment WVs and sentence vectors are shown in Figure 7.

As shown in Figure 7, emotion WV focuses on capturing the EI carried by a single word, such as ‘‘happy’’, ‘‘sad’’ and other emotion words. These emotional words play an important role in understanding the emotional tendency of the whole text. The emotion sentence vector focuses on capturing the EI expressed by the whole sentence, and obtains the emotion representation of the whole sentence by considering the semantic relationship of all the words in the sentence and the sentence structure. In contrast to traditional Word2Vec word embedding vectors, which lack a sentiment factor, the study employed an external sentiment dictionary, namely the HowNet sentiment dictionary, to provide EI for WVs. The lexicon categorizes emotion words into several emotion categories such as happiness, anger, sadness, fear, and worry. The lexicon also provides a detailed analysis of the semantic features of emotion words, which includes semantics, emotion polarity, and emotion intensity. The study classifies all the words in a sentence into positive emotion words, non-emotion words and negative emotion words according to HowNet dictionary, which are represented by 1, 0, and -1 respectively. Finally, Word2Vec word embedding vectors are spliced with sentiment polarity vectors to obtain sentiment WVs. However, the use of sentiment WVs in isolation is sufficient to enhance the emotional richness of the responses. In order for the model to comprehend the global EI of the sentences, the study incorporates sentiment sentence vectors. The study first labels all the texts into prescribed emotion categories. The study classified the dataset into six broad emotion categories, namely like, sad, delayed, angry, happy, and other. After labeling the data sentiment categories, the study represents the sentiment vectors through a discrete One-Hot method. In the model, the user-specified emotion category vectors are first converted into high-dimensional abstract emotion representations, which are then combined with word-level emotion

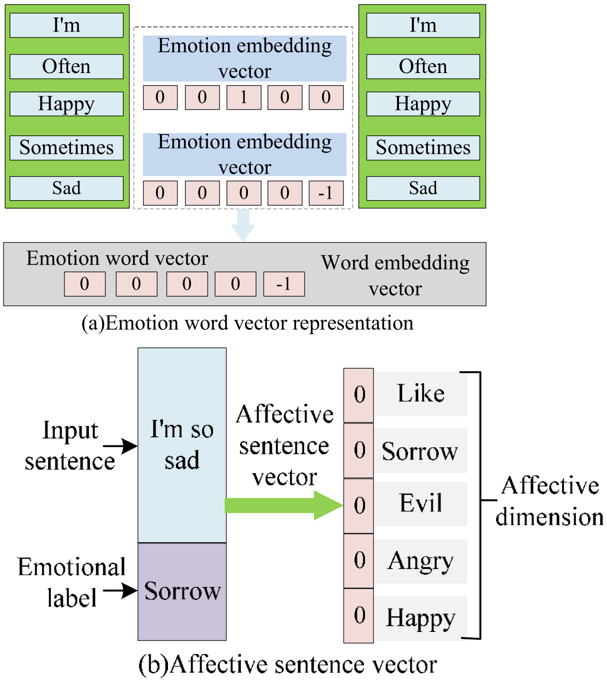


FIGURE 7. Schematic diagram of emotional word vector and emotional sentence vector.

attention vectors to generate hierarchical emotion context vectors carrying the specified emotion categories. The candidate hierarchical emotion context vectors are shown in Equation (10).

$$\tilde{c}_t = relu\left(W_c \cdot c_t + W_E \cdot E_i\right) \tag{10}$$

In Equation (10), $\tilde{c}_t$ is the candidate hierarchical emotion context vector, W_c and W_E are both model parameters, and E_i is the user-specified emotion category vector. The generated hierarchical emotion context vector carrying the specified emotion category is denoted as E_{c_t}, which is computed as shown in Equation (11).

$$E_{c_t} = \alpha \circ c_t + (1 - \alpha) \circ \tilde{c}_t \tag{11}$$

In Equation (11), $\circ$ is the logic operation and α is the gating parameter. The EC of the research design is shown in Equation (12).

$$L\left(\theta\right) = \lambda_1 \cdot L_{s2s}\left(\theta\right) + \lambda_2 \cdot L_{EC}\left(\theta\right) + \lambda_3 \cdot L_{EP}\left(\theta\right) \tag{12}$$

In Equation (12), θ is the model parameter, λ_1, λ_2, λ_3 is the weight coefficient of loss, and $L_{s2s}\left(\theta\right)$, $L_{CE}\left(\theta\right)$ and $L_{EP}\left(\theta\right)$ are the traditional Seq2Seq model loss function, CE loss function and EPT loss function, respectively. The EC loss function is the error between the sentiment categories of the predicted responses and the sentiment categories of the true responses, and the study uses cross entropy as its loss function. The primary objective of EPT is to augment the emotional depth of the model responses. Consequently, the study employs the inverse of the proportion of emotionally polar words as EPT. Combining the above, the study introduces the ECF model as well as HEC and EC into the dialog generation model. As a result, it constructs an intelligent HCID model based on the Seq2Seq model and EI.

III. RESULTS

A series of experiments are designed to test the performance of the research design model and the improvement effect on the technique. Five datasets are used as test sets in the experiments, namely, CMDC, CECG, SST2, SUBI, and IMDB. The CMDC and CECG datasets are sentiment dialog corpus datasets in English and Chinese, respectively. SST2 is commonly used for single-sentence-based classification tasks, and SUBI is a subjective dataset used for sentiment analysis, with 5,000 positive and negative samples each. The IMDB dataset contains 50,000 movie review data that have been labeled with sentiment polarity, with 25,000 positive and negative review samples each.

A. EFFECT OF SEQ2SEQ DIALOG GENERATION MODEL IMPROVEMENT

Aiming at the problem that the traditional Seq2Seq dialog generation model tends to lose the LTD within the text, it is proposed to improve it by utilizing the BERT model. At the same time combining the BERT output vectors with the contextual information by using Fu's method. The study examines the BERT-Seq2Seq-Fu created by the research in contrast with the fundamental Seq2Seq model, using AM-Seq2SEquation, in order to assess the benefits and rationale of the enhanced approach. The comparative analysis experiment is divided into two subjective and objective evaluations. In the subjective evaluation, the study invites five researchers in the natural language processing direction to assess the quality of the responses according to a scale of 0∼5. 0∼1 indicates that the responses are grammatically incorrect and semantically irrelevant. 2∼3 indicates that there is no major problem with the grammar but the content is more generalized. 4∼5 indicates that the responses have high semantic relevance, are grammatically correct, are naturally fluent, and contain a wealth of information. The study also introduces the Kappa score to assess the reliability of the experts' scoring results. The subjective evaluation results of the three models are shown in Figure 8. The objective evaluation indicators include bilingual evaluation understudy -1 (BLEU-1), Distinct-1, Distinct-2, and embedding average score (EAS). The response's quality is determined by how near BLEU-1's value is to 1, which runs from 0 to 1. Indicators called Distinct-1 and Distinct-2 are used to evaluate the diversity of responses. The greater the value, the more information-rich the response content. EAS is used to measure the similarity of content between sentences, and higher values indicate higher correlation between texts. The objective evaluation results are shown in Table 1.

In Figure 8(a), in the Chinese dataset, in terms of the proportion of scores 0∼1, the proportion of the BERT-Seq2Seq-Fu model proposed by the study is lower, with a value of 18.46%, which suggests that the introduction of

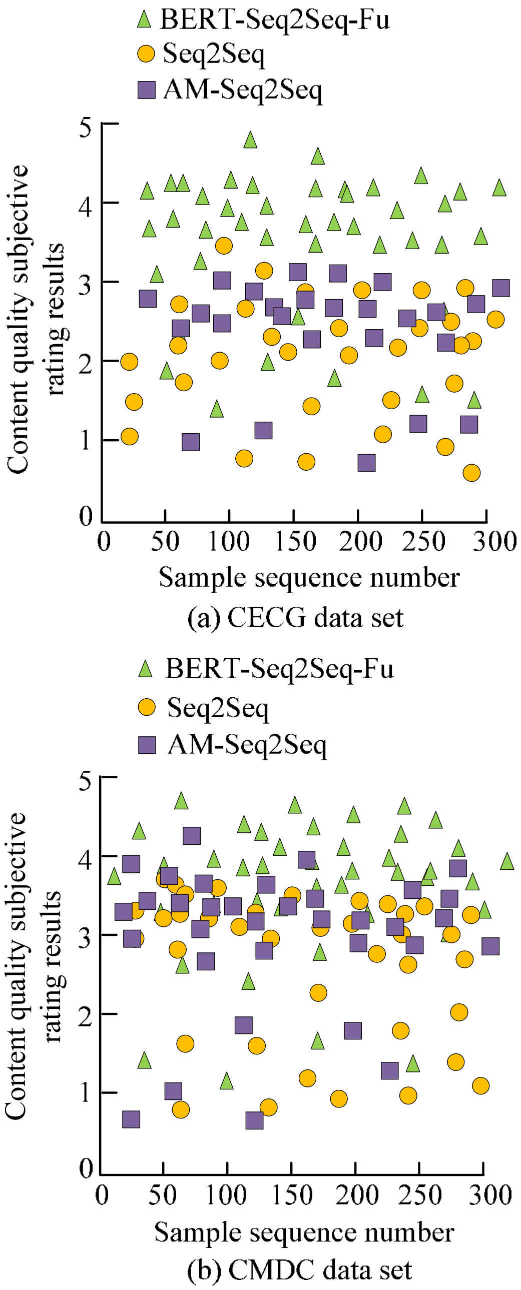


FIGURE 8. Comparison of the subjective evaluation results of the improved Seq2Seq model with the other models.

the BERT model can reduce the generation of boring and generic reply content to some extent. In the interval of score 4 to 5, BERT-Seq2Seq-Fu has the highest proportion with a value of 49.99%, which indicates that the improvement of the research methodology improves the FEC of the encoder. In Figure 8(b), in the English dataset, the scoring results of BERT-Seq2Seq-Fu are concentrated between 3 and 5, which is significantly higher than the other models. This indicates that the research design models all have strong portability and perform well under texts in different languages. In addition, the Kappa scores in both datasets are above 0.5, which indicates that the results of this evaluation are more reliable.

TABLE 1. Comparison of the objective evaluation results of each dialogue generation model.

Data set	Project	Distinc-1	Distinc-2	EAS	BLEU
CECG	BERT-Seq2Seq-Fu	0.070	0.196	0.267	0.297
	Seq2Seq	0.045	0.188	0.255	0.201
	AM-Seq2Seq	0.042	0.185	0.247	0.196
CMDC	BERT-Seq2Seq-Fu	0.068	0.199	0.247	0.286
	Seq2Seq	0.051	0.192	0.220	0.189
	AM-Seq2Seq	0.044	0.190	0.219	0.188

In Table 1, BERT-Seq2Seq-Fu gets significantly higher than the other two models on all metrics. The mean values of Distinc-1, Distinc-2, EAS, and metrics in the Chinese and English datasets are 0.069, 0.198, 0.257, and 0.292, respectively. This further demonstrates that the study utilizes the Fu method to better filter out useless noise and improve the quality of responses. Moreover, Table 1 shows that the research method is better than the single Seq2Seq model in all aspects, which indicates that the research introduces the BERT improvement in the model can effectively improve the quality of responses and avoid generating repetitive, boring and generic responses.

B. PERFORMANCE ANALYSIS OF SENTIMENT CLASSIFICATION MODELS BASED ON AM AND BILSTM

The study constructs an ECF model AT-BiLSTM using AM and BiLSTM, which is applied to the EI embedding process of the dialog model. The study creates two tests to evaluate the effectiveness of this classification model and identify the combination of its hyper-parameters. In the hyper-parameter combination analysis experiment, the study determines the classification performance of the corresponding parameter combination model based on the loss rate and accuracy rate. The results of the hyper-parameter combination analysis experiment are shown in Figure 9. The study carries out a comparison analysis experiment with many benchmark models in order to further assess the performance of the classification model. The comparison models include LSTM, Bi-LSTM, convolutional neural network (CNN), convolutional neural network-recurrent neural network (CNN-RNN). Figure 10 displays the results of the comparison.

When the WV dimension is set to 50, the model's accuracy reaches its maximum of 88.47%, and its lowest loss rate of 0.2674 is observed in Figure 9(a). As the value of Epoch increases, the accuracy of the model tends to increase and subsequently decrease as shown in Figure 9(b). When the training iteration reaches the seventh time, the accuracy achieves its highest at 90.14% and the loss rate is at its lowest at 0.2455. Furthermore, the model in Figure 9(b) takes the least amount of time and has the maximum accuracy when Dropout is set to 0.2. Based on the above content, the study

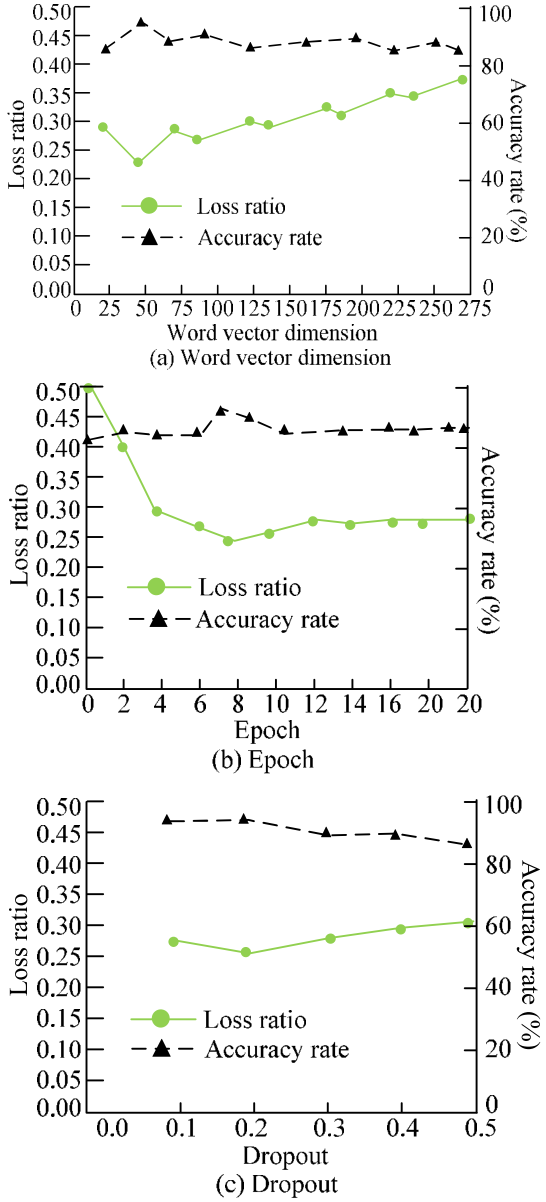


FIGURE 9. Analysis and determination of the hyper-parameter combination of the ECF model.

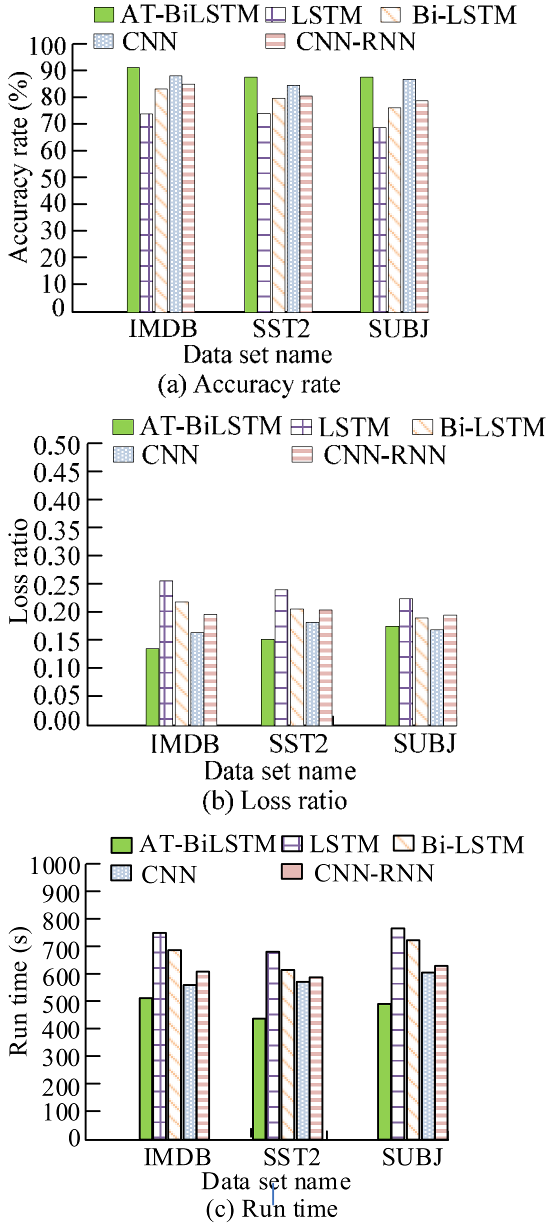


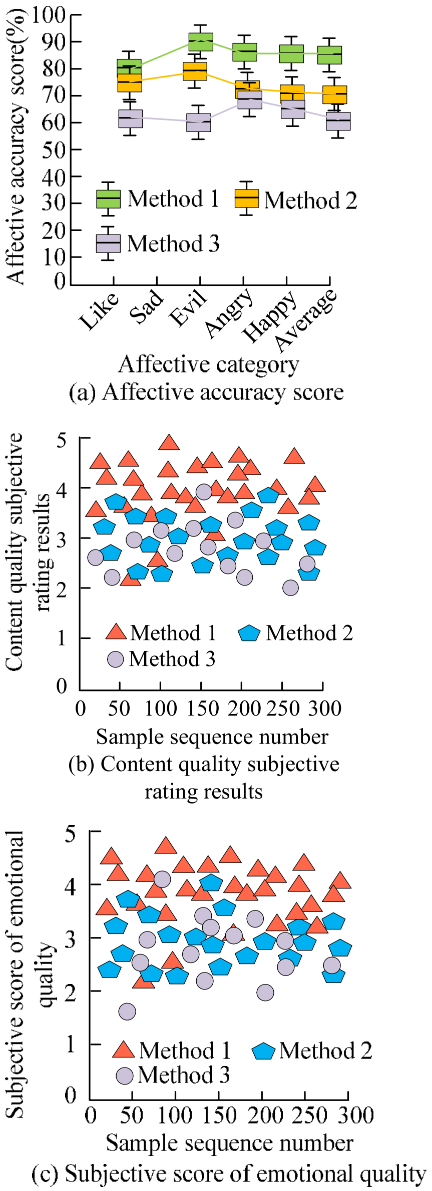
FIGURE 10. Comparison of classification effect of several ECF models.

sets the WV dimension of the classification model to 50, Epoch count to 7, and Dropout to 0.2.

In Figure 10(a), the classification accuracy of AT-BiLSTM on the three datasets are 89.97%, 90.15%, and 91.44%, respectively, and its average accuracy reaches over 90%. Compared with other models, the classification performance of the research model on the three datasets is significantly better. In Figure 10(b), the average loss rate of AT-BiLSTM on the three datasets is 0.25, which is significantly lower compared to other models. The study design model's running time in Figure 10(c) is 689.47 s, which is shorter than that of the other models. This suggests that the addition of AM can successfully increase the model's training speed in the classification job.

C. PERFORMANCE COMPARISON OF EMBEDDING MODELS FOR HIERARCHICAL EMOTION CODING VS. SINGLE-EMOTION CODING

The study combines sentence-level emotion vectors with word-level emotion vectors to obtain HEC and introduces it into the dialog model to realize intelligent emotion expression for HCID. To verify that the research design HEC can effectively solve the problems in single-emotion coding (SEC). The study adds emotional precision (EP) as the evaluation index of emotional level on the basis of the evaluation index applied in the experiments in Section III-A. The comparison models are HEC-based dialog method (Method 1), sentence-level emotion coding-based dialog method (Method 2), and word-level emotion coding-based dialog method (Method 3).



(a) Affective accuracy score

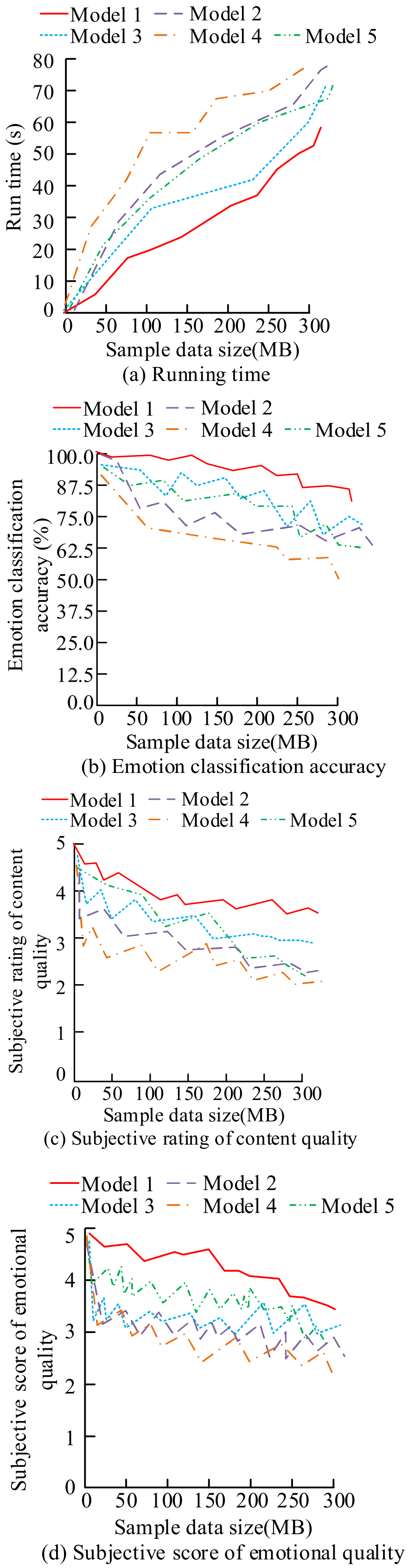
(b) Content quality subjective rating results

(c) Subjective score of emotional quality

FIGURE 11. Comparison of emotional accuracy and subjective evaluation results of single and hierarchical emotion coding.

The study first evaluates the content quality and diversity of responses using Distinct-1, Distinct-2, and EAS. The evaluation results of the three methods for sentiment dialog generation are shown in Table 2. The results of the EP scores of the three methods as the amount of data changes and the subjective evaluation results under each sentiment label are shown in Figure 11.

In Table 2, on the three evaluation indicators, Method 1 has the best overall effect. Compared with Method 2 and Method 3, its average value in the two indicators of reply diversity, Distinct-1 and Distinct-2, is 0.062 and 0.218, respectively. In the EAS indicator, the average value of Method 1 is 0.297. Taken together, it can be noted that the



(a) Running time

(b) Emotion classification accuracy

(c) Subjective rating of content quality

(d) Subjective score of emotional quality

FIGURE 12. Comparative analysis of emotional response performance in the five models.

TABLE 2. The emotional dialogue of the three methods generates the comparison of the objective evaluation results.

Affective	Distinct-1			Distinct-2			EAS		
category	Method 1	Method 2	Method 3	Method 1	Method 2	Method 3	Method 1	Method 2	Method 3
Like	0.066	0.048	0.041	0.227	0.186	0.191	0.288	0.247	0.232
Sad	0.061	0.044	0.039	0.212	0.189	0.192	0.301	0.251	0.231
Evil	0.064	0.047	0.044	0.220	0.192	0.187	0.298	0.252	0.245
Anger	0.060	0.048	0.042	0.213	0.187	0.186	0.300	0.248	0.239
Delighted	0.058	0.049	0.043	0.217	0.193	0.179	0.299	0.249	0.245
Average	0.062	0.047	0.042	0.218	0.189	0.187	0.297	0.249	0.238

study utilizes the replies of the dialogue generation method constructed by HEC to be closer to the real replies in terms of content, and improves the the quality of replies of the dialog method of SEC.

In Figure 11(a), in general, the EP scores of Method 1 have been consistently higher than the other two methods. The average EP score for Method 1 is 80.47%. In Figure 11(b), in the content quality score, the score of Method 1 fluctuates centrally in the range of 3.5 to 4.7, which is a higher average score compared to the other two methods. In Figure 11(c), the score of Method 1 is concentrated to fluctuate in the range of 3.2 to 4.4, with an average score of 4.35. Taken together, it can be seen that the HEC designed by the study can effectively avoid the problems of grammatical errors and low content relevance that are easily found in SEC, while ensuring that the reply content is of high emotional quality.

D. PERFORMANCE ANALYSIS OF HUMAN-COMPUTER INTERACTION DIALOGUE MODEL BASED ON SEQ2SEQ MODEL AND EMOTIONAL INFORMATION

In order to more comprehensively examine the performance of the research-designed HCID model (Model 1) based on Seq2Seq model and EI, the study compared and analyzed it with several of the more popular emotion interaction dialogue models. The comparison models include the emotion interaction dialog model based on generative adversarial network (Model 2), the emotion dialog model based on emotion lexicon and Transformer model (Model 3), the emotion dialog model based on multi-modal fusion (Model 4), and the emotion interaction dialog model based on reinforcement learning (Model 5). The results of the runtime, ECF accuracy, and subjective scoring of response quality for the five models as the amount of response data increases are shown in Figure 12. To compare the models more comprehensively, the study compares the Distinct-1, Distinct-2, EAS, BLEU, and EP scoring results of the five models, and the comparison of the five models for each metric under two different data sets is shown in Table 3.

In Figure 12(a), the running time variation curve of Model 1 has been below the other four models, and its average value is 42.1s, which indicates that Model 1 has high real-time performance and can respond to the user's needs in a timely manner in the dialog replies. In Figure 12(b), the ECF accuracy of Model 1 has been above 80%, which is higher than the other four models. The average subjective score of dialog content quality of Model 1 is 4.35, and the average subjective scores of Models 2, 3, 4, and 5 are 3.58, 3.89, 4.01, and 3.95, which are not as good as that of Model 1. This shows that the subjective quality scores of emotion of Model 1 are significantly higher than those of the other methods.

TABLE 3. Comparison of the objective quality evaluation results of the five model responses.

Project	Distinct-1	Distinct-2	EAS	BLEU	EP(%)
Model 1	0.068	0.204	0.297	0.283	80.452
Model 2	0.058	0.182	0.257	0.266	72.353
Model 3	0.061	0.189	0.269	0.261	70.458
Model 4	0.054	0.176	0.268	0.263	75.821
Model 5	0.052	0.183	0.254	0.259	76.008

As shown in Table 3, the Distinct-1, Distinct-2, EAS, BLEU, and EP scores of Model 1 are 0.068, 0.204, 0.297, 0.283, and 80.452, respectively, which shows that the research design model is in a more ideal state in terms of content quality and emotional quality in dialog responses. Meanwhile, comparing the other four models, it can be noted that Model 1 has a greater advantage in each index, and the values of each index are higher than those of the other four models. As a result, the research-designed model is able to realize more intelligent emotional dialogues and provide users with a good interactive experience.

IV. DISCUSSION AND CONCLUSION

Aiming at the problem that the inclusion of EI in the current HCID system leads to deviations in response syntax as well as sentence relevance, the study proposed a new intelligent HCID model. The study first constructed a dialogue generation model using the improved Seq2Seq model, and introduced EC and ECF models with HEC based on this model to realize human-computer emotional dialogue. In the experiments to analyze the effect of the improvement of the Seq2Seq dialog generation model, the proportion of the BERT-Seq2Seq-Fu model with a value of 0 to 1 in the subjective score was the lowest among all models, with a percentage of 18.46%. The proportion of BERT-Seq2Seq-Fu with a value of 49.99% was the highest in the interval of scores from 4 to 5, which indicated that the improvement of the research methodology improves the FEC of the encoder. The improved method designed by the study can effectively improve the quality of dialog in the dialog generation model.

In the analysis of the ECF model, the classification accuracy of AT-BiLSTM on the three datasets were 89.97%, 90.15%, and 91.44%, respectively, and its average accuracy reached more than 90%. This indicated that the classification model proposed in the study was able to recognize different textual emotions effectively. In addition, the average EP score of HEC was 80.47%, which fluctuated centrally in the range of 3.5 to 4.7 in the content quality score. The values of the indicators of HEC were significantly higher than those of SEC, and the results of Distinct-1, Distinct-2, EAS, BLEU. Moreover, EP scores of Model 1 were 0.068, 0.204, 0.297, 0.283, and 80.452, respectively, which indicated that the research-designed model was in a more satisfactory state in terms of both content quality and emotional quality in dialog responses. In the future research process, the emotional dialog of the machine in the process of multi-round interaction can be further considered to further improve the performance of the model.

REFERENCES

[1] Y. Hu, W. Lu, J. Wei, J. Xu, and M. Ma, "A watermark detection scheme based on non-parametric model applied to mute machine voice," *Multimedia Tools Appl.*, vol. 82, no. 29, pp. 44763–44782, Dec. 2023, doi: 10.1007/s11042-023-15572-x.

[2] A. Joukhadar, N. Ghneim, and G. Rebdawi, "Impact of using bidirectional encoder representations from transformers (BERT) models for Arabic dialogue acts identification," *Ingénierie des systèmes d Inf.*, vol. 26, no. 5, pp. 469–475, Oct. 2021, doi: 10.18280/isi.260506.

[3] B. Mcmillan, J. Myers, A. Nguyen, D. Robinson, and M. Kennard, "Analysis and comparison of natural language processing algorithms as applied to Bitcoin conversations on social media," *J. Investing*, vol. 31, no. 2, pp. 38–59, Jan. 2022, doi: 10.3905/joi.2021.1.213.

[4] S. Kobayashi and P. Hartono, "Eye contact as a new modality for man–machine interface," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 3, pp. 42–49, Apr. 2023, doi: 10.14569/ijacsa.2023.0140306.

[5] Y. T. Li, D. Liang, G. Chen, and W. Tang, "Research on man-machine conversation mode of service robot based on voiceprint recognition and image recognition technology," *CCSB*, vol. 22, no. 2, pp. 89–93, Feb. 2022, doi: 10.1109/ccsb58128.2022.00023.

[6] D. A. A. Dawood and B. A. Hussain, "Machine learning for single and complex 3D head gestures: Classification in human–computer interaction," *Webology*, vol. 19, no. 1, pp. 1431–1445, Jan. 2022, doi: 10.14704/web/v19i1/web19095.

[7] S. Somani, A. J. Russak, F. Richter, S. Zhao, A. Vaid, F. Chaudhry, J. K. De Freitas, N. Naik, R. Miotto, G. N. Nadkarni, J. Narula, E. Argulian, and B. S. Glicksberg, "Deep learning and the electrocardiogram: Review of the current state-of-the-art," *EP Europace*, vol. 23, no. 8, pp. 1179–1191, Aug. 2021, doi: 10.1093/europace/euaa377.

[8] M. H. Saleem, J. Potgieter, and K. M. Arif, "Correction to: Automation in agriculture by machine and deep learning techniques: A review of recent developments," *Precis. Agricult.*, vol. 22, no. 6, pp. 2092–2094, Dec. 2021, doi: 10.1007/s11119-021-09824-9.

[9] Y.-P. Ruan and Z.-H. Ling, "Emotion-regularized conditional variational autoencoder for emotional response generation," *IEEE Trans. Affect. Comput.*, vol. 14, no. 1, pp. 842–848, Jan. 2023, doi: 10.1109/TAFFC.2021.3073809.

[10] T. Law, M. Chita-Tegmark, and M. Scheutz, "The interplay between emotional intelligence, trust, and gender in human–robot interaction: A vignette-based study," *Int. J. Social Robot.*, vol. 13, no. 2, pp. 297–309, Apr. 2021, doi: 10.1007/s12369-020-00624-1.

[11] A. S. Raamkumar and Y. Yang, "Empathetic conversational systems: A review of current advances, gaps, and opportunities," *IEEE Trans. Affect. Comput.*, vol. 14, no. 4, pp. 2722–2739, Oct. 2022, doi: 10.1109/TAFFC.2022.3226693.

[12] M. Firdaus, N. Thakur, and A. Ekbal, "Aspect-aware response generation for multimodal dialogue system," *ACM Trans. Intell. Syst. Technol.*, vol. 12, no. 2, pp. 1–33, Apr. 2021, doi: 10.1145/3430752.

[13] S. Qamar, H. Mujtaba, H. Majeed, and M. O. Beg, "Relationship identification between conversational agents using emotion analysis," *Cogn. Comput.*, vol. 13, no. 3, pp. 673–687, May 2021, doi: 10.1007/s12559-020-09806-5.

[14] Y. Zhang, A. Jia, B. Wang, P. Zhang, D. Zhao, P. Li, Y. Hou, X. Jin, D. Song, and J. Qin, "M3GAT: A multi-modal, multi-task interactive graph attention network for conversational sentiment analysis and emotion recognition," *ACM Trans. Inf. Syst.*, vol. 42, no. 1, pp. 1–32, Aug. 2023, doi: 10.1145/3593583.

[15] C.-Y. Li, Q. Zhao, N. Herencsar, and G. Srivastava, "The design of mobile distance online education resource sharing from the perspective of man-machine cooperation," *Mobile Netw. Appl.*, vol. 26, no. 5, pp. 2141–2152, Oct. 2021, doi: 10.1007/s11036-021-01770-0.

[16] A. V. Savchenko and V. V. Savchenko, "Method for measurement the intensity of speech vowel sounds flow for audiovisual dialogue information systems," *Meas. Techn.*, vol. 65, no. 3, pp. 219–226, Sep. 2022, doi: 10.1007/s11018-022-02072-x.

[17] Z. Chu, "Effects of digital media integrated reciprocal teaching on students' reading ability and motivation," *Revista de Cercetare si Interventie Sociala*, vol. 73, pp. 299–311, Jun. 2021, doi: 10.33788/rcis.73.19.

[18] B. Guo, H. Wang, Y. Ding, W. Wu, S. Hao, Y. Sun, and Z. Yu, "Conditional text generation for harmonious human–machine interaction," *ACM Trans. Intell. Syst. Technol.*, vol. 12, no. 2, pp. 1–50, Apr. 2021, doi: 10.1145/3439816.

[19] X. Li, M. Jiang, Y. Du, X. Ding, C. Xiao, Y. Wang, Y. Yang, Y. Zhuo, K. Zheng, X. Liu, L. Chen, Y. Gong, X. Tian, and X. Zhang, "Self-healing liquid metal hydrogel for human–computer interaction and infrared camouflage," *Mater. Horizons*, vol. 10, no. 8, pp. 2945–2957, Jul. 2023, doi: 10.1039/d3mh00341h.

[20] M. Czerwinski, J. Hernandez, and D. McDuff, "Building an AI that feels: AI systems with emotional intelligence could learn faster and be more helpful," *IEEE Spectr.*, vol. 58, no. 5, pp. 32–38, May 2021, doi: 10.1109/MSPEC.2021.9423818.

[21] G. Tu, J. Wen, C. Liu, D. Jiang, and E. Cambria, "Context- and sentiment-aware networks for emotion recognition in conversation," *IEEE Trans. Artif. Intell.*, vol. 3, no. 5, pp. 699–708, Oct. 2022, doi: 10.1109/TAI.2022.3149234.

[22] S. Z. Razavi, L. K. Schubert, K. van Orden, M. R. Ali, B. Kane, and E. Hoque, "Discourse behavior of older adults interacting with a dialogue agent competent in multiple topics," *ACM Trans. Interact. Intell. Syst.*, vol. 12, no. 2, pp. 1–21, Jun. 2022, doi: 10.1145/3484510.

[23] W.-H.-S. Tsai, Y. Liu, and C.-H. Chuan, "How chatbots' social presence communication enhances consumer engagement: The mediating role of parasocial interaction and dialogue," *J. Res. Interact. Marketing*, vol. 15, no. 3, pp. 460–482, Jul. 2021, doi: 10.1108/jrim-12-2019-0200.

[24] I. Dubovi and I. Tabak, "Interactions between emotional and cognitive engagement with science on Youtube," *Public Understand. Sci.*, vol. 30, no. 6, pp. 759–776, Aug. 2021, doi: 10.1177/0963662521990848.

[25] A. Catala, H. Gijlers, and I. Visser, "Guidance in storytelling tables supports emotional development in kindergartners," *Multimedia Tools Appl.*, vol. 82, no. 9, pp. 12907–12937, Apr. 2023, doi: 10.1007/s11042-022-14049-7.

[26] A. C. da Silveira, E. C. Rodrigues, E. B. Saleme, A. Covaci, G. Ghinea, and C. A. S. Santos, "Thermal and wind devices for multisensory human–computer interaction: An overview," *Multimedia Tools Appl.*, vol. 82, no. 22, pp. 34485–34512, Sep. 2023, doi: 10.1007/s11042-023-14672-y.

[27] T. Kosch, J. Karolus, J. Zagermann, H. Reiterer, A. Schmidt, and P. W. Woúniak, "A survey on measuring cognitive workload in human–computer interaction," *ACM Comput. Surv.*, vol. 55, no. 13s, pp. 1–39, Dec. 2023, doi: 10.1145/3582272.

[28] C. A. Liang, S. A. Munson, and J. A. Kientz, "Embracing four tensions in human–computer interaction research with marginalized people," *ACM Trans. Computer-Human Interact.*, vol. 28, no. 2, pp. 1–47, Apr. 2021, doi: 10.1145/3443686.

[29] H. Liu, T. Liu, Z. Zhang, A. K. Sangaiah, B. Yang, and Y. Li, "ARHPE: Asymmetric relation-aware representation learning for head pose estimation in industrial human–computer interaction," *IEEE Trans. Ind. Informat.*, vol. 18, no. 10, pp. 7107–7117, Oct. 2022, doi: 10.1109/TII.2022.3143605.

[30] H. Zhang, D. Zhang, Z. Wang, G. Xi, R. Mao, Y. Ma, D. Wang, M. Tang, Z. Xu, and H. Luan, "Ultrastretchable, self-healing conductive hydrogel-based triboelectric nanogenerators for human–computer interaction," *ACS Appl. Mater. Interfaces*, vol. 15, no. 4, pp. 5128–5138, Feb. 2023, doi: 10.1021/acsami.2c17904.

[31] M. Gori, S. Price, F. N. Newell, N. Berthouze, and G. Volpe, "Multisensory perception and learning: Linking pedagogy, psychophysics, and human–computer interaction," *Multisensory Res.*, vol. 35, no. 4, pp. 335–366, Apr. 2022, doi: 10.1163/22134808-bja10072.

[32] R. Zhang, C. Jiang, S. Wu, Q. Zhou, X. Jing, and J. Mu, "Wi-Fi sensing for joint gesture recognition and human identification from few samples in human–computer interaction," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 7, pp. 2193–2205, Jul. 2022, doi: 10.1109/JSAC.2022.3155526.

[33] A. Rapp, "In search for design elements: A new perspective for employing ethnography in human–computer interaction design research," *Int. J. Human–Computer Interact.*, vol. 37, no. 8, pp. 783–802, May 2021, doi: 10.1080/10447318.2020.1843296.

[34] T. Kosch, R. Welsch, L. Chuang, and A. Schmidt, "The placebo effect of artificial intelligence in human–computer interaction," *ACM Trans. Computer-Human Interact.*, vol. 29, no. 6, pp. 1–32, Dec. 2022, doi: 10.1145/3529225.

[35] H. Kivijärvi and K. Pärnänen, "Instrumental usability and effective user experience: Interwoven drivers and outcomes of human–computer interaction," *Int. J. Human–Computer Interact.*, vol. 39, no. 1, pp. 34–51, Jan. 2023, doi: 10.1080/10447318.2021.2016236.

[36] S. N. Amin, P. Shivakumara, T. X. Jun, K. Y. Chong, D. L. L. Zan, and R. Rahavendra, "An augmented reality-based approach for designing interactive food menu of restaurant using Android," *Artif. Intell. Appl.*, vol. 1, no. 1, pp. 26–34, Oct. 2022, doi: 10.47852/bonviewaia2202354.

JINHUANG CHEN was born in Guangdong, China, in 1988. He received the bachelor's degree from Guangzhou Xinhua University, in 2012, and the master's degree from Sun Yat-sen University, in 2016.

Since 2012, he has been a Lecturer with the School of Information and Intelligence Engineering, Guangzhou Xinhua University. He has published six papers and holds three patents. His research interests include embedded technology and intelligent control.

PEIQI TU was born in Guangdong, China, in 1993. She received the B.S. degree in biomedical engineering from Guangzhou Medical University, Guangdong, in 2015.

Since 2016, she has been an Experimental Teacher with the School of Biomedical Engineering, Guangzhou Xinhua University. In 2021, she obtained the title of experimenter. She has five articles. Her research interests include electronics, sensing technology, computer-aided diagnosis and treatment technology research, and artificial intelligence.

ZEMIN QIU was born in Guangdong, in 1983. She received the bachelor's degree from Guangdong University of Technology, in 2007, and the master's degree from Sun Yat-sen University, in 2011.

Since 2013, she has been a Lecturer with the School of Information and Intelligent Engineering, Guangzhou Xinhua University. She has published 11 papers and holds two patents. Her research interests include video object tracking and human action recognition.

ZHIJUN ZHENG received the M.Eng. degree in control science and engineering from Guangdong Polytechnic Normal University, Guangzhou, China.

She currently works with the School of Information and Intelligent Engineering, Guangzhou Xinhua University, China. Her research interests include image intelligent perception and processing and the Internet of Things.

ZHAOQI CHEN was born in Guangdong, China, in 1994. He received the bachelor's degree from Xinhua College, Sun Yat-sen University, in 2017.

Since 2017, he has been with the School of Information and Intelligence Engineering, Guangzhou Xinhua University, as a Laboratory Manager. He holds four patents and two papers. His research interest includes the Internet of Things in safety.