AN EFFICIENT DIALOGUE CONTROL METHOD USING DECISION TREE-BASED
ESTIMATION OF OUT-OF-VOCABULARY WORD ATTRIBUTES

Yasuhiro Takahashi, Kohji Dohsaka, and Kiyoshi Aikawa

NTT Communication Science Laboratories, NTT Corporation

ABSTRACT

This paper proposes a dialogue control method for completing a task with a short dialogue even when user utterances include out-of-vocabulary (OOV) words. When the user utterance includes an OOV word, conventional methods try to complete a task by acquiring an understanding of the OOV word through dialogue. However, completing a task does not always require that the system understands the OOV word through dialogue. Our method eliminates the dialogue for acquiring an understanding of the OOV word and tries to complete a task by estimating attributes of OOV words using a decision tree. It does not allow system utterances using those attributes, which stops the user from uttering the same OOV words. A dialogue simulation indicates that our method can complete a task with dialogues that are shorter than those conventional methods need for completing a task.

1. INTRODUCTION

A spoken dialogue system performs tasks through conversation with the user. Since the vocabulary the system can accept is usually limited and the users do not always know that vocabulary, users may include out-of-vocabulary (OOV) words in their utterances. We want to construct a system that can complete a task with as short a dialogue as possible even when user utterances include OOV words. The system we want is one that does not try to acquire an understanding of the OOV word with dialogues when doing so is not useful for completing a task with a short dialogue.

In this paper, we propose a dialogue control method for completing a task with a short dialogue even when user utterances include OOV words. When a user utterance includes an OOV word, our method does not try to acquire an understanding of the OOV word through dialogue. Our method tries to complete a task by estimating the attributes of OOV words using a decision tree and does not allow system utterances using those attributes to prevent the user uttering the same OOV words. Our method avoids dialogues aimed at acquiring an understanding of the OOV word and completes a task with a short dialogue. OOV words in this paper are nouns whose attributes are in the system’s knowledge base.

We assume that the system can segment an OOV word in the user utterance.

Methods of estimating OOV word attributes have been proposed in [1, 2, 3]. In [1], the system collects examples, which are pairs comprising the system’s understanding of the user utterance and the system utterance immediately before the user utterance. It selects the example that is most similar to the pair comprising the system’s understanding of the user utterance including OOV words and the system utterance immediately before the user utterance including OOV words. This method is time consuming because the examples are made from transcriptions of dialogues. In [2], the system applies rules to the system’s understanding of the user utterance including OOV words in order to estimate attributes of OOV words. This method is also troublesome because the rules are handmade. The method in [3] is an example-based one in which the examples are made from system log data automatically, but the attribute estimation accuracy does not increase as the number of examples increases. Efficient dialogue control methods have been proposed in [4, 5]. The method in [4] uses the speech recognition rate and that in [5] is based on reinforcement learning. However, these methods cannot complete a task with a short dialogue when the user utterance includes OOV words because they do not consider that a user utterance may include OOV words.

The estimation of OOV word attributes is a classification problem. Machine learning methods can therefore be applied. In this paper, we apply the decision tree method and compare it experimentally with the example-based one that is based on [3] and whose similarity measure is a value difference metric on memory-based reasoning. To evaluate the efficiency of our dialogue control method, we present a dialogue simulation in which our system is compared with one that does not accept OOV words and one that can segment OOV words in the user utterance and does not target them in understanding and one that can segment OOV words in the user utterance and questions about OOV word attributes.
2. ESTIMATION OF OOV WORD ATTRIBUTES

2.1. Decision Tree-Based Method

In spoken dialogues between a system\(^1\) and user, the system makes an utterance according to the state of its understanding of the user utterance and it changes the state of its understanding of the user utterance according to its understanding of the user utterance that immediately follows its utterance, and it makes an utterance according to the state of its understanding. Then, we introduce the system’s understanding transition, and use it as learning data to make a decision tree. A system’s understanding transition comprises:

**State1:** the state of the system’s understanding

**SysUtt:** the system utterance according to State1

**State2:** the state of the system’s understanding according to the user utterance immediately after SysUtt

An example of such a transition is as follows.

**State1:** (Date: Today)

**SysUtt:** What is the name of the program?

**State2:** (ProgramName: The Sunday News)

This example is from a dialogue between the video recording spoken dialogue system and the user. When the system understands that the user’s request is about today’s program, the system asks the user the name of the program the user wants to record and the system understands that the name of the program is “The Sunday News” by understanding the user utterance immediately after the system utterance.

First, we explain the method for estimating an attribute of an OOV word when State2 includes only one OOV word at the first place of State2. From the learning data, we ascertain what attributes are in State1, the kind of the SysUtt, what attributes are at the first place of State2, and what attributes are in State2 other than those at first place of State2. And, to make a decision tree, we apply the C4.5 algorithm for information collected from learning data.

The method for estimating attributes of OOV words when State2 includes more than two OOV words, we make a decision tree at each place of State2 as above.

2.2. Example-Based Method

Conventional methods that is time consuming for making examples and rules are not feasible for application to real spoken dialogue systems. Thus we compare our method with the example-based one that is based on [3]. The method is not time consuming for making examples and its attribute estimation accuracy is higher than that of [3].

We call an example the system’s understanding transition. To measure the distance between the system’s understanding transition including OOV words and an example, we use a value difference metric on memory based reasoning. This metric is used in the example-based machine translation method [6]. We define the distance \( \Delta \) between the system’s understanding transition \( \tau \) and an example \( \rho \) as follows.

\[
\Delta(\tau, \rho) = \sum_{k=1}^{3} \delta_k^3(D, \tau, f_k, \rho, f_k),
\]

\[
\delta_k(D, \tau, f_k, \rho, f_k) = \sum_{v \in V_o} \frac{|D[f_k = \tau, f_k]|_o = v|}{|D[f_k = \tau, f_k]|} - \frac{|D[f_k = \rho, f_k]|_o = v|}{|D[f_k = \rho, f_k]|}. \]

\( V_o \) represents the set of attributes, and \( f_1, f_2, f_3 \) represents State1, SysUtt, State2 respectively. The \( \tau, f_1, \tau, f_2, \tau, f_3 \) represents \( f_1 \) of \( \tau \), \( f_2 \) of \( \tau \), \( f_3 \) of \( \tau \). The \(|D[f_k = \tau, f_k]|\) represents the number of examples whose \( f_k \) is equal to \( \tau, f_k \) in the example database \( D \), and \(|D[f_k = \tau, f_k]|_o = v|\) represents the number of examples in \( D \) whose \( f_k \) is equal to \( \tau, f_k \) and whose State2 has the attribute \( v \) at the first place.

2.3. Experiments

We collected 600 learning data using the weather information system HUME developed by our research group. There were four attributes: Time, Place, KindOfWarning, KindOfRequest. Since HUME cannot segment an OOV word in the user utterance, we divided the 600 learning data into two groups, one consisting of the 100 learning data and the other of 500 learning data. Using the 100 learning data, we made 100 system’s understanding transitions including OOV words by assuming a part of State2 of each learning data to be OOV words. We obtained Fig. 1 when State2 of the system’s understanding transition included one OOV word only at the first place. The attribute estimation accuracy of the example-based method for 100 examples is 59%. For 400 examples, it is 81%. On the other hand, the attribute estimation accuracy of the decision tree-based method for the 100 learning data is 82%. Thus the attribute estimation accuracy of the decision tree-based method is higher than that of the example-based method based on [3].

3. DIALOGUE USING OOV WORD ATTRIBUTES

3.1. Efficient Dialogue Control Method

Our dialogue control method aims at task completion through a short dialogue between a system and user even when user utterances include OOV words.
Fig. 1. Results on the estimation of OOV word attributes.

When the user utterance includes OOV words, our method:

1. Estimates OOV word attributes
   An attribute of an OOV word attribute is estimated when one OOV word is included only at the first place of State2 and go to the process 2.

2. Decides whether the system should make utterances using the estimated attribute
   It goes to the process 3 when the estimated attribute is not in State1.

3. Restricts system utterances using the estimated attribute
   The system is not permitted to make utterances using the estimated attribute, and it controls dialogue using information about attributes unrelated to the estimated attribute.

We illustrate our method with dialogue between the video recording system and the user. There are three attributes: ProgramName, Date, GenreName. We assume the dialogue between the system and the user is as follows.

SYSTEM: May I help you?
USER: I want to record today’s drama.
SYSTEM: What is the name of the program?
USER: Love Vacation.

The word “Love Vacation” is an OOV word since the true title is “Long Love Vacation”. We assume the word “Long Love Vacation” is in the system’s vocabulary. In following process 1, to estimate the attribute of the word “Love Vacation”, the following system’s understanding transition occurs.

State1: (Date:Today, GenreName:Drama)
SysUtt: What is the name of the program?
State2: (OOV word)

We assume that the OOV word attribute our method estimates is ProgramName. In process 2, the system checks whether the estimated attribute ProgramName is in State1. Since ProgramName is not in State1, the method goes to process 3. In process 3, the system does not make utterances using ProgramName. For example, it cannot ask “What is the name of the program?” The system retrieves programs from the system’s program table on the condition that the user’s request is about today’s dramas. Today, there are two. “Long Love Vacation” and “Love Letters”. The system shows the user these two program names and let the user select the one she wants to record.

3.2. Dialogue Simulation

The dialogue control method was evaluated by dialogue simulation. The dialogue was between the system and a simulated user. We did not use a real user because we want to evaluate various types of dialogues in a short time.

The task was to record the program or to retrieve the program. There were six attributes: Date, Time, GenreName, ProgramName, PersonName, and KindOfRequest. The number of values for each attribute was seven, twelve, ten, fifty, fifty, and four, respectively. We made a program table consisting of 84 programs using these values. The simulated user selected one program from the program table and recorded the program or retrieved the program that had the same condition of the selected program.

The simulated user made utterances based on the randomly selected request that the system could perform. When the simulated user made an utterance, it expressed utterances on the OOV rate. For example, when the simulated user expressed today’s drama and the OOV rate was 0.6, he expressed the word today as the OOV word at the probability of 0.6 and the word drama as well.

We assumed that dialogues between the system and the user could be divided into two phases. One was the confirmation phase, in which the system confirmed the simulated user’s request. The other was the information phase, in which the system conveyed the information the simulated user requested.

We defined dialogue efficiency as the length of the dialogue. The length of the dialogue was the sum of the length of the confirmation phase and that of the information phase. The length of the confirmation phase was the number of attributes included in the system utterance and the simulated user utterance at the confirmation phase. The length of the information phase was the number of programs included in the system utterance at the information phase.
Fig. 2. Results of dialogue simulation.

We compared our system with following systems.

**System A:** Misunderstands an OOV word as one of the words in the system’s vocabulary.

**System B:** Can segment an OOV word in the user utterance and does not target OOV words in understanding.

**System C:** Can segment an OOV word in the user utterance and makes an utterance which attribute is the attribute of the OOV word and processes the OOV word as our system.

**System D:** Can segment an OOV word in the user utterance and estimates the attribute of the OOV word by selecting attributes randomly and processes the OOV word as our system.

We obtained Fig.2. Each point was the average length of 1000 dialogues. To make a decision tree for estimating attributes of OOV words, we collected 700 learning data from dialogues between the system and the simulated user. The attributes estimation accuracy was 71%. When the user made an utterance, attribute values were replaced with the word selected from the system’s vocabulary randomly at the probability of 0.2 to simulate speech recognition errors.

When the OOV rate was high, the simulated user repeated its utterance including OOV words. In this situation, our system was the most efficient. The reason for the inefficiency of the other systems is similar to that when the OOV rate is high. Since the length of the dialogue by paraphrasing OOV words to words in the system’s vocabulary is shorter than that when the OOV rate is high, the difference in the efficiency between our system and the others is smaller than when the OOV rate is high.

4. CONCLUSION

We proposed a decision tree-based method for estimating attributes of OOV words and an efficient dialogue control method that uses the estimated attributes of OOV words. The OOV word estimation accuracy of the decision tree-based method is higher than that of the conventional method. Our dialogue control method is more efficient than typical dialogue control methods for processing OOV words.

Acknowledgments

We thank Dr. Ken’ichiro Ishii, Dr. Hiroshi Murase, and the members of the Dialogue Understanding Group for their helpful comments.

5. REFERENCES


