Research on improved dialogue model

Wei Liu [a], Wen Dong [b]

Beijing Key Laboratory of Network System and Network Culture, Beijing University of Posts and Telecommunications, Beijing 100876, China

[a]twhlw@163.com, [b]wendy5156@163.com

Keywords: Dialogue model, Knowledge base, Chat robot system.

Abstract. This paper presents a novel dialogue model, which is characterized by combining the management, application and understanding of knowledge into a large dialogue model to fit the requirements which goes "the machine is not to try to understand human language, but to try to find the answer meeting human language habits best". Based on this model, this paper constructs a chat robot prototype system, which is validated and the results show the feasibility and effectiveness of the dialogue model that we have developed.

Introduction

The main business application purpose of a chat robot system can be divided into two categories: entertainment and access to information. Although the application purpose of different types of chat robot system is different, they have a common problem: the understanding of user input. Only by understanding of the user's input well can an entertainment chat robot produce a variety of good entertaining answers. For the chat robots for information access, only by understanding the user's intent can they produce the satisfactory output for the user. The dialogue model in this study is an abstract model that can produce answer to the given input question, which has no constraint on implementation.

Linguists discovered and summed up the regulation of human language, but ultimately did not keep pace with the development of human language, because people are always expanding and changing the rules to adapt their communication requirements. Simply speaking, "the rule" of natural language is an objective reality, not changed by anybody, but as the whole society progress and change. From this perspective, discovering and complying with the rule through statistic methods is a positive solution.

With the popularity of high-speed computers, this work has made great progress, the main method is to count the language regulation through the computer. But the regulation grooming and dissemination is to return to the people finally, especially the language based on the context (semantics) whether "consistent with human language habits" will eventually be evaluated by the people. Therefore the chat robot system developed in this paper is based on a rule: the machine is not trying to understand human language, but to find the answers meeting the human language habits best. This rule has two meanings: i. the machine can't understand human language like people; ii. The machine should map the human language to the machine language they can understand, and try to find the answers meeting the human language habits best.

Research on factors effecting the dialogue model

Factors affecting the dialogue model can be divided into two categories: the input questions and knowledge in the knowledge base.

Characteristics of the input questions in the chat robot system. For the chat robot system, the input query can be divided into three categories: short and pithy phrase or concept word, the common questions in daily life and the professional issues.

Different types of user input have very different features. Common questions or statements in daily life usually do not contain any keywords, nor comply with the written rules of grammar, and

have relatively fixed pattern, which led us deal with such issues or statements through pattern matching approach. Successful use of AIML has proved it valid that dealing with these problems or the statement with pattern matching methods. For relatively professional questions or statements, users often ask more specialized and grammatical questions, so for these questions, we can analyze the user's input questions with parser first and then extract keywords from the output of the parse tree and record the objectives of the questions. So for the relatively professional questions, we can use (keywords list, question objectives) on behalf of the questions, as the input of the next stage of processing.

Characteristics of the knowledge base in the chat robot system. There are two forms of knowledge: structured and unstructured. Structured knowledge includes merchant information, (terms, definitions) pair, product information, books or movies and other information, and unstructured knowledge includes (mode, the answer list) pair, (questions, answers) pair, search query and click data and web data. Using structured knowledge is very straightforward, but it's more difficult to use unstructured knowledge.

Dialogue model

A dialogue model may produce two kinds of outputs for user input: i answer; ii computer representation of user input after understanding. These outputs are produced following: i match the user input with the knowledge, including exact and fuzzy matches, which is a way of understanding. For this way of understanding, you can get an answer once the knowledge is matched; ii identify problem types with syntactic analysis and produce the (keyword list, questions, question types, user information) expression, which is also a way of understanding. Output of the second way will be the input of the answers search component (GAE), which will produce the answer successfully.

The user usually doesn't care about how the computer understood, as long as he gets a satisfactory answer. In other words, the purpose of understanding the user input is to produce the correct answer. Our dialogue model is shown in Fig. 1.
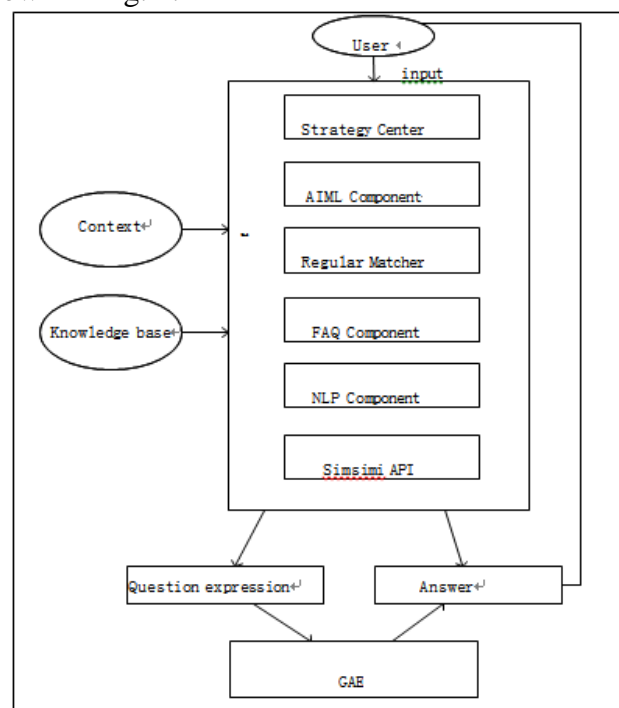

Fig. 1 Dialogue Model

GAE, receiving the problem of dialogue model as input and finding knowledge related to this problem from the knowledge base, generate the candidate answer set shown as Figure 2. After obtaining the candidate answer set, GAE will filter the set and score for each document in the set to confirm the confidence of each candidate. Finally, GAE integrate the candidate answer set from multiple knowledge sources and use top N candidates as answers for users.

When the keywords list in the problem is input into the knowledge search engine, the search system will traverse each knowledge index in turn and generate the candidate answer set. If there are N knowledge index, we will obtain N candidate answer sets. So we need merge N candidate answer sets and mark clearly the source of each answer, after which we need sort the documents in the set to choose $K(K >= 1)$ final answers to return back to users. In the research, we use the rank learning algorithm Listwise to sort the candidate answer set, which uses sorting features including the correlation of search result and query, the characteristics of each search result, the authority of each search result sources and the explicit feedback of users etc.

Experiment

The Background of the Experiment. The performance of the dialogue model is needed to obey the performance of the chat robot system. So we directly use the key performance of the chat robot system as reference to the performance of the dialogue model. The main performance index to evaluate chat robot include the accuracy of the answer, the contextual coherence and the speed of response. In the next, we mainly focus on the accuracy of the answer and the contextual coherence.

The chat robot system we propose is made up of application layer, web server layer and chat server layer. The work procedure is shown as Fig. 2.
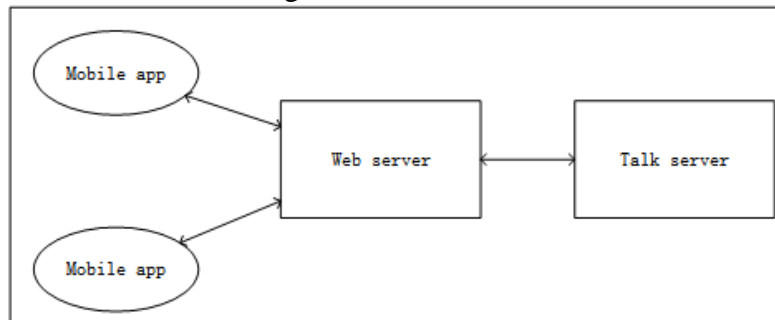

Fig. 2 The whole framework of chat robot system

The data transmission between the application layer and the web server layer is encapsulated by using xml and the web server layer communicates with the chat server layer through protobuf protocol.

The Experiment scheme. We did several experiments, which were mainly used for evaluating the performance of the prototype system. Firstly, one experiment was used for analyzing the validity of the strategy center. Then, we studied that whether the regular verifier component was really worked by one experiment, in which we induced users to input more simple commands or questions. Afterwards, the performance parameter of the FAQ component and the NLP component was checked in detail. Finally, we studied the continuity and smoothness of our chat robot system in the end-to-end chat interaction. These experiments were based on the data, which came from 114,032 chats between 1230 users and the chat robot system after the chat robot system worked six months. All of these experiments can indirectly verify the validity of the dialogue model we propose.

The Analysis of Experimental results. The Validity of the Strategy Center. We get the Table 1 by analyzing the experimental data.

Table 1 the ratio of benefiting from the strategy center

	User percentage
Nearby Finder	5.23%
Movie Recommend	3.10%
News	4.50%
Search	1.02%
Flight info	0.50%
Whether	2.03%
Coupon info	2.02%
ALL	7.10%

The first column of the Table 1 shows that 5.23 percent users used the NearbyFinder strategy of the strategy center and the last column shows that 7.10 percent users got help from the strategy center. Judging from these statistics, we can find that users hope the chat robot to perceive the world and help to solve the problem.

The Validity of Regular Verifier. Through the analysis of the experimental data, we found that the regular verifier component solved 1.10 percent problems in the first two months and it solved 2.01 percent problems in the first four months, which proved that the regular verifier component was effective.

The Performance of the FAQ Component and the NLP Component. To research the performance of the FAQ component and the NLP component, we prepared 1000 problems, which came from Baidu Knows, to analog input from the users. In the experiment, we only used the FAQ component and the NLP component and closed other components.

The result of the experiment is shown as Table 2. We found that 351 problems could find answers in the FAQ component and the NLP component could solve 430 problems, in which 402 answers were from the FAQ index and the others were from the rest of the three indexes. The result proved that it was necessary to query reconstitution. On the whole, 219 problems could not find answers. So, the recalling rate was 78.10 percent.

Table 2 shows that the FAQ component had the highest accuracy, which was up to 79.20 percent and the accuracy of getting answers from other indexes was low, which was only 59.03 percent. On the whole, the accuracy of getting answers was 74.36 percent.

Table 2 the assessment result of the FAQ component of the NLP component

	System returns	Accuracy
FAQ	351	79.20%
No-wiki-NLP	402	71.21%
wiki-NLP	28	59.03%
ALL	781	74.36%

Coherence and fluency of the end-to-end interaction. To study the coherence and fluency of the end-to-end chat interaction, we randomly selected 30 users from 1,230 users as the objects of observation and studied their chat history. Overall, the chat between the robot and the user is coherent, but there are some anomalies: Answers from Simsimi parts appear relatively random, 3% of the answers returned from this part is irrelevant with the actual answer.

Summary

This paper puts forward the views of "the machine is not to try to understand human language, but to try to find the answer meeting human language habits best", on the basis of summarizing the previous theory. The representation method of knowledge determines the methods of using knowledge, so we studied various factors affecting the dialogue models, including the characteristics of the input questions and characteristics of the knowledge base in the chat robot system. Based on this, we propose a novel dialogue model, which is characterized by combining the management, application and understanding of knowledge into a large dialogue model to comply with the requirements. Finally, we verified the effectiveness of our dialogue model with experiments indirectly, the experimental results show that our chat robot system has satisfactory performance, thus proving that the dialogue model is more scalable and has more accurate answers.

Acknowledgement

References

[1] Wirken, D. The Google Goal of Indexing 100 Billion Web Pages (2006). In Editor (Ed.) ^ (Eds.): 'Book the Google Goal of Indexing 100 Billion Web Pages (2006)' (edn.).

[2] Banko, M., Brill, E., Dumais, S., Lin, J. AskMSR: Question answering using the worldwide Web. In Editor (Ed.) ^ (Eds.): 'Book AskMSR: Question answering using the worldwide Web' (2002, edn.): p.7-9.

[3] Mauldin, M.L. Chatterbots, tiny muds, and the turing test: Entering the loebner prize competition. In Editor (Ed.) ^ (Eds.): 'Book Chatterbots, tiny muds, and the turing test: Entering the loebner prize competition' (1994, edn.): p.16-21.

[4] Voorhees, E.M. The Sixth Text Retrieval Conference (TREC-6). Information Processing & Management, 2000, 36 (1): p.1-2.

[5] Zheng, Z. Answer Bus question answering system. In Editor (Ed.) ^ (Eds.): 'Book Answer Bus question answering system' (Morgan Kaufmann Publishers Inc., 2002, edn.): p.399-404.

[6] Schlaefer, N., Gieselmann, P., Schaaf, T., Waibel, A. A pattern learning approach to question answering within the Ephyra framework. In Editor (Ed.) ^ (Eds.): 'Book A pattern learning approach to question answering within the Ephyra framework' (Springer, 2006, edn.): p.687-694.

[7] Kwok, C., Etzioni, O., Weld, D.S. Scaling question answering to the web. ACM Transactions on Information Systems (TOIS), 2001, 19 (3): p.242-262.

[8] Ma, H.Z.T.L.J., Liao, X. Chinese word segmentation with multiple postprocessors in HIT-IRLab. In Editor (Ed.) ^ (Eds.): 'Book Chinese word segmentation with multiple postprocessors in HIT-IRLab' (2005, edn.).

[9] Huang, G.-T., Yao, H.-H. A system for Chinese question answering. In Editor (Ed.) ^ (Eds.): 'Book A system for Chinese question answering' (IEEE, 2003, edn.): p.458-461.

[10] Hang, L.I. A short introduction to learning to Rank. IEICE TRANSACTIONS on Information and Systems, 2011, 94 (10): p.1854-1862.