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

3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

A COOPERATIVE MAN-MACHINE DIALOGUE MODEL FOR PROBLEM SOLVING

Masahiro ARAKI, Taro WATANABE, Felix Merida QUIMBO and Shuji DOSHITA

Department of Information Science, Kyoto University,
Yoshida Hon-machi, Sakyo-ku, Kyoto 606-01, Japan

ABSTRACT

In this paper, we present a constructive model of man-machine dialogue for problem solving. Our spoken dialogue system aims to solve scheduling problems cooperatively by interacting user and system. An important feature of this model is that it separates the knowledge about dialogue structure (conversational space) and about problem (problem solving space) in order to make correspondence to utterances and actions. Conversational space is designed for identifying user's speech act and constructing dialogue segment. Problem solving space represents the structure of the problem and the most appropriate direction for solving problem. For incorporating probabilistic reasoning, each space is represented by network. This time, we adopt this model to keyword based plan recognition and topic prediction of next utterance. In small experiment, this system hit the topics of next user's utterance at 47% at the first candidate, and at 92% within the third candidates.

1. INTRODUCTION

In this paper, we present a constructive model of man-machine dialogue for problem solving. We define the problem solving dialogue as interaction between user (man) and system (machine), both have aim for solving the problem cooperatively.

Our spoken dialogue system aims to solve scheduling problems cooperatively by interacting user and system. For example, setting up a group meeting, negotiating event schedule, etc.

We assume that each member of the group has a personal schedule management system on his/her own workstation. In order to arrange an available time for meeting, each system exchanges their information automatically. Then the user can concentrate on telling system about his/her plans and constraints and decision making.

We collect some dialogue about our task in order to make a dialogue corpus. One participant plays a role of system who knows well our task domain. The other participant who plays a role of user is unfamiliar with the task. The corpus contains 40 dialogues (4 different situation for 10 person, in Japanese). Fig. 1 shows the example of the dialogue in our corpus (translated into English).

User:	Well... tell me an available day... uh, time of spoken dialogue group members from 21 April to 23.
System:	1 P.M. to 5 P.M. at 21, and 11 A.M. to 1 P.M. at 22 are available.
User:	Well, then... tell me the state of reservation of meeting room at 21 and 22.
System:	Which meeting room?
User:	All the meeting room in this department.
	.....

Figure 1: Example of user-system dialogue

Generally, the problem, that is intended to solve by dialogue, is divided into sub-problems. That is represented by AND-OR tree or by other similar representation. Each utterance by user or system can be considered as an action for solving the sub-problems that correspond to the leaf nodes of the AND-OR tree. These assumptions is widely used as a bases of plan recognition. However, there are two major problems in those assumptions to apply dialogue system.

- **Generating response**
  In dialogue system, system's decision about user's plan cannot delay as in story understanding. This is because dialogue system must reply anyway to user in each turn. Therefore, what we need is not to list up possible user's plans, but to select the most plausible plan and to generate system's response. Charniak and Goldman also pointed out that plan recognition must have some probabilistic measure that decide most plausible one [1].
- **Correspondence to utterance and action**
  In many cases, simple one-to-one relation does not stand between utterance and action. One utterance does not always correspond to one elementary action (leaf node). Some utterance may correspond to several actions (e.g. specification of time and specification of place by one utterance). On the other hand, several utterances (by user and system) may consist only one action (e.g. the case that includes question-answering interaction of unspecified slot or the case that includes misunderstanding of user's

intention). Then we need a method of managing utterance and action separately.

In this paper, we propose a dialogue model that can deal with these problems. At the beginning, we consider five step model. It describes the process of user's utterance understanding to system's response generation in five steps. That means, utterances and actions are treated in different step. Therefore, we design two major knowledge representation space: one is conversational space and the other is problem solving space. Conversational space is designed for identifying user's speech act and constructing dialogue segment. Problem solving space represents the structure of the problem and the most appropriate direction for solving problem. For incorporating probabilistic reasoning, each space is represented by network.

In the remainder of the paper, section 2 describes all about five step model of dialogue. Conversational space and problem solving space are explained section 3 and 4 respectively.

2. FIVE STEP MODEL OF PROBLEM SOLVING DIALOGUE

Airenti et. al. proposed the model of dialogue that divide the process of meaning understanding to response generation into five steps [2]. This model is based on the predicate that represent mutual belief or intention and the rules that can apply each step. Our model does not intend to 'smart model to explain phenomena in dialogue' but 'procedure to generate most appropriate hypothesis at any situation of dialogue'. Then, we hire the framework of five step modeling and assign probabilistic reasoning framework to some steps. The outline of our five step model is shown in Fig. 2.

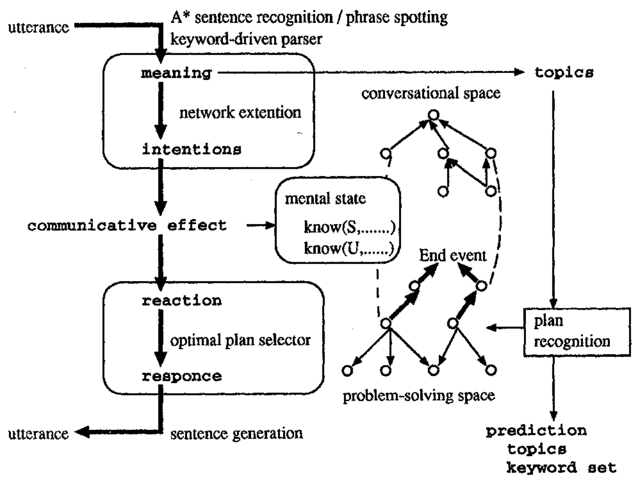

Figure 2: Five step model of dialogue

The procedure of each step is below:

1. **Meaning Understanding**
As a front end of spoken dialogue system, we developed speech recognition subsystem and semantic analysis subsystem. Our speech recognition system is based on A*-admissible context free parsing with word-pair heuristics [3]. Our semantic analyzer is keyword driven parser for sentence hypothesis or word lattice [4]. The output of this step is represented by a pair of propositional content and utterance type.

2. **Intention Understanding**
The knowledge for picking up the user's intention from the semantic representation of the utterance is represented by network structure in conversational space. In this space, there are two types of node. one represents a portion of propositional content of utterance. The other represents a utterance type. These nodes construct a dynamic network by the intention introducing rules and dialogue continuing rules.

3. **Conversational Effect**
System's mental state (that is represented by the predicates; such as believe, intend, etc.) is updated by some kind of user's utterance. And also this mental state has an influence on connecting utterance type node to intention node in conversational space.

4. **Intention Generation**
General structure of problems (scheduling problem, in our task) are represented by problem solving space. This space is prepared static. System's intention is generated by deciding the optimum open node of AND-OR tree of problem solving space. This space is also used in plan recognition by keyword and predication of keyword of next utterance.

5. **Response Generation**
System's response is generated in order to achieve generated system's intention.

3. DIALOGUE MANAGEMENT IN CONVERSATIONAL SPACE

In this section, we describe the dialogue management in conversational space.

The meaning of user's utterance is obtained as a pair of propositional content and utterance type. Fig. 3 shows the semantic representation of utterance taken from the first utterance of Fig. 1 ("Well... tell me an available day... uh, time of spoken dialogue group members from 21 April to 23.").

Fig. 4 shows the example of conversational space that correspond to the semantic representation.

After constructing such network, the method of extracting user's intention and controlling dialogue is below:

```
[ask-value, [content, available_time],
            [time, [94,4,21,0], [94,4,23,24]],
            [member, spoken_dialogue_group]].
```

Figure 3: Sample semantic representation of utterance

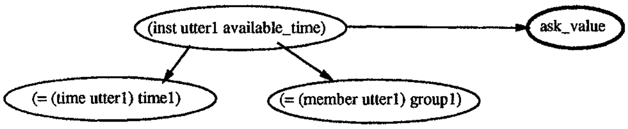

Figure 4: Example of conversational space

1. Make correspondence to the nodes of conversational space (propositional content and utterance type) and system's mental state.
2. If the correspondence is 'valid', that is, there is no contradiction in conversational space and mental state, development of conversational space is done with normal dialogue development rule.
3. If the correspondence is 'invalid', for example, user asks about what he/she knows, supplemental dialogue development rule is used.
4. If necessary, mental state is updated.

The development of conversational space is done by the rule shown in Fig. 5 with the condition concerning with mental state.

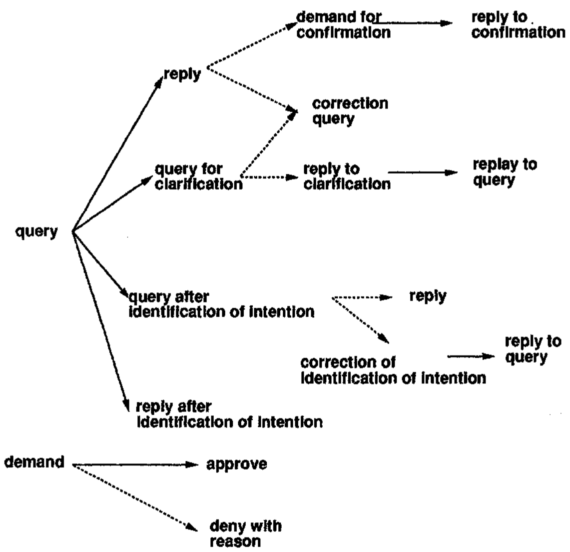

Figure 5: Example of development rule

In the normal case, user's intention of this semantic expression is 'query'. But in ask_value type utterance, there are some cases to interpret it as 'assert'.

Fig. 6 shows the state of conversational space after expanding about system's behavior. Here we assume that system's behavior is correspond to system response in Fig 1 (" 1 P.M. to 5 P.M. at 21, and 11 A.M. to 1 P.M. at 22 are available.").

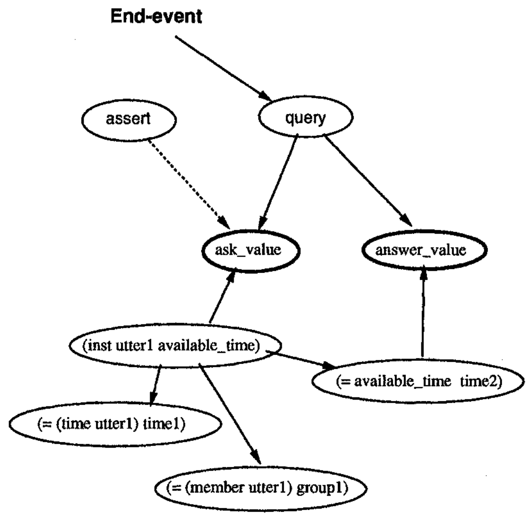

Figure 6: Example of conversational space

Concerning about the problem of response generation, system generates the utterance type part of response by such a way. Even if plan recognition fails in problem solving space, system can decide the utterance type of next response in order to make a fresh start of plan recognition.

4. PLAN RECOGNITION IN PROBLEM SOLVING SPACE

The structure of domain problem is described in problem solving space. It is represented static tree structure. The tree consists of events which have hierarchical functional and abstract relations. These events consist of speech-events and their abstractions. These relations can be expressed using Kautz' Event Hierarchy [5]. This Event Hierarchy has the following components:

- Nodes
  - Speech Event Nodes
    These are nodes which correspond to speech-events in conversational space.
  - Ordinary Nodes
    These are the abstractions of the speech-events
- Links
  - Abstraction Links
    Shows the conceptual relations of two nodes. A→B means that B is an instance of A. In the graph a thick arrow is drawn from A to B.

- Decomposition Links
  Shows the functional relationship between the nodes, when one node can be decomposed into one or more nodes. For example, in order to complete A, Actions B and C must be completed. This is shown on the graph by drawing a thin arrow from B and C to A. B and C are called A's decomposition nodes.

4.1 PLAN RECOGNITION METHOD

This subsection shows the plan recognition and intention generation method in problem solving space.

System's mental state is updated by the interaction of conversational space and mental state. Some cases of this updating include a part of problem solving. When such updating is done, the plan recognition procedure starts. The recognition method is calculating minimal cover of the problem solving space that includes actions already done. After identifying user's plan, this process searches the node that has top priority in order to achieve recognized plan. Fig. 7 shows an example of AND-OR tree in problem solving space.

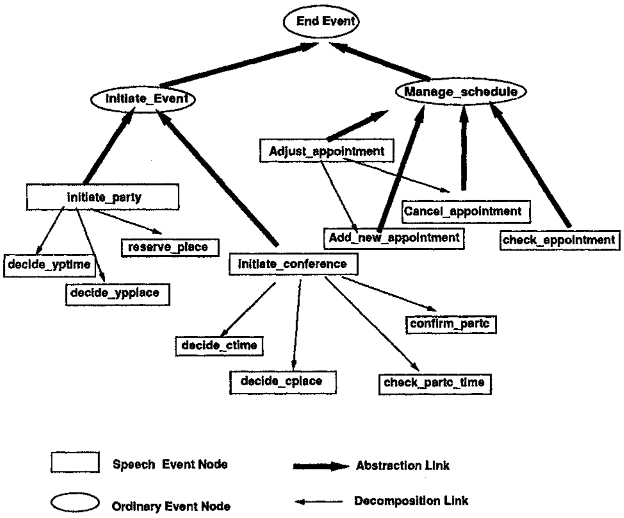

Figure 7: AND-OR tree in problem solving space

If the optimum node are specified, system's intention is generated according to the action that correspond to that node. On the other hand, if any action that correspond to the node in problem solving space is not done, that is, in the case user's utterance is a part of some action, system's intention (and response) is generated in conversational space by conversational development rules.

4.2 PREDICTION OF NEXT TOPICS

Topic identification and prediction of next utterance's topic are done in the same problem solving space. Each node has a keyword set that make the current topics salient. User's topic is extracted by semantic representation at the meaning understanding step. But only by structural information, there exist some current topic candidates. Then we use supplemental information about directive utterance, cue phrase etc.

After identification of current topic, prediction of next utterance's topic is done by searching minimal cover of problem solving space. The keywords that attach all the nodes in minimal cover are candidates of next utterance's topic. They are sorted in probability given from dialogue corpus.

We examine the validity of this model for the field of plan recognition by small experiment. Up to now, only the prediction part is evaluated by dialogue corpus. By this experiment, This system hit the topics of next user's utterance at 47% at the first candidate, and at 92% within the third candidates.

5. CONCLUSION AND FUTURE DIRECTION

In this paper, we presented a cooperative man-machine dialogue model for problem solving. We explained the knowledge representation in conversational space and problem solving space, and the connection of them by mental state.

We will introduce some probabilistic method, say Bayesian network, in each space. And also we intend to integrate subsystem to implement spoken dialogue system.

References

[1] E. Charniak and R. P. Goldman. A bayesian model of plan recognition. *Artificial Intelligence*, 64:53–79, 1993.

[2] G. Airenti, B. G. Bara, and M. Colombetti. Coversation and behavior games in the pragmatics of dialogue. *Cognitive Science*, 17:197–256, 1993.

[3] T. Kawahara, S. Matsumoto, and S Doshita. A*-admissible context-free parsing on HMM trellis for speech understanding. In *PRICAI-92*, pages 1203–1208, 1992.

[4] M. Araki, T. Kawahara, and S. Doshita. A keyword-driven parser for spontaneous speech understanding. In *Int'l Sympo. on Spoken Dialogue*, pages 113–116, 1993.

[5] H. A. Kautz. A circumscriptive theory of plan recognition. In *Intentions in Communication*. Cohen, P. R. and Morgan, J. and Pollack, M. E., 1990.