ISCA Archive
http://www.isca-speech.org/archive

6th International Conference on Spoken Language Processing (ICSLP 2000)
Beijing, China
October 16-20, 2000

An approach to Intelligent Chinese Dialogue System

*Chen Fang, Yuan Baozong **
Motorola China Research Center, Fang.Chen@motorola.com
*Northern Jiaotong University

Abstract: A human-machine natural interactive system, the Intelligent Chinese Dialogue System, is introduced in the paper. In the proposed system, users can ask questions and retrieve right answers in natural speech. The system consists of five modules, i.e., dialogue processor, task processor, natural language interface, user model and knowledgebase. It has the capabilities in dialogue management, table-based knowledge learning of task domain and linguistics domain, and speech generation. It can analyze user' s input in natural language, generate Chinese natural sentences, then to synthesize them as speech output. It is in a domain specific structure, while not designed for a specific domain. Users can edit the tables for knowledge learning, and create their own applications in different domains. The system architecture and the realization procedures are discussed in detail in the paper.

Keywords: Dialogue, Natural language interface, User model, machine learning

1 Introduction

It is the most convenient way for people to communicate with computers in natural ways, such as using natural language, gestures, emotions and etc. The Intelligent Chinese Dialogue System (ICDS) is a human-machine natural interactive system based on our research on dialogue management, language processing and speech processing.

A domain specific application, Park Tour Guide, has been established to demonstrate the concept. Before the system running, the knowledge for the domain is learnt through some tables to create the specific knowledge base. The Guide system accepts user' s natural speech inquiry that falls in 15 information categories, understands his meaning and formalizes it to a sequence of machine scripts. The task processor analyzes the input subsequently and decomposes the task into sub-tasks through inquiring and reasoning in knowledge bases. It passes sub-tasks to the dialogue processor, which is responsible for planning the dialogue based on the user model. The dialogue processor also produces prediction and updates the user model. Finally it gives the result and the priority information to the speech generation module to create a natural speech output. The system can generate 2-15 types of correct answers for each question within a domain. ICDS has capabilities in dialogue processing, task processing, knowledge learning of task domain and linguistics domain, user independent, speech recognition and understanding, speech generation and synthesis. The system concludes five modules, which are dialogue processor, task processor, natural language interface, user model and knowledge base. The flow chart of ICDS is showed as Fig1. The main modules of the system are introduced as the following.

2 Module of Dialogue Processor

The Dialogue Processor is the controller and coordinator in the system, and forms a bridge between domain of task and domain of natural language. It formalized demands in task domain and changed them into semantic forms that are easy to process in natural language domain. It is also responsible for dialogue planning. The dialogue processor actives task processor and gives the demand of task to its accordingly user model. After getting the output from the task processor, the dialogue processor would define priority of the system and the user. Meanwhile, it accepts the next interactive goal (sub-task) supplied by task processor. Then it selects sub-dialogue and produces prediction about user' s inputs to natural language interface, includes indicating the rank of the initiative, providing the formulation of the output to generation part and expecting the semantics of the next input. After that the user' s model is updated.

3 Module of Task Processor

Task Processor is responsible for getting and managing knowledge of domain, analyzing the task and determining the sub-task for interaction, inquiring and reasoning in the knowledge Base, providing the information or the result of the task or showing the multimedia information to the user.

3.1 Represent task

Task representation is the basis of task analysis and reasoning. Generally, a task can be represented as a task tree. The root of the tree shows the destination of the task, every node shows a procedure to reach the destination, while each node can be called sub-task or sub-goal. The relation between nodes is a "part-of" one, says every task can be divided into several sub tasks. The task represented by a sub-node is the condition or one part of the task represented by its parent node. The bottom of the tree is the nodes which act as the basic task elements, cannot be divided further.

3.2 Inquiry and reason in task domain

The task is stored in knowledge base in form of tree structures. The procedure of task analysis is searching the task tree. After determining a node in task tree, its sub-node must be analyzed immediately to choose sub-node in lower level to be processed. Then the sub-node is supplied to dialogue processor to plan dialogue. According to the returned result of dialogue, the task processor will search the task tree again in same way. The main cycle of task processor is:

(1) Accept task from dialogue processor, determine the task node;

(2) Define whether the direct dialogue can be executed according to the processing rules to the current node. If it works, inquire and reason in the knowledge base, then provide needed information to dialogue processor and set up a flag to show the node has been processed, then jump out from parent node to (4). If the node is root node, finish the task and return to (1);

(3) Define the next sub-node to be processed, save the father node and return to (2);

(4) Produce and send predictions to dialogue processor.

3.3 Learn knowledge of task domain

The knowledge type should be chosen of which to be learnt in task domain. Domain knowledge is comprised of touring map and features of scenes. The knowledge is classified into knowledge of graphics, concept and rules. The system can use mouse, keyboard and scanner to mark the line, scenes, street, river, lake and etc. to be introduced in touring map. In graphics knowledge representation, dot represents a scene can be showed as a dot in touring map, written as the form of (locx, locy); Line represent a scene can be showed as open curve, written as the form of(locx1, locy1), (locx2,locy2)…… (locxn,locyn)in which locx1 locxn and locy1 locyn. Domain represents a scene can be showed as closed curves, such as lakes, it can be written as the form of (locx1, locy1), (locx2,locy2)…… (locxn,locyn), in which locx1 locxn and locy1 locyn. In concept and rule knowledge representation, each kind of scenes is corresponded to a kind of frame structure with a set of editing form for learning. After filling the touring map and the features in the forms, the system changes the fillings into related knowledge representations.

4 Module of Natural Language Interface

Module of Natural Language Interface embodies the function to interact with users. It can understand user's question in any sentence pattern in speech, generate and convert the generated text into synthesized speech in high naturalness. Functions of the Natural Language Interface are:

(1) Recognizing and understanding the input based on the expectation.

(2) Generating and synthesizing the output based on the rank of the initiative.

(3) Learning the linguistic knowledge.

4.1 Speech Recognition and understanding

Natural speech inquiring is accepted through speech recognition. A real-time Chinese recognition module is developed based on HMM. The accuracy of 96% is attained in domain related questions. The result of speech recognition must be changed into semantic that computer can understand through language understanding module which consists of grammar analysis module and semantic analysis module. The former gets grammar structure from input natural sentences. It carries on sentence segment, analysis of word attribution, functions and features of each word according to the bi-directional grammar that suited for language understanding and language generation. The voice type and structure type of sentence are also determined. Rank determining divides the input characters into to phrase, sentence or complex sentence.

Semantic analysis assists system to understand the real meaning of sentence. Inquiring sentence is classified into declarative sentence and interrogative sentence. The two sentences have different semantic focal point.

4.2 Text Generation and Speech Synthesis

Natural language generation technique is very important in man-machine natural dialogue system. General system find an exist answer in knowledge base or database to users. We use the technique to generate answer after analyzing question and find related node in knowledge base. At first we go through text planning to extract concrete content related to the semantic. Then text organization produces a linear and correspondent knowledge sequences can be outputted, but they are still internal representation of computer. According to the bi-directional grammar, the transform from internal structure to natural language is completed. The answers become various by using the technique. The system doesn't store every answer while find satisfactory answer in accordance with the demand of bi-directional grammar and output any of the answers by random. To the same question the system can produce 2 to 15 different kinds of output.

After the text of natural language generated, the next step is to convert the text into speech. We build up a speech database with speech of phrases and pronunciation rules. All speech of phrases are cut from natural language speech flow. The database gives completely a description from Chinese characters to speech. An efficient structure of the speech database is also designed. The analyzed text can be read in real time, the synthetic speech has good quality in naturalness and intelligibility.

4.3 Learning Ability

The system can be suited for different domains and all sentence patterns if it's a general system. It's difficult to define a perfect set to comprise all sentence patterns because Chinese sentence is freely in organization. If the input sentence can't be normalized in accordance with the standard type, machine learning is needed. We have designed many editing forms for learning, correspondent to every grammar rank, through which users can fill in new linguistic knowledge. The system can absorb in new sentence pattern and new lexicology that the system doesn't understand while being questioned, and embody the result of learning in sentence generation. The machine improved its intelligent capability through learning.

5 Module of User Model

The user model is a dynamic description about the knowledge level of users. It can make the dialogue more effective. The system can give a suggestion on the content to be outputted to the users according to the user model, which is a cognitive model for analyzing user' s language. The knowledge level of users and the questions to be concerned by users can also be learnt from the user model. The content of the user model

(1) The information the user has already known;
(2) The site the user is interested in;
(3) The feature the user is interested in;
(4) The user' s intention;
(5) The location where the user is;
(6) The action the user can do.

6 Conclusions

The ICDS has been considered to be a general system in a certain level with the capability of knowledge learning in task domain and linguistic domain, which means that it can be transfer to other domain (such as guidance of museum, guidance of city touring and traffic, and so on). It can be acted as real man-machine natural interactive systems in more domains by improving its learning capability.

Reference

1. George Ferguson, James F. Allen. "Generic Plan Recognition for Dialogue Systems". 1993 DARPA Human Language Technology Workshop.
2. Ronnie W. Smith, Steven A. Gordon. "Effects of Variable Initiative on Linguistic Behavior in Human-Computer Spoken Natural Language Dialogue". Association for Computational Linguistics, Vol.23, No.1, 1997.
3. Cecile L. Paris. "The role of the user' s domain knowledge in generation". Computational Intelligence, Vol. 7, No. 2, 1991.
4. Marie W. Meteer. "SPOKESMAN: Data-driven, Object-oriented Natural Language Genetation". Processings of IEEE 7th Conference on AI applications, Vol. 1, Feb., 1991.
5. Spoken Language Systems Group, MIT. "Summary of Research". 1996.
6. Spoken Language Systems Group, MIT. "Summary of Research". 1997.
7. Fang Chen, Baozong Yuan, Intelligent Information Inquiring System Supported by Multimedia. Proceedings of International Conf. on Signal Processing, Beijing, Oct., 1996.
8. Marion Mast, Franz Kummert, Ute Ehrlich, Gernot A. Fink, Thomas Kuhn, Heinrich Niemann, Gerhard Sagerer. "A Speech Understanding and Dialog System with a Homogeneous Linguistic Knowledge Base". IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 16, No. 2, 1994.
9. Eli Goldberg, Norbert Driedger. "Using Natural-Language Processing to Produce Weather Forecasts". IEEE Expert, April 1994.
10. Fang Chen, Baozong Yuan, An Approach to Intelligent Speech Production System. Journal of Computer Science and Technology, Vol.12, No.2, March, 1997.

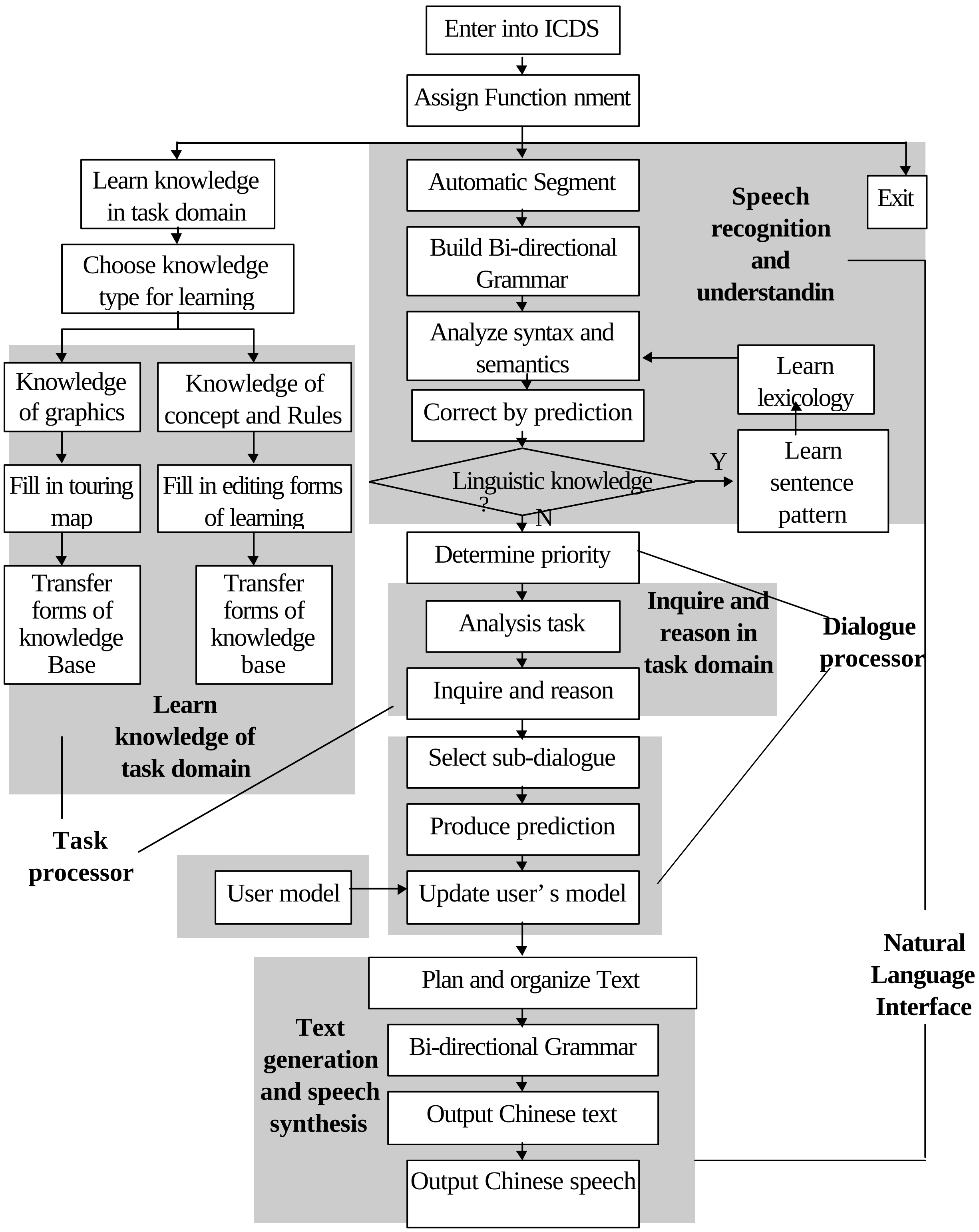

Fig1 Flow Chat of ICDS