Bayes Risk-based Dialogue Management for Document Retrieval System with Speech Interface

Teruhisa Misu [† ‡]

[†]School of Informatics,
Kyoto University
Sakyo-ku, Kyoto, Japan

Tatsuya Kawahara[†]

[‡]National Institute of Information
and Communications Technology
Hikari-dai, Seika-cho, Soraku-gun,
Kyoto, Japan

Abstract

We propose an efficient dialogue management for an information navigation system based on a document knowledge base with a spoken dialogue interface. In order to perform robustly for fragmental speech input and erroneous output of an automatic speech recognition (ASR), the system should selectively use N-best hypotheses of ASR and contextual information. The system also has several choices in generating responses or confirmations. In this work, we formulate the optimization of the choices based on a unified criterion: Bayes risk, which is defined based on reward for correct information presentation and penalty for redundant turns. We have evaluated this strategy with a spoken dialogue system which also has question-answering capability. Effectiveness of the proposed framework was confirmed in the success rate of retrieval and the average number of turns.

1 Introduction

In the past years, a great number of spoken dialogue systems have been developed. Their typical task domains include airline information (ATIS & DARPA Communicator) and bus location tasks. Although the above systems can handle simple database retrieval or transactions with constrained dialogue flows, they are expected to handle more complex tasks. Meanwhile, more and more electronic text resources are recently being accumulated. Since most documents are indexed (e.g., via Web search engines), we are potentially capable of accessing these documents. Reflecting such a situation, in recent years, the target of spoken dialogue systems has been extended to retrieval of general documents (Chang et al., 2002).

There are quite a few choices for handling user utterances and generating responses in the spoken dialogue systems that require parameter tuning. Since a subtle change in these choices may affect the behavior the entire system, they are usually tuned by hand by an expert. It is also the case in speech-baed document retrieval systems. We can make use of N-best hypotheses to realize robust retrieval against errors in automatic speech recognition (ASR). Input queries are often vague or fragmented in speech interfaces, thus concatenation of contextual information is important to make meaningful retrieval. Such decisions tend to be optimized module by module, but they should be done in an integrated way. For example, we could make more appropriate retrieval by rescoring the N-best ASR hypotheses by the information retrieval scores. Even if the target document is identified, the system has several choices for generating responses. Confirmation is needed to eliminate any misunderstandings caused by ASR errors, but users easily become irritated with so many redundant confirmations. Although there are several works dealing with dialogue management in call routing systems (Levin and Pieraccini, 2006), they cannot handle the complex decision making processes in information guidance tasks.

Therefore, we address the extension of conventional optimization methods of dialogue management to be applicable to general document retrieval

Coling 2008: Companion volume – Posters and Demonstrations, pages 59–62
Manchester, August 2008

tasks. In particular, we propose a dialogue management that optimizes the choices in response generation by minimizing Bayes risk. The Bayes risk is defined based on reward for correct information presentation and penalty for redundant turns as well as the score of document retrieval and answer extraction.

2 Task and Knowledge Base (KB)

As the target domain, we adopt a sightseeing guide for Kyoto city. The KBs of this system are Wikipedia documents concerning Kyoto and the official tourist information of Kyoto city (810 documents, 220K words in total).

"Dialogue Navigator for Kyoto City" is a document retrieval system with a spoken dialogue interface. The system can retrieve information from the above-mentioned document set. This system is also capable of handling user's specific question, such as "Who built this shrine?" using the QA technique.

3 Dialogue Management and Response Generation in Document Retrieval System

3.1 Choices in Generating Responses

We analyzed the dialogue sessions collected in the field trial of the "Dialogue Navigator for Kyoto City", and found that we could achieve a higher success rate by dealing with following issues.

1. **Use of N-best hypotheses of ASR**
   There have been many studies that have used the N-best hypotheses (or word graph) of ASR for making robust interpretations of user utterances in relational database query tasks (Raymond et al., 2003). We also improved retrieval by using all of the nouns in the 3-best hypotheses(Misu and Kawahara, 2007). However, the analysis also showed that some retrieval failures were caused by some extraneous nouns included in erroneous hypotheses, and a higher success rate could be achieved by selecting an optimal hypothesis.

2. **Incorporation of contextual information**
   In interactive query systems, users tend to make queries that include anaphoric expressions. In these cases, it is impossible to extract the correct answer using only the current query. For example, "When was it built?" makes no sense when used by itself. We deal with this problem by concatenating the contextual information or keywords from the user's previous utterances to generate a query. However, this may include inappropriate context when the user changes the topic.

3. **Choices in generating responses or confirmations**
   An indispensable part of the process to avoid presenting inappropriate documents is confirmation, especially when the score of retrieval is low. This decision is also affected by points 1 and 2 mentioned above. The presentation of the entire document may also be "safer" than presenting the specific answer to the user's question, when the score of answer extraction is low.

3.2 Generation of Response Candidates

The manners of response for a document d consist of the following three actions. One is the presentation ($Pres(d)$) of the document d, which is made by summarizing it. Second is making a confirmation ($Conf(d)$) for presenting the document d. The last is answering ($Ans(d)$) the user's specific question, which is generated by extracting one specific sentence from the document d.

For these response candidates, we define the Bayes risk based on the reward for success, the penalty for a failure, and the probability of success. Then, we select the candidate with the minimal Bayes risk. The system flow of these processes is summarized below.

1. Make search queries $W_i (i = 1, \ldots, 8)$ using the 1st, 2nd, and 3rd hypothesis of ASR, and all of them, with/without contextual information.
2. For each query W_i, retrieve from the KB and obtain a candidate document d_i and its likelihood $p(d_i)$.
3. For each document d_i, generate presentation $Pres(d_i)$, confirmation $Conf(d_i)$, and answering $Ans(d_i)$ response candidates.
4. Calculate the Bayes risk for 25 response candidates, which are the combination of 4 (N-best hypotheses) $\times$ 2 (use of contextual information) $\times$ 3 (choice in response generation) + 1 (rejection).
5. Select the optimal response candidate that has the minimal Bayes risk.

3.3 Definition of Bayes Risk for Candidate Response

For these response candidates, we define the Bayes risk based on the reward for success, the penalty for a failure, and the probability of success (approximated by the confidence measure). That is, a reward is given according to the manner of response (Rwd_{Ret} or Rwd_{QA}) when the system presents an appropriate response. On the other hand, a penalty is given based on extraneous time, which is approximated by the number of sentences before obtaining the appropriate information when the system presents an incorrect response. For example, the penalty for a confirmation is 2 {system's confirmation + user's approval}, and that of a rejection is 1 {system's rejection}. When the system presents incorrect information, the penalty for a failure $FailureRisk$ (FR) is calculated, which consists of the improper presentation, the user's correction, and the system's request for a rephrasal. Additional sentences for the completion of a task ($AddSent$) are also given as extraneous time before accessing the appropriate document when the user rephrases the query/question. The value of $AddSent$ is calculated as an expected number of risks assuming the probability of success by rephrasal was p[1].

The Bayes risk for the response candidates is formulated as follows using the likelihood of retrieval $p(d)$, likelihood of answer extraction $p_{QA}(d)$, and the reward pair (Rwd_{Ret} and $Rwd_{QA}; Rwd_{Ret} < Rwd_{QA}$) for successful presentations as well as the FR for inappropriate presentations.

- **Presentation of document** d (without confirmation)

$$Risk(Pres(d)) = -Rwd_{Ret} * p(d) + (FR + AddSent) * (1 - p(d))$$

- **Confirmation for presenting document** d

$$Risk(Conf(d)) = (-Rwd_{Ret} + 2) * p(d) + (2 + AddSent) * (1 - p(d))$$

- **Answering user's question using document** d

$$Risk(Ans(d)) = -Rwd_{QA} * p_{QA}(d) * p(d) + (FR + AddSent) * (1 - p_{QA}(d) * p(d))$$

- **Rejection**

$$Risk(Rej) = 1 + AddSent$$

[1]In the experiment, we use the success rate of the field trial presented in (Misu and Kawahara, 2007).

User utterance: When did the shogun order to build the temple?
(Previous query:) Tell me about the Silver Pavilion.

Response candidates:

* With context:
→ $p(\text{Silver Pavilion history}) = 0.4$
→ $p_{QA}(\text{Silver Pavilion history}) = 0.2$: In 1485
- $Risk(Pres(\text{Silver Pavilion history})) = 6.4$
- $\mathbf{Risk(Conf(Silver\ Pavilion\ history))} = \mathbf{4.8}$
- $Risk(Ans(\text{Silver Pavilion history}; \text{In1485})) = 9.7$
...
* Rejection
- $Risk(Rej) = 9.0$

↓

Response: Conf (Silver Pavilion history)
"Do you want to know the history of the Silver Pavilion?"

Figure 1: Example of calculating Bayes risk

Figure 1 shows an example of calculating a Bayes risk (where $FR = 6$, $Rwd_{Ret} = 5$, $Rwd_{QA} = 40$). In this example, an appropriate document is retrieved by incorporating the previous user query. However, since the answer to the user's question does not exist in the knowledge base, the score of answer extraction is low. Therefore, the system chooses a confirmation before presenting the entire document.

4 Experimental Evaluation by Cross Validation

We have evaluated the proposed method using the user utterances collected in the "Dialogue Navigator for Kyoto City" field trial. We transcribed in-domain 1,416 utterances (1,084 queries and 332 questions) and labeled their correct documents/NEs by hand.

The evaluation measures we used were the success rate and the average number of sentences for information access. We regard a retrieval as successful if the system presents (or confirms) the appropriate document/NE for the query. The number of sentences for information access is used as an approximation of extraneous time before accessing the document/NE. That is, it is 1 {user utterance} if the system presents the requested document without a confirmation. If the system makes a confirmation before presentation, it is 3

{user utterance + system's confirmation + user's approval}, and that for presenting an incorrect document is 15 {user utterance + improper presentation (3 = # presented sentences) + user's correction + system's apology + request for rephrasing + additional sentences for task completion} (FR = 6 & $AddSent$ = 8), which are determined based on the typical recovery pattern observed in the field trial.

We determined the value of the parameters by a 2-fold cross validation by splitting the test set into two (set-1 & set-2), that is, set-1 was used as a development set to estimate FR and Rwd for evaluating set-2, and vice versa. The parameters were tuned to minimize the total number of sentences for information access in the development set. We compared the proposed method with the following conventional methods. Note that method 1 is the baseline method and method 2 was adopted in the original "Dialogue Navigator for Kyoto City" and used in the field trial.

Method 1 (baseline)

- Make a search query using the 1st hypothesis of ASR.
- Incorporate the contextual information related to the current topic.
- Make a confirmation when the ASR confidence of the pre-defined topic word is low.
- Answer the question when the user query is judged a question.

Method 2 (original system)

- Make a search query using all nouns in the 1st-3rd hypotheses of ASR.
- The other conditions are the same as in method 1.

The comparisons to these conventional methods are shown in Table 1. The improvement compared with that in baseline method 1 is 6.4% in the response success rate and 0.78 of a sentence in the number of sentences for information access.

A breakdown of the selected response candidates by the proposed method is shown in Table 2. Many of the responses were generated using a single hypothesis from the N-best list of ASR. The result confirms that the correct hypothesis may not be the first one, and the proposed method selects the appropriate one by considering the likelihood of retrieval. Most of the confirmations were generated using the 1st hypothesis of ASR. The Answers to questions were often generated from the search queries with contextual information. This result suggests that when users used anaphoric expressions, the appropriate contextual information was incorporated into the question.

Table 1: Comparison with conventional methods

	Success rate	# sentences for presentation
Method 1 (baseline)	59.2%	5.49
Method 2	63.4%	4.98
Proposed method	65.6%	4.71

Table 2: Breakdown of selected candidates

	w/o context			with context		
	Pres	Conf	Ans	Pres	Conf	Ans
1st hyp.	233	134	65	2	151	2
2nd hyp.	140	43	28	2	2	6
3rd hyp.	209	50	46	1	6	5
merge all	75	11	3	18	0	91
rejection	111					

5 Conclusion

We have proposed a dialogue framework to generate an optimal response. Specifically, the choices in response generation are optimized as a minimization of the Bayes risk based on the reward for a correct information presentation and a penalty for redundant time. Experimental evaluations using real user utterances were used to demonstrate that the proposed method achieved a higher success rate for information access with a reduced number of sentences. Although we implemented only a simple confirmation using the likelihood of retrieval, the proposed method is expected to handle more complex dialogue management such as the confirmation considering the impact for the retrieval (Misu and Kawahara, 2006).

References

Chang, E., F. Seide, H. Meng, Z. Chen, Y. Shi, and Y. Li. 2002. A System for Spoken Query Information Retrieval on Mobile Devices. *IEEE Trans. on Speech and Audio Processing*, 10(8):531–541.

Levin, E. and R. Pieraccini. 2006. Value-based Optimal Decision for Dialog Systems. In *Proc. Spoken Laguage Technology Workshop (SLT)*, pages 198–201.

Misu, T. and T. Kawahara. 2006. Dialogue strategy to clarify user's queries for document retrieval system with speech interface. *Speech Communication*, 48(9):1137–1150.

Misu, T. and T. Kawahara. 2007. Speech-based interactive information guidance system using question-answering technique. In *Proc. ICASSP*.

Raymond, C., Y. Esteve, F. Bechet, R. De Mori, and G. Damnati. 2003. Belief Confirmation in Spoken Dialog Systems using Confidence Measures. In *Proc. Automatic Speech Recognition and Understanding Workshop (ASRU)*.