Dialogue Management Based on Inferred Behavioral Goal
— Improving the accuracy of understanding
by dialogue context —

Keisuke Watanabe Yasushi Ishikawa

Information Technology R&D Center, MITSUBISHI Electric Corporation
5-1-1 Ofuna, Kamakura, Kanagawa, 247-8501, Japan
{ksuke, yasushi}@isl.melco.co.jp

Abstract

A dialogue management method for a speech-based interactive system is described. The construction of an effective spoken dialogue system that has unrestricted input requires user-initiative dialogue management techniques. In order to realize user-initiative dialogue, however, the system must basically accept all sentences included in its language model for recognition. That creates the problem of lower recognition accuracy, because of the weaker constraints with respect to speech understanding. We have thus proposed a method for scoring the appropriateness for the dialogue context to each of the N-best speech understanding results. In this method, the user’s behavioral goal is inferred from a history of utterances and a dialogue context score is calculated for each of the N-best candidates based on transition probabilities of the behavioral goals. According to the results of preliminary evaluation experiments using a dialogue system focusing on a task of making hotel reservations, we succeeded in reducing the error rate (misunderstandings in acoustic scores only) to 66% for 15 sentences.

1 INTRODUCTION

With the improvements in continuous speech recognition technologies in recent years, the possibilities have opened up substantially for practical use of spontaneous speech interfaces. The greatest advantage of spontaneous speech input is that the user can input his thoughts directly, allowing an extremely wide range of applications including searches for telephone numbers and other information, reservations of hotels or airplane tickets, and transactions involving stocks or investment securities. The development of an effective speech dialogue system that makes use of the merits of such spontaneous speech input cannot be achieved with system-initiative dialogue management in which the system regulates the dialogue transitions based on preset scenarios; what is required is a user-initiative dialogue management technology that allows the dialogue to proceed while essentially leaving the initiative with the user.

Nevertheless, in order to realize user-initiative dialogue, generally all sentences included in its language model must be available for recognition at all times. With respect to speech understanding, this leads to a problem in which the constraints are weakened and the accuracy of understanding is reduced.

Several approaches to solving this problem include dialogue strategies, such as permitting a certain degree of recognition error while employing confirmation dialogue based on the reliability of the recognition results[1] and attempts to increase the recognition rate using dialogue context or the status of an application[2]. The former method has the disadvantage of basically increasing the amount of interaction. The latter approach requires the design of detailed rules for identifying task-dependent contexts.

We have thus proposed a method for scoring the appropriateness for the dialogue context to each of the N-best understanding results. In this method, the relations between the overall behavioral goal — the motivation for the utterance — and the various subordinate goals are represented as a hierarchical structure, and attach utterance intentions to each of the behavioral goals. The user’s behavioral goal is inferred from the degree of activity that is calculated to each goal from the history of intentions. Furthermore, by assigning a transition probability between the various behavioral goals, we can establish a dialogue context score, which is determined from the occurrence probability of the behavioral goal and the inferred behavioral goal transition probability for each candidate. By combining this score with the acoustic and language score, we can obtain the most appropriate speech understanding result from among several speech understanding candidates.

In Section 2, we describe the dialogue management methods based on behavioral goals, which are the basic framework for our user-initiative dialogue management methods. In Section 3, we present the methods of extracting the most appropriate candidate from N-best results based on dialogue context. Results of evaluation experiments with a dialogue system are presented in Section 4. In Section 5 we discuss these results and Section 6 concludes the paper.
2 USER-INITIATIVE DIALOGUE MANAGEMENT METHODS

When a user-initiative dialogue is being conducted, in order for a user to efficiently achieve his dialogue goals, there must be appropriate interactions from the system consistent with the dialogue situation. For this to happen, it is important to infer the user’s dialogue goals which change in keeping with the progress of the dialogue and ensure that the system’s operational goals are consistent with the user’s dialogue goals.

2.1 Behavioral Goals

The user’s global dialogue goal, which becomes the motivation for an utterance, is defined as the behavioral goal. The behavioral goals are established depending on the task, and the relationships between goals are defined by the following hierarchical structures:

1. Essential vertical relationship
   a relationship in which the lower level behavioral goal must be achieved in order to achieve the higher level behavioral goal

2. Vertical relationship
   a simple relationship of higher and lower level goals

Fig.1 illustrates an example of behavioral goals in the case of the hotel reservation task. The solid lines indicate essential vertical relationships, and the dotted lines indicate vertical relationships. The starting point of an arrow indicates the lower level behavioral goal, while the end point indicates the higher level goal.

2.2 Inference of Behavioral Goals

Behavioral goals are inferred based on the activity of the behavioral goals as calculated from the sequence of user’s utterances. For the various behavioral goals $G$ illustrated in Fig.1, we define the sets of intentions $I_G = I_1, \ldots, I_N$ that activate those goals and the change rate $\delta(I_t)$ of the activity.

Each time an utterance $U_t$ is input, the activity $A_G^{(t)}$ for each behavioral goal $G$ is updated using (1). $I_t$ is the intention for utterance $U_t$. The activity for the lower-level goal $G_L$ is transferred to the higher level goal based on a fixed ratio $\alpha$. In the case of behavioral goals for which updates were not carried out consecutively $N_a$ times or more, the activity is reduced by multiplication with the constant $\beta(<1)$. The behavioral goal for which the activity exceeds the threshold is taken as the inference result.

$$A_G^{(t)} = \begin{cases} \delta(I_t) \cdot A_G^{(t-1)} + \sum_{G_L \in G_L} \alpha \cdot A_{G_L}^{(t)}, & \text{if } I_t \in I_G \\ \beta \cdot A_G^{(t-1)}, & \text{if } t - t_0 > N_a \\ 0, & \text{otherwise} \end{cases}$$

(1)

2.3 Initiative Control

The control of initiative is carried out using the following strategy, depending on the inferred behavioral goal.

**strategy1** When there is no inference result

Continue dialogue with user initiative

**strategy2** When no inference result status continues for a prescribed number of turns

Ask the user his behavioral goal directly

**strategy3** When there are multiple inference results

Confirm with the user the behavioral goal having the highest activity

**strategy4** When there is one inference result, and the essential lower-level behavioral goal for that result has not yet been achieved

Set the achievement of the essential lower-level goal as the system’s goal

**strategy5** When there is one inference result, and the essential lower-level behavioral goal for that result has been achieved

Set the achievement of the inferred behavioral goal as the system’s goal

3 SCORING BASED ON DIALOGUE CONTEXT

This section describes a method by which the knowledge bank for behavioral goals is made more detailed and used as top-down knowledge to improve the accuracy of speech understanding.
S1: Hello, this is the Reservation Center for hotels in Yokohama.
U1: I’m looking for a hotel near Shin-Yokohama Station that has a room available for one night on December 16th.
S2: We have three hotels meeting those requirements; the Shin-Yokohama International Hotel, the Fuji View Hotel, and the Shin-Yokohama Prince Hotel.
U2: Does a single room at the Shin-Yokohama International Hotel cost less than 10,000 yen?
S3: Yes, a single room at the Shin-Yokohama International Hotel is 9,300 yen.

Figure 2: Example dialogue

<table>
<thead>
<tr>
<th>Rank</th>
<th>Recognition Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Jaa, yon ko onegaisimasu (Okay, I’d like four please.)</td>
</tr>
<tr>
<td>2</td>
<td>Jaa, ee ni ko onegaisimasu (Okay, I’d like – um – two please.)</td>
</tr>
<tr>
<td>3</td>
<td>Jaa, yoyaku onegaisimasu (Okay, Please make the reservation.)</td>
</tr>
</tbody>
</table>

### 3.1 Appropriateness of Dialogue Context

Fig.2 shows an example dialogue for the hotel reservation task. In this dialogue context, when there are three recognitions such as making reservations or retrieving information, the system’s action is calculated using the following algorithms.

#### step1
The activity $A_{G}^{(t-1)}$ of behavioral goal $G$ is updated with the $t-1$-th utterance. The activity for all behavioral goals is normalized, and this becomes the probability value $p_{t-1}(G)$ for the behavioral goal by (3).

$$
\begin{align*}
    p_{t-1}(G) &= \frac{A_{G}^{(t-1)}}{\sum_{G} A_{G}^{(t-1)}} && (3)
\end{align*}
$$

#### step2
Assume the hypothesis $H_t$ to be the $t$-th input, the activity $A_{G}^{(t)}$ is calculated for all behavioral goals and the behavioral goal is inferred. The activity for all behavioral goals is normalized, and this becomes the probability value $p_t(G|H_t)$ by (4).

$$
\begin{align*}
    p_t(G|H_t) &= \frac{A_{G}^{(t)}}{\sum_{G} A_{G}^{(t)}} && (4)
\end{align*}
$$

#### step3
The dialogue context score $S_c(H_t)$ is calculated by (5).

$$
S_c(H_t) = \max_{G, G'} \{ p_t(G'|H_t) \cdot p(G'|G, S) \cdot p_{t-1}(G') \} && (5)
$$

The score $S(H_t)$ for the hypothesis $H_t$ is calculated as the sum of the dialogue context score $S_c(H_t)$ and the acoustic score $S_a(H_t)$ by (6). The $H_t$ with the highest $S(H_t)$ is selected as the recognition result.

$$
S(H_t) = S_a(H_t) + w \cdot S_c(H_t) && (6)
$$

### 3.2 Transition Probability for Behavioral Goals

We will consider transition probabilities for the behavioral goals, in order to assign scores for evaluating the appropriateness of the dialogue context to each hypothesis of the N-best understanding results.

The probability for transition from behavioral goal $G$ to behavioral goal $G'$ is defined as in (2).

$$
p(G'|G, S) = \{ s(G), s(v_1), \ldots, s(v_M) \} && (2)
$$

Transition probability is dependent not only on behavioral goal $G$, but also on the status of behavioral goals and the parameters required for the system’s action. In (2), $s(G)$, which represents the status of behavioral goal $G$, has one of three values: achieved, failed, or initial status. $v$ represents the parameters required for system operations such as making reservations or retrieving information, while $s(v)$ represents the status of parameter $v$. For example, parameters include such items as “hotel name” and “date,” for which there are three values of parameter status: determined, undetermined, and initial status. Transition probabilities between all behavioral goals are defined beforehand in the knowledge bank for behavioral goals.

### 3.3 Scoring Algorithms

For the N-best recognition results $H_1, \ldots, H_N$ of $t$-th utterance, the dialogue context score $S_c(H_t)$ is calculated using the following algorithms.

#### step1
The activity $A_{G}^{(t-1)}$ of behavioral goal $G$ is updated with the $t-1$-th utterance. The activity for all behavioral goals is normalized, and this becomes the probability value $p_{t-1}(G)$ for the behavioral goal by (3).

#### step2
Assume the hypothesis $H_t$ to be the $t$-th input, the activity $A_{G}^{(t)}$ is calculated for all behavioral goals and the behavioral goal is inferred. The activity for all behavioral goals is normalized, and this becomes the probability value $p_t(G|H_t)$ by (4).

#### step3
The dialogue context score $S_c(H_t)$ is calculated by (5).

The score $S(H_t)$ for the hypothesis $H_t$ is calculated as the sum of the dialogue context score $S_c(H_t)$ and the acoustic score $S_a(H_t)$ by (6). The $H_t$ with the highest $S(H_t)$ is selected as the recognition result.
Table 2: Speech understanding results

<table>
<thead>
<tr>
<th></th>
<th>CS*</th>
<th>1st</th>
<th>2nd or lower</th>
<th>None</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed dialogue</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Without</td>
<td>48 (83%)</td>
<td>6 (10%)</td>
<td>4 (7%)</td>
<td></td>
</tr>
<tr>
<td>With</td>
<td>50 (86%)</td>
<td>4 (7%)</td>
<td>4 (7%)</td>
<td></td>
</tr>
<tr>
<td><strong>Free dialogue</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Without</td>
<td>254 (81%)</td>
<td>18 (6%)</td>
<td>43 (13%)</td>
<td></td>
</tr>
<tr>
<td>With</td>
<td>257 (82%)</td>
<td>15 (5%)</td>
<td>43 (13%)</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>302 (81%)</td>
<td>24 (8%)</td>
<td>47 (13%)</td>
<td></td>
</tr>
<tr>
<td>With</td>
<td>307 (82%)</td>
<td>19 (5%)</td>
<td>47 (13%)</td>
<td></td>
</tr>
</tbody>
</table>

* context score

a specified dialogue context; and 2) a completely free dialogue experiment in which the dialogue goal was to make a hotel reservation.

For simplicity, concerning the knowledge on the relations between behavioral goals and intentions, we pre-selected the respective intentions that result from the behavioral goals and made their probabilities equal. Also, for the goal transitions, the destinations are classified into three categories: 1) no transition, 2) transition is possible, and 3) transition is highly probable. The probability value for the first category is zero and the probability value for the third category is twice that of the second category. Elements of 2) or 3) make transitions with equal probabilities. In the inference of the goal, the goal that has the highest probability is selected as the unique goal at the time of the utterance.

Context scores were calculated for the 10 best semantic candidates, which were selected from among the 300 best recognition candidates, which in turn where obtained by a speech recognition system that employs a finite-state automaton language model and 1091 words vocabulary. To confirm the effect of the context processing, the acoustic scores were taken to be the same. In the case of equal context scores, the candidate that has highest acoustic score was chosen.

4.1 Evaluation Experiments with Prescribed Dialogue Context

We conducted experiments using fixed dialogue context to evaluate cases in which the behavioral goal is clear at the time of input. We presented the subject with a dialogue script; after gaining a clear understanding of the dialogue context, we had the subject input a free utterance that continued from the dialogue context presented. We thus obtained 58 utterances which could be accepted by the language model. Table 2 (Fixed dialogue) shows a breakdown of the speech understanding results for the 58 utterances.

4.2 Evaluation Experiments using Free Dialogue

We conducted experiments using a dialogue system to carry out evaluations in the case of utterances in a free dialogue context. We had the subjects engage in free dialogue with a dialogue goal of making a hotel reservation. 315 utterances which could be accepted by the language model are obtained. Table 2 (Free dialogue) shows a breakdown of the speech understanding results for the 315 utterances.

5 DISCUSSION

We discuss the causes of errors for nineteen utterances in the “2nd or lower” category. There were nine utterances (47%) in which dialogue context scores were the same as a result of identical intention. In the current framework the system is unable to make a distinction between two candidates which have the same intention and parameters (ex. “date” or “the number of nights”) but the different values of the parameters. One possible approach for dealing with this type of error would be to make the relationship between the utterance and the behavioral goal more detailed, to include the parameter value.

In terms of other causes of errors, nine utterances (47%) were processing errors in the speech understanding module, and one utterance (6%) was inference error of behavioral. The former situations can be corrected through improvements in the accuracy of translation processing knowledge.

In the current preliminary evaluations, the transition probabilities for the behavioral goals were allotted manually. In future, it will be necessary to consider methods of acquiring probability automatically, as well as methods for deciding the weight coefficient \( w \) for the dialogue context score.

6 CONCLUSION

We have proposed a method of improving speech understanding rates by evaluating the appropriateness of N-best understanding results in a dialogue context. Context score is determined from the occurrence probability of the behavioral goal and the inferred behavioral goal transition probability for each candidate. After conducting preliminary evaluation experiments using a spoken dialogue system, we confirmed that this method operated effectively in reference to five utterances from among 15 utterances for which the correct answers resided in rank 2 or lower of the N-best speech understanding results.

References
