A ROBUST DIALOGUE MODEL FOR SPOKEN DIALOGUE PROCESSING

MasahiroARAKI[*] *and ShujiDOSHITA*[†]

∗ Center for Information and Multimedia Studies, Kyoto University
† Department of Information Science, Kyoto University

ABSTRACT

In this paper, we propose a robust processing model of spoken dialogue. Our dialogue model is a cognitive process model (1) which integrates stepwise processing from utterance understanding to response generation, (2) which specifies the interactions between the processing of each steps and two level dialogue management mechanism, and (3) which identifies the possible errors caused by speech recognition error and specifies the method of recovering from the error. Also, We examined the validity of this model using new evaluation paradigm: system-to-system dialogue with linguistic noise. By this evaluation, the robustness of proposed cognitive process model is shown in relatively low recognition error situation.

1. INTRODUCTION

In order to make an interactive dialogue system, we need two management processes: (1) understanding process which manages the subprocesses from utterance understanding to response generation, and (2) dialogue management process which aggregates the utterances into the discourse segment, and manages focus and intentions of dialogue. Furthermore, in applying such a dialogue model to spoken dialogue systems, we need (3) an error correction mechanism in dialogue processing which deal with input errors caused by speech recognition errors. In previous researches of spoken dialogue, these three aspects are treated independently.

In the research of understanding process, there are two major approaches: one is parallel multi-agent with distributed databases [1], the other is sequential processing combining some module [2], [3]. From the viewpoint of the usage of various constraints and cognitive modularized mechanism, multi-agent approaches are good model of dialogue processing of human being. However, constraints satisfaction mechanism in multi-agents is difficult to implement, and hard to control. Then, sequential processing is widely hired in implementing understanding process.

In the research of dialogue management, two major management methods are widely used: stack structure [4], [5] and AND-OR tree structure [6], [7]. Stack structure is easy to implement and has simple relation to the attentional state. However, it is hard to manage the hopping to subdialogue, to realize variable initiative, and difficult to make collaborative response form the task level. On the other hand, many of AND-OR tree structure, which confuses linguistic structure and intentional structure, cannot deal with deviated subdialogue from problem structure, e. g. clarification dialogue, meta dialogue about system's ability etc.

In treating speech recognition errors in previous researches, the main point was implementing robust parsing. Major limit of robust parser is the phenomena of replacing a word by the same syntactic/semantic categorical word, e.g. if *Monday* is replaced by *Sunday*, robust parser cannot find out the replacement by its syntactic/semantic knowledge. In addition, the lack of selectional case word(s) cannot be found out by robust parser. Therefore, we have to give consideration the speech recognition error management in dialogue model.

From the above discussion, we decided that the major points in constructing dialogue model are closely combining sequential module, distinguishing linguistic structure and intentional structure, and constructing robust dialogue manager. Our dialogue model is a cognitive process model (1) which integrates the specified phased processing from utterance understanding to response generation, (2) which specifies the interactions between the processing of each steps and dialogue management mechanism, and (3) which identifies the possible errors caused by recognition error and the method of recovering the error.

2. COGNITIVE PROCESS MODEL OF DIALOGUE

2.1. Five Steps Process Model

We have specified the process from utterance understanding to response generation based on Airenti's cognitive process model [3]. Our extension is to deal with whole dialogue (Airenti's model treats only one turn) and to specify enough to implement spoken dialogue systems. We redefine the steps as (1) meaning understanding, (2) intention understanding, (3) communicative effect, (4) reaction generation, and (5) response generation (see Figure 1). Also, we specified the interaction between cognitive process and dialogue management subsystems. By these extensions, the model can deal with errors which occur at each steps in processing.

```
1. Meaning understanding
   if shared_bel(S, U, do(U, express(S, int(U, do(S, E))))) = true ∨ shared_bel(S, U, do(U, express(S, bel(U, P)))) = true
   then goto Intention understanding;
   else goto Response generation
2. Intention understanding
   if shared_bel(S, U, cint(U, S, int(U, do(U, S, G)))) = true ∨ (shared_bel(S, U, do(U, S, G)) ∧
      (shared_bel(S, U, cint(U, S, int(U, do(S, E)))) ∨ shared_bel(S, U, cint(U, S, P)))) = true
   then goto Communicative effect;
   else goto Response generation
3. Communicative effect
   if shared_bel(S, U, cint(U, S, int(U, do(U, S, G)))) = true      then try(int(S, do(S, U, G)));
   if shared_bel(S, U, cint(U, S, int(U, do(S, E)))) = true      then try(int(S, do(S, E)));
   if shared_bel(S, U, cint(U, S, P)) = true      then try(bel(S, P));
   goto Reaction generation
4. Reaction generation
   if shared_bel(S, U, cint(U, S, int(U, do(U, S, G)))) = true
      then (cint(S, U, int(S, do(S, U, G))) ∧ cint(S, U, int(S, do(S, E)))) ∨
         (cint(S, U, ¬ int(S, do(S, U, G))) ∧ cint(S, U, bel(S, P)))
   if shared_bel(S, U, cint(U, S, int(U, do(S, E)))) = true
      then cint(S, U, done(S, E)) ∨ (cint(S, U, ¬ int(S, do(S, E))) ∧ cint(S, U, bel(S, P)))
   if shared_bel(S, U, cint(U, S, P)) = true
      then cint(S, U, int(S, do(S, U, do(S, E)))) ∨ (cint(S, U, ¬ bel(S, P)) ∧ cint(S, U, bel(S, P')))
   goto Response generation
5. Response generation
   Ask back ∨ Generation by surface interaction rule ∨ Generation following the generated intention
```

S: spoken dialogue system, U: user, E: action, P : preposition, G :goal, try: predicate which tries to make given proposition true, express: expressing communicative intention.

Figure 2: Cognitive process in spoken dialogue

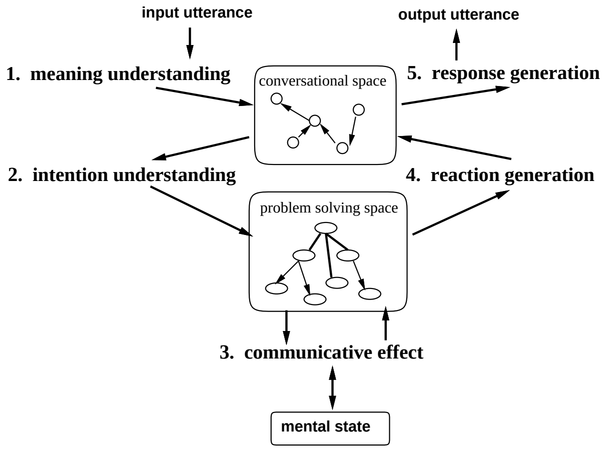

Figure 1: Five steps model of dialogue understanding

Figure 2 shows the information flow of overall processing and processing in each steps.

2.2. Conversational Space

In order to understand illocutionary act of utterance, to deal with elliptical expression, to generate proper response to other participant's utterance, the cognitive process model needs to have a management mechanism of the progress of conversation. In our model, we presuppose the existence of an interaction unit as the minimal unit of dialogue and it is managed in conversational space.

An interaction unit consists of *initiation*, *response* and *followup*. Initiation appears at the top of interaction unit. Response may appear after initiation successively. In some cases, before response utterance or at the place of it, another interaction unit may be inserted. Sometimes, interaction unit ends up by followup.

The role of conversational space is to maintain the pattern of interaction unit and develop dialogue by exchanging the information to the process model (mainly, in the intention understanding step).

Conversational space is a kind of dynamic network growing with the development of dialogue. In conversational space, there are three types of nodes: phrase node, instance node, and slot-filling node. The definition of the node is almost same as the definition in [8]. The relation of elements in conversational space is shown in Figure 3.

In this conversational space, not only objects, attributes, and discourse segment purpose, which Grosz et al. treated as the elements of focus, but also all the elements in interaction unit can use in processing of elliptical and referential expression according to the distance from present focus part of space. Also in this space, we can deal with surface interaction, e. g. clarification subdialogue, without consulting higher level knowledge.

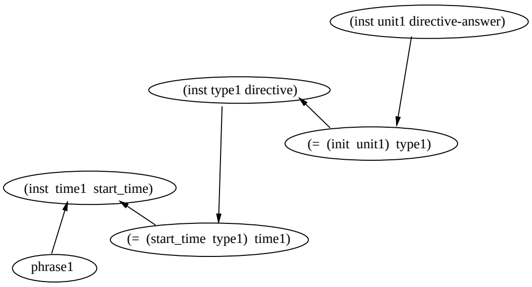

Figure 3: Relation of elements in conversational space

2.3. Problem Solving Space

We will feel spoken dialogue system as 'cooperative' if spoken dialogue system make proper answer and/or good suggestion. In order to generate such response, spoken dialogue system must recognize user's plan and select proper speech act as system's response. Furthermore, understanding process needs some information from higher level of processing, e.g. the function that answers whether recognized illocutionary act is a proper move in current user's plan. For these purpose, we need some planning mechanism in dialogue management. We decided to use *Event hierarchy* [9] as a method of representing the plan. It is suitable for plan recognition as a process of gathering observed actions into an end plan. We call this network *Problem Solving Space* (PSS).

PSS is a static network that represents relationships between plan and subplans, and between plan and actions (Figure 4). This space is used in intention understanding step and reaction generation step.

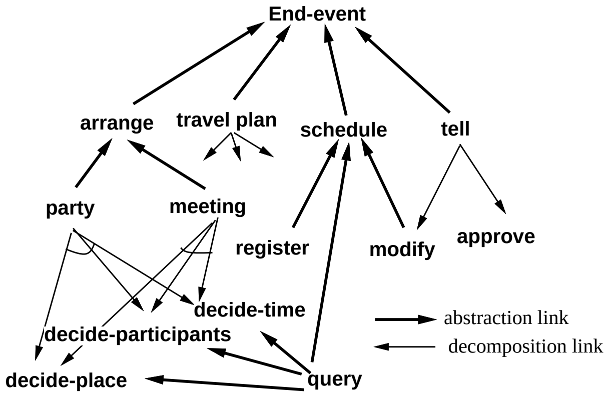

Figure 4: Problem Solving Space (part)

We apply *minimal covering* method [9] for plan recognition in PSS. The basic point of this procedure is to find forest that covers all the subplans and actions previously achieved.

3. EXAMPLE OF DIALOGUE PROCESSING

In this section, we show an example of system behavior. Figure 5 shows the example dialogue between user and personal schedule management system.

U1: "Register a meeting from 2 P.M."
S2: "Until what time?"
U3: "Please modify until 5."
(misrecognition: register → modify)
S4: "Do you want to modify it?"
U5: "Please register it."

Figure 5: Example dialogue between user and personal schedule management system

As a result of robust parsing, we assume we can get following surface semantic representation of U1.

shared_bel(S, U, do(U, lit-illoc(S, do(S, register(
[[start_time, 2], [obj, meeting]])), directive)))

In meaning understanding step, the surface semantic representation is interpreted as a initiation of turn, because there is no element in conversational space and literal illocutionary force of the surface semantic representation is directive. Then we can get following shared belief.

shared_bel(S, U, do(U, express(S, int(U, do(S,
register([[start_time, 2], [obj, meeting]]))))))

Next, in intention understanding step, as there is no shared plan between user and system, possible plan hypothesis, which user's action can be one of steps, is searched in problem solving space. The result of plan recognition is register_meeting_plan. Then we can get following two shared beliefs.

shared_bel(S, U, cint(U, S, int(U, do(U, S,
register_meeting_plan)))) ∧
shared_bel(S, U, cint(U, S, int(U, do(U, S,
register([[start_time, 2], [obj, meeting]])))))

In communicative effect, the validity of the recognized plan is checked in problem solving space and current mental states. If there is no problem both, the system has following two intentions.

int(U, do(U, S, register_meeting_plan)))) ∧
int(U, do(U, S, register([[start_time, 2],
[obj, meeting]])))))

In reaction generation step, the dialogue system finds out that user's action is valid but it needs another information to register a meeting. Then the system search for another action of complementing register_meeting_plan.

cint(S,U,int(S, do(U, inform_ref([end_time,S]))))

Finally, in response generation step, we use sentence template for inform_ref, in this case *motivateByInterrogative*, to make system's response (S2).

From the viewpoint of interaction unit, dialogue context is initiation (U1) followed by initiation (S2). In the pattern of interaction unit, U3 must be the response to S2, or initiation which relates to S2. However, the recognition result of U3 does not suit both hypotheses.

shared_bel(S, U, do(U, express(S,
  int(U, do(S, modify([[end_time, 4]]))))))

Then, supposing verb misrecognition, the system makes recovering sub dialogue S4. According to U5, the system replaces the verb (modify → register) at U3 in conversational space, deletes the interaction unit of recovering sub dialogue, and continues on dialogue.

4. EVALUATION

In evaluating the dialogue model, we used the environment for the evaluation of automatic system-to-system dialogues [10]. Figure 6 shows the concept of the environment.

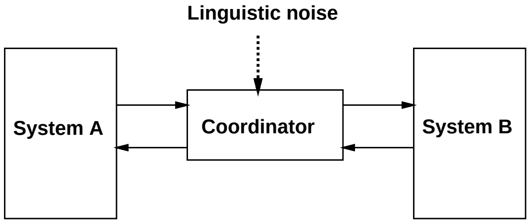

Figure 6: Concept of Evaluation environment

Random linguistic noise is put into the communication channel by the dialogue coordinator program. This noise is designed for simulating speech recognition errors. The performance of a system is measured by the task achievement rate (ability of problem solving) and by the average number of turns needed for task completion (conciseness of dialogue) under a given recognition error rate.

As a task domain of this experiment we selected personal schedule management described as dialogue example (Figure 5). We set the error rate of speech recognition 10 % / 25 % / 40 %. Each experiment is done 16 or 17 times. We simulate speech recognition errors by replacing one content word at given rate. This type of error reflects the errors often occurring in case of template matching in robust parsing.

We define the breakdown of dialogue in two pattern considering the rational user's behavior to present computer systems. The first pattern of breakdown is a failure in confirmation. If there are more than two errors in confirmation utterance, we assume that user gives up the dialogue because user may select another way of communication instead of uncertain speech input. The second pattern of breakdown is an excess of number-of-turns limit. We defined the limit as twice as error free dialogue.

Table 1 shows the results of this experiment.

Table 1: Robustness of dialogue model

error rate(%)	0	10	25	40
task achievement (%)	-	100	47	19
average turns (all)	7.0	7.7	9.7	10.0
average turns (success)	7.0	7.7	10.3	11.7

Under 10interaction caused by recognition error is 10robustness of this method in relative low recognition error situation. However, under 25redundant interaction yields 39found out information redundant utterance can be helpful in many situation.

5. CONCLUSION

We have propose a robust processing model of spoken dialogue and examined the validity of this model by system-to-system dialogue with linguistic noise evaluation. As a future research, we plan to evaluate the effect of plan recognition in the situation of misunderstanding.

References

[1] J. Peckham. Speech understanding and dialogue over the telephone: An ovrview of progress in the sundial project. In *Proc. of the 2nd European Conference on Speech Communication and Technology*, pages 1469–1472, 1991.

[2] A. Jonsson. A dialogue manager using initiative-response units and distributed control. In *Proc. of 5th Conference of the European Chapter of the Association for Computational Linguistics*, pages 233–238, 1991.

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

[4] B. J. Grosz and C. L. Sidner. Attention, intention and the structure of discourse. *Computational Linguistics*, 12:175–204, 1986.

[5] J. F. Allen, B. W. Miller, E. K. Ringger, and T. Sikorski. A robust system for natural spoken dialogue. In *Proc. of 34th Meeting of the Assoc. for Computational Linguistics*, pages 62–70, 1996.

[6] S. R. Young. The minds systems: using context and dialogue to enhance speech recognition. In *Proc. of DARPA Speech and Natural Language Workshop*, pages 131–136, 1989.

[7] R. W. Smith, D. R. Hipp, and A. W. Biermann. An architecture for voice dialog systems based on prolog-style theorem proving. *Computational Linguistics*, 21(3):281–320, 1995.

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

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

[10] M. Araki and S. Doshita. Automatic evaluation environment for spoken dialogue systems. In E. Mayer, M. Mast, and S. LuperFoy, editors, *Dialogue Processing in Spoken Language Systems*, pages 183–194. Springer, 1997.