Spoken document retrieval method combining query expansion with continuous syllable recognition for NTCIR-SpokenDoc

Satoru Tsuge
Daido University
10-3 Takiharu-cho, Minami-ku,
Nagoya, Aichi 457-8539 Japan
tsuge@daido-it.ac.jp

Hiromasa Ohashi
Nagoya University
Furo-cho, Chikusa-ku,
Nagoya, Aichi 464-8603 Japan
hiromasa.ohashi@g.sp.m.is.nagoya-u.ac.jp

Norihide Kitaoka
Nagoya University
Furo-cho, Chikusa-ku,
Nagoya, Aichi 464-8603 Japan
kitaoka@nagoya-u.jp

Kazuya Takeda
Nagoya University
Furo-cho, Chikusa-ku,
Nagoya, Aichi 464-8603 Japan
kazuya.takeda@nagoya-u.jp

Kenji Kita
The University of Tokushima
2-1 Minamijosanjima-cho,
Tokushima, Tokushima
770-8506 Japan
kita@is.tokushima-u.ac.jp

ABSTRACT

In this paper, we propose a spoken document retrieval method which combines query expansion with continuous syllable recognition. The proposed method expands a query by using words from the web pages collected by a search engine. It is assumed that relevant document vectors exist on the plane which is constructed from the query vector and the extended vector. The weight parameter between a target document vector and a query vector is calculated for query expansion. In addition, target documents are mapped not only to space constructed by continuous word speech recognition results, but also to space constructed by syllable speech recognition results. Then, the proposed method calculates a distance between the query vector and the document vector for each space and combines these distances. For evaluating the proposed method, we conducted spoken document retrieval experiments on the SpokenDoc task of the NTCIR-9 meeting. The experimental results showed that the proposed method improved the mean average precision score from the baseline provided by the meeting organizer of 0.392784 to 0.406085 when running the formal run of SpokenDoc task.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

General Terms

Algorithms, Experimentation, Performance

Keywords

NTCIR-9, Spoken Document Retrieval

1. INTRODUCTION

With the rapid growth of online information, e.g., the World Wide Web (WWW), a large collection of full-text documents are now available, and opportunities for retrieving a useful information have increased. Information retrieval is now becoming one of the most important issues for handling large amounts of text data. In addition, there are also many other kinds of media data, such as pictures, movies, music, speech, and so on, on the Internet. Opportunities to retrieval these media data are also increasing.

Two spoken document retrieval tasks were added to the NTCIR-9 meeting, Spoken a Document Retrieval task (SDR) and a Spoken Term Detection task (STD). Our group, whose name is "Team Big Four Dragons (TBFD)", will participate in the SDR task at NTCIR-9. To perform this task, we need to the requested information from the target documents, which are spoken documents translated by an automatic speech recognition system (ASR).

Typical SDR systems first transforms speech into word or sub-word sequences using an automatic speech recognizer. The system then builds indexes of these words for text retrieval. Because there are some recognition errors and out-of-vocabulary (OOV) terms in the target documents, it is difficult to retrieve documents using only the keyword retrieval method.

Some methods have been proposed for retrieving spoken documents. One method is to use sub-word units instead of words for the index[4]. This method avoids the OOV problem. In addition, a method using a combination of phone-based and word-based recognition results has been proposed for speaker term detection[2]. For dealing with recognition errors, indexes have been constructed using of words in a lattice/confusion matrix[3].

In this paper, we propose a novel spoken document retrieval method combining query expansion with continuous syllable recognition. In our method, the target documents and queries are represented as vectors by a vector space model (VSM). The VSM is an approach for mapping documents into a space associated with the index terms in the documents. Our method maps documents to multiple spaces

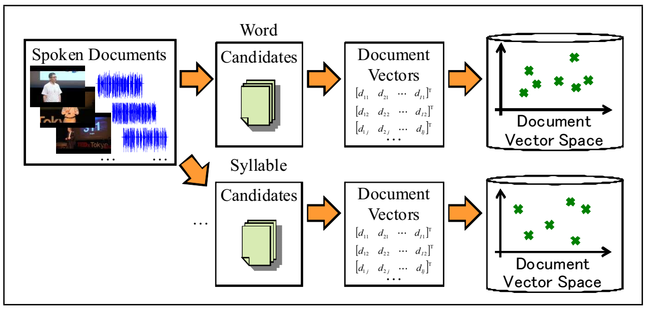

Figure 1: Mapping documents to vector spaces using VSMs

which are constructed of words and sub-words. We believe that the proposed method can find relevant documents by using large amount of information of documents to represent the target documents in multiple spaces. It is difficult to retrieve relevant documents using the query vector if the number of index terms in the query is small. Hence, the proposed method expands the query vector by using web pages collected from the Internet. For query expansion, we propose a method which calculates a weight parameter between the original query vector and the expanded vector.

Because the proposed method represents the document in multiple spaces, we can calculate a distance between the document vector and the query vector in each space. We believe the combination of these distances can improve retrieval performance because the distances calculated in multiple spaces represent different levels of similarity, which can be used for retrieval. Therefore, in this paper, we propose a distance combination method.

This paper is organized as follows. In the following section, we introduce the VSM and describe the proposed method. Section **3** shows spoken document retrieval results of a dry run and a formal run. In Section **4**, we discuss our findings and future plans for using the proposed method. Finally, in Section **5**, we summarize our conclusions.

2. PROPOSED METHOD

2.1 Vector Space Model

A VSM maps the target document and query to vector spaces constructed using the index terms. The vector spaces used for the proposed method are constructed as follows (the flow of this construction is shown in Fig.1.) In this paper, we use target documents which are speech recognition results provided by the NTCIR-9 organizer. Hence, it is not necessary to perform morphological analysis because speech recognition results are divided into words.

1. Index terms

   Nouns, katakana sequences, and alphabet sequences are used for the index terms of word-level VSMs. In the pronunciation dictionary, the morpheme numbers of nouns, katakana sequences, and alphabet sequences are from 4 to 32, 84, and 88, respectively.

   The index terms of syllable-level VSMs are syllable 3-grams.

2. Index term weighting

By using the index terms, every documents and query are represented as a feature vectors. In this paper, we used three kinds of term weighting methods, which are described as follows:

- Term Frequency (TF)

  For the word-level VSM, TF is calculated by using the N-best speech recognition candidates, as shown in the following equation:

  $$d_{ij} = \sum_{n=1}^{N} \frac{\mathrm{TF}_{ij}^{n}}{n}, \tag{1}$$

  where d_{ij} indicates the value of the jth index term in the ith document. TF_{ij}^{n} is the term frequency of the jth index term in the ith document which is the nth speech recognition candidate.

- Term Frequency - Inverse Document Frequency (TF-IDF)

  In addition to TF, inverse document frequency is used for term weighting. The TF-IDF is calculated using the following equation:

  $$d_{ij} = \sum_{n=1}^{N} \frac{\mathrm{TF}_{ij}^{n}}{n}\mathrm{IDF}, \tag{2}$$

  $$\mathrm{IDF} = \log\left(\frac{D}{D_j}\right), \tag{3}$$

  where D and D_j indicate the total number of documents and the number of documents in which the jth term index appears.

- Binary Weight

  In this weighting method, the term weighting is 1 when the term appears in the document. If the term does not appear in the document, the term weighting is 0. The term weight is calculated using the following equation:

  $$d_{ij} = 1 \ \ if \ \sum_{n=1}^{N} \frac{\mathrm{TF}_{ij}^{n}}{n} > 0, \tag{4}$$

  $$or \tag{5}$$

  $$d_{ij} = 0 \ \ if \ \sum_{n=1}^{N} \frac{\mathrm{TF}_{ij}^{n}}{n} = 0. \tag{6}$$

  $$\tag{7}$$

2.2 Query Expansion

Although documents which are speech recognition results are divided into words, queries are in the form of a sentences. Hence, we first perform a morphological analysis of the query sentence. Then, using the morphological analysis results, we construct the query vector using the same procedure described in the previous section. If the number of index terms in the query is very small, it is difficult to retrieve relevant documents using the query vector. The average number of index terms in a query was only 4.9 under the conditions of the dry run of SpokenDoc task. Therefore, we expand the query in order to better retrieve relevant documents. Query expansion is an effective technique used to add more useful words to a query. Next, we will describe the query expansion method used in this paper.

We begin query expansion by performing a morphological analysis of the query sentence and selecting nouns from the morphological analysis results. Then, using the selected nouns, we search for web pages from the Internet using a search engine. We perform a morphological analysis of these web pages and calculate the expanded vectors in the same way as we did during VSM construction.

The expanded query vector, $\hat{\boldsymbol{q}}$, is calculated using Equ.(8):

$$\hat{\boldsymbol{q}} = (1-\alpha)\boldsymbol{q}_o + \alpha \boldsymbol{q}_e, \tag{8}$$

where $\hat{\boldsymbol{q}}$ is the expanded query vector which is used for spoken documents retrieval. $\boldsymbol{q}_o$ and $\boldsymbol{q}_e$ are the original query vector constructed from the query sentence and the expanded vector constructed from the collected web pages, respectively. α is the weight parameter between the original query vector and the expanded vector.

Although Equ.(8) is similar to the Rocchio-based relevant feedback method, our query expansion method uses information from web pages which are searched with the query instead of utilizing user feedback. In the Rocchio-based feedback method, it is important to determine the weight parameter in order to improve retrieval performance. Therefore, in this paper, we propose a novel weight parameter calculation method.

We believe that the optimal value of the weight parameter of query expansion is different for each document. In order to determine the optimal value of the weight parameter of query expansion, we assume that there are relevant documents in the space which is constructed between the original query vector and the expanded vector (we will call this "relevant document modeling" hereafter) shown in Fig.2. This method determines the weight parameter by minimizing the distance between the expanded query vector and each document vector. The weight parameter of query expansion, $\alpha(\hat{\boldsymbol{q}}_m, \boldsymbol{d}_i)$, is calculated using the following formula.

$$\begin{aligned} \alpha_{(\hat{\boldsymbol{q}}_m, \boldsymbol{d}_i)} &= argmax_\alpha(cos(\boldsymbol{d}_i, \hat{\boldsymbol{q}}_m)) & (9) \\ &= argmax_\alpha\left(\frac{\boldsymbol{d}_i \cdot \hat{\boldsymbol{q}}_m}{|\boldsymbol{d}_i||\hat{\boldsymbol{q}}_m|}\right) & (10) \\ &= argmax_\alpha\left(\frac{\boldsymbol{d}_i \cdot (\alpha\boldsymbol{q}_o + (1-\alpha)\boldsymbol{q}_e)}{|\boldsymbol{d}_i||(\alpha\boldsymbol{q}_o + (1-\alpha)\boldsymbol{q}_e)|}\right). & (11) \end{aligned}$$

$\alpha_{(\hat{\boldsymbol{q}}_m, \boldsymbol{d}_i)}$ is calculated using following equation:

$$\begin{aligned} \alpha_{(\hat{\boldsymbol{q}}_m, \boldsymbol{d}_i)} = {} & -(2((\boldsymbol{q}_o \cdot \boldsymbol{q}_o)(\boldsymbol{d}_i \cdot \boldsymbol{d}_i))((\boldsymbol{q}_e - \boldsymbol{q}_o) \cdot \boldsymbol{d}_i) - \\ & (2(\boldsymbol{q}_o \cdot (\boldsymbol{q}_e - \boldsymbol{q}_o))(\boldsymbol{d}_i \cdot \boldsymbol{d}_i))(\boldsymbol{q}_o \cdot \boldsymbol{d}_i)) \\ & (2(\boldsymbol{q}_o \cdot (\boldsymbol{q}_e - \boldsymbol{q}_o))(\boldsymbol{d}_i \cdot \boldsymbol{d}_i)(\boldsymbol{q}_e - \boldsymbol{q}_o) \cdot \boldsymbol{d}_i - \\ & (\boldsymbol{q}_e - \boldsymbol{q}_o) \cdot (\boldsymbol{q}_e - \boldsymbol{q}_o)(\boldsymbol{d}_i \cdot \boldsymbol{d}_i)(\boldsymbol{q}_o \cdot \boldsymbol{d}_i)))^{-1}. \end{aligned} \tag{12}$$

2.3 Combination of distance

In our method, there are different types of VSMs (word-level VSMs with or without query expansion, and syllable-level VSMs), and different types of term weighting methods (TF, TF-IDF, and Binary Weight), for spoken document retrieval. The retrieval documents produced using each method are different because each method uses different information from the documents. Therefore, we combine these methods for retrieving more relevant documents. In this paper, we specifically propose a spoken document retrieval method combining query expansion with continuous syllable recognition.

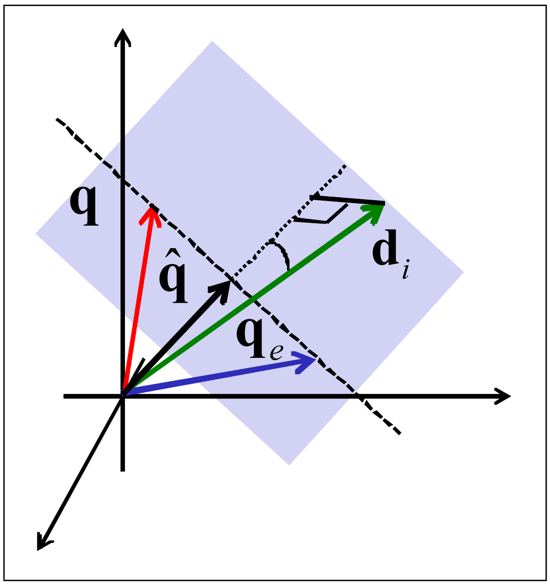

Figure 2: Relevant document modeling

The proposed method combines the distances calculated by the VSMs as shown in the following equation:

$$\text{dist} = \sum_k \beta_k cos_k(\boldsymbol{d}_i, \boldsymbol{q}), \tag{13}$$

where $cos_k()$ indicates the cosine distance of VSM$_k$ and β_k is the weight parameter of each cosine distance. Using the distance calculated by Equ.(13), the proposed method retrieves the documents. We believe that the proposed method obtains more relevant documents than conventional methods because the proposed method uses a large amount of information, namely word and syllable recognition results, for spoken document retrieval.

3. EXPERIMENTS

In order to evaluate the proposed method, we conducted the spoken document retrieval experiment under the conditions of the SpokenDoc task of NTCIR-9[1].

3.1 Experimental conditions

For the target documents, we use the speech recognition results provided by the NTCIR-9 organizer. In this experiment, we focus on a lecture retrieval of SpokenDoc task.

For index terms, we use the following words:

- Index terms of word-level VSM:

  From a vocabulary list used in word-based ASR system, we choose nouns, alphabet sequences, and katakana sequences for the index terms of our word-level VSM. The morpheme numbers of the nouns is from 4 to 32, the morpheme number of katakana sequences is 84, and the morpheme number of alphabet sequences is 88 in this vocabulary list. As results, the total number of index terms of our word-level VSM is 14,716.

- Index terms of syllable-level VSM:

Table 1: Retrieval results of each method

Term weight	VSM level	N	MAP
TF	word	2	0.1870
TF-IDF	word	5	0.3412
Binary	word	1	0.2305
TF-IDF	syllable	1	0.2924

For index terms of our syllable-level VSM, we choose the syllable 3-grams from the syllable-based transcription.

The total number of index terms of our syllable-level VSM is 208,541.

For index term weighting, we use the following method:

- TF for word-level VSM
- TF-IDF for word-level VSM
- Binary Weight for word-level VSM
- TF-IDF for syllable-level VSM

For the queries, morphological analysis is performed, then the index terms are chosen. For the syllable-based VSM, the queries are translated into katakana sequences.

Using the nouns of each query, we obtain the web pages from Wikipedia using Google's search-engine. The index terms selected from these web pages are used for query expansion ($\boldsymbol{q}_e$ in Equ.(8)).

Mean Average Precision (MAP) is used for the evaluation. The MAP is calculated as follows:

$$\mathrm{MAP} = \frac{1}{|Q|}\sum_{q \in Q} AveP_q, \tag{14}$$

where $AveP_q$ indicates the average precision of query, q.

3.2 Dry Run Results

This section describes the experimental results of the dry run. In the following sections, we describe

- Results of each VSM,
- Efficiency of query expansion,
- Evaluation of the proposed method,
- Evaluation of distance combinations.

3.2.1 *Results of each method*

First, we compare the MAP of each term weighting method of each VSM. Table 1 shows the MAP of each method. We also show the number of speech recognition candidates used for VSM construction (column N), which showed the highest MAP for each method.

Table 1 shows that the use of multiple speech recognition candidates improves MAP under the condition of word-level VSM with TF-IDF weighting. These results imply that the index terms which are important for retrieval are lacking if there are recognition errors within the best candidate. Moreover, we can confirm that TF-IDF is more suitable for spoken document retrieval compared to TF. Therefore, we will use TF-IDF as our weighting method for word-level VSMs in the following experiments. This table also shows that the MAPs of binary weighted word-level VSM and TF-IDF weighted syllable-level VSM are lower than that of TF-IDF weighted word-level VSM.

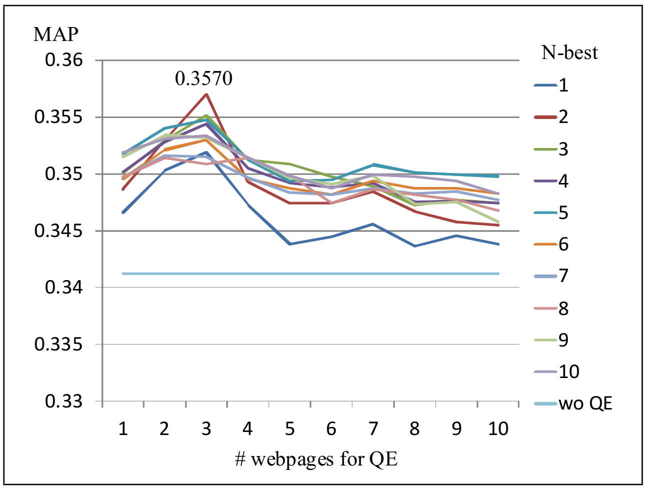

Figure 3: Experimental results of query expansion(QE)

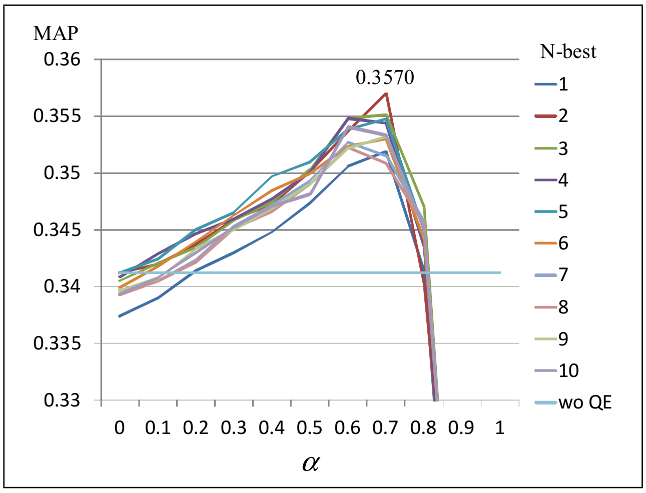

Figure 4: MAP scores with different α

3.2.2 *Results of query expansion*

In this section, we describe experimental results of the query expansion. We conduct our query expansion under the condition at word-level VSM with TF-IDF weighting, the method which showed the highest MAP.

Figure 3 shows the MAP scores with different speech recognition candidates and different numbers of web pages used for query expansion. In this experiment, the value of the weighting parameter of query expansion, which is α in Equ.(8), is 0.6. From this figure, we can see that query expansion improves MAP for spoken document retrieval. The figure also shows that MAP with different speech recognition candidates is slightly different when the number of candidates (N) is more than two. Based on this result, we believe that the number of index terms in the expanded query vector does not increase, even if the number of candidates increases to two or more.

On the other hand, the figure shows that MAP is the highest when the number of web pages used for query expansion

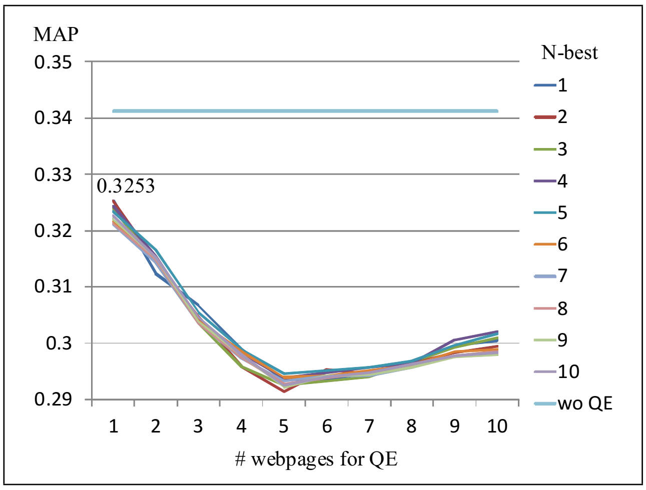

Figure 5: Experimental results of proposed method without restriction

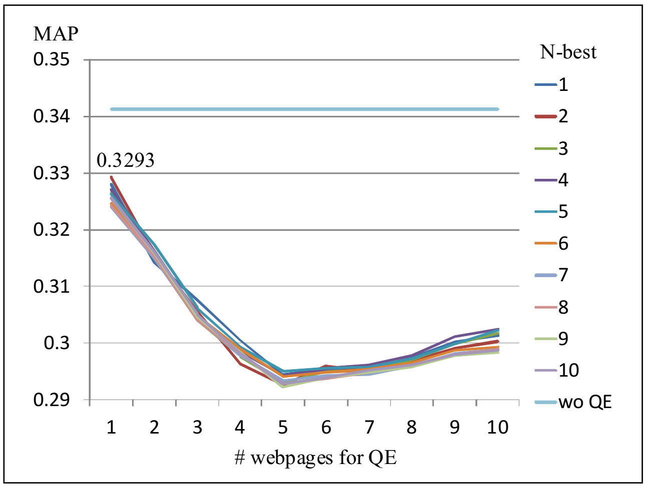

Figure 6: Experimental results of proposed method with restriction

is three. Because the web pages used for the query expansion are collected from Wikipedia by Google's search engine, the amount of information unnecessary for spoken document retrieval is much less as compared to unnecessary information collected from all web-sites. However, from this figure, we can also see that when web pages whose search rank was four or more were used, there was a negative impact on retrieval performance. From this result, the number of web pages and the number of speech recognition candidates for query expansion are set to three and two, respectively, in the following experiments.

Fig.4 shows the MAP with different query expansion weight parameters (α in Equ.(8)). From this figure, we can see that MAP is the highest when α is 0.7. In this experiment, the number of index terms in the original query vector is significantly smaller (the average number of index terms in a query vector is 4.9). Therefore, the weight parameter of the query expansion is larger than 0.5 because the expanded vector, which is constructed from searched web pages has more useful information compared to the original query vector.

3.2.3 *Results of the proposed method*

In this section, we describe the experimental results of the proposed method in the dry run experiment. Figure 5 shows the experimental results of the proposed method with different speech recognition candidates and different numbers of web pages used for query expansion when the weight parameter of query expansion, $\alpha_{(\boldsymbol{d}_i, \hat{\boldsymbol{q}}_m)}$ in Equ.(9), is not restricted. Figure 6 shows the experimental results of the proposed method with different speech recognition candidates and different numbers of web pages used for query expansion when the weight parameter, $\alpha_{(\boldsymbol{d}_i, \hat{\boldsymbol{q}}_m)}$ in Equ.(9), is restricted from 0.0 to 1.0.

From these figures, we can see that the highest MAP score occurred when the number of web pages used for query expansion is one. From investigation of these results, we can see that the weight parameter of the query expansion tends to be large, i.e., the expanded vector is more important than the original query vector for retrieval. However, MAP scores of the proposed method do not surpass those of the best method without query expansion (word-level VSM with TF-IDF term weighting).

By comparing Fig.5 and Fig.6, we can see that the MAP scores of the proposed method with restriction are slightly higher than those of the proposed method without restriction. The proposed method determines the value of the weight parameter of the query expansion where the distance between the document vector and the expanded query vector becomes the smallest. Hence, the expanded query vector has negative elements if the value of weight parameter is more than 1.0 or less than 0.0. We believe that this causes the degradation of the MAP score. We will investigate the details of these experimental results in the future.

3.2.4 *Combination of distances*

In this section, we combine the distances which are calculated using the binary weight of the word-level VSM, TF-IDF weighting of the word-level VSM with query expansion, and TF-IDF weighting of the syllable-level VSM.

The distance is calculated using the following equation:

$$\begin{aligned} \text{dist} &= (1-\beta_1-\beta_2)\cos_{wb}(\boldsymbol{d}_i, \boldsymbol{q}) \\ &\quad +\beta_1 \cos_{we}(\boldsymbol{d}_i, \boldsymbol{q}) + \beta_2 \cos_{syl}(\boldsymbol{d}_i, \boldsymbol{q}), \end{aligned} \tag{15}$$

where the $cos_{wb}()$, $cos_{we}()$, and $cos_{syl}()$ indicate the cosine distances calculated by binary weight of the word-level VSM, TF-IDF weighting of the word-level VSM with query expansion, and TF-IDF weighting of the syllable-level VSM, respectively. β_1 and β_2 are weight parameters for the distance combination.

Fig.7, Fig.8, and Fig.9 show the experimental results of the distance combination method with different weight parameters for distance combination. In addition, Tab.2 shows the experimental results of the proposed distance combination method with the highest MAP score. Comparing the previous results, which are the results of the individual methods shown in table 1, Fig.3, Fig.4, Fig.5, and Fig.6, the distance combination method improves MAP scores over those of individual methods. Therefore, it is concluded that the distance combination method is the most efficient for spoken document retrieval.

By comparing distance combination methods, we can see a similar tendency in the weight parameters of distance com-

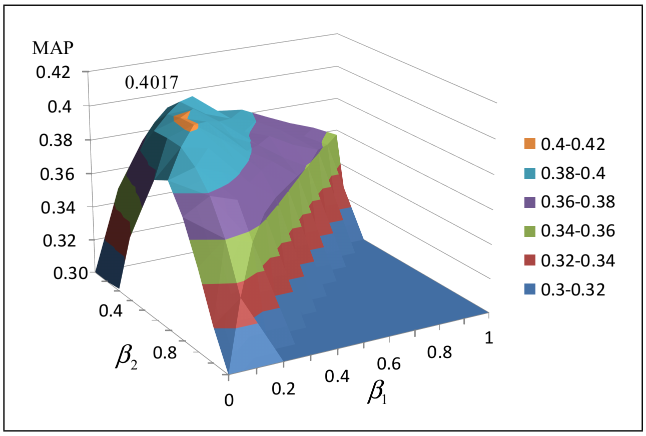

Figure 7: MAP scores of distance combination with different weight parameters ($\alpha = 0.6$)

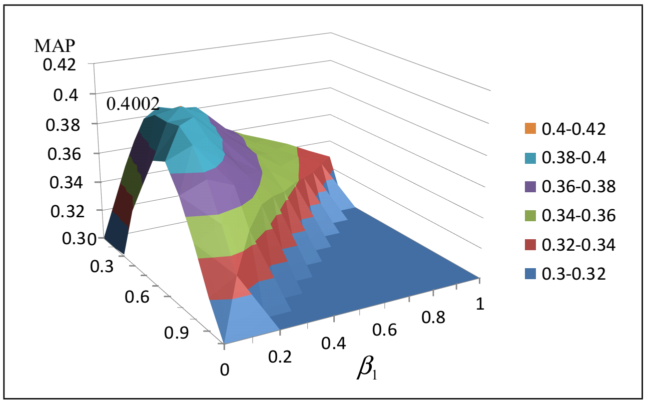

Figure 8: MAP scores of distance combination with different weight parameters (weight parameter of query expansion w/o restriction)

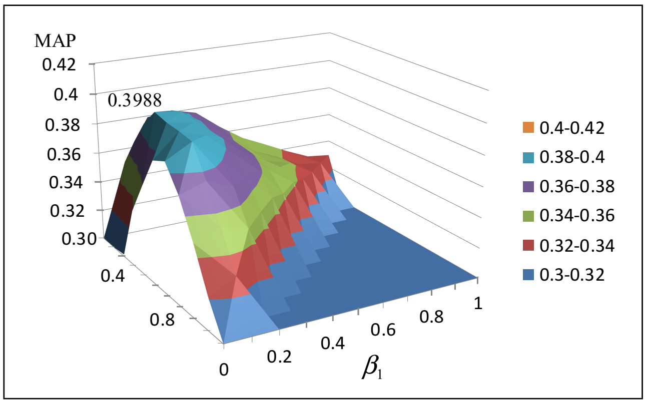

Figure 9: MAP scores of distance combination with different weight parameters (weight parameter of query expansion with restriction)

Table 2: Retrieval results of distance combination

α	β_1	β_2	MAP
0.7	0.2	0.4	0.4017
Proposed with RES	0.1	0.3	0.3988
Proposed w/o RES	0.2	0.3	0.4002

Table 3: Formal run results

Method 1	0.426962
Method 2	0.405018
Method 3	0.406085
Baseline	0.392784

bination (β_1 and β_2) for the three methods. Moreover, the MAP score of the fixed weight parameter of query expansion shown in table 2 is higher than those of the proposed method.

3.3 Formal Run Results

In this section, we describe the spoken document retrieval results of the formal run. The experimental conditions are the same as the dry run. The details of the experimental conditions and the queries are described in the organizer's overview paper[1]. We made a minor improvement in the use of morphological analysis results and thus the optimal weights are slightly different from the previous submission.

In this experiment, we evaluate the following three methods which are described in **3.2.4**.

- Method 1

  For the weight parameter of query expansion, α is set to 0.6. For the weight parameter of distance combination, β_1 and β_2 are 0.5 and 0.2, respectively.

- Method 2

  The weight parameter of query expansion is calculated using the proposed method with parameter restriction ($0.0 \leq \alpha \leq 1.0$). For the weight parameter of distance combination, β_1 and β_2 are 0.4 and 0.1, respectively.

- Method 3

  The weight parameter of query expansion is calculated using the proposed method without parameter restriction ($0.0 \leq \alpha \leq 1.0$) For the weight parameter of distance combination, β_1 and β_2 are 0.4 and 0.1, respectively.

Using these conditions for each method showed the highest MAP scores during the dry run experiment.

Table 3 shows the experimental results of the formal run. In this table, we also show the baseline result which was provided by the NTCIR-9 organizer. From this table, we can see that all methods improve the MAP score over that of the baseline.

4. DISCUSSION

According to various research including ours in this paper, some directions for improvement of spoken document retrieval are suggested.

To tackle the unknown word problem, we have to use not only word-level recognition but also sub-word level recognition. Our results show that syllable recognition methods can improve retrieval performance. More accurate subword modeling, such as bi-syllable modeling or a breakdown of word-level recognition results, are promising approaches. Syllable sequence matching may be an effective way to reduce mis-recognition. Needless to say, the improvement of continuous word recognition accuracy heavily affects the final retrieval results.

Query expansion is also effective and we can easily imagine that the quality and quantity of the expanded query relate to retrieval performance. We used Wikipedia as the source of query expansion in the experiments, but we also conducted a pilot experiment using a general Google search. The search results may contain a lot of garbage, but effective words must occur much more often.

The results using a Google search for the query expansion are shown in Fig.10. This figure can be compared with Fig.3. We obtained an improvement of more than 0.01 of MAP score by using a general Google search over using Wikipedia alone. The best score was with the 17-best documents used for the expansion (much more than with Wikipedia). A Google search can extract a wider variety of documents than Wikipedia, with the result that the query could discover the relevant document more robustly. Combining this with binary-weighted vector space and continuous syllable recognition-based vector space, we obtained a MAP score of 0.4077, which is better than using Wikipedia alone.

We believe that "relevant document modeling" must be important. As mentioned in Section **2.2**, we proposed a new query expansion method using automatic determination of a set of weights for the query vector and the vector corresponding to the documents used for the expansion (we'll call it the "expansion vector" hereafter) for each spoken document. Mathematically speaking, the combination of these vectors has one degree of freedom and we calculate the cosine distance between the combined vector and a document vector. From the geometrical viewpoint, this is equivalent to calculating the cosine of the angle between a document vector and the plane spanned by the query and expansion vectors. We can interpret that we model the set of relevant documents by the plane, which consists of the combined vectors with arbitrary weights for the query vector and expansion vector. This plane has more freedom than the vector defined by a fixed weight, but the freedom is very limited. We then give more freedom to the model. We treat a vector of each word included in the query (that is, a vector whose elements are all set to zero except only one element to one) as a basis and freely combine the bases to make a document subspace, which we consider as a model of the relevant documents. To evaluate a document, we calculate the cosine between the document vector and the vector projected on the space. This is a kind of subspace method in the pattern recognition research fields. As for the expansion vector, the term frequencies of the expansion document are more reliable than those of the query because the expansion documents, extracted from a huge number of web pages, are much larger than the query. So we retain the relative frequencies among the words included in the document except for the words also included in the query. We set the elements of the expansion vector corresponding to the words in the query to zeros and then the vector is normalized to the unit length. This operation makes the vector orthogonal with the vectors of single words and thus we can apply subspace

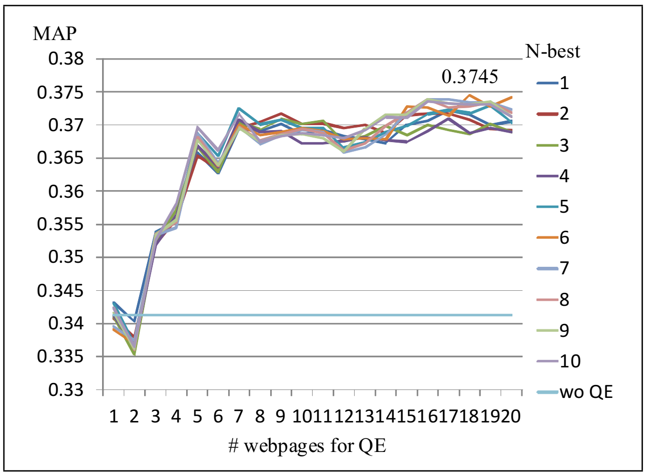

Figure 10: Experimental results of query expansion(QE) using Google search

methods to calculate similarities such as cosine distance. So far, we have not obtained good results (MAP scores of up to only 0.2630), so we will investigate and refine this subspace-based approach to achieve better results.

5. SUMMARY

In this paper, we have proposed a spoken document retrieval method combining query expansion with continuous syllable recognition. The proposed method maps documents to multiple spaces constricted by index terms of composed words and syllables. In addition, at the word level, the proposed method expanded a query vector using text from web pages and calculated an expanded query vector. The proposed method calculates a weight parameter of the query expansion for each document, i.e., the weight parameter of each document is different even if the query is the same. Then, the distances between a query and different documents are calculated on each space, and these distances are combined to achieve spoken document retrieval.

To evaluate the proposed method, a spoken document retrieval experiment was conducted using the SpokenDoc task from the NTCIR-9 meeting. Experimental results showed that the proposed method improved the mean average precision score (MAP) above the baseline provided by the organizer of NTCIR-9 meeting, although the MAP score of the proposed method was slightly lower than methods which used a fixed weight value against the query.

In future work, we plan to investigate and refine the subspace-based approach, described in the Discussion section to achieve better results.

6. REFERENCES

[1] T. Akiba, H. Nishizaki, K. Aikawa, T. Kawahara, and T. Matsui. Overview of the ir for spoken documents task in ntcir-9 workshop. *Proceedings of the 9th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-lingual Information Access*, 2011.

[2] K. Iwata, K. Shinoda, and S. Furui. Robust spoken term detection using combination of phone-based and

word based recognition. *Proc. of Interspeech*, pages 2195–2198, 2008.

[3] M. Saraclar and R. Sproat. Lattice-based search for spoken utterance retrieval. *HLT-NAACL*, pages 129 —–136, 2004.

[4] V. T. Turnen. Reducing the effect of oov query words by using morph-based spoken document retrieval. *Proc. of Interspeech*, pages 2158–2161, 2008.