A Rule Based Approach to Extraction of
Topics and Dialog Acts in a Spoken Dialog System

Yasuhisa NIIMI, Tomoki OKU, Takuya Nishimoto and Masahiro ARAKI

Kyoto Institute of Technology
Matsugasaki, Sakyou-ku, Kyoto 6068585, Japan
{niimi, nishi, araki}@d.j.kit.ac.jp, tomoki@vox.d.j.kit.ac.jp

Abstract
This paper presents a rule based approach to extraction of
dialog acts and topics from utterances in a spoken di-
alog system, SDSKIT-3, with a task-independent dialog
controller which is based on an extension of the frame-
driven method. We demonstrated it could control dialogs
in several different task domains, only given a set of topic
frames and a set of rules manually designed for the dis-
course analysis which were both task-dependent. In this
paper we report an approach to semiautomatic derivation
of a set of rules for the discourse analysis from informa-
tion needed to specify a task, for example, a set of topic
frames, a conceptual tree of those words and case frames
of those verbs which are likely to be used in the task do-
main. This method was examined with a corpus of twenty
dialogs. Correct extraction rates were 82% for the topic
and 80% for the dialog act.

1. Introduction
In spoken dialog systems, the analysis of dialog struc-
tures plays important roles in interpreting utterances.
First it is capable of context dependent interpretation of
utterances. Second the discourse history resulted from
the discourse analysis makes it possible to predict what
a user will speak next, which leads to dynamic switch-
ing of language models in speech recognition. The dis-
course 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 an indi-
vidual utterance, and the other is to describe the transition of
topics and dialog acts through a dialog.

There have been rule based approaches and corpus
based approaches to the discourse analysis. Rule 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]. These rules are designed mainly
based on expertise of linguists. On the other hand cor-
pus 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 utter-
ance. An amount of precisely annotated corpora would
estimate reliable dialog models although it takes tremen-
dous efforts to collect such corpora.

This paper presents a rule based approach to extrac-
tion of dialog acts and topics from utterances in a spoken
dialog system, SDSKIT-3 (Spoken Dialog System at Ky-
oto Institute of Technology) we have been developing. It
works under a task-independent dialog controller which
is based on an extension of the frame-driven method. We
demonstrated it could control dialogs in several different
task domains, only given a set of topic frames and a set of
rules manually designed for the discourse analysis which
were both task-dependent[5],[6].

In this paper we report an approach to semiautomatic
derivation of a set of rules for the discourse analysis from
information needed to specify a task, for example, a set
of topic frames, a conceptual tree of those words and case
frames of those verbs which are likely to be used in the
task domain. This method was examined with a corpus of
twenty dialogs including 186 utterances. Correct extrac-
tion rates were 82% for the topic and 80% for the dialog
act. Informal tests conducted in other three different task
domains showed comparable results, although a small
amount of corpus was used in each task domain. This
shows the proposed rule based method is quite promis-
ing.

We will outline the spoken dialog system, SDSKIT-3,
in section 2. Then we will describe the discourse analysis
and how to derive rules used therein in section 3, and
experiments carried out to test the proposed method in
section 4. Lastly section 5 concludes the paper.

2. Overview of SDSKIT-3

2.1. An Example of Dialogs

Fig.1 illustrates an example of dialogs which the system
would have with users in a domain, information service
for sightseeing. 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.

SDSKIT-3 is composed of a speech interface and a dialog controller. The speech interface has two streams; speech input and speech output. In the speech input stream utterances are converted into string of words and then into semantic representations, while an inverted process is performed in the output stream. The dialog controller extracts a topic and a dialog act from a semantic representation of an utterance, performs some necessary actions, for example, retrieving information from a database and creates an abstract response to a user.

Figure 1: An example of the dialog.

2.2. Semantic Analysis

We use three kinds of knowledge sources for the semantic analysis of utterances; a simple thesaurus, a set of case frames, and a set of syntactic rules for acceptable sentences.

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.

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. In SDSKIT-3 case frames and syntactic rules are integrated in a framework of the definite clause grammar (DCG).

2.3. Topic Frames

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 a task domain. A topic frame is a set of slots. A slot is described by a slot name, a range of values to fill the slot, a method for filling the slot, and an action of the controller after the slot has been filled. The range of values are usually described by concept words. In addition, the slots of a topic frame are given the order of filling themselves. There are four ways to fill an unfilled slot; (1) to ask a user, (2) to retrieve instances from a database, (3) to use the default value attached to the slot, and (4) to link one of other topic frames. The dialogue controller takes one of three actions after a slot has been filled; these are (1) to move the control to the slot with the next highest priority, (2) to move the control to the topic frame just linked, and (3) to present the information a user has asked.

Since slots of a frame can take other frame as their value, a set of topic frames forms an implicit tree as whole, which we call a topic tree. We assume topics develop along this topic tree during a dialog. This forms a subtree of the topic tree, which we call a dynamic topic tree. The topics in the dialog illustrated in Fig. 1 are specialized along the structure as shown in Fig. 2.

Figure 2: A part of the dynamic topic tree.

2.4. 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. The dialog acts shown in Table 1 are upper categories of the dialog act, and each is divided into 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. Since we use the same set of dialog acts for several different task domains, this transition network is task-independent.

### 2.5. Frame-based Dialog Control Scheme

As mentioned previously, SDSKIT-3 represents the dialog history on the topic with a dynamic topic tree and retains a pointer to the focussed slot of the node. This is the situation that SDSKIT-3 is about to interpret user utterances or to speak to the user. Fundamentally the behaviors of SDSKIT-3 are subjected to the structure of the dynamic topic tree. However, it acts differently in interpreting user utterances and in speaking to the user.

First we will explain how it works in the interpretation of user utterances. As outlined in the next section, the bottom-up discourse analysis produces candidates for the topic, and the top-down discourse analysis searches for some of the bottom-up candidates in the dynamic topic tree equidistantly from the focussed node. The distance between two nodes is defined by the number of edges which must be passed to move one node to another. When nodes without a child node are found in the search, they are expanded by copying corresponding subtrees of the topic tree.

Next we will explain how the dialog controller works in speaking to the user. Its behaviors are quite simple in this case. It searches for an unfilled slot in the dynamic topic tree in the depth-first way from the focussed node, and tries to fill that slot. When the topic corresponding to the current node has been terminated by the utterance the user just issued, the dialog controller first backtracks on the dynamic topic tree to the first unterminated node, and makes it the new focussed node. Then the depth-first search restarts from that node to find an unfilled slot. This is the principle of the topic shift by the dialog controller.

### 3. Discourse Analysis

#### 3.1. General Framework

The discourse analysis proposed in this paper 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, by applying to it rules derived from task-dependent knowledge sources. 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 dynamic topic tree and a state of the transition network on dialog acts. These top-down candidates are ordered in a heuristic way.

The bottom-up hypotheses are 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.

#### 3.2. Extraction of Topics

We have two tables for the bottom-up analysis on the topic; topic-slot table and verb-focus table. These tables are made up from task-dependent knowledge sources. The topic-slot table is automatically made up from the set of topic frames. As stated in section 2.3, concept words for values which can fill that slot are given to each slot of a topic frame. Relations among a topic, a slot and concept words for that slot are organized as a table in which a key is a concept word. 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.

The verb-focus table is formed from case frames of verbs. In several task domains a specific case filler of a verb tends to be focussed as a topic. The verb-focus table contains the relation between a verb and its case which might be focussed.

Bottom-up candidates for topics are decided by using these two tables. The bottom-up analysis, looking up each case filler of the semantic representation of an utterance in the frame-slot table, first finds frames some of whose slots can be filled with that case filler. Then it merges the pairs with the same topic into a topic frame with two or more slots filled, and lastly decides a focussed slot by consulting the verb-focus table.

The top-down analysis searches for some of bottom-up candidates in the dynamic topic tree equidistantly from the focussed node. The best matches are the candidates with the shortest distance.

#### 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 task-oriented actions of a user.
like 'visit' and 'stay', for example in 'information service for sightseeing.' (2) verbs to demand information like 'tell me', (3) verbs to express user’s decisions in a considered task domain like 'add ... to my plan' and (4) predicates to express responses like 'I see' and 'That's fine'. This classification should be done by referring to dialog corpus on a specific task domain. 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.

### Table 2: Classification of Dialog Acts

<table>
<thead>
<tr>
<th>Predicate Types</th>
<th>Modality</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>AI, CN</td>
</tr>
<tr>
<td>(2)</td>
<td>AI, CN</td>
</tr>
<tr>
<td>(3)</td>
<td>AI, CN</td>
</tr>
<tr>
<td>(4)</td>
<td>AI, GI, RC, AK</td>
</tr>
</tbody>
</table>

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. The task domain selected for the test is 'route guidance'. Conversing with the Oz, users can find a route from one place to another. The task they must achieve is to decide what transportation to take, how much time to take to go from one place to another and how much money to spend. Twenty dialogs were collected from twenty subjects. A trained student except the authors manually annotated a topic and a dialog act to each utterance in this corpus. All the knowledge sources necessary to define the task, for example, a set of topic frames were designed mainly by general knowledge of authors and adjusted by using ten dialogs randomly selected out of twenty. The tables explained in section 3 were semiautomatically derived from these knowledge sources.

The proposed algorithm was tested using utterances (of which the number is 186) in the remaining dialogs. The extraction rate was 81.7% for the topic and 80.1% for the dialog act. Most errors were due to that the definition of case frames of verbs was imperfect and that test utterances included verbs out of vocabulary. We also conducted informal tests for other different task domains, like 'guide for personal computer buyers', 'restaurant guide' and 'trouble checking of cars'. For these tasks extraction rates were around 80% for the topic and the dialog act, although the numbers of dialogs collected in these task domains were small, that is, about ten for each. These results show the proposed rule based method is quite promising although it was examined with a small amount of corpus.

### 5. Conclusions

This paper has reported a rule 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 rules to the semantic representation of an utterance. These rules are semiautomatically derived from the definition of a specific task. 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 twenty dialogs including 186 utterances. Correct extraction rates were 82% for the topic and 80% for the dialog act. Informal tests conducted in other three task domains showed comparable results. These results shows the proposed method is promising. We should examine the proposed algorithm with a larger corpus of dialogs.

### 6. References