Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

# Research on cross-cultural empathy expression and adaptive interaction patterns in human-computer communication by integrating deep learning algorithms

Xinruo Zhang[1,✉]

[1] *School of Literature and Media, Xi'an Institute of Translation, Xi'an, Shaanxi, 710015, China*

ABSTRACT

This paper investigates human-computer communication within the framework of deep learning and identifies three key features of such interaction. A cross-cultural empathy feature aliasing model based on Graph Neural Network-Attention Mechanism-Bi-directional Gating Unit (GCN-Attention-BiGRU) is proposed, with categorical cross-entropy and L2 regularization as the loss function. By integrating IoT and deep learning, an adaptive interaction model is developed and evaluated through experiments. Results show high mean scores for empathy (4.537), relevance (4.447), and fluency (4.499) across 60 samples, indicating effective empathy feature extraction. Additionally, the proposed model demonstrates greater efficiency and adaptability compared to traditional interaction models, enhancing cross-cultural empathy in human-computer communication.

*Keywords:* graph neural network, attention mechanism, two-way gating unit, adaptive interaction mode, cross-cultural empathy expression

## 1. Introduction

Language is the key to cross-cultural communication. Speech recognition and translation technologies in human-computer interaction can help people avoid language barriers and improve communication efficiency in cross-language communication. Cross-cultural communication is not only about language learning and knowledge transfer, but also about mutual interaction and understanding [7, 11, 5, 12]. The application of human-computer interaction (HCI) technology can promote the interactivity of cross-cultural communication, and the application of HCI technology can provide rich cultural

✉ Corresponding author.
*E-mail address:* zhangxinruo21@163.com (X. Zhang).

Received 10 April 2024; Revised 22 May 2024; Accepted 21 August 2024; Published Online 20 May 2025.

DOI: 10.61091/jcmcc126-07

knowledge and play an important guiding role in cross-cultural communication [6, 19, 21].

Cross-cultural communication usually includes two needs: linguistic environment and social environment. Due to cultural background and language differences, cross-cultural communication is difficult to expand, so human-computer interaction (HCI) technology has become the main tool to help cross-cultural communication [9, 16, 17, 2]. Human-computer interaction (HCI) is a discipline that integrates computer technology with psychology, sociology and other disciplines. There are various human-computer interfaces and visualization techniques applied to enable meaningful communication between computer systems and humans, improve understanding between humans and machines, reduce the cognitive load on humans, and expand outlets [3, 20, 8, 10]. The most important feature of HCI technology is that it can better realize the communication and interaction between human and computer. Human-computer interaction technology on the one hand focuses on how to make the computer better respond to human commands [4, 1], on the other hand, it pays more attention to how to help human beings better communicate with computers directly through these tools, which has certain characteristics of humanization and intelligence, and therefore has been widely used in many fields [18, 15, 14].

Taking the dialog utterance and the dialog context information as the input of the task, and the emotion category of the current speaker as the output, a human-computer communication cross-cultural empathy feature aliasing model based on graph neural network-attention mechanism-bidirectional gating unit (GCN-Attention-BiGRU) is proposed, and the main principle of which is to utilize the natural structure of the graph to simulate the emotion propagation characteristics of the dialog in order to improve the human-computer communication The main principle is to use the natural structure of graphs to simulate the emotion propagation characteristics in dialogues to improve the cross-cultural empathy feature recognition ability of human-computer communication. Deep learning and IoT each have many advantages, and the organic integration of the two realizes the construction of adaptive interaction model. Confirm the training environment, dataset, and evaluation index, and explore the cross-cultural empathy features and adaptive interaction model in human-computer communication under deep learning theory.

## 2. Cross-cultural expressions of empathy and adaptive interaction models

### *2.1. Relevant theoretical foundations*

2.1.1. Deep learning. Deep learning theory is a complex and evolving field, which is closely related to the knowledge of multiple disciplines, and through continuous research and practice, it brings tools and methods to all walks of life and promotes the progress of science. With the improvement of big data technology and computing power, deep learning technology is also developing rapidly [13, 22]. Although deep learning is a special field of machine learning, it has a deeper and more complex hidden layer structure than traditional machine learning, so deep learning methods can more accurately learn the hidden information behind massive data. With these advantages, deep learning is widely used in image processing, language analysis, speech recognition and other fields. In recent years, it has also been gradually applied to numerous studies on air quality prediction and achieved good results.

2.1.2. Cross-cultural overview. All human behavioral activities can find their cultural origins, and the similar mental programs produced by human beings in the course of the same educational background and practical life are culture, which is an important element in distinguishing the pluralistic world in contemporary times. The globalization of the economy and the mobility of society provide the conditions for the intermingling of different cultures, changing traditional and existing cultures and creating new cultures through cultural collisions, and it is the increasing cultural exchanges and cultural collisions that have brought about the new social phenomenon of "interculturality". Globalization has brought about an intricate sociological and phenomenological situation in which social and cultural differences are strengthened or dissolved, and the ideologies and cultures of different social groups are independent of, interact with, and complement each other." Therefore, interculturality is the close integration of the best parts of a foreign culture with the main culture, thus forming an organic combination of great compatibility.

2.1.3. Empathy theory. Empathy is a complex, multidimensional and higher-order social intelligence skill that manifests itself in the form of emotional empathy [23]. During intercultural communication, empathy is often perceived and expressed spontaneously, but for computers this process is a very complex task. Empathy analysis is a technique that simulates the process of human empathy, and its purpose is to give computers the ability to express and adapt to human emotions on the basis of recognizing and understanding human emotions, thus allowing computer systems not only to perceive human emotions, but also to provide decision support services in response to individual emotional states and purposeful intentions.

### *2.2. Exploration of human-computer communication under deep learning theory*

2.2.1. Changes in the roles of communication actors. Robots are non-organic, and people instinctively hold a compartmentalized and vigilant mentality towards machine-mediated media. The current machine-mediated technology has been able to deeply learn and imitate human language, voice and tone through real-time video analysis, input algorithms, text analysis and other means, and even qualitatively improve the analysis and learning of human emotions and attitudes, etc., so that the communication barriers between humans and machines are gradually dissolved, and the power of machine discourse has been enhanced, and the massage effect is thus revealed. When robots begin to learn the content, mood, and tone of other robots, machine-to-machine "interpersonal communication" becomes possible, and this interpersonal communication may become another "source" of influence on human-robot communication. With the expansion of the boundaries of "interpersonal" communication, a new type of communication has become a closed loop, that is, human-machine-machine-human communication. The result may be a direct change in roles, with the human being no longer being the subject of communication but becoming an accessory to it, which in the long run may lead to the gradual marginalization of the human being.

2.2.2. Mixed-flow interactions for human-computer communication. Robots can be instructed to generate content that is highly compatible with human linguistic expressions, and physical robots are able to perform specific actions in addition to feeding back textual content, functioning as chats and collaborations. It has been shown that the degree of anthropomorphization and mechanization of a robot affects the level of trust in the robot, with people having a higher level of trust in social robots with a high degree of anthropomorphization, as this motivates them to subconsciously view the robot as a “partner”. Low anthropomorphism can lead to feelings of discomfort and alienation, which can reduce trust. Human-computer communication is based on the interaction logic between humans and "non-humans", but when machines begin to be good at learning and imitating everything about people, and can even become "companions", "perfect assistants", "expert consultants" and other characters after training, the humanization embodied in the communication content will become an imperfect embodiment in human-computer interaction. The robot will look for imperfect content and "correct" it in the process of being trained, so that it tends to correct in the direction that the robot thinks is "perfect". As a result of this interaction, interactive content and communication topics can be proposed by robots, and people will become not only vassals of roles, but also vassals of content, and the balance of human-computer symbiosis will be broken. Under the continuous "perfect" feedback and correction of robots, human personality may be annihilated, and human value rationality may become invalid, but in fact, this value rationality is the key element to distinguish why people are human and why machines are machines.

2.2.3. Emotional alienation of human-computer communication. As brain-computer interface technology continues to mature, human-computer symbiosis will also bring human-computer communication into a deeper and wider realm. At the same time, “human-computer communication” may even be transformed into “interpersonal communication”. People are able to communicate naturally with robots in the dual world of reality and virtual reality, exchange information, and develop feelings, etc. Robots get the enhancement of emotional ability in input and spontaneous learning, and synchronize to produce a high degree of initiative, and are able to carry out emotional guidance and emotional interactions in the exchange of information with human beings, which results in the alienation of emotional roles, and even become the active propagator of emotions. Therefore, robots must be made to comply with various legal obligations and ethical and moral regulations of human society, and assume specific social roles and responsibilities, so as to prevent the collapse or alienation of values caused by such role alienation. Such alienation should be based on the foundation that the civil rights of robots are improved and guaranteed. This requires the concerted efforts of the government, technology developers, and society at large to break down legal restrictions, ethical disputes, and anthropocentrism, and to develop machine technology that is good, harmonious, and rationally coordinated.

### *2.3. Cross-cultural empathy feature recognition model for human-computer communication*

Under the cross-cultural empathy of human-machine communication, multi-round textual dialog empathy feature recognition, as an important research component to enhance machine emotional intelligence, aims to categorize the emotions of the current discourse based on the information of the speaker, dialog context, and so on. The current dialog utterance as well as the dialog context information is taken as the input of the task, and the emotion category of the current speaker is taken as the output. On the basis of deep learning algorithms, a cross-cultural empathy feature aliasing model for human-machine communication based on Graph Neural Network-Attention Mechanism-Bi-Directional Gating Unit (GCN-Attention-BiGRU) is proposed, whose main purpose is to simulate the characteristics of emotion propagation in dialogues by using the natural structure of graphs, so as to improve the ability of cross-cultural empathy feature recognition for human-machine communication.

2.3.1. Input modules. The input module mainly processes the input dialog to provide data support for the next layer of the network. It mainly includes feature extraction of the dialog and the construction of word set, entity set and lexical annotation set. In the CNN, a 300-dimensional pre-trained GloVe word vector model is used to initialize the embedding layer, and the convolutional layer adopts convolutional kernel sizes of 3, 4, and 5, with 50 feature mappings per convolutional kernel, and then maximally pools the obtained convolutional features with a window size of 2, followed by activation through ReLU, and then finally splices the obtained results into the 100-dimensional fully-connected layer to obtain context-independent discourse feature representations through activation. After that, the context-independent discourse feature representation is obtained. The whole CNN network is trained on emotionally labeled discourse level. Context-independent discourse feature representations are extracted for each round of discourse $U_t$:

$$u_1, u_2, \ldots, u_t, \ldots, u_n = CNN\left(U_1, U_2, \ldots U_t, \ldots, U_n\right), \tag{1}$$

where $(U_1, U_2, \ldots U_t, \ldots, U_n)$ is the dialog history and $U_t$ is the discourse at moment $t$, which consists of $K$ words, i.e., $U_t = (w_1, w_2, \ldots, w_k)$. $u_1, u_2, \ldots, u_t, \ldots, u_n$ is a context-independent representation of the discourse vector obtained after feature extraction by the CNN.

2.3.2. Network layer. Graph neural networks in multi-round conversations are usually represented as nodes of a graph model with the state of each moment, and the connections between conversation rounds are represented by edges between nodes in the graph model, where the definition and encoding of graph nodes, the construction of graphs, the feature extraction of graphs, and the fusion of features are performed in the model.

1) Definition and encoding of graph nodes. Word-level graph nodes. The word-level graph itself is denoted as $G_\tau = \{V_\tau, A_\tau\}$, where $\tau$ denotes the word-level graph type, $V_\tau$ denotes the set of graph nodes, and $A_\tau \in \mathbb{R}^{|V_\tau| \times |V_\tau|}$ denotes the adjacency matrix of the word-level graph, about which the construction of the adjacency matrix will be introduced in the next section on graph construction. The nodes of the word-level graph come from the word set, entity set and lexical annotation set in the input layer, for the $i$th node $v_\tau^i \in V_\tau$ the node feature representation is $x_\tau^i$. The word graph and lexical annotation graph are initialized with one-hot $x_\tau^i$. In the entity graph the feature vector of the entity nodes is used $x_e^i$. The set of feature representations of all the nodes is $X_\tau$. The nodes of different types in $G_\tau$ do not affect each other. The word level graph mainly uses two layers of GCN to get the embedding representation of the graph nodes $H_\tau$. During the model training process, $H_\tau$ the update process is represented as:

$$H_\tau = \tilde{A}_\tau \cdot ReLu\left(\tilde{A}_\tau X_\tau W_\tau^1\right) W_\tau^2, \tag{2}$$

$$\tilde{A}_\tau = D_\tau^{-\frac{1}{2}}\left(I + A_\tau\right) D_\tau^{\frac{-1}{2}}, \tag{3}$$

$$\left[D_\tau\right]_{ij} = \sum_j \left[A_\tau\right]_{ij}, \tag{4}$$

$$\left[ReLu\left(x\right)\right]_i = \max\left(\left[x\right]_i, 0\right), \tag{5}$$

where $I$ is the unit matrix, $W_\tau{}^1$, $W_t^2$ are the parameters to be trained, and $[A]_{ij}$ denotes the $(i, j)$th element of matrix $A$.

Sentence graph node. The sentence graph is represented as $G_s = \{V_s, A_s\}$, where $s$ denotes the type of sentence graph, sentence graph node $v_s^i \in V_s$ denotes the single discourse text feature vector of the speaker in a multi-round conversation $S$ and $A_s$ is the corresponding adjacency matrix of the sentence graph. The sentence graph and word-level graph is a hierarchical graph, and the network layer structure is shown in Figure 1.

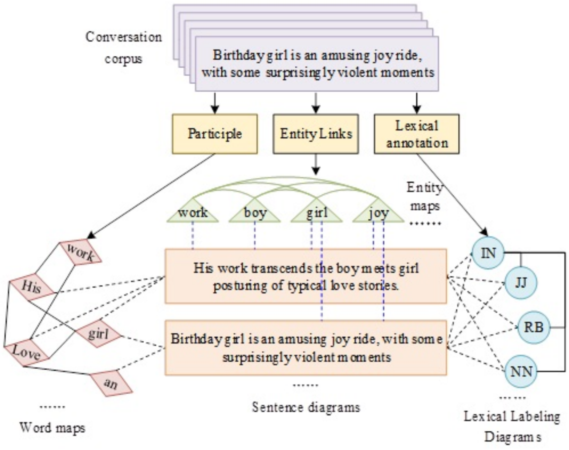

Conversation corpus
Birthday girl is an amusing joy ride, with some surprisingly violent moments
Participle
Entity Links
Lexical annotation
Entity maps
work
boy
girl
joy
......
His work transcends the boy meets girl posturing of typical love stories.
Birthday girl is an amusing joy ride, with some surprisingly violent moments
......
Sentence diagrams
work
His
girl
Love
an
......
Word maps
IN
JJ
RB
NN
......
Lexical Labeling Diagrams

**Fig. 1.** The relationship between sentence graph and word level graph

In the hierarchical graph, the sentence graph $G_s$ node is obtained by word-level graph hierarchical pooling and then feature splicing. Specifically, after first obtaining $H_\tau$ through Eq. (2), hierarchical pooling is applied to the word-level graph to extract graph node features $\widehat{x}_\tau^i$, and the specific calculation process is as follows:

$$\hat{x}_\tau^i = u\left(H_\tau^T s_\tau^i\right), \tau \in w, p, e, \tag{6}$$

$$u(x) = \frac{x}{\|x\|_2}, \tag{7}$$

where $x \| y$ denotes the first place splice of the two vectors and $\|\cdot\|_2$ denotes the L2 norm of the vectors, even if the elements in $x$ are normalized to the L2 norm. $(\cdot)^T$ denotes the transpose of the matrix.

Dialog graph nodes. Dialogue graph $G_h$ contains all the information of the whole dialogue, $h$ denotes the dialogue graph type, $V_h$ denotes the dialogue graph node, and for each node $v_i$ is initialized using the corresponding context-dependent sequential coding features $g_i$. And in this paper, bidirectional GRU, i.e. BiGRU, is used to encode context-independent representations $u_t$ to obtain sequential context-aware features $g_i$:

$$g_i = BiGRU\left(g_{i(+,-)1}, u_i\right), i = 1, 2, \ldots, N. \tag{8}$$

When the neighborhood-based transformation process is applied to encode the context of the speaker hierarchy, the vertex features change accordingly.

2) Construction of the graph. After the initialization and construction of graph nodes is completed, it is necessary to define the relationships between nodes and nodes, i.e., the edges of the graph, so as to complete the construction of the graph.

In this paper, we use a graph attention mechanism based approach to compute the adjacency matrix of a conversation graph $A_h$. The values in the adjacency matrix represent the weight values of the neighboring nodes. For any two graph nodes $v_h^i$, $v_h^j \in V_h$, consider that the context window size of the past conversation is $b$ and the context window of the future conversation is $a$, then the conversation graph adjacency matrix is calculated as follows:

$$A_h^{ij} = soft\max\left(g_i^T W_e \left[g_{i-b}, \ldots, g_{i+a}\right]\right), \tag{9}$$

where $W_e$ is the trainable parameter, the reason for performing softmax normalization is to ensure that the weights of the edges do not exceed 1. In fact, the final value obtained is the value of the weight between the most relevant nodes within the window to the current node $v_h^i$. The adjacency matrix, i.e., the weight value of each edge, computed by the attention function $A_h^{ij}$ is constant and does not change during the learning process.

3) Feature transformation of graph. After the completion of the graph construction can not directly obtain the emotional state of the interlocutor, it is also necessary to carry out deeper feature extraction of the entire conversation from multiple dimensions, that is, it is necessary to convert the features of the nodes and edges of the constructed graph, so as to extract the dependence of the interlocutor's emotions on the context of the conversation, the persistence of the interlocutor's own emotional features, and the emotional characteristics of the interlocutor's emotions, to provide a basis for the identification of the next step of the emotion of the discernment. The word level graph is a hierarchical graph that is used as a basis for the recognition of emotions. The word-level graph is spliced into the sentence graph by extracting features through hierarchical pooling of the graph, and there is no need for feature conversion of the word-level graph, so only the sentence graph and the dialog graph need to be feature extracted.

2.3.3. Output modules. After extracting all the features of the graph, a fully connected network is used for dialog emotion recognition. Specifically, firstly, the final features extracted from the dialog graph are put into a layer of fully connected layer, then ReLU activation is used, the activated result is spliced with the final features of the sentence graph and then fed into the fully connected layer, and finally the emotion classification probability distribution is obtained by softmax, the process is as follows. That is:

$$l_i = ReLu\left(W_l \tilde{h}_t + b_l\right), \tag{10}$$

$$P_i = soft\max\left(W_{s\max}\left(l_i \left\| y_s^i\right.\right) + b_{\max}\right), \tag{11}$$

$$\tilde{y}_l = \arg\max_k \left(P_i[k]\right), \tag{12}$$

where $y_s^i \in Y_s^2$, $b_l$, and $b_{\max}$ are the biases of the fully connected layer, $P_i$ denotes the probability distribution of each emotional state corresponding to dialog $i$, $\|$ in Eq. (11) denotes the vector splicing, and $W_l$ and $W_{s\max}$ are the weights learned during the training process. Eq. (12) represents the label corresponding to the maximum value of probability in $P_i$, which is the emotional state of the interlocutor corresponding to the $i$th sentence recognized by the model.

2.3.4. Model training. After the model definition is completed, it is time to train a good model, and the judgment of a model is based on the robustness of the model. As an important part of model training, the definition of loss function is directly related to the final model, the smaller the loss function, the better the model robustness. The GCN-Attention-BiGRU model proposed in this paper adopts the categorical cross-entropy and L2 regularization as the loss function during model training, which can be expressed as:

$$L = -\frac{1}{\sum\limits_{s=1}^{N} c(s)} \sum_{i=1}^{N} \sum_{j=1}^{c(i)} \log P_{i,j}\left[y_{i,j}\right] + \lambda \left\|\theta\right\|_2, \tag{13}$$

where $N$ is the number of conversations, $c(i)$ denotes the number of discourses in sample $i$, $P_{i,j}$ is the probability distribution of the emotional labels corresponding to the $j$th round of conversations in conversation $i$, $y_{i,j}$ is the expected emotional state of the $j$th round of conversations in conversation $i$, $\lambda$ is the L2 regularization weights, and $\theta$ is the training parameters.

### *2.4. Adaptive interaction model*

Deep learning and the Internet of Things each has many advantages, the organic integration of the two can realize the traditional Internet of Things in the sensing device sensing information and receiving instructions connected to the Internet, the real realization of the network, and through the cloud computing technology to achieve massive data storage and computing, greatly simplify the delivery process of the application, reduce the cost of delivery, and have a higher benefit of the application.

2.4.1. Definition of internet of things. Deep learning definition has been given in detail above, this subsection will define the Internet of Things (IoT), the Internet of Things is an information sensing device that connects any object to the Internet through radio frequency identification RFID technology, global positioning system, laser scanner, two-dimensional code, and infrared sensor, etc., in accordance with predefined protocols, to exchange information and communicate, and to realize communication and dialogue between human and object, and between objects, i.e., objects with a The information network connecting objects and objects with comprehensive sensing ability and reliable transmission and intelligent processing ability for information.

2.4.2. Construction of the model. In the interactive application of IoT and deep learning, two key factors need to be grasped in order to ensure that users ultimately receive accurate and complete application services:

1) Reliability of acquired IoT information (QoI) described as the degree of correctness, authenticity and real-time degree of acquired IoT information, etc.

2) Reliability of service provided by the system (QoS) described as the degree of timeliness and quality of service provided by the whole IoT system to the user application services.

The adaptive interaction model is shown in Figure 2. The IoT information collected by various intelligent sensing devices is transmitted to the cloud computing center through the network, and after the preliminary screening and calculation of the cloud computing center, the reliability of IoT information (QoI) and the reliability of services provided by the system (QoS) are obtained respectively, and then through the information analyzing processor, the QoI and QoS will be compared and analyzed with the intelligent threshold set by the system respectively, and the adaptive interaction processor will give different execution programs.

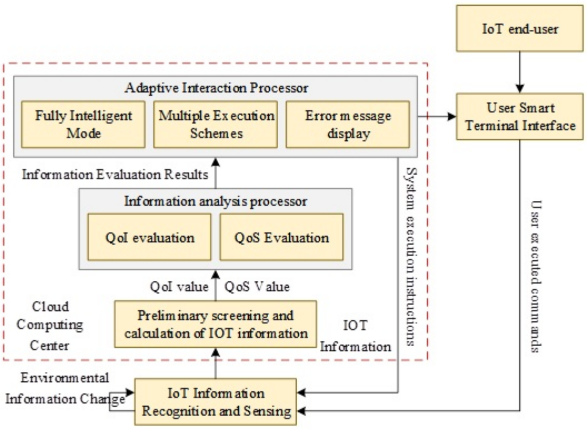

IoT end-user
Adaptive Interaction Processor
Fully Intelligent Mode
Multiple Execution Schemes
Error message display
User Smart Terminal Interface
Information Evaluation Results
Information analysis processor
QoI evaluation
QoS Evaluation
System execution instructions
User executed commands
QoI value
QoS Value
Cloud Computing Center
Preliminary screening and calculation of IOT information
IOT Information
Environmental Information Change
IoT Information Recognition and Sensing

**Fig. 2.** Adaptive interaction mode

## 3. Example analysis of cross-cultural empathy and interaction patterns

### *3.1. Analysis of empathic feature recognition models*

3.1.1. Setting up the training environment. This experiment uses PyTorch deep learning framework to construct a graph neural network-attention mechanism-bidirectional gating unit (GCN-Attention-BiGRU) human-computer communication cross-cultural empathy feature recognition model for experimental analysis, the experimental equipment used is the CPU for the 11th Gen Intel(R) Core(TM) i5-11260H @ 2.60GHz, 8GB of RAM, the graphics device is NVIDIA GeForce RTX 3060 Laptop GPU, the platform used is Windows 10 operating system, and the software used is Pycharm, python3.6, CUDA11.1 and CUDNN8.0.5.

3.1.2. Loss function analysis. To ensure an adequate empathy dataset, a publicly accessible and annotated Reddit dataset was used as the data source for this study. And it was divided into training and validation sets in the ratio of 8:2. The training process is optimized using Adam optimizer, and the loss function classification cross entropy and L2 regularization are the most loss function of the model in this paper. The training is carried out for a total of 20 cycles, taking into account the effect of the memory, the training is divided into two phases, the batch size of the first phase is set to 10, and the training is carried out for 10 cycles. The second stage of the batch size is set to 20, training for 10 cycles, after each training cycle to verify the model of this paper, the results of the loss function during the training process are shown in Table 1. Based on the data in the table, it can be seen that the training set and the validation set with the growth of the training cycle, and eventually the loss value of the two stabilized at 0.025, 0.016, at this time, the accuracy of this paper's model reaches 93.43%, maximizing the reduction of the model of the training process of the loss generated.

3.1.3. Assessment of indicators. The main purpose of constructing a recognition model is to be able to label each discourse in a multi-round conversation with sentiment and intent, so this paper is designed to use the accuracy of multi-round conversation recognition in addition to the recognition accuracy of individual discourse. Classification accuracy of individual discourse:

$$Acc_{single} = \frac{N_{corr}}{N}, \tag{14}$$

where $N_{corr}$ denotes the number of correct classifications and $N$ denotes the number of desired utterances. And the classification accuracy of multi-round dialog is shown below:

$$Acc_{multi} = \frac{N_{corr}}{N_{multi}}, \tag{15}$$

where $N_{corr}$ denotes the number of multi-round conversations in which each utterance is correct, and $N_{multi}$ denotes the number of conversations containing more than two rounds.

**Table 1.** Loss function results during training

| Epoch | Stages | Training loss | Proof loss | Accuracy rate (%) |
|---|---|---|---|---|
| 1 | Stages1 | 1.299 | 0.751 | 64.01 |
| 2 | | 1.123 | 0.683 | 64.17 |
| 3 | | 1.115 | 0.612 | 65.95 |
| 4 | | 0.785 | 0.586 | 66.94 |
| 5 | | 0.747 | 0.576 | 72.03 |
| 6 | | 0.692 | 0.562 | 74.84 |
| 7 | | 0.681 | 0.544 | 75.14 |
| 8 | | 0.647 | 0.404 | 78.71 |
| 9 | | 0.534 | 0.337 | 78.77 |
| 10 | | 0.447 | 0.323 | 79.22 |
| 11 | Stages2 | 0.399 | 0.321 | 79.66 |
| 12 | | 0.372 | 0.171 | 80.41 |
| 13 | | 0.324 | 0.157 | 82.78 |
| 14 | | 0.245 | 0.141 | 82.88 |
| 15 | | 0.205 | 0.113 | 86.83 |
| 16 | | 0.182 | 0.099 | 87.26 |
| 17 | | 0.178 | 0.085 | 88.42 |
| 18 | | 0.148 | 0.065 | 90.33 |
| 19 | | 0.072 | 0.049 | 92.86 |
| 20 | | 0.025 | 0.016 | 93.43 |

3.1.4. Analysis of ablation experiments. In order to verify the effectiveness of the model in this paper, based on the dataset and evaluation indexes, the ablation experiments are analyzed from four aspects: single discourse emotion, multi-round conversation emotion, single discourse intention, and multi-round conversation intention, and the results of the ablation experiments of the recognition model are shown in Table 2. According to the data performance in Table 2, it can be seen that the comprehensive recognition rate of the graph convolutional neural network as the baseline model is 0.756, and with the addition of the attention mechanism, it improves by 0.125 compared to the baseline model (GCN), and when the bidirectional gated recurrent unit (BiGRU) is introduced to the baseline model (GCN), at this time, the comprehensive recognition rate reaches 0.905, and finally the bidirectional gated recurrent unit ( BiGRU) and the attention mechanism are introduced to the benchmark model (GCN) at the same time, the model's comprehensive recognition rate of single discourse emotion, multi-round dialog emotion, single discourse intention, and multi-round dialog intention is 0.924, which can well meet the needs of cross-cultural empathic expression in human-computer communication.

**Table 2.** Identify the ablation results of the model

| Model | Individual utterance emotion | Multiple rounds of conversations about emotions | Individual discourse intention | Multiple rounds of dialogue intent |
|---|---|---|---|---|
| GCN | 0.738 | 0.739 | 0.772 | 0.775 |
| GCN+ Attention mechanism | 0.872 | 0.878 | 0.881 | 0.893 |
| GCN+BiGRU | 0.893 | 0.897 | 0.904 | 0.924 |
| GCN+ Attention mechanism+ BiGRU | 0.912 | 0.923 | 0.929 | 0.932 |

3.1.5. Analysis of model applications. This subsection is based on the empathy, relevance, and fluency perspectives, and 60 volunteers with bachelor's degrees were invited to manually evaluate the efficacy of the model application. Evaluators assessed the empathy, fluency, and relevance of each response by assigning a score from 1 to 5 to each response. To clarify the scoring criteria, this paper provides explanations for each metric:

1) Empathy. Empathy encompasses emotional empathy and cognitive empathy. It refers to the ability to reply to be able to understand the speaker's emotions and respond accordingly, as well as the ability to understand the mental state of others and predict their thoughts based on the situation.

2) Relevance. Whether the human-computer reply accurately addresses the speaker's words, provides insightful suggestions, effectively grasps the topic of the conversation, etc.

3) Fluency. Whether a reply is fluent or not does not necessarily require strict adherence to a specific writing format, as long as grammatical and lexical errors do not impede comprehension.

The results of the model application analysis are shown in Table 3. The data show that the mean values of empathy, relevance, and fluency for the 60 samples are 4.537, 4.447, and 4.499, indicating that the users have a positive attitude towards the model application effect of this paper. The empathy feature extraction ability that the model of this paper has, it is a pleasant thing to reason smoothly. And the model fully considers the questioning strategy to be expressed in the reply to express joy in the questioning action to further resonate with the speaker emotionally, and further questioning the follow-up, thus prompting more rounds of interaction with the speaker. Therefore, the responses formed by the model proposed in this paper further enhance the distance with the speaker, produce emotional resonance, and guide the speaker to express more information, promoting further interaction between the two sides.

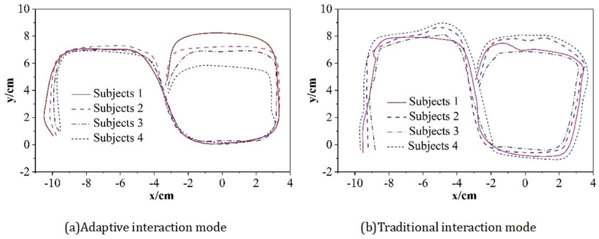
### *3.2. Interaction model application analysis*

In order to verify the adaptive performance of the designed model, 4 subjects were selected from the above 60 samples, and these 4 subjects used 2 interaction modes (this paper: adaptive interaction mode, control: traditional interaction mode) to control the motion trajectory of the intelligent wheelchair at different time periods. With the extension of the subjects' interaction time, muscle fatigue began to appear in the masticatory muscles, and the degree of muscle fatigue gradually deepened with the growth of time, the amplitude of EMG signals increased significantly, and the signal characteristics changed, which reduced the action recognition rate of the system and led to

**Table 3.** Analysis results of model application

| N | Empathy | Correlation | Fluency | N | Empathy | Correlation | Fluency |
|---|---|---|---|---|---|---|---|
| 1 | 4.201 | 4.541 | 4.242 | 31 | 4.641 | 4.659 | 4.124 |
| 2 | 4.445 | 4.008 | 4.756 | 32 | 4.901 | 4.311 | 4.642 |
| 3 | 4.4 | 4.738 | 4.815 | 33 | 4.085 | 4.212 | 4.139 |
| 4 | 4.504 | 4.099 | 4.989 | 34 | 4.426 | 4.844 | 4.919 |
| 5 | 4.349 | 4.842 | 4.632 | 35 | 4.452 | 4.101 | 4.488 |
| 6 | 4.42 | 4.714 | 4.544 | 36 | 4.914 | 4.162 | 4.373 |
| 7 | 4.221 | 4.028 | 4.404 | 37 | 4.925 | 4.028 | 4.808 |
| 8 | 4.726 | 4.365 | 4.763 | 38 | 4.157 | 4.978 | 4.019 |
| 9 | 4.619 | 4.116 | 4.34 | 39 | 4.79 | 4.561 | 4.193 |
| 10 | 4.491 | 4.554 | 4.269 | 40 | 4.008 | 4.158 | 4.194 |
| 11 | 4.324 | 4.665 | 4.883 | 41 | 4.87 | 4.34 | 4.261 |
| 12 | 4.355 | 4.927 | 4.918 | 42 | 4.095 | 4.454 | 4.982 |
| 13 | 4.939 | 4.889 | 4.229 | 43 | 4.938 | 4.8 | 4.747 |
| 14 | 4.418 | 4.928 | 4.554 | 44 | 4.661 | 4.067 | 4.31 |
| 15 | 4.044 | 4.961 | 4.299 | 45 | 4.589 | 4.331 | 4.603 |
| 16 | 4.93 | 4.201 | 4.234 | 46 | 4.004 | 4.255 | 4.103 |
| 17 | 4.475 | 4.059 | 4.357 | 47 | 4.598 | 4.088 | 4.542 |
| 18 | 4.729 | 4.989 | 4.465 | 48 | 4.77 | 4.435 | 4.709 |
| 19 | 4.528 | 4.069 | 4.136 | 49 | 4.763 | 4.186 | 4.801 |
| 20 | 4.723 | 4.312 | 4.595 | 50 | 4.092 | 4.406 | 4.288 |
| 21 | 4.762 | 4.912 | 4.379 | 51 | 4.967 | 4.786 | 4.194 |
| 22 | 4.559 | 4.962 | 4.901 | 52 | 4.459 | 4.188 | 4.347 |
| 23 | 4.283 | 4.952 | 4.35 | 53 | 4.723 | 4.239 | 4.065 |
| 24 | 4.721 | 4.446 | 4.717 | 54 | 4.352 | 4.424 | 4.694 |
| 25 | 4.371 | 4.663 | 4.581 | 55 | 4.805 | 4.458 | 4.959 |
| 26 | 4.826 | 4.297 | 4.814 | 56 | 4.734 | 4.272 | 4.741 |
| 27 | 4.227 | 4.686 | 4.599 | 57 | 4.249 | 4.501 | 4.793 |
| 28 | 4.629 | 4.011 | 4.742 | 58 | 4.965 | 4.921 | 4.698 |
| 29 | 4.996 | 4.734 | 4.481 | 59 | 4.474 | 4.681 | 4.082 |
| 30 | 4.386 | 4.323 | 4.014 | 60 | 4.209 | 4.781 | 4.114 |

the system being uncontrolled. Figure 3 gives the movement trajectory of the wheelchair in different time periods, where (a)∼(b) are the adaptive interaction mode and traditional interaction mode, respectively. Curves 1∼4 in the figure are the training time: 1-0∼15min, 2-15∼30min, 3-30∼45min, 4-45∼60min. Table 4 then gives the four subjects in different time periods to control the intelligent wheelchair to complete the the specified route in different time periods. After the subjects continue to operate the traditional interaction mode for 45min, the smart wheelchair starts to lose control, the traditional interaction mode loses the ability to recognize muscle movements and control the wheelchair, and the time consumed to complete the specified route is obviously more than the adaptive human-computer interaction mode designed in this paper, which also verifies that the adaptive human-computer interaction mode based on the Internet of Things and deep learning has a better stability and adaptive ability.



(a)Adaptive interaction mode

(b)Traditional interaction mode

**Fig. 3.** The movement of the wheelchair in different time periods

**Table 4.** The time it takes to complete a specified route

| N | Traditional interaction mode | | | | Adaptive interaction mode | | | |
|---|---|---|---|---|---|---|---|---|
| | 0∼15min | 15∼30min | 30∼45min | 45∼60min | 0∼15min | 15∼30min | 30∼45min | 45∼60min |
| 1 | 74.8 | 116.5 | 228.7 | 342.3 | 54.1 | 54.6 | 59.9 | 62.7 |
| 2 | 76.7 | 104.8 | 232.2 | 353.7 | 57.2 | 58.2 | 68.7 | 75.3 |
| 3 | 83.5 | 108.6 | 258.1 | 357.4 | 65.8 | 66.1 | 75.1 | 76.8 |
| 4 | 89.7 | 105.3 | 269.2 | 366.1 | 67.6 | 67.3 | 74.8 | 81.9 |

## 4. Conclusion

On the basis of relevant theories, this paper constructs a cross-cultural empathy aliasing model for human-computer communication and an adaptive interaction model respectively, and analyzes the application effects of the two.

1) The bidirectional gated recurrent unit (BiGRU) and the attention mechanism are introduced into the baseline model (GCN) at the same time, and the model's combined recognition rate of single discourse emotion, multi-round conversation emotion, single discourse intention, and multi-round conversation intention is 0.924. In addition, the mean values of empathy, relevance, and fluency for the 60 samples are 4.537, 4.447, and 4.499, which indicates that the user hold a positive attitude toward the application effect of the model in this paper, which is of great practical significance for advancing the development of cross-cultural empathy expression in human-computer communication.

2) The time consumed by the traditional interaction model is also significantly more than that of the adaptive human-computer interaction model designed in this paper, indicating that the adaptive human-computer interaction model based on the Internet of Things and deep learning has better stability and adaptive ability.

## Funding

This research was supported by the Research on the high-quality development of the gymnasium (Project Number: 20240745).

## References

[1] H. Bansal and R. Khan. A review paper on human computer interaction. *International Journal of Advanced Research in Computer Science and Software Engineering*, 8(4):53, 2018. https://doi.org/10.23956/IJARCSSE.V8I4.630.

[2] P. Bourges-Waldegg. *Handling Cultural Factors in Human-Computer Interaction.* University of Derby (United Kingdom), 2021. https://doi.org/10.48773/92z39.

[3] J. Brejcha. *Cross-Cultural Human-Computer Interaction and User Experience Design: A Semiotic Perspective.* CRC Press, 2015. https://doi.org/10.1201/b18059.

[4] A. Dix. Human–computer interaction, foundations and new paradigms. *Journal of Visual Languages & Computing*, 42:122–134, 2017. https://doi.org/10.1016/j.jvlc.2016.04.001.

[5] R. Heimgärtner. *Cultural Differences in Human-Computer Interaction: Towards Culturally Adaptive Human-Machine Interaction.* Oldenbourg Wissenschaftsverlag Verlag, 2012. doi.org/10.1524/9783486719895.bm.

[6] R. Heimgärtner. Reflections on a model of culturally influenced human–computer interaction to cover cultural contexts in hci design. *International Journal of Human-Computer Interaction*, 29(4):205–219, 2013. https://doi.org/10.1080/10447318.2013.765761.

[7] M. G. Helander. *Handbook of Human-Computer Interaction.* Elsevier, 2014.

[8] L. Hunyadi. Multimodal human-computer interaction technologies. *Argumentum*, 7:240–260, 2011.

[9] B. J. Hurn, B. Tomalin, B. J. Hurn, and B. Tomalin. *What is Cross-Cultural Communication?* Springer, 2013. https://doi.org/10.1057/9780230391147_1.

[10] T. Issa, P. Isaias, T. Issa, and P. Isaias. Usability and human computer interaction (hci). *Sustainable Design: HCI, Usability and Environmental Concerns*:19–36, 2015. https://doi.org/10.1007/978-1-4471-6753-2_2.

[11] J. Lazar, J. H. Feng, and H. Hochheiser. *Research Methods in Human-Computer Interaction.* Morgan Kaufmann, 2017.

[12] L. Li. Addressing cross-cultural design challenges in social media platforms: a human-computer interaction perspective. In *International Conference on Human-Computer Interaction*, pages 75–88. Springer, 2024. https://doi.org/10.1007/978-3-031-60901-5_6.

[13] X. Li, Z. Song, B. Zhi, J. Pu, and C. Meng. Intelligent identification of rock mass structural based on point cloud deep learning technology. *Construction and Building Materials*, 456:139340, 2024. https://doi.org/10.1016/j.conbuildmat.2024.139340.

[14] A. G. Lopes. Using research methods in human computer interaction to design technology for resilience. *JISTEM-Journal of Information Systems and Technology Management*, 13:363–388, 2016. https://doi.org/10.4301/S1807-17752016000300001.

[15] Z. Lyu. State-of-the-art human-computer-interaction in metaverse. *International Journal of Human–Computer Interaction*, 40(21):6690–6708, 2024. https://doi.org/10.1080/10447318.2023.2248833.

[16] R. Mead and C. J. Jones. Cross-cultural communication. *The Blackwell Handbook of Cross-Cultural Management*:283–291, 2017. https://doi.org/10.1002/9781405164030.ch14.

[17] J. Miehle, K. Yoshino, L. Pragst, S. Ultes, S. Nakamura, and W. Minker. Cultural communication idiosyncrasies in human-computer interaction. In *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 74–79, 2016. https://doi.org/10.18653/v1/W16-3610.

[18] S. O'Brien. Translation as human–computer interaction. *Translation Spaces*, 1(1):101–122, 2012. https://doi.org/10.1075/ts.1.05obr.

[19] M. Pikhart. Cognitive and computational aspects of intercultural communication in human-computer interaction. In *International Conference on Human-Computer Interaction*, pages 367–375. Springer, 2020. https://doi.org/10.1007/978-3-030-49913-6_31.

[20] L. C. d. C. Salgado, C. S. de Souza, C. M. Ferreira, and C. F. Leitão. Characterizing intercultural encounters in human-computer interaction. In *Cross-Cultural Design: 8th International Conference, CCD 2016, Held as Part of HCI International 2016, Toronto, ON, Canada, July 17-22, 2016, Proceedings 8*, pages 108–119. Springer, 2016. https://doi.org/10.1007/978-3-319-40093-8_12.

[21] S. Sayago. *Cultures in Human-Computer Interaction*. Springer, 2023. https://doi.org/10.1007/978-3-031-30243-5.

[22] C. Shen, D. Tang, P. Wang, Z. Lyu, M. Zhang, B. Liu, C. Yang, and L. Yu. Fiber distribution in uhpc under different influencing factors evaluated with a novel method based on deep learning. *Construction and Building Materials*, 457:139350, 2024. https://doi.org/10.1016/j.conbuildmat.2024.139350.

[23] M. Zhang and X. Lu. Application of empathy theory in the study of the effectiveness and timeliness of information dissemination in regional public health events. *Frontiers in Public Health*, 12:1388552, 2024. https://doi.org/10.3389/fpubh.2024.1388552.