International Journal on Cybernetics & Informatics (IJCI) Vol. 11, No.1/2, April 2022

APPLICATION OF EMOTIONAL VOICE USER INTERFACE IN SECURITIES INDUSTRY

Ma Xueming, Chen Yao, Zhang Qian and Jiang Yun

Orient Securities Co., Ltd, Shanghai, China

ABSTRACT

Through the combination of sentiment algorithm and emotional design, the application and prospect of the intelligent service robot in the securities industry were explored. Based on user research methods, an emotional voice user interface system was constructed, which enabled the robot with the ability of sentiment recognition and feedback. As a result, the sentiment algorithm model achieved 97.18%accuracy. The Comparative experimental results showed that after optimization, the customers' investment inclination and satisfaction with the robot have been significantly improved. This will help the securities business department to effectively reach more customers while optimizing the service experience.

KEYWORDS

Voice user interface, intelligent service robot, sentiment algorithm, emotional design, securities industry.

1. INTRODUCTION

In the era of big data, artificial intelligence has become an important way to accelerate the rapid implementation of smart finance. In recent years, the intelligent transformation of business departments was first applied in the banking industry, while its application in the securities industry is still in the initial stage.

In 2021, Orient Securities introduced the intelligent service robot "Nai" as a window for the business department to communicate with customers, in order to provide customers with business consultation, investment teaching, and financial products introduction (Figure 1).However, through a year's data tracking, it was found that the intelligent robot lacked some anthropomorphic attributes, which made customers' cognition of it still in the contradictory state of "a tool" or "an intelligent service robot". In order to make customers have more natural and fluent experience in the dialogue, it is vital to make the robot have consistent personality and speech style, and give corresponding feedback to different emotions of customers, so as to generate real social interaction with customers.

Figure 1. Intelligent Service Robot "Nai"

David C. Wyld et al. (Eds): NLDML, CCITT-2022
pp. 15-29, 2022. IJCI – 2022 DOI:10.5121/ijci.2022.110202

The emotions generated by customers in the process of interaction will affect their impression and evaluation of the brand image, and then affect their desires and feelings of trading. According to the ABC Theory of Emotion proposed by psychologist Ellis, people's Consequences (C) are not directly induced by antecedents (A) but caused by false beliefs (B) generated by incorrect cognition and evaluation of it[1]. Good communication can make customers have positive impressions on financial brands and further stimulate investment enthusiasm and trust. On the contrary, negative emotion will arouse customers' suspicions and make them resist the follow-up operation. Therefore, emotional experience is an important factor that can not be ignored in the service industry.

Through the combination of Natural Language Processing (NLP) with the theoretical methods of Cognitive Psychology and Personality Psychology, a closed-loop "emotional" voice user interface system is established (Figure 2). The main purpose of the system is to build a complete personality system for the robot and optimize the design of "dialogue scripts". With the intelligent service robot ‘Nai’ as the front-end hardware device, the customers' voices will be converted into text by Automatic Speech Recognition (ASR). The end-to-end method based on deep neural network uses Hidden Markov Model (HMM) recognition to input-encode-decode-output voice data. After answering the questions by natural language processing technology, the corresponding text enters the emotional feedback stage. Text is converted into human speech by Text To Speech (TTS), so that the robot can identify the customer's emotional state, and use corresponding speech to answer.

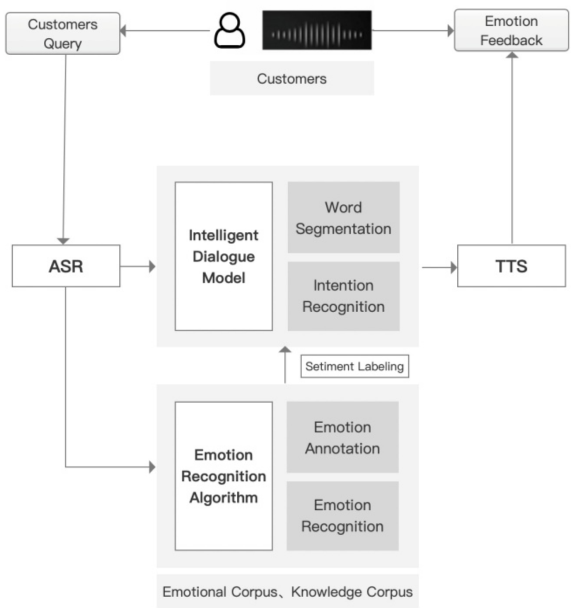

Figure 2. Emotional Voice User Interface System

The emotional design system is carried out from three aspects: emotion annotation, emotion recognition, and emotion feedback, focusing on the following objectives (Figure 3):

(1) A sentiment algorithm model of the intelligent robot is constructed, which fills the blank of corpus in securities industry.
(2) Using the theory of personality psychology, innovatively put forward the robot emotional system, which endows the robot with more personality attributes.

(3) The emotional dialogue scripts are designed and applied in the process of emotional feedback to establish the closed-loop "emotional" voice user interface system.
(4)

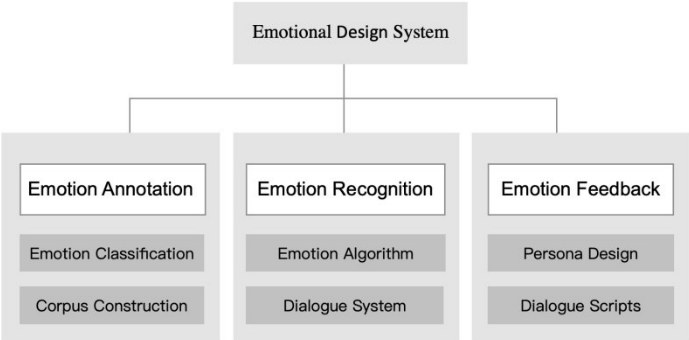

Figure 3. Emotional Design System

Making the robot have the ability of emotion recognition and feedback during the dialogue with customers is helpful to change the dialogue scene from passive input to active feedback, thus bringing active, smooth, and caring services to them.

2. EMOTION RECOGNITION ALGORITHM

2.1. Emotion Annotation

The effectiveness of the algorithm model mainly depends on the quality of the training set. In order to make the model recognition more suitable for real and abundant scenarios, two parts of data were selected as training sets for the emotional model. The original corpus was taken from the real customers' conversations of the background data. The other part was the chat data from Weibo, which complemented the richness of answers and modal particles.

Through clustering analysis of background data, the intelligent voice interaction in the securities business department is divided into two categories, namely, task-based conversation and chat-based conversation. The task-based conversation occurs in financial business scenarios, including company introduction, business department introduction, stock market consultation, business consultation, investment teaching, account opening, and fund products introduction, etc. While the chat-based conversation take places in customer guidance scenarios, including basic greeting, robot self-introduction, robot function, robot personification, and chatting (Figure 4).

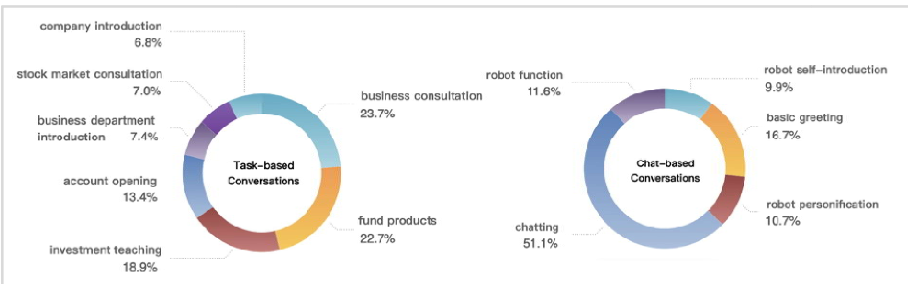

Figure 4. Conversations Classification of the Securities Business Department

After the original corpus is established, it is necessary to label the corpus with different emotional categories. In the research of sentiment algorithms, emotion tagging corpus can be roughly divided into two categories, one is one-dimensional emotion tendency corpus, and the other is multi-dimensional composite emotion corpus. For example, Mishne marked 132 kinds of English emotions on the Live Journal blog system [2]. Tagging forms are also more diverse, such as emotional symbols of Japanese blogs tagging [3], emoji expressions tagging [4], and so on.

Intelligent conversations in different scenes meet different users' needs, so the range of emotion recognition is also different. For example, the main function of the chat robot is companionship, so it has the most extensive emotional recognition. Tmall Elf subdivides emotions with a three-dimensional model of emotions (pleasure, arousal and domination)[5]. The main functions of home robots are linking and manipulating, so there are relatively few requirements for emotion recognition. At present, there is no research on emotional tagging in the securities industry, so it is very necessary to build a corresponding emotional corpus.

Emotional conversations of the securities business departments are mainly aimed at reducing customers' negative emotions and enhancing their goodwill towards the corporate financial brands. Therefore, this survey uses emotional disposition to divide the questions into positive, neutral, and negative labels, so as to construct the emotional corpus. On the one hand, if the customers have negative emotions, the robot needs to appease and guide them in time. On the other hand, if the customers show positive emotions, the robot needs to respond positively to further stimulate their interest.

In order to reduce the error and control the consistency, the corpus was annotated by the master of psychology and the master of algorithms at the same time. After the two sides' exchanges were marked, the different results were discussed and determined in combination with the opinions of another algorithm doctor. Finally, a corpus containing 10, 0516 emotional disposition labels were formed for algorithm training of emotion recognition.

2.2. Emotion Recognition Algorithm

With the help of the pre-training method, emotion recognition has made remarkable progress. However, prior knowledge of emotion, such as emotion words and polarity of emotion words, is neglected in most pre-training processes, although in fact they are widely used in traditional emotion recognition methods.

In this emotion recognition algorithm, the sentiment knowledge enhanced pre-training for sentiment analysis (SKEP) based on emotion knowledge enhancement is used. Through unsupervised learning, the emotion prior knowledge is automatically mined, and then the pre-training target is constructed by using the prior knowledge so that the machine can better identify emotions [7].

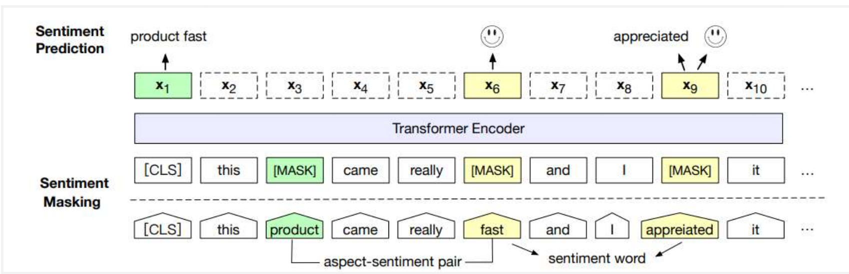

Figure 5. SKEP [0]

As shown in figure 5, SKEP consists of two parts: (1) Sentiment masking identifies the emotional information of the input sequence based on automatically mined emotional knowledge, and generates a masked version by deleting this information. (2) Sentiment prediction requires a transformer encoder to restore the masked version of its information. Three prediction targets are jointly optimized: sentiment word (SW) prediction (X9), word polarity (SP) prediction (X6 and X9) and viewpoint collocation (AP) prediction (X1).

The concrete steps of building SKEP are as follows: Firstly, based on statistical methods, emotional knowledge is automatically mined from a large number of unmarked data, including emotional words (like emotional words fast and implied), emotional word polarity (like positive polarity corresponding to fast) and opinion collocation (like binary group composed of < product, fast >).

Then, based on the emotion knowledge automatically mined, SKEP masks some words in the original input sentence, that is, replaces them with special characters [MASK]. In addition to masking words or continuous fragments like traditional pre-training, SKEP also masks such goals as opinion collocation.

Finally, SKEP designs three emotional optimization objectives, which requires the model to restore the masked emotional information, including: the viewpoint matching prediction based on multi-label optimization (As shown in Figure5, X1 predicts < product, fast > emotional collocation.); sentiment prediction (X6 predicts fast); sentiment polarity classification (X6 and X9 predict the sentiment polarity of this position).

In this way, by pre-training the sentiment-oriented optimization goal, the automatically mined sentiment knowledge is effectively embedded into the semantic representation of the model, and finally, the sentiment-oriented semantic representation is formed.

After the SKEP model training is completed, the classification layer is added to the transformer encoder, and the sentiment probability is calculated according to the output semantic representation, so as to identify the sentiment of the text [8][9]. Finally, the sentiment algorithm model is evaluated, and the accuracy of the data set is as follows 97.18%, which indicates that the sentiment recognition effect of this model works well.

2.3. Intelligent Dialogue System

The typical structure of an intelligent dialogue system includes four key components. They are natural language processing, dialogue state tracking, dialogue strategy learning, and natural language generation [10].

Firstly, in the aspect of natural language processing, it parses and maps the text language input by users to semantic slots, which are predefined according to different scenes. Figure 6 shows an example of inputting an English natural language representation of "show restaurant at New York tomorrow" to the machine, where "New York" is the location designated as the slot value, and the domain and intention are specified respectively. There are two typical expression types, and one is discourse-level category, such as user's intention and discourse category. The other is text information extraction, such as named entity identification, slot value filling, and so on. Dialogue intention detection is the segmentation of discourse into a predefined meaning.

Sentence	Slot Value	Intention	Domain
show	O	Find Restaurant	Order
restaurant	O		
at	O		
New	B-desti		
York	I-desti		
tomorrow	B-date		

Figure 6. Example of Natural Language Representation

Dialogue status tracking is the core component to ensure the normal operation of the dialogue system [11]. It will estimate the user's goal in each round of conversation, manage the input and conversation history of each round, and output the current conversation state [11]. This typical state structure is often referred to as slot filling or semantic framework. Traditional methods have been widely used in most businesses, and manual rules are usually used to select the most possible output results. However, these rule-based systems are prone to frequent mistakes, so the results are not always satisfactory. In this survey, deep learning and transfer learning methods are used to realize dialogue tracking, and the effect can basically reach the level that can be accepted.

According to the state representation of the state tracker, strategy learning is to generate the next available system operation. Both supervised learning and reinforcement learning can be used to optimize strategic learning. Supervised learning is aimed at the behaviors generated by rules. In the online shopping scene, if the dialogue status is "recommended", then the "recommended" operation will be triggered, at which time the system will retrieve the products from the product database. In this survey, the reinforcement learning method is used to further train the dialogue strategy, which guides the system to formulate the final strategy. In the actual experiment, the effect of the reinforcement learning method is better than that of rule-based and supervision-based methods.

Natural language generation refers to mapping the selected operation and generating a response. A good generator usually depends on such factors as appropriateness, fluency, readability, and variability[10]. The traditional natural language processing method is usually to execute a sentence plan, which maps the input semantic symbols to the intermediary form representing the language, such as tree or template structure, and then transforms the intermediate structure into the final response through surface implementation. It uses the encoder-decoder form of deep learning long-term and short-term memory network to combine question information, semantic slot values, and dialogue behavior types to generate correct answers. At the same time, it also uses an attention mechanism to deal with the key information of the current decoding state of the decoder and generates different responses according to different behavior types.

3. EMOTIONAL FEEDBACK DESIGN

3.1. Personalization of Intelligent Robot

3.1.1. Personality Construction Theory

After recognizing the sentiment input by customers, the robot needs to output emotional feedback to form an effective closed loop. It needs to have the same personality to keep a consistent conversational style. Therefore, based on the theory of personality psychology, this survey builds a complete personality system for the robot and optimizes the design of the "dalogue script" according to this system.

Personality is a unique integration model that constitutes a person's thoughts, emotions, and behaviors, with uniqueness, stability, integration, and functionality. Personality can't be directly observed, but it can be inferred from external behavior patterns. The ever-changing behavior and the stable inner quality constitute personality [12]. A robot's personality can be embodied in many ways, such as being seen through visual interaction (appearance and expression); being heard through voice interaction (voice and conversation); and being felt through gesture interaction (action, emotion, and multi-modality).

Personality theory can be roughly divided into two categories. Personality typology theory is to divide personality into several types according to a certain standard, and each individual belongs to only one type. On the contrary, the theory of personality trait does not divide personality into absolute types but holds it that it is composed of many traits as a lasting and stable behavior tendency.

In personality typology theory, Myers–Briggs Type Indicator (MBTI) is widely used. Myers divides personality into four dimensions, and each dimension was divided into two opposite tendencies. Sixteen different personality types can be formed when all tendencies were combined [13]. Personality trait theory originated from an experimental study put forward by Allport in 1915. He believes that personality is composed of many trait units, each trait forms a trait curve, and individuals are in a certain position on the trait curve [14]. For individuals, personality includes cardinal traits, central traits, and secondary traits.

In order to establish an intuitive and abundant personality system for the robot, this survey combines Myers–Briggs Type Indicator and Allport Personality Traits Theory as the foundation of the personification system.

3.1.2. Establishment of Personification System

Based on the personality theory, combined with a variety of user research methods, a complete personalized design methodology of the intelligent robot is established (Figure 7). Firstly, the role elements of robot characters are refined through in-depth interviews and desktop research. Then, in the form of an intelligent robot persona workshop, 17 experienced experience designers and algorithm engineers are recruited as expert user representatives. Questionnaires, personality tests, and storyboards are used to fill in personality information for robots and conversation scripts designing. Through algorithm mapping, the problem is combined with emotional conversation packages, so that the robot can give corresponding feedback to different emotions of users.

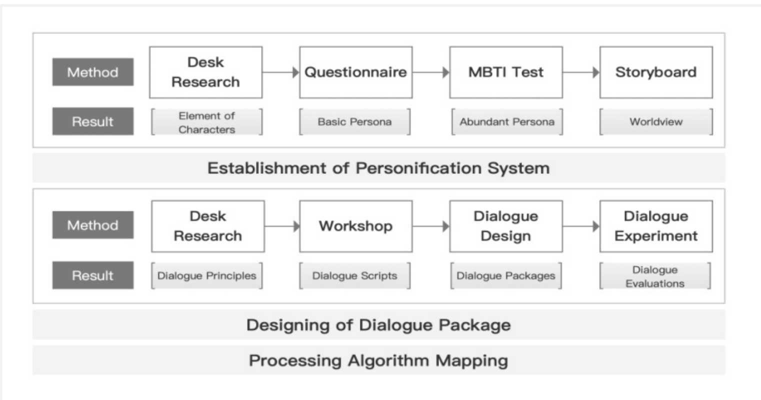

Figure 7. Design Method of Personification System

Referring to the process of Alibaba's IP design, the personification system including basic persona, abundant persona, and worldview was established [15]. Element of characters is an important direction for the establishment of persona. On the one hand, it serves as an image tone to assist the extension, on the other hand, it is used to gather divergent thinking and ensure the unity of robot personality. The robot design should not only meet the positioning of the business scene but also meet the needs and expectations of customers. In order to better locate the demand, an in-depth interview is conducted with the investment consultants of the business department. Combined with the preliminary insight into the background dialogue data of the robot, the elements of the intelligent service robot in the securities business department are refined from the following five aspects: friendly, professional, energetic, positive, accompanied.

The elements of basic persona are mainly the natural attributes and social attributes of robots, which are used to convey the basic external image of robots and answer the question "How to be like a person". Through the online questionnaire and semi-structured discussion, the understanding and cognition of expert participants about robots are collected, including gender, age, education, occupation, birthday, constellation, blood type, and other information.

Abundant persona answers the question "How to have an interesting soul". First of all, the experts were asked to answer the MBTI personality questionnaires by substituting the robot role. According to the results of the questionnaire, the core personality characteristic of the robot is "ESFJ executive". This personality is altruistic, helpful, and responsible, which fits well with the service attributes of the intelligent robot. Then, based on Allport's personality trait theory, the personality of "ESFJ executive" was diverged, and word cloud clustering analysis was used to conclude the traits. The result shows that the cardinal trait of the robot is "trustworthy", and the central characteristics include intelligence, efficiency, initiative, friendly, logical, well-mannered, and so on.

People's psychological demands are divided into five levels from low to high by psychologist Maslow: physiology, safety, belonging and love, respect and self-realization. The needs of robots can also be layered. The basic needs are living habits, and the intermediate demands include hobbies and interpersonal relationships. The advanced requirements are self-fulfilling such as ideals and mottos. According to the different demands, the story versions with five themes were designed to collect the clues of the robot's worldview and were generated into the character stories, so as to complete the objective world construction of the robot. Finally, all the personality attributes have been summarized and the personality portrait of the intelligent service robot in the securities business department has been established.

3.2. Emotional interactive dialogue design

3.2.1. Dialogue Principles

Like the way of constructing the personification system, before the dialogue design, it is necessary to refine the principles first, so as to make the dialogue unified and natural. The principles of the voice user interface of smart speakers are regarded as the main reference because of their most extensive interaction scenes with users, such as Google Conversation Design, Amazon Conversation Design, Baidu DuerOS system principles, etc. In this survey, the DuerOS system principle which is most appropriate to the scene of the intelligent service robot in the securities business department is selected as the main reference. This principle gives consideration to both rationality and sensibility in conversations. The rational principle means "usefulness", emphasizing goal-centered, accurate, and concise, while the perceptual principle is the "pleasantness" of dialogues, emphasizing nature, friendly, and individuated [16].

In addition to the basic dialogue principles, conversations also need to meet the transformation goals of targeted business scenarios. During the question-and-answer process, robots sometimes can't cover all personalized questions, and can't help users to handle business directly. If the robot can't find the answer, guiding can be used as an alternative. This is also reflected in the speech experiment test of the DuerOS system. The results show that the strategy of intelligent voice conversation at this time should not be to end but to tell the users what to do next or provide relevant alternatives at the same time [16]. Therefore, the "guidance principle" has been added to the dialogue principle system of intelligent robots in the securities business department, which means that speech can arouse customers' desire to continue the dialogue and also provide solutions to problems, such as telling customers to download App by themselves or looking for professional consultants for help. Finally, 10 indicators are determined as: Usefulness (accuracy, effectiveness, naturalness, and coherence); Pleasantness (fault tolerance, friendliness, and emotionality); and Guidance (continuity, reliability, and personalization). The indicators are used as the topic of the evaluation questionnaire for data collection.

3.2.2. Emotional Dialogue Design

In order to make the robot's speech conform to the unified personality setting, conversational characteristics of natural language, and communication habits in dialogue scenes, based on the robot's personality system and dialogue design principles, this survey designs corresponding dialogue packages for different scenes including pick-up, bottom-up, question-and-answer and chat dialogue scripts.

In addition, in order for the robot to make corresponding feedback on the identified emotions, it is necessary to superimpose different emotional dialogue scripts on the basic dialogue package. For customers' questions with positive emotions, the robot will use pleasant answers, such as "Congratulations on opening my hiding skills!". On the contrary, if the robot recognizes that customers have negative emotions, it will use warm words like "Hey, Nai is always there" to slow down the customers' avoidance behavior caused by negative emotions. By matching the sentiment algorithm with the identified emotion labeling problem, the robot can automatically trigger the switching of the corresponding emotional corpus when it receives the corresponding emotion of the dialogue. Before giving regular answers, emotional feedback will be served first, so as to improve the experience of emotional voice interaction.

4. RESULTS EVALUATION

In order to evaluate the effect of the sentiment algorithm and dialogue design, besides the accuracy of the algorithm (97.18%), the optimization effect was also evaluated by a comparative experiment. A structured questionnaire was used to compare the experience index scores before and after optimization, and the promotion scores and significance were calculated.

4.1. Participants

36 non-research-related users were recruited for the evaluation experiment. The age of users ranged from 25 to 35 years old, with a balanced gender distribution (47.2% males). A total of 36 valid questionnaires were collected, and the effective rate of the questionnaires reached 100%.

4.2. Materials

The evaluation material is the dialogue process between the customers of the securities business department and the intelligent service robot "Nai". Firstly, three different portraits of customers were designed, including gender, age, occupation, investment experience, and investment amount. Then, the dialogue scenes were generated according to different customers' demands. Each scene was made for pre-test (A) and post-test (B) materials separately. Finally, the evaluation materials of six dialogue scripts were formed. Code each scene, namely S1-A, S1-B, S2-A, S2-B, S3-A, and S3-B.

4.3. Procedure

After introducing the test background and requirements, participants were asked to read the dialogue scripts of each scene in turn and fill in the dialogue evaluation questionnaire.

In order to avoid the influence of individual differences among participants, the experiment adopted the method of within-group design, and each participant completed two evaluation questionnaires of pre-test and post-test in turn during the experiment. At the same time, in order to eliminate the sequential effect, half of the participants evaluated the pre-test materials first and the other half evaluated the post-test materials first.

Each participant was assigned two different pre-test and post-test scenarios. Because the scene was an irrelevant variable, the Latin square balance method was used to balance the six kinds of materials.

After participants finished the first scene topic, they were asked to watch a 5-minute comedy short film as a blank insert, and then evaluate the next scene. Finally, the purpose of the experiment was introduced and the participants were thanked.

4.4. Indicators

By means of the questionnaire survey, the pre-test and post-test comparison indexes of robot dialogue scenes were collected. Firstly, the evaluation indexes of intelligent voice conversations were collected by literature retrieval. Then, screened according to the scene requirements of the securities business department, and defined the meaning of each indicator. The task-based dialogue focused on the dimension of usefulness and guidance. The chat-based dialogue focused on the dimension of pleasure and guidance.

The questionnaire was divided into four parts. The first part and the fourth part are open topics, the main purpose of which is to collect users' original impressions of intelligent service robot and their feelings after the experiment. The second part is about the customers' willingness of downloading the App and asking for further consultation. According to the principles of dialogue scripts design (usefulness, pleasantness, and guidance) determined in Chapter 3, the third part is the intelligent dialogue experience scores, including 10 items of experience evaluation indicators and the satisfaction score, all of which are scored by Likert 5-point scale.

4.5. Results

The reliability of 36 questionnaires was tested by SPSS19.0, and the Cronbach coefficient of 13 items was 0.93, which indicated that the result had high reliability. Descriptive statistics, paired sample T-test, and regression analysis were carried out on the pre-test and post-test data, and the results were as follows.

4.5.1. Investment Inclination

The customers' investment inclination is divided into two dimensions, including the willingness to download the App and to enter the business department for further consultation. As shown in Table 1, before optimization, the customers' willingness to download App is low (M = 1.94, SD = 0.79), which may be due to the insufficient guidance to download the App. Customers' willingness to enter the store for consultation is at a medium level (M = 3.06, SD = 0.89), which indicates that they have the idea of asking for help but lack the motivation to enter the business department.

After optimization, customers' willingness to download App is significantly improved (M = 3.67, SD = 0.86). Paired sample t-test is conducted on the pre-test and post-test data, and the result shows that there is a significant difference between the two groups ($t = -10.3$, $p < .001$). It shows that the robot's speech plays a good role in promoting the download rate of the App, as shown in Table 1.

After optimization, customers also have a higher willingness to enter the business department for consultation (M = 3.65, SD = 0.96), and it is significantly higher than that before the optimization ($t = -2.85$, $p <. 01$), which indicates that the robot-guided speech has become an effective booster for customers' behaviour. From the difference between the two transformation dimensions, customers' willingness to download App before the optimization is significantly lower than their willingness to enter the business department for consultation ($t = -7.80$, $p <.001$). After optimization, both of them are promoted to a higher level with no significant difference, as shown in Table 1.

Table1. Difference test of customers' intention and satisfaction

	Stage	*M*	*SD*	*t*	*p*
Intentionto download app	*Pre-test*	*1.94*	*0.79*	*-10.30*	*0.000*
	Post-test	*3.67*	*0.86*		
Intention to consult	*Pre-test*	*3.06*	*0.89*	*-2.85*	*0.007*
	Post-test	*3.65*	*0.96*		
Satisfaction	*Pre-test*	*2.47*	*0.91*	*-7.08*	*0.000*
	Post-test	*3.69*	*1.06*		

4.5.2. Satisfaction

Before optimization, the overall score of customers' satisfaction is low (M = 2.47, SD = 0.91), and the proportion of satisfied customers (Top2 options) is 13.9%. After optimization, the overall score becomes higher (M = 3.69, SD = 1.06), and the proportion of satisfied customers reaches 63.9%, which is 50% higher than before. The paired sample t-test of pre-test and post-test data shows that the customers' satisfaction has been significantly improved ($t =-7.80$, $p <.001$), as shown in Table 1.

After optimization, it shows that there is a significant positive correlation between customers' satisfaction and intention. Among them, the willingness of downloading App is positively correlated with the satisfaction score ($r = 0.63$, $p <.01$), and the willingness of consulting is positively correlated with the total satisfaction score as well (r = 0.73, p <.01), indicating that the more satisfied customers are, the stronger their willingness to transact, as shown in Table 2.

Table 2. Description statistics of customers' intention and satisfaction

	M	*SD*	1	2	3
1.Satisfaction	3.69	1.06	1		
2.Intentionto download app	3.67	0.86	0.63^{**}	1	
3.Intention to consult	3.65	0.96	0.73^{**}	0.61^{**}	1

4.5.3. Indicators Evaluation

The evaluation scores on all dimensions of intelligent conversation experience are consistent before and after optimization. The dimensions with the highest scores are friendliness, emotionality, and coherence. The lowest dimensions are personalization, reliability, and guidance, as shown in Figure 8. After optimization, there is no significant difference between pre-test and post-test in each dimension, but the improvements of accuracy, personalization, and naturalness are the highest, which shows that by enriching the dialogue contents, the robot has improved the speech strategy for different customers to some extent. Moreover, the addition of colloquial chat corpus also makes the process of dialogue experience more natural and coherent, as shown in Figure 8.

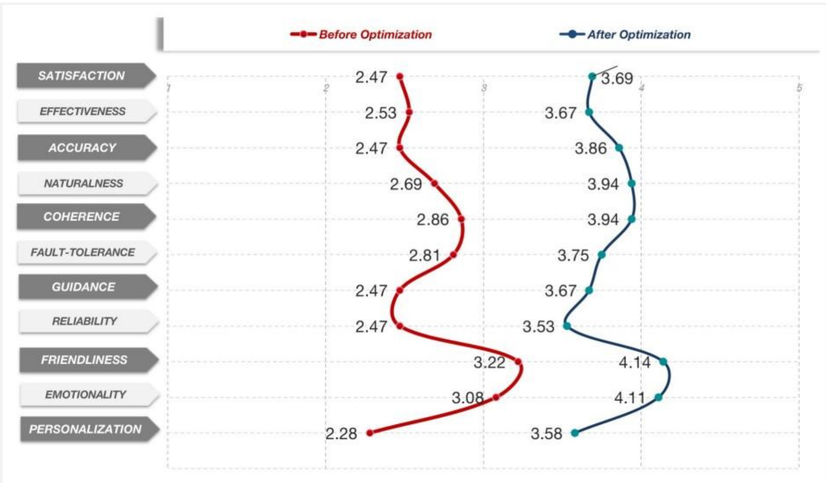

Figure 8. Descriptive statistics of experience indicators

Regression analysis of the total score of satisfaction with each index dimension shows that "consistency" and "reliability" can significantly predict the satisfaction index, which shows that improving the experience of intelligent dialogue in these two aspects has a more obvious effect on improving customers' satisfaction.

5. CONCLUSIONS

This survey focuses on the scenes of the securities business department, and optimizes the speech function of the intelligent service robot "Nai" from the aspects of emotion annotation, emotion recognition, and emotion feedback design, so as to enhance the experience value of customers. The main conclusions of this survey are as follows：

(1)The sentiment algorithm model of the intelligent robot in the securities business department is carried out, and the accuracy of the data set is as follows 97.18%.
(2)With the theory of personality psychology, a closed-loop "emotional" voice user interface design system is established, which enables the robot with the ability of emotional recognition and feedback.
(3)The emotional dialogue scripts are designed and applied in the robot. Paired sample T-test and regression analysis were used on the pre-test and post-test data. The comparative experimental results show that after optimization, the customers' investment inclinations have been significantly improved($t = -10.30$, $p < .001$), and the customers' satisfactions with the robot have also been significantly improved ($t = -7.80$, $p < .001$).

6. DISCUSSION

This research is a preliminary innovative attempt at intelligent service robots in the securities industry. The emotional algorithm model in this paper is based on the SKEP model, which can automatically learn the relationship between word meaning and word order, with strong generalization ability and high accuracy of emotion recognition. Through the emotional algorithm model, the robot can effectively gain insight into the customer's positive and negative emotions, and map them to the corresponding speech package, thus giving customers a more "temperature" answer.

However, due to the lack of corpus data in the real scenes involved in this survey, there is still room for optimization of the natural language processing model. In the future, it is necessary to continuously collect real data to optimize the NLP model in order to improve its accuracy. By collecting the corpus of the corresponding field, the pre-trained model can be customized and applied to the more specified vertical fields.

In addition, this survey mainly uses language and text information for identification, and the dimension of emotion recognition is insufficient. The hardware upgrade of the robot can collect more extensive environmental information around it. Subsequent research can help robots identify the user's expression and posture through AI vision algorithm, identify the user's intonation and tone through ASR speech recognition technology algorithm, and identify the user's heartbeat and blood pressure through various sensors, so as to recognize the user's mood more comprehensively and accurately. The latest research shows that the electroencephalogram-based emotion recognition has become crucial in enabling the Human-Computer Interaction (HCI) system to become more intelligent [17]. For instance, the one-dimensional EEG data can be converted to Pearson's Correlation Coefficient (PCC) featured images and be fed into the CNN model to recognize emotion [18].

The integration of algorithms and psychology will also be more and more widely used in the human-computer interaction. In this survey, the emotion recognition algorithm is applied to the intelligent dialogue system, which helps the dialogue robot to select the text that matches the user's emotions better. Along this line of consideration, the emotion recognition algorithm can be applied to more scenes. In the case of intelligent customer consultation service, it can be used to inspect and monitor the customers' negative dissatisfaction so that the manual customer service intervention can be triggered. In the case of manual customer service, it can also be used to monitor the service attitude of customer service personnel. It is believed that more business scenarios are waiting to be explored in the future.

ACKNOWLEDGEMENTS

We would like to express our gratitude to all those who helped us during the process of researching and writing this thesis. Special acknowledgements should be shown to the engineers in System R&D Headquarters, who gave us kind encouragement and useful instructions all through the thesis. We also want to thank the UED teammates for their friendly assistance and punctual information that helped us through the whole process of the experiments. Last but not least, sincere gratitude should also go to the reviewers, whose valuable suggestions have greatly helped us in the process of revising this paper better.

REFERENCES

[1] Ellis A. Overview of the Clinical Theory of Rational-emotive Therapy, In:R. Grieger (eds) Rational-emotive Therapy: A Skills-based Approach, New York, Nostrand Reinhold Company, 1980.
[2] Mishne G. Experiments with mood classification in blog posts [C] // Proceedings of ACM SIGIR 2005 Woekshop on Stylistic Analisis of Text for Information Acess. 2005.
[3] Ptaszynski M, Rzepka R, Araki K et al. Automatically annotating a five-billion-word corpus of Japanese blogs for sentiment and affect analysis [J]. Computer Speecj & Language, 2014, 28 (1) : 38-55.
[4] Jiaxiong Hu, Qianyao Xu, et al. Emojilization: An Automated Method For Speech to Emoji-Labeled Text [C]. CHI 2019, May 4–9, 2019, Glasgow, Scotland, UK.
[5] Mehrabian, A. and Russell, J.A. (1974) An Approach to Environmental Psychology. The MIT Press, Cambridge.
[6] Tian H , Gao C , Xiao X , et al. SKEP: Sentiment Knowledge Enhanced Pre-training for Sentiment Analysis [C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020.
[7] Peter D Turney. 2002. Thumbs up or thumbs down?:semantic orientation applied to unsupervised classification of reviews. In ACL 2002.
[8] Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel Bowman. 2018. GLUE: A multi-task benchmark and analysis platform for natural language understanding. In Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP.
[9] Zhengyan Zhang, Xu Han, Zhiyuan Liu, Xin Jiang, Maosong Sun, and Qun Liu. 2019. ERNIE: Enhanced language representation with informative entities. In ACL 2019.
[10] A. Bordes, Y. L. Boureau, and J. Weston. Learning end-to-end goal-oriented dialog. In ICLR, 2017.
[11] D. Goddeau, H. Meng, J. Polifroni, S. Seneff, and S. Busayapongchai. A form-based dialogue manager for spoken language applications. In Spoken Language,1996. ICSLP 96. Proceedings., Fourth International Conference on, volume 2, pages 701–704. IEEE, 1996.
[12] Dengfeng Wang, Hong Cui. The definition of personality and the differences between China and the West[J]. Psychological Research, 2008, 1(1): 3-7.
[13] Myers IB, McCaulley MH. MBTI Manual: A guide to the development and use of the Myers-Briggs Type Indicator. Consulting Psychologists Press, 1998:9-10.
[14] Allport, G.W. Personality: Apsychological International[M] . New York: Holt, Rinehart & Winston, 1961.
[15] Xiandou. Alibaba Design Ucan 2020: The evolution of brand image IP in Z Era. https://v.youku.com/v_show/id_XNDc0NzgyNjU1Ng==.html?spm=a2hzp.8253869.0.0
[16] Yanyan Sun, Shiyan Li. Emotional Voice Interaction Design: Human computer interaction research map and design case of Baidu AI user experience department[J]. Decoration: Artificial Intelligence and Innovation Design, 2019, 319:22-27.
[17] Islam, M. R., Moni, M. A., Islam, M. M., Rashed-Al-Mahfuz, M., Islam, M. S., Hasan, M. K., ... &Lió, P. (2021). Emotion recognition from EEG signal focusing on deep learning and shallow learning techniques. IEEE Access, 9, 94601-94624.
[18] Islam, M. R., Islam, M. M., Rahman, M. M., Mondal, C., Singha, S. K., Ahmad, M., ... & Moni, M. A. (2021). EEG channel correlation based model for emotion recognition. Computers in Biology and Medicine, 136, 104757.

AUTHORS

Name: Ma Xueming

Education Background: Master Degree of Applied Psychology, Fudan University, Shanghai, China

Affiliation: Orient Securities Co., Ltd, Shanghai, China

Occupation: Design Engineer

Name: Chen Yao

Education Background: Master Degree of Engineering, Huazhong University of Science and Technology, Wuhan, China

Affiliation: Orient Securities Co., Ltd, Shanghai, China

Occupation: Algorithm Engineer

Name: Zhang Qian

Affiliation: Orient Securities Co., Ltd, Shanghai, China

Occupation: Chief Design Engineer

Name: Jiang Yun

Affiliation: Orient Securities Co., Ltd, Shanghai, China

Occupation: Design Director