Scene Understanding Technology of Intelligent Customer Service Robot Based on Deep Learning

Jianfeng Zhong*
Department of Business Administration, School of Management, Shenzhen University, Shenzhen, Guangdong, 518000, China
*Corresponding author e-mail: zhangjf@ma.szu.edu.cn

Abstract. As a value-added service that improves the efficiency of online customer service, customer service robots have been well received by sellers in recent years. Because the robot strives to free the customer service staff from the heavy consulting services in the past, thereby reducing the seller's operating costs and improving the quality of online services. The purpose of this article is to study the intelligent customer service robot scene understanding technology based on deep learning. It mainly introduces some commonly used models and training methods of deep learning and the application fields of deep learning. Analyzed the problems of the traditional Encoder-Decoder framework, and introduced the chat model designed in this paper based on these problems, that is, the intelligent chat robot model (T-DLLModel) obtained by combining the neural network topic model and the deep learning language model. Conduct an independent question understanding experiment based on question retelling and a question understanding experiment combined with contextual information on the dialogue between online shopping customer service and customers. The experimental results show that when the similarity threshold is 0.4, the method achieves better results, and an F value of 0.5 is achieved. The semantic similarity calculation method proposed in this paper is better than the traditional method based on keywords and semantic information, especially when the similarity threshold increases, the recall rate of this paper is significantly better than the traditional method. The method in this article has a slightly better answer sorting effect on the real customer service dialogue data than the method based on LDA.

Keywords: Deep Learning, Intelligent Robot, Scene Understanding, Customer Service Robot

1. Introduction
Online shopping is a popular shopping method in recent years. Major Internet shopping malls and logistics companies are in the ascendant, and registered users are also increasing year by year [1]. But at the same time, poor communication between consumers and sellers has always been an obstacle in the online shopping model. The main reason is that sellers provide buyers with consulting services through online customer service in order to reduce costs [2]. A customer service staff often needs to
deal with several buyers at the same time. This form of service is destined to be cumbersome and stressful [3, 4]. At this time, if there is a robot question answering system based on customer service dialogue to provide customer service personnel with answers, it will greatly improve their work efficiency. Therefore, customer service robots have a good application prospect in the online shopping environment [5, 6].

In recent years, the research and application of customer service chat robots in China has also made great progress [7, 8]. Dickinson B proposes a guided tour customer service robot project based on WeChat platform for tourist guide services in scenic spots, which can provide tourists with online tour guide consultation and rescue services. It is of great value for alleviating the pressure of tour guide customer service in peak tourist season, improving passenger experience, and reducing customer service costs [9]. Hadi R introduced a PRINTEPS-based flow reasoning and robot operating system (ROS) PRINTEPS application in the customer reception service of a robot coffee shop, and combined image perception with knowledge processing. Based on this platform, we proved that the robot's behavior in the coffee shop can be modified by changing the applicable rule set [10].

This paper takes robot scene understanding as the research object and conducts research on the above challenges. Scene understanding refers to the process by which a robot recognizes its environment based on perception data. It not only obtains the geometric information of the scene, but also needs to obtain the semantic information of the objects in the scene (such as tags and positions). Around this problem, this article mainly studies the dialogue generation process of intelligent chat robots based on deep learning, and studies some of the main problems that appear in the chat, and makes full use of some framework model knowledge principles of deep learning to train the chat model and generate dialogue. A chatbot model of Attention+LSTM that supports an open environment is proposed, and the key mechanism under this architecture is studied.

2. Research on Scene Understanding Technology of Intelligent Customer Service Robot Based on Deep Learning

2.1 Deep Learning Language Model

The use of the Encoder-Decoder framework is superficially understood as: input a paragraph or an article, and generate another paragraph or another article through the model method. Input sequence X, we hope to use Encoder-Decoder to generate its corresponding sequence Y. Regardless of whether the input language and the output language are the same, X and Y are composed of their respective word sequences:

\[ X = \{x_1, x_2, ..., x_m\} \] (1)

\[ Y = \{y_1, y_2, ..., y_m\} \] (2)

Convert it into an intermediate semantic vector, representing C:

\[ C = F(x_1, x_2, .., x_m) \] (3)

The formula for generating the target sequence in the recurrent neural network can be abbreviated as:

\[ y_i = g(Cy_{i-1}, s_i) \] (4)

Where s represents the hidden layer of the output RNN, C represents the semantic vector generated by the encoding, yi-1 represents the output of the previous time period, which is the input of the current time period, and g represents a nonlinear multi-layer neural network, generate the probability that each word in the dictionary belongs to yi.

2.2 Attention+LSTM Chat Robot Model

Deep learning is a method of automatically learning features. In essence, it is a machine learning model that has built a large amount of training data and a large number of hidden layers to learn more useful
features to achieve the purpose of improving classification or prediction accuracy. To achieve the purpose of feature learning. Different from traditional shallow learning, deep learning: 1) The depth of the model structure, usually 5 layers, 6 layers, or even 10 layers of hidden nodes; 2) The importance of feature learning is clearly emphasized, that is, through the layer-by-layer feature transformation transforms the feature representation in the original space to the new feature space, making model classification and prediction easier.

In this paper, the word2vec tool in the wordembeding language model is used to obtain the word vector, and the word vector obtained by the word2vec language tool has a richer and more meaningful semantic representation. The characteristic of this vector is that similar words have similar vector representations. Use the LSTM network model as the coding model, and then combine with Attention to generate the semantic vector C. Also use the LSTM network model, take the semantic vector output from the encoding stage as the input of the decoding stage, and finally get the output target sequence. In summary, the Encoder-Decoder model of Attention+LSTM is a language model based on deep learning that can be used to generate sentence pairs.

3. Experiment of Scene Understanding Technology of Intelligent Customer Service Robot Based on Deep Learning

This paper constructs the ontology manually through traditional methods, and the ontology description language uses OWL. First of all, determine the scope of the ontology to be the dialogue between online shopping customer service and customers. Secondly, research the general form of content and structure in the field of online shopping customer service. Extract the information skeleton based on related websites and manual customer service history records, determine the hierarchical structure of the ontology, construct concept classes, and the relationship between concept classes. 150 sentences were obtained from the merchant’s historical customer service records, divided into 15 groups, and each group expressed the same semantics. Among these 15 groups, some groups have similar expressions but different semantics. The evaluation corpus is derived from real dialogue scenarios. During the dialogue process, it is considered that the first k answers obtained after each question sentence are retrieved for relevance score. The relevance is divided into five levels, with a weight of 0 to 4, and a weight. The smaller the correlation, the lower the correlation. Based on the answer relevance, the experimental corpus is divided into 10 dialogue scenarios, a total of 10 groups of dialogues, and the average length of each group of dialogues is 10 sentences.

4. Experimental Analysis of Scene Understanding Technology of Intelligent Customer Service Robot Based on Deep Learning

4.1 Independent Question Comprehension Experiment Based on Question Retelling

Question paraphrase refers to the language change of a question, but the changed question still expresses the same semantics. In order to verify the effectiveness of the question comprehension method in this article, for the independent question sentence, this article adopts the method based on question retelling. Based on the experiment of question retelling, each question is repeated, and the accuracy and recall rate of each sentence after repetition are counted. Finally, the accuracy and recall rates of all questions are averaged as the accuracy and recall rate of the question understanding method.

The experiment based on question paraphrase uses the semantic information of question comprehension to calculate the similarity, and the sentence greater than the similarity threshold is regarded as the paraphrase of the current sentence. In the experiment, the method in this paper is based on the question topic, object and semantic block, combined with the similarity calculation formula to carry out a retelling experiment on the question sentence in the corpus, and compare the method based on LDA. The experimental results of the method in this paper are shown in Table 1 and the experimental results of the method based on LDA are shown in Table 2.
Table 1. Experimental results of this method

| Similarity threshold | Mean accuracy | Average recall rate |
|----------------------|---------------|---------------------|
| 0.1                  | 0.19          | 0.99                |
| 0.2                  | 0.27          | 0.97                |
| 0.3                  | 0.41          | 0.95                |
| 0.4                  | 0.53          | 0.91                |

It can be seen that with the increase of the similarity threshold, the number of times the question is repeated is less and less, but the accuracy of the question repetition has been greatly improved, but the recall rate is also lost. As the similarity increases, the accuracy rate also improves, which shows that the question understanding method proposed in this paper can effectively recall similar questions when used in similarity calculation, which shows the effectiveness of question understanding. But when the similarity increases, the recall rate also decreases, which shows that there are still some questions that cannot be recalled by this method. One of the reasons is that the complexity of the data itself makes it difficult to recall all, and it also shows that the method has room for improvement. In addition, when the similarity threshold is 0.4, the method achieves better results, and an $F$ value of 0.5 is achieved.

Table 2. Experimental results based on LDA method

| Similarity threshold | Mean accuracy | Average recall rate |
|----------------------|---------------|---------------------|
| 0.1                  | 0.05          | 0.94                |
| 0.2                  | 0.12          | 0.87                |
| 0.3                  | 0.18          | 0.75                |
| 0.4                  | 0.21          | 0.67                |

The experimental results based on the LDA method for question retelling are similar to the results of the method in this paper. On the whole, with the increase of the similarity threshold, the accuracy rate is improved and the recall rate decreases. This reflects from another aspect that the data itself is very complicated, and it is difficult to ensure high accuracy without missing similar questions. In addition, it can be seen that the results in Table 1 are better than those in Table 2 in terms of various indicators.

In order to better show the comparative effect of the experiment, the accuracy results of the two methods are shown in Figure 1:

![Figure 1](image_url)

Figure 1. Accuracy comparison of independent question repetition experiment

It can be seen from the figure that the accuracy of the two methods increases with the increase of the similarity threshold, and the trend is basically the same, indicating that the two methods can
achieve roughly the same effect in the measurement of similarity. However, the method in this paper is slightly better than the method based on LDA in each similarity threshold, which shows the effectiveness of the method in the online shopping environment for question understanding. The change of recall rate shows that with the increase of similarity, the recall of similar questions becomes less and less, which reflects the complexity of the corpus from the side. At the same time, it can be seen that the method in this paper is slightly better than the LDA-based recall effect.

It can be seen from the results that the method in this article is better than the traditional effect on the three indicators. Therefore, it can be known that the semantic similarity calculation method based on this article is better than the traditional method based on keywords and semantic information, especially when the similarity threshold increases, the recall rate of this article is significantly better than the traditional method.

4.2 Question Understanding Experiment Combined with Contextual Information

The experiment of question understanding combined with contextual information mainly evaluates the effectiveness of the question understanding method in this article in an interactive context. The experiment evaluated the average nDCG value of the first N answers fed back by the system when each user asked a question during each dialogue dynamic process, and compared the LDA-based method. The MAP value and nDCG value results of the related answers are shown in Table 3:

| Method                  | MAP   | NDCG@10 | NDCG@11 | NDCG@12 |
|-------------------------|-------|---------|---------|---------|
| Based on LDA method     | 0.54  | 0.52    | 0.54    | 0.55    |
| Method of this paper    | 0.6   | 0.58    | 0.63    | 0.59    |

It can be seen from the above table that as the number of evaluation answers increases, the NDCG value of the answer also increases to a certain extent, indicating that as the number of answers increases, more related answers are recalled. The MAP of the method in this paper is higher than that of the comparison method, indicating that the related answers rank higher in the results of the method in this paper. The comparative effect of the two methods is shown in Figure 2:

![Figure 2](image)

**Figure 2.** Experimental results of question comprehension with context information

When the number of evaluation answers increases from 10 to 13, the nDCG values of the two methods will increase to a certain extent, which also means that as the number of answers increases, more relevant answers will be extracted. At the same time, the method in this article is better than the method based on LDA in the overall index. The experimental results show that the calculation of
question similarity in this paper is better than the method based on keyword retrieval. The method in this paper has a slightly better answer ordering effect on real customer service conversation data than the method based on LDA, which shows the effectiveness of this method on customer service corpus.

5. Conclusions
With the rapid development of robotics, people have higher and higher expectations for robots to understand the surrounding environment and perform tasks autonomously. However, the local computing capabilities and knowledge of robots have limitations, resulting in their presence in complex environments, especially open and dynamic. In the unstructured environment, precision and accuracy are severely limited. Therefore, this topic researches and analyzes the robot scene understanding problem of the deep learning language model: this paper studies the knowledge of deep learning and the application of deep learning in the creation of intelligent chat robots, including the preparation of Chinese corpus and word vectors Training, syntactic analysis, semantic disambiguation, creating and training models, and effect prediction. And using the Encoder-Decoder framework, how to automatically generate responses from the input to realize a generative chatbot in the development field. By describing the technical principles of the Encoder-Decoder framework, the main issues that need to be considered in the field of chatbot research and their corresponding solutions are explained.

Acknowledgments
Research and application of key technologies of Intelligent Enhancement (IA) in channel application for tens of millions of customers

References
[1] Hafi L E, Isobe S, Tabuchi Y, et al. System for augmented human–robot interaction through mixed reality and robot training by non-experts in customer service environments. Advanced Robotics, 2019, 34(4):1-16.
[2] Stephen, Bennett. Customer Service Representatives: On the Front Lines. Fuel Oil News, 2018, 83(8):25-26.
[3] Belanche D, Luis Vicente Casaló Ario, Flavian C, et al. Service robot implementation: a theoretical framework and research agenda. Service Industries Journal, 2020, 40(3-4):203-225.
[4] Joerling M, Boehm R, Paluch S. Service Robots: Drivers of Perceived Responsibility for Service Outcomes. Journal of Service Research, 2019, 22(4):404-420.
[5] Rivolta A, Lunghi P, Lavagna M. GNC & robotics for on orbit servicing with simulated vision in the loop. Acta astronautica, 2019, 162(Sep.):327-335.
[6] Meyer P, Jonas J M, Roth A. Frontline Employees' Acceptance of and Resistance to Service Robots in Stationary Retail - An Exploratory Interview Study. Journal of Service Management Research, 2020, 4(1):21-34.
[7] Moinian S. 5 Tips for Improving Your Customer Service. Plumbing Engineer, 2018, 46(7):128-128.
[8] Bower P. The Key Role of Customer Service in Sales and Operations Planning. APICS, 2019, 29(1):20-21.
[9] Dickinson B. Providing Customer Service Versus Creating a Customer Experience. Air Conditioning, Heating & Refrigeration News, 2018, 265(16):22-22.
[10] Hadi R. When Humanizing Customer Service Chatbots Might Backfire. NIM Marketing Intelligence Review, 2019, 11(2):30-35.