SPEECH-BASED INTERACTIVE INFORMATION GUIDANCE SYSTEM USING QUESTION-ANSWERING TECHNIQUE

Teruhisa Misu Tatsuya Kawahara

School of Informatics, Kyoto University, Kyoto 606-8501, Japan

ABSTRACT

This paper addresses an interactive framework for information navigation based on document knowledge base. In conventional audio guidance systems, such as those deployed in museums, the information flow is one-way and the content is fixed. In order to make an interactive guidance system, we propose the application of question-answering (QA) techniques. Since users tend to use anaphoric expressions in successive questions, we investigate appropriate handling of contextual information based on topic detection, together with the effect of using N-best information in ASR output. Moreover, we apply the QA technique to generation of system-initiative information recommendation. A navigation system on Kyoto city information was implemented. Effectiveness of the proposed techniques was confirmed through a field trial by a number of real novice users.

Index Terms— spoken dialogue system, question-answering, information guidance

1. INTRODUCTION

The target of spoken dialogue systems is being extended from simple databases such as flight information to general documents including manuals[1] and newspaper articles[2]. In such systems, the automatic speech recognition (ASR) result of the user utterance is matched against a set of target documents using the vector space model, and documents with high matching scores are presented to the user. We have developed "Speech Dialogue Navigator", which can retrieve information from a large-scale software support knowledge base (KB) with a spoken dialogue interface[3].

Most of these types of dialogue systems assume that a display is available as an output device, and thus a list of matched documents can be presented. However, this is not the case when only speech interface is available, for example, using phones and audio guidance systems. Considering user's easiness of comprehension, the amount of the content presented at a time should be limited. But simply summarizing the retrieved document may cause a loss of the important portion the user intended to know or may be interested in. Actually, in the conventional audio guidance systems deployed in museums and sightseeing spots, users cannot ask questions on the missed portion. We therefore propose a more interactive scheme by incorporating the question-answering (QA) technique to follow up the initial query enabling random access to any part of the document.

There are some problems with QA in such situations. One important issue is contextual analysis. In dialogue session, users tend to make questions that include anaphoric expressions. In these cases, it is impossible to extract the correct answer using the current question only. (For example, "When was it built?" makes no sense with this sentence alone.) In many conventional database query tasks, this problem is solved by using the task domain knowledge such as the semantic slots of the backend database [4, 5]. Whereas the majority of the conventional QA tasks, such as TREC QA Track[6], have dealt with independent questions that have respective answer for each, there have been only a few works that addressed successive questions[7]. But they have basically hand-crafted questions rather than collecting real dialogues. In this work, we address the QA task in a real interactive guidance system using a topic tracking mechanism.

Furthermore, we introduce generation of system-initiative information recommendation. In spoken dialogue systems, users often have a difficulty in making queries because of unsureness of the list of information the system possesses. Moreover, the system-initiative guidance is also useful in navigating users in the tasks without definite goal, such as sightseeing guidance. In order to make an interactive guidance, we propose the application of the QA technique to generate system-initiative recommendations.

Based on the above concepts, we have designed and implemented an interactive guidance system of "Dialogue Navigator for Kyoto City", and conducted a field trial for about three months. Key evaluation results of the QA function are presented in this paper.

2. FRAMEWORK OF THE SYSTEM

The proposed guidance system prepares two modes of a user-initiative retrieval/QA mode (pull-mode) and a system-initiative recommendation mode (push-mode), and switches them according to the user's state. When a query or a question is uttered by a user, the system switches to the retrieval/QA mode and generate a respective response. When the system detects the silence of the user, it switches to the system-initiative recommendation mode and presents information re-

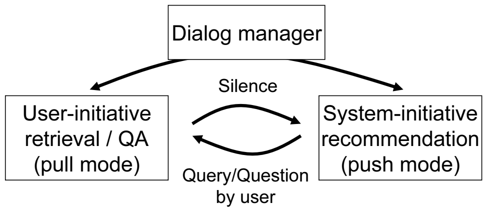

Fig. 1. System overview

Table 1. Specification of knowledge base (KB)

	# documents	# sections	# words
Wikipedia	269	678	150K
Tourist information	541	541	70K
Total	810	1,219	220K

lated to the current topic. The flow of this process is shown in Figure 1.

As the target domain, we adopt a sightseeing guidance of Kyoto city. The KBs of this domain are Wikipedia[1] documents concerning Kyoto and the official tourist information of Kyoto city. Table 1 lists the size of these KBs.

3. USER-INITIATIVE INFORMATION RETRIEVAL AND QUESTION-ANSWERING

The user utterances are classified into two categories. One is an information query, such as "Please explain Golden Pavilion". For such queries, the system retrieves from the KB by section unit, and the document section with the largest matching score is presented to the user. The other is a question, such as "When was it built?". The system extracts the sentence from the KB that includes the answer to the question and presents it to the user. This procedure is shown in Figure 2.

3.1. Contextaul Analysis based on Topic Detection

In dialogue systems, incorporation of contextual information is an important issue to generate a meaningful query for retrieval. As the deterministic anaphora resolution[8] is not easy and always error-prone, and stochastic matching is used in information retrieval, we adopt a strategy to concatenate contextual information or keywords in the user's previous utterances to generate a query. The simplest way is to use all utterances of the current user. However, it might add inappropriate context because the topic might have been changed in the session. We therefore determine the length of context (number of previous utterances) used for retrieval by tracking the topic of the dialogue.

Whereas De Boni[9] proposed semantic similarity techniques to detect contextual questions with typed-input, it would be difficult to adopt such an approach for dialogue systems with speech-input, in which queries tend to be short and reference words are often omitted[2]. As a topic, we therefore

[1] http://wikipedia.org/
[2] especially in Japanese

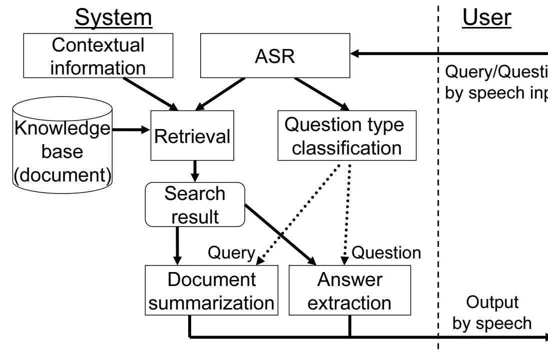

Fig. 2. Overview of document retrieval and QA

use metadata of the KB or a title of the document. Thus, the topic can be tracked by keeping the current focused documents, which usually correspond to sightseeing spots or Wikipedia entries.

3.2. Document Retrieval

We adopt an orthodox vector space model to calculate a matching score (degree of similarity) between user query and the document in the KB. That is, the vector of the document is made based on the occurrence counts of nouns in the document by section unit. The vector for the user query is also made by merging N-best hypotheses of the ASR result of the current utterance and previous utterances about the current topic as a context. We also use the ASR confidence measure (CM) as a weight for the nouns. Matching score is calculated by the product of these two vectors.

For the retrieved document, a summary is generated by extracting important sentences for concise presentation.

3.3. Answer Extraction

We have implemented a general answer extraction module. For each named entity (NE) in the retrieved document that matches the question type (who, when, . . .), a score is calculated using following features.

- Degree of similarity between the user utterance and the document (3.2)
- Number of matched content words in the sentence including the NE
- Number of matched content words included in the clause that depends on / depended by the clause that includes the NE

The system then selects the NE with the highest score as an answer to the question.

4. SYSTEM-INITIATIVE RECOMMENDATION

For interactive information recommendation, we propose to generate system-initiative questions. They are semi-automatically made from the current document using the

QA technique. This is complemented by conventional information recommendation techniques based on the document structure and document similarity.

4.1. Generation of System-Initiative Questions (Method 1)

This method is intended to successively present more details of the target topic, after the initial summary presentation. The user may be interested in the part that was not included in the summary. Although it is possible to prompt such as "Would you like more details?", we propose a more interactive method by generating system-initiative questions in order to attract interest of the user.

A set of possible questions is prepared using the following procedure. It is almost reverse to the process to find an answer to the user's question.

1. Pick up the NE which may attract user's interest based on $tf * idf$ criterion.
2. Substitute the NE with the corresponding interrogative.
3. Delete the subordinate clause using a syntactic parser.
4. Transform the sentence into interrogative form

Figure 3 shows an example of transforming a sentence in the KB into a question using the above mentioned procedure.

Original: By the way, **Queen Elizabeth** praised this stone garden very much, when ...

$\Downarrow$ (Substitute target NE into the corresponding interrogative)

- By the way, **who** praised this stone garden very much, when ...

$\Downarrow$ (Delete subordinate clause)

- Who praised this stone garden very much?

$\Downarrow$ (Transform into interrogative)

Question: Do you know who praised this stone garden very much?

Fig. 3. Example of system-initiative quetion generation

4.2. Recommendation based on Document Structure and Similarity

We have also implemented two conventional recommendation techniques based on the document structure and document similarity.

- **Recommendation based on document structure (Method 2)**
  Wikipedia documents are described hierarchically using section structure. Thus, another section of the current document can be picked up for presentation.

U1:	Please explain Golden Pavilion.
S1:	Golden Pavilion is one of the buildings in the Rokuon-ji in Kyoto, and is the main attraction of the temple sites. The entire pavilion except the basement floor is covered with pure gold leaf.
U2:	When was it built?
S2:	Golden Pavilion was originally built in 1397 to serve as a retirement villa for Shogun Ashikaga Yoshimitsu.
(Silence)	
S3:	Well then, do you know what was awarded to this temple in 1994?
U3:	No, please tell me.
S4:	It was awarded as listing on the UNESCO World Heritage in 1994.
U4:	How can I get there?
......	

Fig. 4. Example dialogue

- **Recommendation based on document similarity (Method 3)**
  We can select a document that has a large similarity with the current document. This technique is adopted in information recommendation of Web pages.

5. SYSTEM EVALUATION

We implemented a guidance system "Dialog Navigator for Kyoto City". An example dialogue of the system using the QA technique is shown in Fig. 4. We carried out a field trial at our university museum. Users are in a wide variety of ages from children to senior people and apparently have few experiences in using spoken dialogue systems. No instructions on the system were given. In total 2,500 dialogue sessions (20,000 utterances) were collected. In this paper, we evaluated using 427 dialogue sessions chosen from a particular time period. For the ASR system, a trigram language model was trained using the KB, a dialogue corpus of different domain, and Web texts[10]. The average word accuracy was 70.6%.

5.1. Evaluation in Question-Answering Performance

First, we evaluated the performance of QA in terms of success rate[3] using 366 questions. We regarded QA as successful when the system made an appropriate response to the question. That is, if an answer to the question exists in the KB, we regarded QA as successful when the system presented the answer. On the other hands, if there is no answer in the KB, we regarded QA as successful when the system told it. The QA success rate was 60.7% (62.9% in the case correct answers exist in the KB, and 47.2% in the case when they do not).

We also evaluated the effect of ASR confidence measure (CM) for QA performance. The system used the CM as a

[3]Though the QA performance is usually evaluated using mean reciprocal rank (MRR), we adopt the simple success rate, because it is not possible to present alternative candidates via speech.

Table 2. Effect of using ASR confidence measure

Use of CM	Success rate(%)
Yes	60.7 (62.9, 47.2)
No	55.7 (54.0, 66.0)

Table 3. Effect of using N-best hypotheses

Use of N-best hypotheses	Success rate(%)
Merge 3-best hypotheses (proposed)	60.7 (62.9, 47.2)
1-best only (baseline)	57.9 (61.0, 39.6)
Optimal hypothesis (reference)	63.1 (65.8, 47.2)

Table 4. Contextual effect for QA

Use of context	Success rate(%)
Current topic (proposed)	60.7 (62.9, 47.2)
No context	36.9 (30.4, 75.5)
Previous one utterance	54.6 (54.3, 56.6)
All utterances	55.5 (56.5, 49.1)

weight in the matching between user query and the document in the KB. We compared with the case where the CM was not used. Table 2 lists these results, and confirms the effect of the CM.

Next, we evaluated the effect of using N-best hypotheses of the ASR result. In our system 3-best hypotheses of the ASR result were used for making a query and extracting an answer. We compared with the case where only the first hypothesis was used (baseline). We also investigated the case where an optimal hypothesis was selected manually (reference). Table 3 lists these results. The effect of using 3-best hypotheses is clearly confirmed, compared with the case to using only the first hypothesis. However, it was shown that higher success rate could be obtained if an optimal hypothesis was selected. This success rate could be achieved by introducing the confirmation strategy[3].

We then evaluated the effect of the context length (= number of previous utterances) used for the retrieval. This result is shown in Table 4. Without context, the success rate is significantly degraded. But using all previous utterances has adverse effect. It was shown that incorporation of appropriate context information by topic tracking effectively improved the performance.

5.2. Evaluation of System-Initiative Recommendation

In order to confirm the effect of the proposed system-initiative question, the system was set to make possible recommendations randomly. The number of recommendations presented by the system during the 427 dialogue sessions was 319 in total. We regarded a recommendation as accepted when the user positively responded[4] to the proposal given by the system. The acceptance rate of each presentation technique is shown in Table 5. The acceptance rate by the system-initiative question (method 1) is much higher than that of other methods.

[4] by human judgment

Table 5. Comparison of recommendation method

Recommendation method	Acceptance rate(%)
Question (Proposed method 1)	74.7
Document structure (Method 2)	51.1
Document similarity (Method 3)	30.8

The result suggests that recommendations using the question form are more interactive and attractive.

6. CONCLUSIONS

We have proposed an interactive scheme for information guidance using question-answering techniques. In order to make interactive guidance, we incorporated question-answering techniques into both user-initiative information retrieval and system-initiative information presentation. We have implemented a sightseeing guidance system and evaluated with respect to QA-related techniques. It was shown that the QA-based technique worked well in improving the system performance.

7. REFERENCES

[1] K. Komatani, T. Kawahara, R. Ito, and H. G. Okuno, "Efficient dialogue strategy to find users' intended items from information query results," in *Proc. COLING*, 2002, pp. 481–487.

[2] E. Chang, F. Seide, H. M. Meng, Z. Chen, Y. Shi, and Y. C. Li, "A system for spoken query information retrieval on mobile devices," *IEEE Trans. on Speech and Audio Processing*, vol. 10, no. 8, pp. 531–541, 2002.

[3] T. Misu and T. Kawahara, "Dialogue strategy to clarify user's queries for document retrieval system with speech interface," *Speech Communication*, vol. 48, no. 9, pp. 1137–1150, 2006.

[4] D. Bohus and A. I. Rudnicky, "RavenClaw: Dialog management using hierarchical task decomposition and an expectation agenda," in *Proc. Eurospeech*, 2003.

[5] K. Komatani, N. Kanda, T. Ogata, and H. G. Okuno, "Contextual constraints based on dialogue models in database search task for spoken dialogue systems," in *Proc. Interspeech*, 2005.

[6] NIST and DARPA, "The twelfth Text REtrieval Conference (TREC 2003)," in *NIST Special Publication SP 500–255*, 2003.

[7] T. Kato, J. Fukumoto, and and N. Kando F. Masui, "Are open-domain question answering technologies useful for information access dialogues? – an empirical study and a proposal of a novel challenge," *ACM Trans. of Asian Language Information Processing*, vol. 4, no. 3, pp. 243 – 262, 2005.

[8] M. Matsuda and J. Fukumoto, "Answering question of iad task using reference resolution of follow-up questions," *Proc. the Fifth NTCIR Workshop Meeting on Evaluation of Information Access Technologies*, pp. 414–421, 2006.

[9] M. De Boni and S. Manandhar, "Implementing clarification dialogues in open domain question answering," *Natural Language Engineering*, vol. 11, no. 4, pp. 343–361, 2005.

[10] T. Misu and T. Kawahara, "A bootstrapping approach for developing language model of new spoken dialogue systems by selecting Web texts," in *Proc. Interspeech*, 2006, pp. 9–12.