A NEW METHOD FOR DIALOGUE MANAGEMENT IN AN INTELLIGENT SYSTEM FOR INFORMATION RETRIEVAL

Kenji Abe¹, Kazushige Kurokawa¹, Kazumari Taketa¹, Sumio Ohno² and Hiroya Fujisaki¹

¹ Science University of Tokyo, 2641 Yamazaki, Noda, 278-8510, Japan
² Tokyo University of Technology, 1404-1 Katakura, Hachioji, 192-0982, Japan

ABSTRACT

This paper proposes a new method for dialogue management to be used in an intelligent system for information retrieval. The user and the system are modeled as two separate finite-state automata exchanging information through dialogue, and the system manages the dialogue efficiently by estimating the internal states of the user model. These models are constructed on the basis of analysis of a corpus of simulated dialogues, and the proposed method is evaluated from the point of view of perplexity reduction of language models.

1. INTRODUCTION

With the rapid and widespread use of information networks in our society, it has become increasingly important to search and retrieve the information that is truly relevant. Existing search engines are designed to be easy to use even for inexperienced users, but they are too simple and too inefficient for many purposes. On the other hand, conventional systems for information retrieval using more sophisticated procedures are not always easy to use for inexperienced users since they presuppose certain amount of knowledge on the structure of databases and on the method for constructing the search formula. In many cases, it is difficult for the user to identify and express his/her intention precisely, and it is difficult also for the system to infer the user’s intention correctly. These difficulties can be greatly reduced by introducing spoken dialogue between the user and the system.

Based on these considerations, we have proposed an intelligent system for information retrieval using concept-based search and human-machine dialogue through spoken language as the main medium [1-3]. The present paper first gives an overview of the user interface of the system, which adopts a new method for dialogue management based on user and system modeling. In this method, the user and the system are modeled as two separate finite-state automata exchanging information through dialogue, and the system manages the dialogue efficiently by estimating the internal states of the user model. This paper then describes the collection and annotation of simulated dialogues, as well as the acquisition of models on the basis of analysis of the annotated corpus. Finally, the proposed method is evaluated from the point of view of perplexity reduction of language models.

2. DIALOGUE MANAGEMENT BASED ON USER AND SYSTEM MODELING

Figure 1 shows a schematic diagram of the user interface in the information retrieval system. One of its essential features is dialogue management through user and system modeling. In conventional dialogue systems, dialogue management is performed by modeling the dialogue itself. This is done by analyzing actual dialogues and constructing a state-transition diagram for representing possible exchanges between a user and the system. Such an approach is not ideal since it does not describe the user and the system separately, and therefore leads to complexity and inflexibility. In the present system, we construct models of the user and the system as separate finite-state automata which exchange information through dialogue [4]. In particular, the states in the user model are meant to represent the internal states of the user’s intention.

It should be noted, however, that the user model is not the user’s mind itself, but is only an approximate representation to help the system to make inference on the true internal state of the user’s mind and to predict the user’s next utterance. This approach has the following advantages over the conventional approach:

1. Clearer and simpler description of the dialogue
2. Possibility of separately modifying models of the user and the system
3. Improving the performance of speech recognition/understanding by reducing the perplexity.

The first point is obvious. The second point is important since each user, exactly speaking, is different in his/her background, knowledge, interest, query, etc., and adapting the dialogue to suit each user will require adjustment or modification of the user model. It is also easier to modify the system for a new or an improved service if we have a separate model for the system.

The third point is important for recognition/understanding of user’s utterances. Current technology for speech recognition/understanding utilizes a fixed language model which is derived from the average statistics of a language, and thus is not optimum since it does not take into account the fact that the probability of occurrence of linguistic units such as letters vary widely from context to context. This fact can be taken into account only by having a user model, in which each state constrains the range of utterances to be generated.
3. BUILDING AND USING MODELS

3.1. Collection of Simulated Dialogues

For the purpose of investigating the states of the user and the system and their transitions in dialogues for information retrieval, 100 simulated dialogues between a user and the system, consisting of a total of 3417 utterances, were collected by the method of ‘Wizard of Oz.’ These simulated dialogues were carried out between a speaker who plays the role of the user and another speaker who plays the role of the system. The speaker playing the role of the user is aware of the purpose on information retrieval but his utterances are unconstrained. The speaker playing the role of the system is aware of the functions and limitations of the system, and his utterances are selected from a pre-determined manual. The number of speakers who played the user’s role was 10, and the number of speakers who played the role of the system was 7.

3.2. Acquisition of Models

All the utterances that were considered to be meaningful for the purpose of information retrieval were transcribed and classified. Furthermore, certain actions that were not accompanied by any utterance but were considered to be significant for information retrieval were also written down along with the utterances. As a result, the user’s utterances/actions were found to fall into 15 types, and the system’s utterances into 14 types.

By inferring the processing stage of information retrieval at which each of these utterance/action types would occur, 8 states were identified as the user’s states, and 9 states were identified as the system states. A tagged corpus was then constructed by tagging every meaningful utterance/action with its type and the associated state of the user or the system.

Based on the tagged corpus, the probability that a transition from a state \( S_i \) to another state \( S_j \) is elicited by an input \( I_k \) while giving an output (either in the form of an utterance or in the form of an action) \( O_l \) was calculated for each state of the user and the system. The resulting sets of stochastic rules for the user’s states and the system’s states constitute the user model and the system model, respectively.

3.3. Constructed User Model and System Model

Figure 2 shows the user model and the system model constructed. In this figure, \( S_i \) \((i = 1 \sim 8)\) refers to a state of the user model and \( u_j \) \((j = 1 \sim 12)\) indicates an output from a state of the user model in the form of an utterance produced at the time of state transition. Likewise, \( S_i \) \((i = 1 \sim 9)\) refers to a state of the system model and \( s_j \) \((j = 1 \sim 14)\) indicates an output from a state of the system model. The two models are interacting; the output from the user model is given to a specific state of the system model and elicits a state transition. Likewise, an output from the state model is given to the user model and elicits a state transition. Hence the state transition rules are described in terms of the following four elements: (1) the current state, (2) input to the current state, (3) output from the current state, and (4) the next state, and the dialogue is carried out by exchanging information between the two models.

Table 1 shows the set of state transition rules and their occurrence probabilities in the user model. In this table, “Null” means that there is no input nor output in the transition rule.

The user model thus constructed does not represent any particular user but represents the average of a number of users. However, this average model can be adapted automatically to a specific user by using the record of his/her interactions with the system.

3.4. Identification and Prediction of User States and Utterances

The user model serves not only to identify the current utterance type and the current state of the user, but also to predict the next state and the next utterance type. Here the ‘current’ state means the state of the user model from which the state transition starts, and the ‘next’ state means the state to which the transition ends. Namely, upon receiving a user’s utterance, the system can identify its utterance type on the basis of its linguistic content and the prediction by the model. The result, in turn, is used to identify the current state, as well as to make statistical prediction on the user’s next utterance and the next state.
Figure 2: The user model and the system model constructed on the basis of analysis of the dialogue corpus.
### 4. PERPLEXITY REDUCTION OF LANGUAGE MODELS

In order to evaluate the quantitative advantage that can be gained by having a separate language model for each user state, the following perplexities are calculated and compared for letter-based bigrams.

1. Perplexity for utterances in the entire dialogue corpus.
2. Perplexities for utterances originating from each of the seven user states and their arithmetic mean.

As shown in Table 2, perplexities for all the user states as well as their mean are always smaller than the perplexity of the corpus as a whole. Although it is not shown here, the same tendency exists for phone-based bigrams. Hence the use of a separate language model for each user state is expected to increase the rate of correct recognition of user utterances.

### 5. SUMMARY AND CONCLUSION

This paper has proposed a new method for dialogue management based on separate modeling of the user and the system, and described its implementation in a human-machine dialogue system for academic information retrieval. The use of models for the user and the system as two separate finite-state automata, exchanging information through spoken dialogue, leads to clearer and simpler description of the dialogue, flexibility of adapting the user model to individual users, ease of predicting user’s states and utterances, and reduction of perplexities in recognition of user’s utterances. The method is being implemented in an intelligent system for academic information retrieval based on human-machine dialogue, which will be reported elsewhere [5].

### REFERENCES


