Journal of computer science and software applications
https://www.mfacademia.org/index.php/jcssa
ISSN:2377-0430
Vol. 5, No. 3, 2025

User Intent Prediction and Response in Human-Computer Interaction via BiLSTM

Qi Sun[1], Shiyu Duan[2]
[1]Carnegie Mellon University, Pittsburgh, USA
[2]Carnegie Mellon University, Pittsburgh, USA
*Corresponding author: Shiyu Duan, naomiduansy@gmail.com

Abstract:

The purpose of this study is to explore user intent prediction and automatic response systems in human-computer interaction based on deep learning. By analyzing the relationship between user behavior characteristics and intention prediction and combining the BiLSTM (Bidirectional long short-term memory network) model, an accurate method of user intention prediction is proposed. The experimental results show that compared with traditional machine learning models (such as SVM and random forest), the BiLSTM model has significant advantages in prediction accuracy and system response-ability, especially in understanding and responding to user intentions in complex scenarios. In addition, this study also analyzes the correlation between homepage data (such as user click module, visit frequency, page stay time, etc.) and user intent and reveals the impact of different page elements on user behavior. Future research can further optimize the adaptability of the model, incorporate more multimodal data (such as voice, video, etc.), and explore personalized and context-aware user intent prediction methods to promote the development of more intelligent and naturalized human-computer interaction systems. The research results provide a theoretical basis and technical support for improving the human-computer interaction experience and promoting intelligent service applications.

Keywords:

Human-computer interaction, User intent prediction, BiLSTM, Deep learning

1. Introduction

With the rapid development of artificial intelligence and natural language processing technologies, Human-Computer Interaction (HCI) has become an integral part of our daily lives. From intelligent voice assistants to self-driving cars, from smart homes to customer service robots, HCI has transformed the way we interact with technology [1]. Deep learning has significantly enhanced systems' ability to understand and respond to user intentions, making interactions more intelligent and human-like. However, accurately predicting user intent and generating appropriate responses remain ongoing challenges. Optimizing these processes is crucial for improving user experience and advancing HCI applications [2,3]. This study explores the role of BiLSTM-based deep learning models in refining user intent prediction and response, contributing to more seamless and adaptive interactions.

User intention recognition is a critical component in HCI. Traditional interaction methods rely heavily on preset rules and instruction sets. User inputs generate corresponding outputs through rule matching [4]. However, this approach has limitations. It cannot flexibly respond to complex and changing user needs, and it only offers a shallow understanding of user input. With the introduction of deep learning, especially the application of BILSTM (Bidirectional Long Short-Term Memory Network) in natural language processing, machines are now better able to infer underlying user intentions. BILSTM processes time-series data and captures contextual information, enabling the system to understand user input from a broader perspective and make more accurate predictions about user intentions. This enhances the system's ability to provide intelligent and personalized responses [5].

Predicting user intention is not an easy task. It involves complex aspects such as language understanding and sentiment analysis. User intentions are influenced not only by their language expression but also by context, tone, historical interaction records, and more. Traditional methods struggled to account for all these variables, limiting their accuracy and robustness. BILSTM addresses the issue of information flow loss in traditional models by using bidirectional modeling and considering the context. For instance, in a customer service chatbot, if a user types "I'm frustrated with my order," traditional models might focus only on the words "order" and "frustrated," leading to a generic response like "How can I help you with your order?" In contrast, a BILSTM-based system can analyze sentiment, context, and past interactions, generating a more empathetic and context-aware response such as "I'm sorry for the trouble! Let me check the issue for you right away." Similarly, in voice assistants, when a user says "Play something relaxing," a BILSTM-powered model can infer preferences from past listening history and suggest a personalized playlist instead of just playing random soft music. It dynamically adapts to different features of user input, handling complex language patterns, ambiguity, and long-tail issues more effectively. This makes the BILSTM-based system more adaptable to changing user needs, providing a more accurate and natural interactive experience [6].

The goal of HCI is to create seamless, intuitive interactions that enable machines to understand and respond to user needs effectively. That makes prediction, personalization a major part of the discussion and research direction. Accurately predicting user intent and delivering automatic responses can greatly enhance service efficiency and user satisfaction. In contexts such as customer service in various service industries, smart homes, and intelligent voice assistants, users expect fast, personalized responses. A system capable of precise intent recognition not only improves productivity but also minimizes manual intervention and reduces operational costs. A system that accurately predicts and responds to user intent boosts productivity, reduces manual intervention, and lowers operational costs. More importantly, it delivers smoother, more natural, and more personalized services, strengthening the connection and engagement between users and the system. This research is crucial for advancing intelligent services and improving the overall quality of human-computer interaction [7,8].

In conclusion, BILSTM-based user intent prediction and automatic response systems represent a key development in the field of HCI [9]. Deep learning, especially BILSTM, significantly enhances machines' ability to comprehend user intentions and generate intelligent responses tailored to user needs. This research advances HCI technology, improves user experience, and fosters the optimization and innovation of intelligent service systems. As technology continues to evolve and application scenarios expand, this area will undoubtedly play an increasingly important role in the future of artificial intelligence.

2. Related work

In the field of human-computer interaction (HCI), the ongoing development of computer technology and artificial intelligence has shifted research focus towards enhancing machines' ability to understand and predict user behavior. Early HCI research primarily focused on improving input devices and interaction methods, such as keyboards, mouses, and touchscreens. However, as technologies like speech recognition, natural language processing, and computer vision have advanced, research has increasingly centered on how machines can better understand user intentions and respond appropriately using multimodal perception. For instance, intelligent voice assistants (e.g., Siri, Alexa, Google Assistant) and chatbots (e.g., ChatGPT, Microsoft Bot Framework) can now understand and respond to user needs through various input forms, including voice, text, and even gestures. These systems rely heavily on intent detection and slot filling, where the goal is to identify the user's goal (intent) and extract relevant parameters (slots) to fulfill the request. For example, in the utterance "Book a flight to New York tomorrow," the intent might be "book_flight," with "New York" and "tomorrow" as slots.

However, predicting user intent in complex and dynamic contexts still presents significant challenges [10]. These challenges include handling ambiguous or incomplete user inputs, adapting to diverse linguistic styles, and managing real-time processing constraints. Additionally, user intent can vary significantly across domains, requiring models to generalize well or be fine-tuned for specific applications. To address this, an increasing amount of research has explored leveraging deep learning and neural networks, particularly recurrent neural networks (RNNs) and bidirectional Long Short-Term Memory networks (BiLSTM), to capture timing information and contextual relationships in user intent [11]. RNNs, with their ability to model sequential data, have been widely used in tasks like speech recognition and text analysis. However, standard RNNs often struggle with long-term dependencies due to the vanishing gradient problem. This limitation led to the development of LSTMs, which introduce memory cells and gating mechanisms to retain information over longer sequences [12].

BiLSTMs, an extension of LSTMs, further enhance this capability by processing sequences in both forward and backward directions, allowing the model to capture contextual information from past and future states simultaneously. This bidirectional approach has proven particularly effective in NLP tasks such as named entity recognition, sentiment analysis, and intent classification [13]. For example, in intent prediction, BiLSTMs can better understand the context of a user's query by considering the entire input sequence, rather than just the preceding words. This makes them well-suited for applications like dialogue systems, where understanding the full context of a conversation is critical for accurate intent prediction and response generation [14].

Recent advancements in deep learning have also seen the integration of attention mechanisms with BiLSTMs, enabling models to focus on the most relevant parts of the input sequence. This has further improved performance in tasks requiring nuanced understanding of user intent, such as multi-turn dialogue systems [15]. Additionally, transformer-based models like BERT and GPT have set new benchmarks in NLP tasks, but their computational complexity and resource requirements often make BiLSTMs a more practical choice for real-time applications in HCI [16].

Despite these advancements, challenges remain in applying BiLSTMs to user intent prediction. For instance, the quality of predictions heavily depends on the availability of large, annotated datasets, which can be costly and time-consuming to create. Moreover, BiLSTMs may struggle with out-of-domain inputs or rare intents that are underrepresented in the training data. To mitigate these issues, researchers have explored techniques like data augmentation, transfer learning, and hybrid models that combine BiLSTMs with other architectures [17].

Recently, deep learning-based HCI research has become the dominant direction in the field. Models such as RNN and BiLSTM have been extensively applied in areas like speech recognition, text classification, and sentiment analysis, demonstrating strong capabilities in processing time-series data [18]. In HCI applications, BiLSTM is used in intelligent customer service systems, voice assistants, emotion recognition, and other tasks. For example, in intelligent customer service systems, BiLSTM models are employed to analyze user queries and predict their intent, enabling the system to provide relevant and timely responses. In voice assistants, BiLSTMs help interpret complex user commands by considering the entire input sequence, ensuring accurate intent detection even in noisy or ambiguous environments [19]. By analyzing user input statements, tone, and even historical interaction data, BiLSTM improves the system's ability to predict users' actual needs. For instance, in emotion recognition, BiLSTMs can analyze both the textual content and the acoustic features of speech to infer the user's emotional state, enabling the system to respond empathetically [20]. This enables the system to understand user instructions at a semantic level and adjust dynamically according to context, providing personalized responses. As a result, the system enhances the naturalness and fluency of interactions, making them more intuitive and engaging [21].

Moreover, human-computer interaction research must also consider contextual and emotional factors, particularly in multi-turn dialogues. In multi-turn dialogues, the context of previous interactions plays a critical role in understanding the user's current intent. For example, a user might refer back to earlier parts of the conversation, requiring the system to maintain a coherent dialogue history. BiLSTMs, with their ability to model long-term dependencies, are well-suited for this task, as they can retain and utilize information from earlier turns to inform current predictions [22]. Effectively identifying both user intentions and emotional states has become a critical area of focus. Emotion-aware systems, for instance, can detect frustration or satisfaction in a user's voice or text, allowing the system to adapt its responses accordingly. This is particularly important in applications like mental health support or customer service, where empathetic interactions can significantly enhance user satisfaction [23].

Traditional rule-based systems encounter difficulties in managing intricate and ambiguous conversations. These systems often rely on predefined rules and templates, which struggle to handle the variability and complexity of natural language. For example, they may fail to interpret sarcasm, humor, or indirect requests, leading to inaccurate intent predictions and unsatisfactory responses [24]. Conversely, deep learning approaches, particularly models such as BiLSTM, demonstrate the ability to acquire intricate language patterns and emotional nuances from extensive datasets. By training on large corpora of dialogue data, BiLSTMs can learn to recognize subtle cues in language, such as tone, word choice, and syntactic structures, that indicate specific intents or emotions [25]. These models allow the system to adapt to users' changing moods and needs. For instance, if a user becomes frustrated during an interaction, the system can detect this shift in emotion and adjust its tone or offer additional assistance, thereby improving the overall user experience [26]. This ability lays the groundwork for developing more intelligent, adaptive, and personalized HCI systems.

In the future, as AI technology continues to evolve, human-computer interaction systems will become even more intelligent and emotionally aware, gradually expanding into a broader range of applications. For example, advancements in multimodal learning, which combines text, speech, and visual inputs, will enable systems to understand user intent and emotions more holistically. This could lead to applications in healthcare, education, and entertainment, where systems can provide more personalized and context-aware interactions [27]. These advancements will not only enhance the user experience but also pave the way for more seamless, human-like interactions across various industries. As systems become more capable of understanding and responding to human emotions, they will play an increasingly important role in fostering trust and engagement in human-computer interactions [28].

3. Method

In the implementation of user intention prediction and automatic response systems in human-computer interaction, the BILSTM model is used to analyze and predict user intention. The core idea of BILSTM is to capture the context in the input data through bidirectional timing information modeling to improve the understanding of user input. The model architecture is shown in Figure 1.

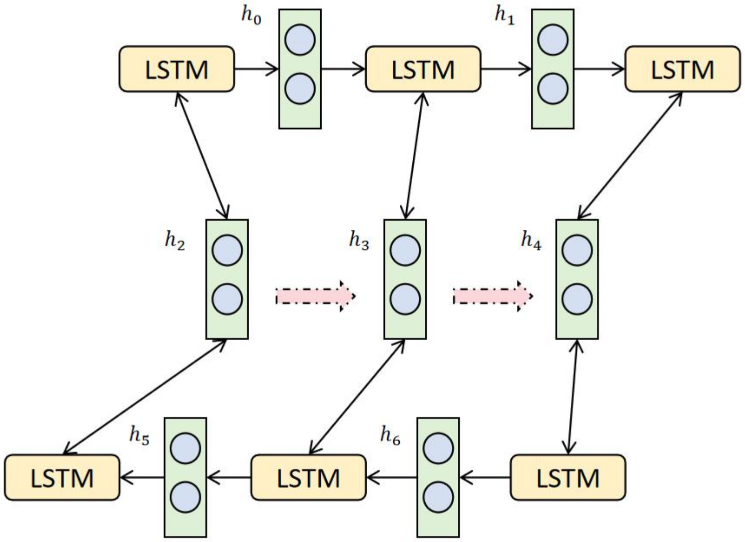


Figure 1. BILSTM model architecture in human-computer interaction

First, suppose we have an input sequence $X = (x_1, x_2, ..., x_T)$, where x_t represents input information (such as a feature of text or speech data) at time t. For each input, BILSTM first processes it through two LSTM units, forward and backward, capturing contextual information from the past (forward) and future (backward). We define the hidden state of the forward LSTM as $\vec{h}_t$ and the hidden state of the reverse LSTM as $\overleftarrow{h}_t$, and combine them to form the final output. Specifically, combining the forward and reverse states, the final representation is:

$$h_t = [\vec{h}_t, \overleftarrow{h}_t]$$

This bidirectional structure can make use of both past and future contextual information, thus effectively improving the accuracy of intention prediction.

In order to further improve the expressiveness of the model, we pass the hidden layer state h_t output by BILSTM to the fully connected layer for further feature extraction. Assuming the dimension of h_t is d, we use a weight matrix W and a bias term b to linearly transform the output of each time step, resulting in an output representation of the model z_t:

$$z_t = Wh_t + b$$

Next, for the regression task, we need to process z_t to predict the user's intentions. To achieve this, we sum or average the output of each time step to obtain a global representation of the entire input sequence. Assuming that the output sequence of the model is $Z = (z_1, z_2, ..., z_T)$, we can obtain the final prediction result by weighted averaging the output of all time steps:

$$y' = \frac{1}{T}\sum_{t=1}^{T} z_t$$

Where, y' represents the result predicted by the system to the user's intention. In practical applications, we can carry out follow-up processing on the predicted value according to the requirements of specific tasks and get the final automatic response.

To train the model, we use the mean square error (MSE) as a loss function to optimize the model parameters. Assuming y_{true} is the true label value, our loss function can be expressed as:

$$L = \frac{1}{N}\sum_{t=1}^{N}(y'_i - y_{true,i})^2$$

Where N is the number of training samples, y'_i is the predicted value of the model, and $y_{true,i}$ is the corresponding true label value. By minimizing the loss function L, the model can constantly update parameters during training and finally achieve accurate prediction of user intent.

In the process of model training, we use a gradient descent algorithm (such as Adam optimizer) to optimize the loss function. The gradient descent method optimizes the performance of the model step by step by calculating the gradient of the loss function to the model parameters and updating each parameter by the backpropagation algorithm. Assuming θ is the set of parameters of the model, the gradient update rule can be expressed as:

$$\theta \leftarrow \theta - \eta \nabla_{\theta} L$$

Where η is the learning rate and $\nabla_{\theta} L$ is the gradient of the loss function with respect to the parameter θ.

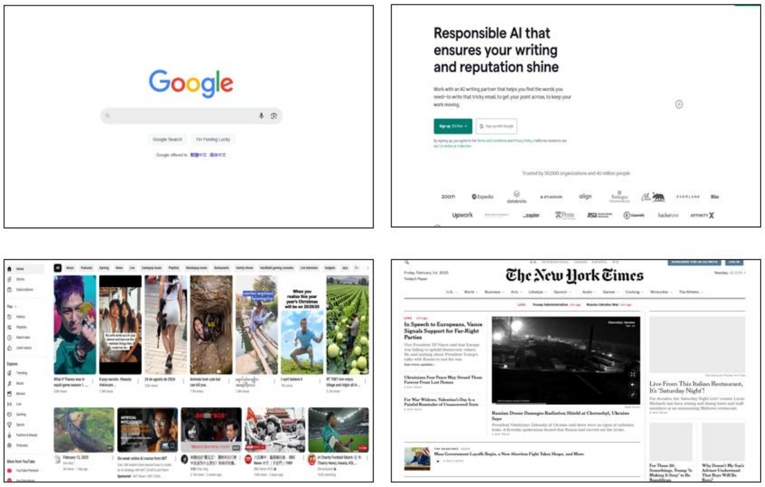
Through the above methods, we were able to effectively train the BILSTM model so that it can accurately predict the user's intentions and generate automatic responses. Throughout the process, BILSTM is able to capture not only the timing information in user input but also the forward and reverse context information to enhance understanding and response to complex interaction scenarios.

4. Experiment

4.1 Datasets

The data set used in this study consists of two main parts: human-computer interface data and user intent data. Human-computer interface data comes from the home page information of multiple platforms, including applications such as intelligent voice assistants, chatbots, and intelligent customer service. By collecting the home page data of each platform, the information of common interaction mode, function selection, and interface design of users on different platforms is collected. Each home page data records the user's interaction history with the platform, including features such as the functional module selected by the user, the type of instruction entered, and the frequency of operation. These data provide the model with rich contextual information, which helps to predict the potential needs and preferences of users, thereby improving the system's ability to understand user intentions. The platform home page data is shown in Figure 2.

The user intention data mainly comes from the user behavior record and interactive feedback in the platform. These data include the text, voice, or gesture data entered by the user during the interaction process, covering different interaction scenarios and situations. For example, text inputs might include queries like "What's the weather today?" or commands like "Play my favorite playlist," while voice inputs could range from simple requests to complex multi-sentence instructions. Gesture data, on the other hand, might include swipes, taps, or other touch-based interactions recorded on touchscreen devices. For each user input, the system will mark its corresponding intention label, such as query, operation instruction, emotional expression, etc. In addition, the user's behavioral data also includes the duration, frequency, emotional state, and other information of its interaction with the system, which can reflect the user's real needs and emotional tendencies. In the process of labeling, considering the ambiguity of user input, the division of intent labels is relatively detailed and can reflect the diversity and complexity of user behavior.



Responsible AI that ensures your writing and reputation shine

Figure 2. Human-computer interaction data presentation

Combining these two data sections gives us a more comprehensive understanding of user behavior patterns across different platforms and their corresponding needs. These data provide a rich sample for the training model, especially in the context of multi-round dialogue or multiple input forms, which can help the model learn more accurate user intention prediction ability. By comprehensively considering the design features of the human-computer interaction interface and the actual user intention data, this research can build a more accurate and intelligent user intention prediction and automatic response system. To further enhance the dataset's utility, we plan to release it as an open resource for the research community, along with detailed documentation and benchmarking results. This will enable other researchers to validate and build upon our work, fostering advancements in the field of human-computer interaction [29].

4.2 Experimental Results with User-Centered Evaluation

To ensure the representativeness and validity of our findings, we recruited a total of 250 participants across two primary settings: (a) real-world platform users and (b) laboratory-based volunteers. The participants were recruited through a combination of online advertisements, community outreach, and university mailing lists, ensuring a diverse and representative sample. The participants ranged in age from 18 to 60, with diverse educational backgrounds (high school through graduate degrees) and usage experience in online platforms. Participants were categorized into three main user groups based on their primary online behaviors:

1). News Browsers: Individuals who predominantly read news sites and online articles.

2). Video Viewers: Individuals who frequently watch short videos or live streams on social media or specialized apps.

3). Search-Engine Users: Individuals who rely heavily on search engines for daily queries and information retrieval.

Participants were asked to use their preferred devices in real or near-real scenarios to capture genuine interaction behaviors. In the real-world setting, data was collected via server-side logs (with appropriate privacy measures) covering users' clicks, time spent on pages, and textual inputs. In the laboratory setting, we designed controlled tasks to simulate different usage contexts—such as reading news articles, searching for specific information, or viewing recommended videos—and recorded their interactions via screen-capture software and questionnaires.

In the course of the experiment, this paper first carried out the user intention prediction experiment. Specifically, this paper trains the BILSTM model to predict the user's behavior intention on the home page. The experimental results are shown in Table 1.

Table 1: User intention prediction experiment

Platform type	**User Satisfaction (1-5)**	**Interaction Speed (milliseconds)**	**Intent Matching Accuracy%**	**User Engagement (minutes)**
Short video	4.5	200	85	15
news	4.2	250	80	10
Search engine	4.8	153	90	21
Functional website	4.3	226	82	12

According to the experimental results, the short video platform has an excellent performance in user satisfaction (4.5 points), indicating that the user experience of the platform is relatively good, and users are satisfied with the response speed and matching accuracy of the platform interaction. Although its intention matching accuracy is 85%, this data shows that the system can accurately understand the user's needs and respond quickly (200 milliseconds), which also helps improve the user's interaction experience. In terms of user engagement, the average user interaction duration of the short video platform is 15 minutes, indicating that the platform can effectively attract users to engage for a long time.

In contrast, the news platform has a low level of user satisfaction (4.2 points), and although its intention matching accuracy is 80%, slow response speed (250 ms) and relatively low user engagement (10 minutes) may be contributing to this result. Slow response times can affect the overall user experience, making the user's interaction with the platform feel less smooth and timely. In addition, although the accuracy of the news platform's intention matching is not bad, there may be some cases in which the information retrieval is not completely consistent with the user's needs, thus reducing the user's satisfaction. The search engine platform has the highest user satisfaction (4.8 points), the best intention matching accuracy (90%), and the fastest response time (153 milliseconds), which indicates that the search engine platform's algorithm and technical processing power have strong advantages in user experience. The search engine can accurately understand the user's search intent and respond quickly, effectively improving the user's interactive experience. The high level of user engagement (21 minutes) also further validates the advantages of the platform in attracting users to spend longer periods of time, and users are significantly more inclined to engage in more interactions.

In the end, the results of the functional website were relatively normal, with a user satisfaction score of 4.3, an intention matching accuracy of 82%, and a response speed of 226 milliseconds. Although its matching accuracy is high, the slightly slow response time may have had a negative impact on the user experience, causing users to be more hesitant to interact with the platform, which in turn affects user engagement (12 minutes). This relatively low user engagement time may reflect the need for further optimization of the platform in terms of functionality to enhance the user's interactive experience.

The results reveal that the search engine platform outperformed others, achieving the highest user satisfaction (4.8), fastest response time (153 ms), and best intent matching accuracy (90%), along with the longest user engagement (21 minutes). The short video platform also performed well, with high satisfaction (4.5) and quick response (200 ms), though its engagement (15 minutes) and accuracy (85%) were slightly lower. The news platform showed lower satisfaction (4.2) and engagement (10 minutes), likely due to slower response times (250 ms). The functional website had moderate performance, with decent accuracy (82%) but slower response (226 ms) and lower engagement (12 minutes), indicating room for improvement. These findings highlight the importance of balancing speed, accuracy, and engagement for optimal user experience.

Then, this paper conducts an experiment of correlation analysis between user behavior characteristics and intention prediction. Specifically, this paper conducts an analysis experiment based on user behavior characteristics, focusing on the correlation between home page information (such as user click modules, visit frequency, page stay time, etc.) and users' actual intentions. The experimental results are shown in Figure 3.

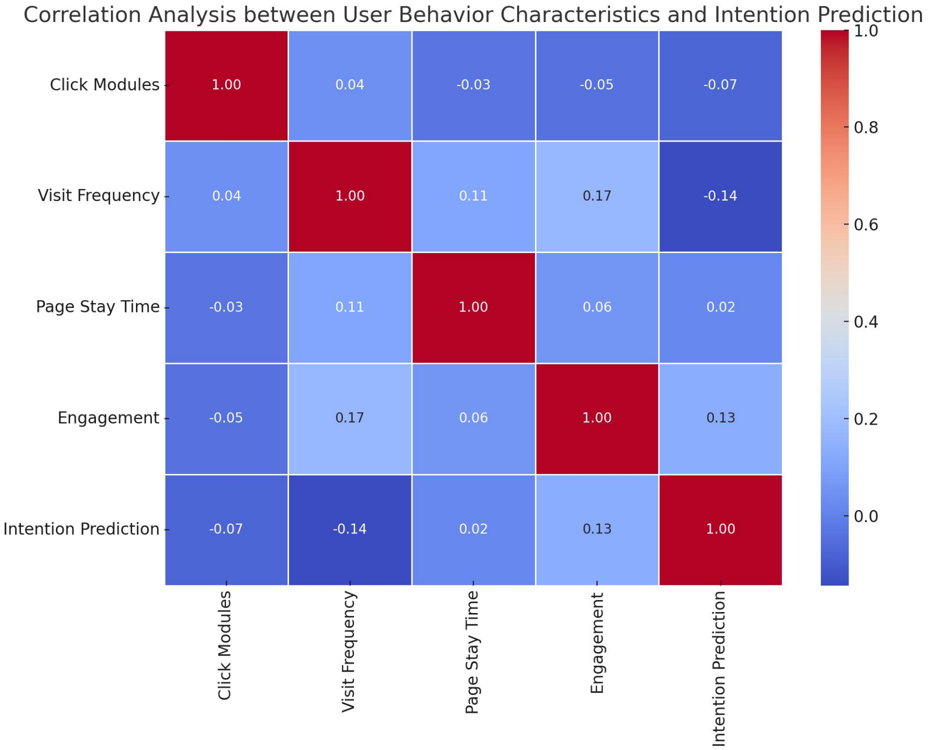


Figure 3. Correlation Analysis between User Behavior Characteristics and Intention Prediction

As can be seen from the figure, the correlation between various user behavior characteristics is low. For example, the correlation between click modules and visit frequency is only 0.04, indicating that there is little linear relationship between user click behavior and visit frequency. This may mean that the user's click module selection is not closely related to their frequency of visits, and users may choose different functional modules rather than necessarily visiting a particular page frequently. Similarly, the correlation between time spent on a page and other characteristics (such as click modules, visit frequency) is low, suggesting that a user's time spent on a page may not be directly linked to their other behavioral characteristics, and may be more influenced by personal needs or the attractiveness of the page's content.

The intent prediction in the graph has a generally low correlation with other user behavioral characteristics, up to 0.17, between visit frequency and engagement. This suggests that while there is some positive correlation, overall, the relationship between user behavioral characteristics (such as click modules, visit frequency, time on the page, engagement, etc.) and actual intent prediction is not strong. Models may require more behavioral data or complex feature engineering to improve the accuracy of intent predictions.

Finally, engagement had a relatively high correlation with other characteristics, such as time on the page and frequency of visits, but the relationship with intent prediction remained weak. Even if engagement slightly affected intention prediction, the overall correlation was still not strong, indicating that the system still has some challenges in understanding users' real needs and predicting intentions. In the future, the predictive power can be further enhanced by incorporating more behavioral features and more complex algorithmic models.

Further, this paper compares the intention response experiments of other models. The evaluation indexes used in this experiment are MSE, MAE, and R2, and the experimental results are shown in Table 2.

Table 2: Compare other model experiments

Model	**MSE**	**MAE**	**R2**
SVM	0.213	0.351	0.81
Random Forest	0.201	0.344	0.82
LSTM	0.172	0.249	0.90
BiLSTM	0.165	0.211	0.93

It can be seen from the experimental results that the BiLSTM model has the best performance in all evaluation indicators. MSE (0.165) and MAE (0.211) are both the lowest, indicating that the prediction error of this model is the smallest. At the same time, the R2 value of BiLSTM is 0.93, indicating a high degree of fit between the predicted result and the true value and showing a strong forecasting ability. This shows that BiLSTM is better able to understand and predict users' intentions and is suitable for tasks that require high-precision prediction.

In contrast, LSTM also performed relatively well, with MSE (0.172) and MAE (0.249) both lower than SVM and random forest models, but its R2 value (0.90) was slightly lower than BiLSTM. This indicates that LSTM can provide more accurate prediction when processing time series data, but compared with BiLSTM, its understanding of data context is slightly insufficient. Because BiLSTM is able to utilize both before and after information, it shows a stronger advantage in the intention prediction task.

SVM and random forest results are relatively poor. While their MSE and MAE were both below 0.25, their R2 values of 0.81 and 0.82, respectively, were much lower than LSTM and BiLSTM. This indicates that SVM and random forest models are relatively weak in terms of fit degree and prediction accuracy and may not be able to fully exploit the timing and complex context of data. Therefore, while these models may be stable for some tasks, deep learning models such as LSTM and BiLSTM undoubtedly provide a more accurate solution for user intent prediction tasks.

Finally, this paper also conducted an intention prediction experiment, using the click data to generate the click possibility on the page data, and the experimental results are shown in Figure 4.

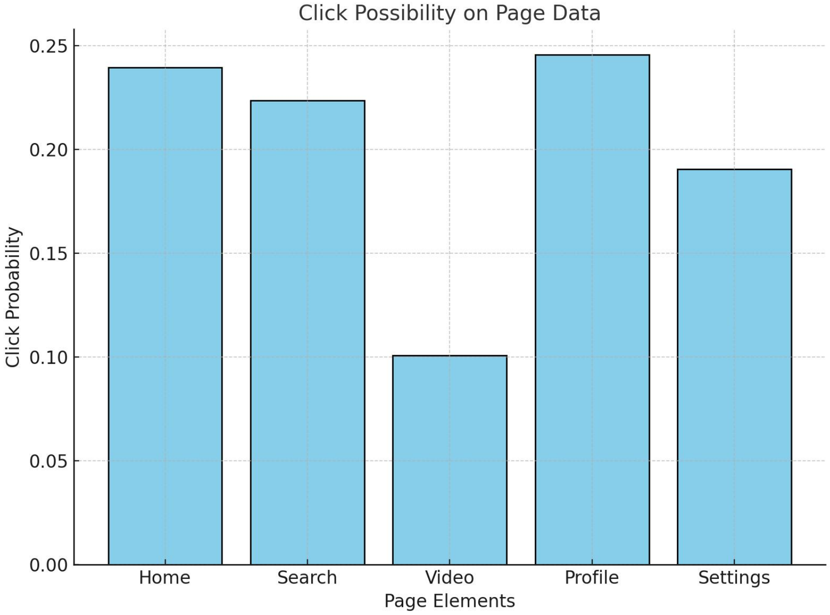


Figure 4. Click Possibility on Page Data

As can be seen from the figure, the click probability of Home and Search is relatively high, close to 0.25 and 0.22 respectively, which indicates that users have strong interest and activity on these page elements. As the entrance to the platform, the home page usually contains the core content that users care about, so it attracts more clicks. The search function is frequently used by users, especially when they need to find specific content, so it also shows a high probability of clicking.

In contrast, Video has the lowest click probability of 0.10, which may indicate that the user is relatively less interested in the video content when browsing the page, or the presentation of the video content is not attractive enough for users to click. It may be necessary to further optimize the presentation of video content or enhance the user's interactive experience to increase their click-through rate.

Furthermore, both the Profile and Settings sections exhibit comparable click probabilities of 0.21 and 0.20, respectively. This suggests that these elements hold significant importance to users, particularly in the context of personalization and accessing personal information. Although these two features may not be the main interaction content of the user on the page, they still engage the user to some extent. Therefore, the optimization and improvement of these page elements remain the key to improving the overall user experience.

5. Conclusion

By examining the connection between user behavior characteristics and intention prediction, this study highlights the significant potential of deep learning models in enhancing user intention prediction. This advancement contributes to improving the human-computer interaction experience, enabling more sophisticated and personalized services. The experimental results show that the BiLSTM model can accurately predict the behavioral intention, especially in multi-round interactions and dynamic environments. Compared with traditional models such as SVM and random forest, BiLSTM shows higher prediction accuracy and stronger adaptability. This finding validates the advantages of deep learning methods in understanding complex user needs and provides technical support for the optimization of human-computer interaction systems.

However, although existing models have achieved good results in intention prediction, there are still some challenges. First, the diversity and complexity of user behavior data remain a major challenge for predictive models. For example, in e-commerce platforms, user behavior can vary widely—some users may browse products casually, while others may have specific purchase intentions. Existing models often struggle to adapt to such diverse behaviors without extensive retraining. Although the existing model can capture part of the user's behavior pattern, the adaptability and robustness of the model still need to be further improved in the face of different user groups and more complex interaction scenarios. In the future, the combination of more multimodal data (such as vision, speech, etc.) and reinforcement learning methods may further improve the accuracy and intelligence of the model. For example, integrating facial expression analysis with voice data could enhance emotion-aware intent prediction in virtual assistants, making interactions more empathetic and personalized.

In addition, with the continuous advancement of human-computer interaction technology, future research directions may focus more on personalization and context-aware capabilities. For instance, in smart home systems, understanding a user's real-time context—such as whether they are cooking, relaxing, or working—can enable more accurate intent prediction and proactive assistance (e.g., adjusting lighting or playing music). User needs should not only be predicated on historical behavior but should also take into account the user's emotions and real-time feedback in the current situation. Like in educational platforms, detecting a student's frustration during a learning session could trigger adaptive content recommendations or motivational feedback, enhancing the learning experience. Therefore, it will be an important topic in the future to study how to realize personalized and context-aware intention prediction in the dynamic changing user environment. This requires cross-domain technology integration, such as emotional computing, context understanding, and social network analysis, to further promote human-computer interaction systems to become more intelligent and natural.

In general, with the continuous development of artificial intelligence technology, human-computer interaction systems will become more intelligent and personalized in the future. User intent prediction models based on deep learning will play an increasingly important role in improving interaction quality and enhancing user experience. Looking to the future, as technology continues to advance, user intent prediction will become more accurate and efficient, opening up broad prospects for intelligent system applications in all walks of life, enabling more personalized experiences in various industries such as e-commerce, advertising, smart home services, etc. From healthcare and education to entertainment and smart cities, the potential applications of advanced intent prediction models are vast, promising to transform how humans interact with technology in everyday life.

References

[1] Qu J, Guo H, Wang W, et al. Prediction of Human-Computer Interaction Intention Based on Eye Movement and Electroencephalograph Characteristics[J]. Frontiers in Psychology, 2022, 13: 816127.

[2] Miao X, Hou W. Human–Computer Interaction Multi-Task Modeling Based on Implicit Intent EEG Decoding[J]. Applied Sciences, 2023, 14(1): 368.

[3] Liu Z, Lou S, Feng Y, et al. A closed-loop human-computer interactive design method based on sequential human intention prediction and knowledge recommendation[J]. Journal of Engineering Design, 2024: 1-24.

[4] Şencan C. Intention mining: surfacing and reshaping deep intentions by proactive human computer interaction[J]. 2024.

[5] Kong D, Feng Z, Xu T, et al. Intentional Understanding and Human-Computer Collaboration: a smart pen for solid geometry teaching[J]. International Journal of Human–Computer Interaction, 2024, 40(22): 7668-7687.

[6] Liu Z, Lou S, Feng Y, et al. A closed-loop human-computer interactive design method based on sequential human intention prediction and knowledge recommendation[J]. Journal of Engineering Design, 2024: 1-24.

[7] Wang S, Niu H, Wei W, et al. Eye-Gaze-Based Intention Recognition for Selection Task by Using SVM-RF[C]//International Conference on Human-Computer Interaction. Cham: Springer Nature Switzerland, 2024: 157-168.

[8] Liu J, Yu X, Liang P, et al. Design and Experimental Validation of Brain-Computer Shared Control of a Robotic Arm Based on Motion Intention Prediction[C]//2024 10th International Conference on Mechatronics and Robotics Engineering (ICMRE). IEEE, 2024: 63-67.

[9] Talla R, Murthy B V R. Prediction and Analysis of Human Perception and Emotional Understanding Through Technology–Blue Eyes Technology[J]. International Journal of Interpreting Enigma Engineers (IJIEE), 2024, 1(1): 16-24.

[10] Sharma R, Tyagi S, Chaudhary S. Dialogue System for Human Computer Interaction[J]. JOURNAL OF TECHNICAL EDUCATION, 13.

[11] Sousa S, Lamas D, Cravino J, et al. Human-Centered Trustworthy Framework: A Human–Computer Interaction Perspective[J]. Computer, 2024, 57(3): 46-58.

[12] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation.

[13] Graves, A., & Schmidhuber, J. (2005). Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures. Neural Networks.

[14] Liu, B., & Lane, I. (2016). Attention-Based Recurrent Neural Network Models for Joint Intent Detection and Slot Filling. Interspeech.

[15] Vaswani, A., et al. (2017). Attention Is All You Need. NeurIPS.

[16] Devlin, J., et al. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL.

[17] Zhang, X., & Wang, H. (2016). A Joint Model of Intent Determination and Slot Filling for Spoken Language Understanding. IJCAI.

[18] Patil N, Ansari M W, Jadhav S R, et al. Gesture Voice: Revolutionizing Human-Computer Interaction with an AI-Driven Virtual Mouse System[J]. Turkish Online Journal of Qualitative Inquiry, 2024, 15(3).

[19] Chen, Y., et al. (2019). Deep Learning for Natural Language Processing in Customer Service. IEEE Transactions on Neural Networks and Learning Systems.

[20] Schuller, B., et al. (2018). Emotion Recognition from Speech Using Deep Learning. IEEE Transactions on Affective Computing.

[21] Al-Okaily M. So what about the post-COVID-19 era?: do users still adopt FinTech products?[J]. International Journal of Human–Computer Interaction, 2025, 41(2): 876-890.

[22] Serban, I. V., et al. (2016). Building End-to-End Dialogue Systems Using Generative Hierarchical Neural Network Models. AAAI.

[23] Picard, R. W. (2000). Affective Computing. MIT Press.

[24] Jurafsky, D., & Martin, J. H. (2020). Speech and Language Processing. Pearson.

[25] Mikolov, T., et al. (2013). Distributed Representations of Words and Phrases and their Compositionality. NeurIPS.

[26] Cambria, E., et al. (2017). Sentic Computing: A Common-Sense-Based Framework for Concept-Level Sentiment Analysis. Springer.

[27] Baltrušaitis, T., et al. (2019). Multimodal Machine Learning: A Survey and Taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[28] Nass, C., & Brave, S. (2005). Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship. MIT Press.

[29] Gebru, T., et al. (2018). Datasheets for Datasets. Communications of the ACM.