Error-Tolerant Question Answering for Spoken Documents

Tomoyosi Akiba, Hirofumi Tsujimura

Deptartment of Information and Computer Sciences
Toyohashi University of Technology
1-1 Hibarigaoka, Tenpaku-cho, Toyohashi, 441-8580, JAPAN
akiba@cl.ics.tut.ac.jp

10.21437/Interspeech.2007-177

Abstract

This paper proposes an error-tolerant question answering method for spoken documents. Though the question answering system for written documents can be directly applied to the transcribed spoken documents by using a LVCSR system, the recognition errors significantly degrade the QA performance. Especially, it is often the case that the answer itself is miss-recognized and in that case it becomes quite difficult to find the answer. To cope with such a problem, instead of conventional NE extraction, the proposed method utilizes named entity detection that decides only whether a section of speech, i.e. an utterance, contains named entities of a specific type. Because the NE detection is much easier task and utilized wider context than the NE extraction, it is expected to work robustly for erroneous transcribed speech data. The experimental results showed that the proposed method outperformed the baseline methods with respect to the spoken document with recognition errors.

Index Terms: spoken document retrieval, question answering, named entity detection

1. Introduction

Open-domain Question Answering (QA) was first evaluated extensively at TREC-8 [7]. The goal in the QA task is to extract words or phrases as the answer to a question from an unorganized document collection, rather than the document lists obtained by traditional information retrieval (IR) systems.

Speech technology, i.e. LVCSR, can enhance QA systems in several ways. Speech-driven QA [1, 2], in which spoken questions are used as inputs, have promise for improving the utility of QA systems. On the other hand, spoken documents can replace the written documents as the target document collection for QA. It can also enhance the utility of QA by targeting the increasing number of spoken materials exchanged through the Internet.

When we enhance the text-based QA system to dealing with spoken documents, one of the most common problems arises from the recognition errors. Especially, the answer of the input question itself can be miss-recognized. In this case the QA system cannot find the correct answer string from the recognized text, while the corresponding speech data is correct. Unfortunately, it is often the case because the answer of the question is often likely to be a named entity, and many of the named entities are possibly a rare or new terms so that they are not included in the dictionary of the LVCSR system.

In this paper, we propose a robust QA system for spoken documents. The system uses the named entity detection instead of the conventional named entity extraction. The NE detection detects the existence of a specific type of NE in the rather longer context, i.e. an utterance, than the exact NE string. Support vector machine is used to implement the subsystem, and the word and the POS n-gram are used for the features.

Section 2 describes the detail of our proposed method that exploiting the named entity detection subsystem and its integration to the question answering system for spoken documents. Section 3 describes the experimental results for evaluating both the NE detection individually and the total performance of QA. Section 4 describes our conclusion.

2. Named Entity Detection for Spoken Documents

The goal of named entity detection is to decide whether a section of speech, i.e. an utterance, contains named entities of a specific type, while the goal of named entity extraction requires also to identify the location of its appearance in the speech. Levit et al. [4] applied NE detection for pre-processing of NE extraction for speech data and reported that it improved the performance of the total NE extraction process.

We applied the NE detection to question answering for spoken documents. One of the advantages of the NE detection is that it is expected to work robustly with respect to speech recognition errors on automatically transcribed spoken documents.

Sudoh et al. [6] also investigated the named entity extraction on speech data. It incorporates the confidence score of speech recognition into NE extraction process. However, the goal of the method is to extract the correct NE string from the speech data. The miss-recognized NEs, which is included in the speech data but transcribed wrongly by speech recognition, should be excluded in their task definition. It is quite different from our goal, where all NEs in the speech data should be identified for question answering.

2.1. Detection Method and Features

The coarse seven NE types, PERSON, LOCATION, ORGANIZATION, DATE, TIME, MONEY, RATE, are used for the NE detection, which are defined in [5].

The NE detection subsystem was implemented as a set of seven binary classifiers, each of which decides whether an input utterance includes the named entity of one of the seven types of NEs. Support Vector Machine was used for the classifiers.

The features are word uni-grams (referred as W1), word bi-grams (referred as W2), POS uni-grams (referred as P1), and POS bi-grams (referred as P2), which are obtained by applying a Japanese morphological analyzer to the automatically transcribed utterance. All possible combinations of the features were investigated in the experiment.

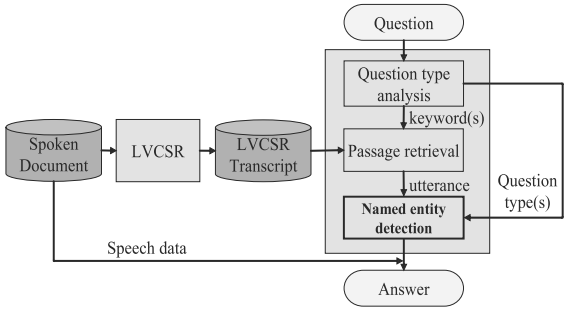

Figure 1: The configuration of the QA system.

2.2. Integration to Question Answering

The configuration of our question answering system for spoken documents is illustrated in Figure 1. It is basically a simple extension of the question answering system for text, excepting the following two. Firstly, the target documents of QA are replaced by the automatically transcribed text obtained from spoken documents by using a LVCSR system. Secondly, the named entity detection subsystem replaces the NE extraction subsystem in the original QA system. It receives an expected answer type from the question type analysis subsystem and invokes the corresponding binary classifier to get the SVM score.

In order to incorporate the SVM classifier into the question answering system, the SVM score is interpolated with the original IR score obtained by the passage retrieval subsystem. The SVM score is normalized by applying sigmoid function. The candidate answer utterances are rescored by the following equation.

$$\text{score} = \alpha \cdot \text{passage\_score} + \frac{1-\alpha}{1+e^{-\text{SVM\_score}+\beta}}$$

The output of the system is a ranked list of utterances rather than a ranked list of exact answer strings, in order to avoid locating the exact answer and to improve the robustness to the erroneous automatic transcription of spoken documents.

3. Experiments

3.1. Data

For the evaluation of the performance of our proposed method, one hundred and seventeen programs collected from Japanese broadcast news were used. They were sent on the air from June 1st to July 14th in 1996 and contain 1076 articles of 8720 utterances. All of them are transcribed manually. The speech recognition accuracy with respect to this data was about 63 %.

From them, thirty programs (including 284 articles of 1959 utterances), which were sent on the air from July 1st to 12th, were manually annotated with NE tags of the seven types. We used automatically transcribed text for both learning and testing. (Section 3.2)

The other eighty seven programs (including 792 articles of 6761 utterances) were used for the target documents of the QA system used for evaluating the QA performance. (Section 3.3)

3.2. NE detection

Firstly, the experimental evaluation for investigating the individual performance of the NE detection subsystem was conducted. The seven classifiers were implemented as described in section 2.1 by using the NE annotated thirty programs for the training data for the SVM. 12-fold cross validation, in which we split the training data sequentially with the time-line, was performed for the evaluation. The precision and the recall were controlled by introducing the threshold on the SVM score, in that point positive and negative class of the NE type was distinguished.

To see the performance of the NE detection directly on the miss-recognized speech data, we divided the test data for each NE type into three groups: the group of the utterances that include at least one correctly recognized NE of the specific type (referred as **NE correct**), the group of the utterances that include the NE of the specific type but all of them are miss-recognized (referred as **NE error**), and the others (referred as **no NE**). Then we excluded the **NE correct** group from the test data for evaluation of each NE type. Table 1 summarizes the number of the utterances used for the test data.

We compared our method with the Japanese rule-based NE extraction tools for baseline method [8]. The performance of the baseline method for the manually transcribed text is 84.6, 86.6 and 78.6 in F-measure (%) for PERSON, LOCATION and ORGANIZATION, respectively.

Figure 2 shows the precision-recall curves for each NE type. On the whole, the NE detection outperformed the rule-based NE extraction. It also shows that the effective feature varies for the NE types. For example, the word uni-gram feature is effective for PERSON, DATE and RATE, while word bi-gram is effective for ORGANIZATION and MONEY. For LOCATION, DATE and TIME, the combination of the uni-gram and bi-gram features is better than using them individually.

Table 2 and 3 summarize the best performed results obtained by selecting the best threshold and features for each NE type.

3.3. Question Answering

Next, the question answering performance by using our method was investigated. The fifty factoid questions that asked for the short term answers appeared in the target broadcast news was used for the evaluation.

The NE detection subsystem implemented in previous section was integrated to the existing QA system [3] by using the method described in section 2.2. The target spoken documents of the QA system was the other eighty seven programs of the spoken documents described in section 3.1 than those used for learning the SVM.

Additionally, in order to simulate the speech recognition errors in automatic transcription of the spoken documents, from the dictionary of the LVCSR system used for the automatic transcription, we excluded the entries that appeared in the answer strings of the fifty questions. In other words, the automatically transcribed text for the spoken documents did not include the correct answer strings for the fifty questions used for our evaluation.

The system outputs ten ranked answers $a_1..a_{10}$ for each question q. Each answer is scored on the inverse number of its order, called Reciprocal Rank (RR). The score of the question q, $RR(q)$, is the highest score of its all answers.

$$rr(a_i) = \begin{cases} 1/i & \text{if } a_i \text{ is a correct answer} \\ 0 & \text{otherwise} \end{cases}$$

$$RR(q) = \max_{a_i} rr(a_i)$$

The mean RR (MRR) for all fifty questions was used as the

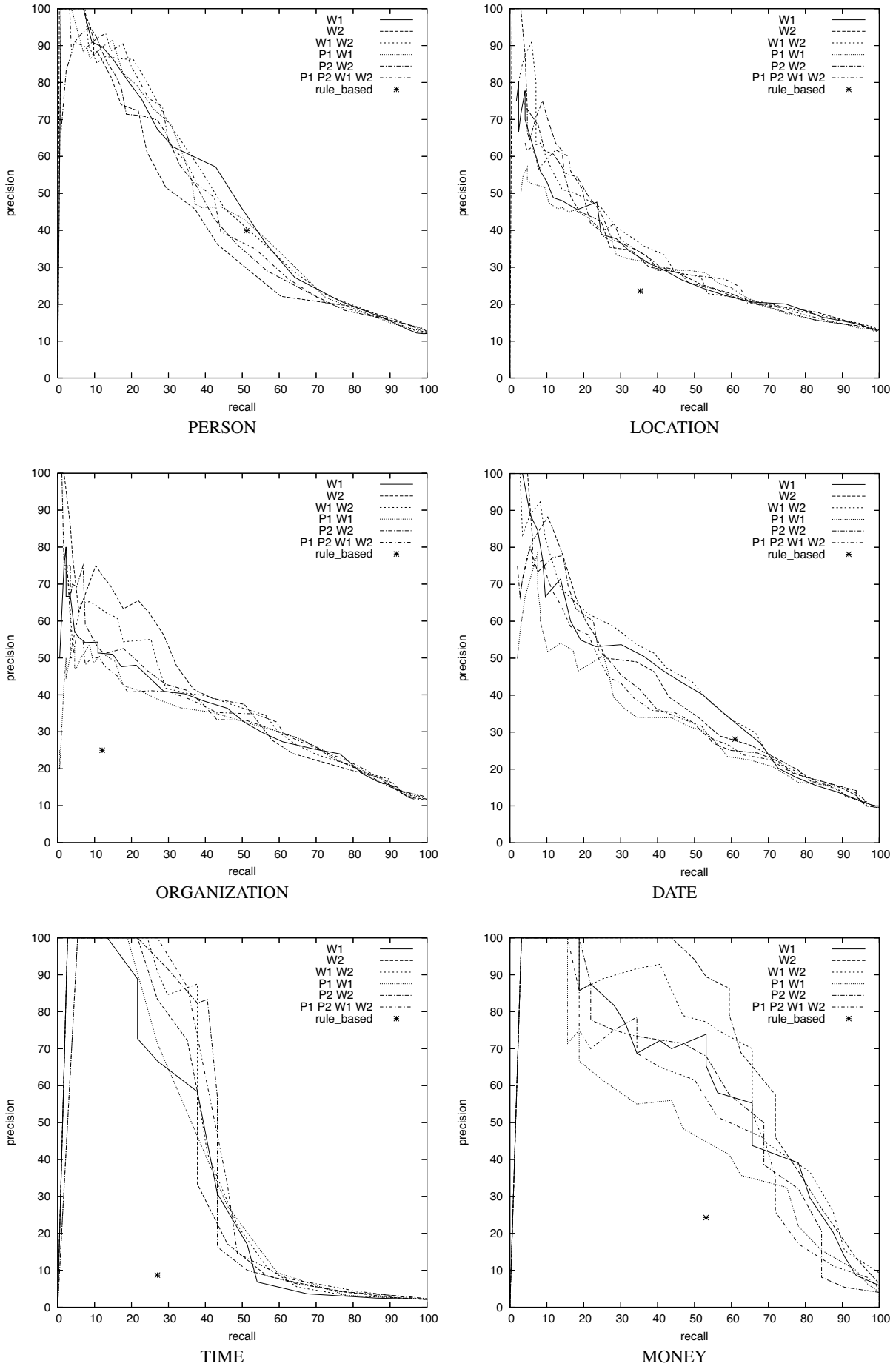

Figure 2: The precision-recall curves for six NE types.

Table 1: The number of the utterance with and without recognition errors of each NE type.

NE type	PER	LOC	ORG	DATE	TIME	MONEY	RATE
NE correct	170	613	476	430	55	65	27
NE error	215	168	171	146	37	32	20
no NE	1574	1178	1312	1383	1867	1862	1912

Table 2: The NE detection results (%) by adjusting the parameters for name expressions.

Method	PERSON			LOCATION			ORGANIZATION		
	P	R	F	P	R	F	P	R	F
baseline	39.9	51.2	44.8	23.5	35.3	28.2	25.0	12.1	16.3
SVM	57.2	42.8	48.9	33.3	41.8	37.1	37.5	50.6	43.0
(features)		(W1)			(W1,W2)			(W2)	

Table 3: The NE detection results (%) by adjusting the parameters for numerical expressions.

Method	DATE			TIME			MONEY			RATE		
	P	R	F	P	R	F	P	R	F	P	R	F
baseline	28.1	61.0	38.4	8.7	27.0	13.2	24.3	53.1	33.3	0	0	0
SVM	43.6	49.3	46.3	83.3	40.5	56.6	86.4	59.4	70.4	58.3	35.0	43.8
(features)		(W1,W2)			(W2,P2)			(W2)			(W1)	

Table 4: The question answering result measured by MRR.

method	MRR	*MRR with error simulation*
passage retrieval	0.404	0.351
NE filtering	0.364	0.321
proposed	0.441	0.362

evaluation metric for question answering.

We compared three results: the result by scoring only from passage retrieval (referred as **passage retrieval**), the result by applying the rule-based NE extraction and filtering the utterance with inconsistent NE types (referred as **NE filtering**), and the result by our method of incorporating the SVM score described in section 2.2 (referred as **proposed**). Table 4 shows the results.

It was shown that **NE filtering** degraded the performance of baseline **passage retrieval**, because it over-rejected the correct utterances. The **proposed** method could select the correct utterances robustly and improved the baseline. Note that it consistently outperformed the baseline whether the answer was correctly recognized or not (as shown at the row labeled *MRR with error simulation* in Table 4).

4. Conclusion

In this paper, we proposed error-tolerant question answering for spoken documents. The named entity detection was applied to text-based question answering process instead of NE extraction. The experimental results showed that it worked robustly for erroneous automatic transcription of spoken documents and improved the total question answering performance, especially in the case that the answer of the question itself was miss-recognized. The other error-tolerant techniques, including exploiting the N-best list of recognized text and utilizing parallel text materials of the targeting speech data, should be investigated in the future work.

5. References

[1] T. Akiba and H. Abe. Exploiting passage retrieval for n-best rescoring of spoken questions. In *Proceedings of International Conference on Speech Communication and Technology (Eurospeech)*, pages 65–68, 2005.

[2] T. Akiba, A. Fujii, and K. Itou. Effects of language modeling on speech-driven question answering. In *Proceedings of International Conference on Spoken Language Processing*, pages 1053–1056, 2004.

[3] T. Akiba, A. Fujii, and K. Itou. Question answering using "common sense" and utility maximization principle. In *Proceedings of The Fourth NTCIR Workshop*, 2004. `http://research.nii.go.jp/ntcir/workshop/OnlineProceedings4/QAC/NTCIR4-QAC-AkibaT.pdf`.

[4] M. Levit, P. Haffner, A. Gorin, H. Alshawi, and E. Nöth. Aspects of named entity processing. In *Proceedings of International Conference on Speech Communication and Technology (Eurospeech)*, pages 672–675, 2004.

[5] S. Sekine and Y. Eriguchi. Japanese named entity extraction evaluation – analysis of results. In *Proceedings of International Conference on Computational Linguistics*, pages 25–30, 2000.

[6] K. Sudoh, H. Tsukada, and H. Isozaki. Incorporating speech recognition confidence into discriminative named entity recognition of speech data. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 617–624, 2006.

[7] E. Voorhees and D. Tice. The TREC-8 question answering track evaluation. In *Proceedings of the 8th Text Retrieval Conference*, pages 83–106, Gaithersburg, Maryland, 1999.

[8] I. Watanabe, F. Masui, and J. Fukumoto. NExT – a named entity extraction tool. `http://www.ai.info.mie-u.ac.jp/~next/`.