EXTRACTION OF THE DIALOG ACT AND THE TOPIC FROM UTTERANCES IN A SPOKEN DIALOG SYSTEM

Y. Niimi *N. Takinaga* *T. Nishimoto*

Dept. of Electronics and Information Science, Kyoto Institute of Technology
Matsugasaki, Sakyo-ku, Kyoto, 6068585 JAPAN

ABSTRACT

This paper presents an approach to extraction of dialog acts and topics from utterances in a spoken dialog system. Two knowledge sources are used to describe the dialog history. One is a transition network of dialog acts and the other is a tree of topics which might appear in domain communications. Dialog acts and topics are extracted through bottom-up and top-down analyses. Bottom-up candidates are decided by applying a set of specially designed rules to the semantic representation of an utterance, and top-down candidates by using the current state of the dialog history. The logical ANDs between bottom-up and top-down candidates are taken to decide the dialog act and topic of an utterance. This method was examined with a corpus of fourteen dialogs including 335 utterances. Correct extraction rates were 85% for the topic and 82% for the dialog act.

1. INTRODUCTION

The recent advance of speech technology has made it possible to build continuous speech recognition systems working in real time. Using such systems as an interface, we can construct human-machine dialog systems. In spoken dialog systems, the analysis of dialog structures plays important roles in interpreting utterances. First it is capable of context dependent interpretation of utterances, that is, complement of ellipses, and resolution of anaphoric expressions. Second the discourse history resulted from the discourse analysis makes it possible to predict what a user will speak next, which leads to dynamic switching of language models in speech recognition, thus resulting in increase of recognition accuracy. The discourse history can be described in terms of topics and dialog acts[1]. The discourse analysis has two aspects. One is to extract a topic and a dialog act from individual utterances, and the other is to describe the transition of topics and dialog acts through a dialog.

There have been knowledge based approaches and corpus based approaches to the discourse analysis. Knowledge based approaches use a set of rules to extract topics and dialog acts from utterances and to describe the transition of these through dialogs[2],[3]. The set of these rules are designed mainly based on expertise of linguists. On the other hand corpus based approaches use two probabilities $P(T|W)$ and $P(D|W)$; $P(T|W)$ is a conditional probability of a topic T and $P(D|W)$ is a probability of a dialog act D given a set of words W included in an utterance[4]. These two probabilities are estimated from annotated corpora and used to extract a topic and a dialog act from an utterance. Moreover, n-grams model or HMMs are used to describe the transition of topics and dialog acts through dialogs[5]. An amount of precisely annotated corpora would estimate reliable dialog models although it takes tremendous efforts to collect such corpora.

This paper presents a knowledge based approach to extraction of topics and dialog acts from utterances of a user in a spoken dialog system. In this paper the task of a spoken dialog system is information service for sightseeing. Thus, topics and dialog acts processed in the system are domain specific, but most of dialog acts are common to those processed in other approaches. Two knowledge sources are used to describe the dialog history. One is a transition network of dialog acts which describes possible transitions from a dialog act to another. The other is a tree of topics which might appear in sightseeing dialogs. We assume topics develop along this topic tree. Thus topics developed while a dialog goes on form a subtree of this tree, which we call a dynamic topic tree.

Dialog acts and topics are extracted through bottom-up and top-down analyses. Bottom-up candidates for dialog acts and topics are decided by applying a set of specially designed rules to the semantic interpretation of an utterance. Top-down candidates for dialog acts and topics are decided by using the current state of the dialog history which is described by a trace in the transition network on the dialog act and the dynamic topic tree. Then, the logical ANDs between the bottom-up and top-down candidates are taken to decide the dialog act and topic of an utterance of the user.

This method was examined with a corpus of fourteen dialogs including 335 utterances. Correct extraction rates were 85% for the topic and 82% for the dialog act. This result shows the proposed knowledge based method is quite promising although it was examined with a small amount of corpus.

S101	This is the tourist information service in Kyoto. Can I help you?
U101	I'd like to do a day tour in Kyoto.
S102	What are you interested in?
U102	I'd like to visit gardens.
S103	I see. Gardens in what eras would you like to visit?
U103	Muromachi and Momoyama.
S104	Among famous gardens in Muromachi era are the gardens of Ryoanji temple, Kinkakuji temple and Ginkakuji temple.
U104	The garden of Ryoanji is a famous stone garden, isn't it?
S105	That's right.

Figure 1: An example of the dialog.

2. REPRESENTATION OF THE DIALOG HISTORY

2.1. Topics

Fig.1 illustrates an example of dialogs which the system would have with users. First the system, repeating questions to a user, elucidates the specification of his sightseeing tour, that is, a period of the tour, a hotel to stay, places to visit and so forth, and then offers some candidates for these items. Then the user, inquiring detail information on these items, decides what are worth to involving in his plan.It has been known that topics in a goal-oriented dialog move according to a task-dependent tree structure. In fact the topics in the illustrated example are developed along the structure as shown in Fig.2.

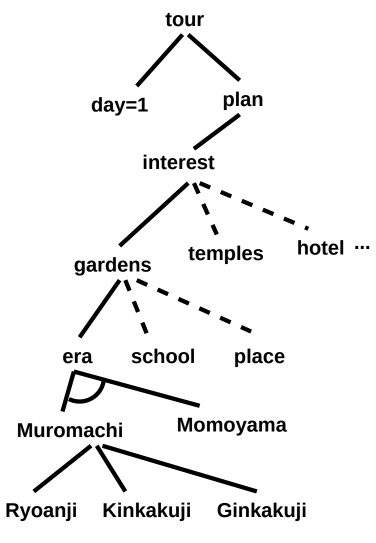

Figure 2: A part of Topic Table.

Our spoken dialog system has a set of 'topic frames' as a knowledge source on topics. A topic frame forms mutually related topics into a frame which might appear in sightseeing dialogs. For example, the name of a hotel, the room charge, the location and so forth form a hotel frame. Since slots of a frame can take other frame as their value, a set of topic frames forms a tree as whole, which we call a topic tree. We assume topics develop along this topic tree. Thus topics mentioned while a dialog goes on form a subtree of the topic tree, which we call a dynamic topic tree.

As we have reported in [3],however, an AND-OR tree is more suited for representing movements of topics than a simple tree. In the AND-OR tree, AND-nodes represent topics introduced by a user, and OR-nodes represent topics introduced by the system. If the user inquires about two or more sights (each assumed to be a topic), the system must offer information on all of them. On the other hand, even if the system proposes two or more candidates for a visit, the user is not interesting in all of them, but might move to the other topics. An AND-OR tree is suited to reflect this difference.

Nodes of a dynamic topic tree has another distinction, prototype and instance. A prototype node is a node derived from the topic tree and an instance node is created from data retrieved from the database on sightseeing spots. Instance nodes are leaf nodes of a dynamic topic tree and thus cannot be expanded. Nodes of a dynamic topic tree are also tagged by one of four labels: focused, suspended, closed and unvisited. This information is used in focus-shift control.

2.2. Dialog Acts

Each utterance in a dialog has its own purpose, that is, an intention a user wants to convey to the dialog system. In this paper we call this a dialog act. Table 1 lists a set of dialog acts used for domain communications together with their corresponding instances from Fig.1. Confirmation in the list is for a user to confirm his/her knowledge or belief, or for the system to confirm its inference, for example, based on the default features, but is not used as means of meta communications to clarify ambiguity of speech recognition. The dialog acts shown in Table 1 are upper categories of the dialog act, and each is divided into domain specific subcategories. The spoken dialog system has as a knowledge source on dialog acts a state transition network in which a state corresponds to a dialog act. This network describes possible transitions of dialog acts through dialogs. Thus, the discourse history on dialog acts is represented by some state in this network.

Table 1: Dialog Acts

Dialog Acts	Examples
GR : Greeting	S101
AI : Ask-information	S102, S103
GI : Give-information	U101, U102, U103 S104
CN : Confirmation	U104
RC : Response-to-confirmation	S105
AK : Acknowledgement	—

3. DISCOURSE ANALYSIS

3.1. General Framework

Fig.3 shows the flow of the discourse analysis proposed in this paper. It involves bottom-up and top-down analyses. Given a semantic interpretation of an input utterance, the bottom-up analysis produces bottom-up hypotheses, that is, candidates for topics and dialog acts. The top-down analysis predicts topics and dialog acts to likely appear in the current utterance referring to the dialog history which is represented by the AND-OR tree of topics and a state of the transition network on dialog acts. These top-down candidates are ordered in a heuristic way.

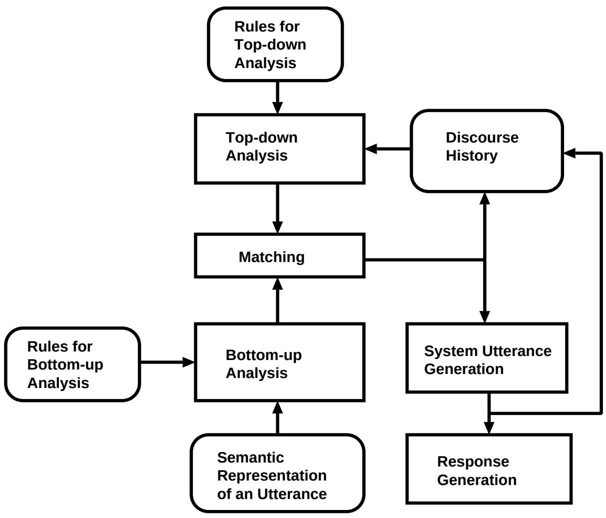

Figure 3: An Overview of the Discourse Analysis.

The bottom-up hypotheses are then matched against the top-down predictions. The best match gives the topic and the dialog act of an utterance under consideration, which are in turn preserved into the dialog history and also passed to the system utterance generation module. The topic and the dialog act of the next utterance issued by the dialog system are decided by this module, and also preserved into the dialog history.

The semantic interpretation of an utterance is produced based on the case grammar. In the case grammar the meaning of a sentence is represented by a case frame associated with a main verb of that sentence. A case frame is described by a set of slots, each indicating one of such relations between a verb and a noun phrase, like an agent, object and instrument. Noun phrases included in an utterance are assigned to some slot of the case frame based on semantic markers of the noun phrases. Thus, the semantic interpretation of an utterance is represented by a list of three terms, a main verb, a case frame with slots filled, modality information including the style of an utterance.

3.2. Extraction of Topics

All the words our spoken dialog system can accept are semantically grouped and organized as a tree structure like a thesaurus. The words corresponding to upper nodes (close to the root) in this structure are called concept words, and play roles of semantic markers in interpretation of utterances. As stated in section 2, topics which are mutually related are formed into a topic frame. To each slot of a topic frame are given semantic markers (concept words) of words which can fill that slot. Relations among a topic, a slot and semantic markers of that slot are used as a knowledge source for the bottom-up analysis on the topic. These relations are organized as a table called a topic_slot table. That is, a topic and a slot meeting a relation mentioned above can be retrieved by looking up a concept word in the topic_slot table.

There is another knowledge source called a verb_focus table which is formed from case frames of verbs. In our task domain a slot filler of a verb with a specific case marker tends to be focused as a topic. The verb_focus table contains this relation.

Bottom-up candidates for topics are decided by using these two tables. First for each slot filler of the case frame of an utterance, pairs of a topic and a slot are found by looking up in the frame_slot table, and then pairs with the same topic are merged into a topic frame with two or more slots filled. Then remaining topic frames are pruned and a focused slot is decided by consulting the verb_focus table.

The top-down analysis on the topic is conducted based on the discourse history; the dynamic topic tree and the focused node, and the dialog act of the utterance issued immediately before by the dialog system. Furthermore, there is a knowledge source for the top-down analysis, called focus_shift rules. These rules describe what a node in the dynamic topic tree the focus moves to according to a dialog act of a system utterance, and are ordered based on the extent to which each focus shift can occur. For example, if a dialog act of a system utterance is 'ask_information', then the item asked could be assumed to be focused, and a user is expected to answer that question, which means that the focus does not move. If a dialog act of a system utterance is 'give_information', then a user is expected to have obtained an answer to his/her question, which could be assumed to be focused, and can thus move the focus to any topic so far as coherency of a dialog is kept. Thus, by applying focus_shift rules to the currently focused node in the dynamic topic tree, topics which can appear in the next utterance of a user can be decided together with their priorities.

3.3. Extraction of Dialog Acts

The dialog system has a table for the bottom-up analysis on the dialog act. The table describes relations of a dialog act to a predicate and its modality information of an utterance. In this table predicates are grouped into four types: (1) verbs to express sightseeing actions of a user like 'visit' and 'stay', (2) verbs to demand information like 'tell me', (3) verbs to express user's actions to sightseeing plans like 'add ... to my plan' and (4) predicates to express responses like 'I see' and 'That's fine'. Modality information is grouped into question and non-question. The dialog acts listed in Table 1 are classified as shown in Table 2. The case frame of an utterance and some clue words, if any, are used to resolve ambiguities in Table 2 and to classify those dialog acts into task dependent subcategories.

Table 2: Classification of Dialog Acts

Predicate Types	Modality	
	question	non-question
(1)	AI, CN	GI
(2)	AI, CN	AI
(3)	AI, CN	GI
(4)		AI, GI, RC, AK

As stated in section 2, the discourse history on the dialog act is represented by a current state in the state transition network describing possible transitions of dialog acts. The top-down prediction on the dialog act can be obtained simply as dialog acts which can be reached by a step from the current state.

4. PERFORMANCE TEST

In order to test the proposed method for automatic extraction of topics and dialog acts, we collected a corpus of dialogs through Oz method. A subject was given a simple scenario for sightseeing and asked to make a sightseeing plan by consulting the information service system. Twenty two dialogs were collected from twenty two subjects. One of the authors manually annotated a topic and a dialog act to each utterance in this corpus. All the rules explained in section 3 have been designed mainly by general knowledge of authors and adjusted by using eight dialogs randomly selected out of twenty two.

The proposed algorithm has been tested using user's utterances (of which the number is 355) in the remaining dialogs. Table 3 shows extraction rates in the bottom-up analysis only and in combination of the bottom-up and top-down analyses. Column 'UQ' indicates rates of the cases a unique candidate is proposed for each utterance and coincides with manually annotated one, column 'ML' rates of the cases multiple candidates are proposed and either of them coincides with annotated one, and column 'ER' rates of the cases any candidate proposed does not coincide with annotated one or no candidates are proposed.

The table shows only the bottom-up analysis has attained correct extraction rates of 65% for the topic and 71% for the dialog act. The combination of the bottom-up and top-down analyses drastically increases correct extraction rates to 85% for the topic and 82% for the dialog act. This result shows the proposed knowledge based method is quite promising although it was examined with a small amount of corpus. The reason ML rates do not vanish in the combination of the two analyses is that top-down candidates with the same priority are produced in the ordering process, and the reason ER rates increase in the combination is that the intersection of bottom-up and top-down candidates happens to become empty.

Table 3: Results of Extraction of the Topic and the Dialog Act

	Bottom-up(%)			Combination(%)		
	UQ	ML	ER	UQ	ML	ER
Topic	64.5	29.3	6.2	84.8	8.2	7.0
Dialog Act	70.7	23.9	5.4	82.1	12.1	5.9

UQ : A candidate is uniquely decided.
ML : Multiple candidates are produced.
ER : None of candidates coincide with manual tags or are not produced.

5. CONCLUSION

This paper has reported a knowledge based approach to automatic extraction of topics and dialog acts from utterances in a spoken dialog system. The proposed method involves the bottom-up and top-down analyses. The bottom-up analysis proposes candidates for topics and dialog acts by applying a set of specially designed rules to the semantic representation of an utterance. The top-down analysis proposes candidates for both referring to the discourse history. The bottom-up candidates are matched against the top-down ones. The best matches give the topic and the dialog act of an utterance under consideration.

This method was examined with a corpus of fourteen dialogs including 335 utterances. Correct extraction rates were 85% for the topic and 82% for the dialog act. Most of incorrect extractions were due to that a set of rules for extraction and the coverage of topics were incomplete. We should improve these points and examine the proposed algorithm with a larger corpus of dialogs.

6. REFERENCES

[1] Grosz, J.B. and Sidner, C."Attentions, Intentions, and the Structure of Discourse," Computational linguistics, Vol.12, 175- 204, 1986.

[2] Iwadera, T., Ishizaki, M., and Moromoto, T. "Recognizing An Interactional Structure and Topic-Oriented Dialogues," Proc. of ESCA Workshop on Spoken Dialogue Systems, 41-44, 1995.

[3] Niimi, Y. and Kobayashi, Y. "A Top-Down Discourse Analysis In A Speech Dialogue System," Proc. of Fifth European Signal Processing Conference, 1275-1278, 1990.

[4] Garner, P.N., Browning, S.R., Moor, R.K. and Russell, R.J. "A Theory of Word Frequencies and Its Application to Dialogue Move Recognition," Proc. of Int. Conf. on Spoken Language Processing, 1880-1883, 1996.

[5] Kita, K., Fukui, Y., Nagata, M. and Morimoto, T. "Automatic Acquisition of Probabilistic Dialogue Models," Proc. of Int. Conf. on Spoken Language Processing, 196-199, 1996.