Extractive Chinese Spoken Document Summarization Using Probabilistic Ranking Models

Yi-Ting Chen[1], Suhan Yu[1], Hsin-min Wang[2], and Berlin Chen[1]
[1]National Taiwan Normal University, Taipei, Taiwan
[2]Academia Sinica, Taipei, Taiwan

g93470070@csie.ntnu.edu.tw

Abstract. The purpose of extractive summarization is to automatically select indicative sentences, passages, or paragraphs from an original document according to a certain target summarization ratio, and then sequence them to form a concise summary. In this paper, in contrast to conventional approaches, our objective is to deal with the extractive summarization problem under a probabilistic modeling framework. We investigate the use of the hidden Markov model (HMM) for spoken document summarization, in which each sentence of a spoken document is treated as an HMM for generating the document, and the sentences are ranked and selected according to their likelihoods. In addition, the relevance model (RM) of each sentence, estimated from a contemporary text collection, is integrated with the HMM model to improve the representation of the sentence model. The experiments were performed on Chinese broadcast news compiled in Taiwan. The proposed approach achieves noticeable performance gains over conventional summarization approaches.

Keywords: hidden Markov model, probabilistic ranking, relevance model, speech recognition, spoken document summarization.

1 Introduction

Due to the ever-increasing storage capability and processing power of computers, vast amounts of multimedia content are now available to the public. Clearly, speech is one of the most important sources of information about multimedia content, such as radio broadcasts, television programs, and lecture recordings, as it provides insight into the content. Therefore, multimedia access based on associated spoken documents has received a great deal of attention in recent years [1]. However, unlike text documents, which are structured with titles and paragraphs and are thus easier to retrieve and browse, associated spoken documents of multimedia content are only presented with video or audio signals; hence, they are difficult to browse from beginning to end. Even though spoken documents are automatically transcribed into words, incorrect information (resulting from recognition errors and inaccurate sentence or paragraph boundaries) and redundant information (generated by disfluencies, fillers, and repetitions) prevent them from being accessed easily. Spoken document summarization, which attempts to distill important information and remove redundant and incorrect content from spoken documents, can help users review spoken documents efficiently and understand associated topics quickly [2].

Although research into automatic summarization of text documents dates back to the early 1950s, for nearly four decades, research work has suffered from a lack of funding. However, the development of the World Wide Web led to a renaissance of the field and summarization was subsequently extended to cover a wider range of tasks, including multi-document, multi-lingual, and multi-media summarization [3]. Generally, summarization can be either extractive or abstractive. Extractive summarization selects indicative sentences, passages, or paragraphs from an original document according to a target summarization ratio and sequences them to form a summary. Abstractive summarization, on the other hand, produces a concise abstract of a certain length that reflects the key concepts of the document. The latter is more difficult to achieve, thus recent research has focused on the former. For example, the vector space model (VSM), which was originally developed for ad-hoc information retrieval (IR), can be used to represent each sentence of a document, or the whole document, in vector form. In this approach, each dimension specifies the weighted statistics associated with an indexing term (or word) in the sentence or document. The sentences with the highest relevance scores (usually calculated as the cosine measure of two vectors) to the whole document are included in the summary. To summarize more important and different concepts in a document, the indexing terms in the sentence with the highest relevance score are removed from the document and the document vector is reconstructed accordingly. Then, based on the new document vector, the next sentence is selected, and so on [4]. The latent semantic analysis (LSA) model for IR can also be used to represent each sentence of a document as a vector in the latent semantic space of the document, which is constructed by performing singular value decomposition (SVD) on the “term-sentence” matrix of the document. The right singular vectors with larger singular values represent the dimensions of the more important latent semantic concepts in the document. Therefore, the sentences with the largest index values in each of the top L right singular vectors are included in the summary [4]. In another example, each sentence in a document, represented as a sequence of terms, is given a significance score, which is evaluated using a weighted combination of statistical and linguistic measures. Sentences are then selected according to their significance scores [5]. In the above cases, if a higher compression ratio is required, the selected sentences can be further condensed by removing some less important terms. A survey of the above extractive summarization approaches and other IR-related tasks in spoken document understanding and organization can be found in [1].

The above approaches can be applied to both text and spoken documents. However, spoken documents present additional difficulties, such as recognition errors, problems with spontaneous speech, and the lack of correct sentence or paragraph boundaries. To avoid redundant or incorrect content when selecting important and correct information, multiple recognition hypotheses, confidence scores, language model scores, and other grammatical knowledge have been utilized [2, 6]. In addition, prosodic features (e.g., intonation, pitch, energy, and pause duration) can be used as important clues for summarization; however, reliable and efficient ways of using these prosodic features are still under active research [7, 8]. Summaries of spoken documents can be presented in either text or speech format. The former has the advantage of easier browsing and further processing, but it is subject to speech recognition errors, as well as the loss of the speaker's emotional/prosodic information, which can only be conveyed by speech signals.

In contrast to conventional approaches, we address the issue of extractive summarization under a probabilistic modeling framework. We investigate the use of the hidden Markov model (HMM) [9] for spoken document summarization, whereby each sentence of a spoken document to be summarized is treated as an HMM for generating the document, and the sentences are ranked and selected according to their likelihoods. In addition, the relevance model (RM) [10, 11] of each sentence, estimated from a contemporary text collection, is integrated with the HMM model for better representation of the sentence model. The experiments were performed on Chinese broadcast news compiled in Taiwan.

The remainder of the paper is organized as follows. Section 2 explains the structural characteristics of the hidden Markov model and the relevance model used in this paper. Section 3 presents the experiment setup and the evaluation metric used for spoken document summarization. The results of a series of summarization experiments are discussed in Section 4. Finally, in Section 5, we present our conclusions.

2 Proposed Summarization Models

2.1 Hidden Markov Model (HMM)

In an ad-hoc IR task, the relevance measure of a query Q and a document D_i can be expressed as $P(D_i \mid Q)$; i.e., the probability that the document D_i is relevant given that the query Q was posed. Based on Bayes' rule and some assumptions, the relevance measure can be approximated by $P(Q \mid D_i)$. That is, in practice, the documents are ranked according to $P(Q \mid D_i)$. Each document D_i can be interpreted as a hidden Markov model (HMM) composed of a mixture of n-gram probability distributions for observing a query Q [9]. Meanwhile, the query Q is considered as observations, expressed as a sequence of indexing terms (or words, or syllables), $Q = w_1 w_2 ... w_j ... w_J$, where w_j is the j-th term in Q and J is the length of the query, as illustrated in Fig. 1. The n-gram distributions for the term w_j, for example the document unigram and bigram models, $P(w_j \mid D_i)$ and $P(w_j \mid w_{j-1}, D_i)$, are estimated directly from the document D_i and linearly interpolated with the collection's unigram and bigram models, $P(w_j \mid C)$ and $P(w_j \mid w_{j-1}, C)$, estimated from a large text collection C. Then, the relevance score of a document D_i and a query Q is calculated by

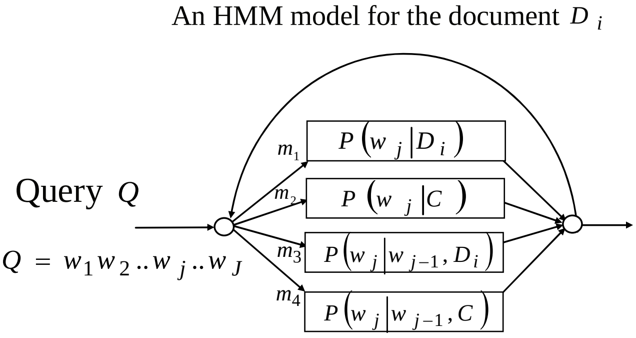

Fig. 1. An illustration of the HMM-based retrieval model

$$\begin{aligned} P(Q|D_i)_{HMM} &= \left[m_1 \cdot P(w_1|D_i) + m_2 \cdot P(w_1|C)\right] \\ &\quad \times \prod_{j=2}^{J} \left[m_1 \cdot P(w_j|D_i) + m_2 \cdot P(w_j|C) + m_3 \cdot P(w_j|w_{j-1}, D_i) + m_4 \cdot P(w_j|w_{j-1}, C)\right], \end{aligned} \tag{1}$$

which can be viewed as a combination of information from a local source (i.e., the document) and a global source (i.e., the large text collection). The unigram and bigram models of the documents and the collection are usually estimated using the maximum likelihood estimation (MLE). The weighting parameters, $m_1,..,m_4$, can be optimized by the expectation-maximization (EM) or minimum classification error (MCE) training algorithms, given a training set of query examples with corresponding query-document relevance information [9].

When the HMM is applied to extractive spoken document summarization, each sentence $S_{i,g}$ of a spoken document D_i is treated as a probabilistic generative model (or HMM) consisting of *n*-gram distributions for predicting the document, and the terms (or words) in the document D_i are taken as an input observation sequence. In this paper, we only investigate unigram modeling for HMM; thus, the HMM model for a sentence can be expressed as:

$$P_{HMM}(D_i \mid S_{i,g}) = \prod_{w_j \in D_i} \left[\lambda \cdot P(w_j \mid S_{i,g}) + (1-\lambda) P(w_j \mid C)\right]^{c(w_j, D_i)}, \tag{2}$$

where λ is a weighting parameter and $c(w_j, D_i)$ is the occurrence count of a term w_j in D_i. In the HMM, the sentence model $P(w_j \mid S_{i,g})$ and the collection model $P(w_j \mid C)$ for each sentence are simply estimated from the sentence itself and a large external text collection, respectively. The weighting parameter λ can be further optimized by taking the document D_i as the training observation sequence and using the following EM training formula:

$$\hat{\lambda} = \frac{\sum_{w_j \in D_i} c(w_j, D_i) \cdot \frac{\lambda \cdot P(w_j \mid S_{i,g})}{\lambda \cdot P(w_j \mid S_{i,g}) + (1-\lambda) \cdot P(w_j \mid C)}}{\sum_{w_l \in D_i} c(w_l, D_i)}. \quad (3)$$

Once the HMM models for the sentences have been estimated, they are used to predict the occurrence probability of the terms in the spoken document. The sentences with the highest probabilities are then selected and sequenced to form the final summary according to different summarization ratios.

2.2 Relevance Model (RM)

In the sentence HMM, as shown in Eq. (2), the sentence model $P(w_j \mid S_{i,g})$ is linearly interpolated with the collection model $P(w_j \mid C)$ to have some probability of generating every term in the vocabulary. However, the true sentence model $P(w_j \mid S_{i,g})$ might not be accurately estimated by MLE, since the sentence only consists of a few terms, and the portions of the terms in the sentence are not the same as the probabilities of those terms in the true model. Therefore, we explore the use of the relevance model (RM) [10, 11], which was originally formulated for IR, to derive a more accurate estimation of the sentence model. In the extractive spoken document summarization task, each sentence $S_{i,g}$ of the document D_i to be summarized has its own associated relevant class $R_{S_{i,g}}$, which is defined as the subset of documents in the collection that are relevant to the sentence $S_{i,g}$. The relevance model of the sentence $S_{i,g}$ is defined as the probability distribution $P(w_j \mid RM_{i,g})$, which gives the probability that we would observe a term w_j if we were to randomly select some document from the relevant class $R_{S_{i,g}}$ and then pick a random term from that document. Once the relevance model of the sentence $S_{i,g}$ has been constructed, it can be used to replace the original sentence model, or it can be combined with the original sentence model to produce a better estimated model. Because there is no prior knowledge about the subset of relevant documents for each sentence $S_{i,g}$, a local feedback-like procedure can be employed by taking $S_{i,g}$ as a query and posing it to the IR system to obtain a ranked list of documents. The top K documents returned by the IR system are assumed to be relevant to $S_{i,g}$, and the relevance model $P(w_j \mid RM_{i,g})$ of $S_{i,g}$ can therefore be constructed by the following equation:

$$P(w_j \mid RM_{i,g}) = \sum_{D_l \in \{\mathbf{D}\}_{\text{Top } K}} P(D_l \mid S_{i,g}) P(w_j \mid D_l), \quad (4)$$

where $\{\mathbf{D}\}_{\text{Top } K}$ is the set of top K retrieved documents; and the probability $P(D_l \mid S_{i,g})$ can be approximated by the following equation using the Bayes' rule:

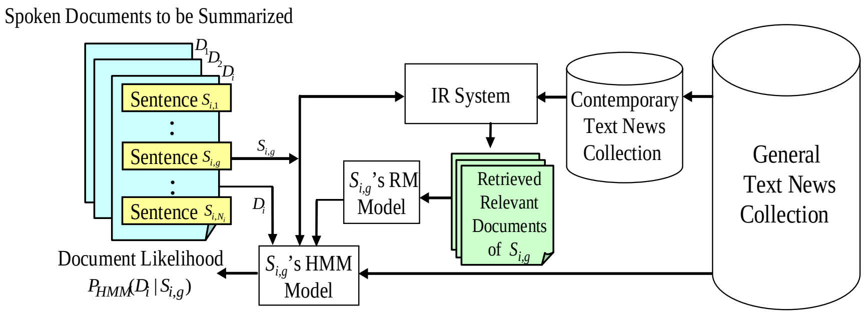
Spoken Documents to be Summarized
D_1
D_2
D_i
Sentence $S_{i,1}$
Sentence $S_{i,g}$
Sentence S_{i,N_i}
$S_{i,g}$
D_i
IR System
Contemporary Text News Collection
General Text News Collection
Retrieved Relevant Documents of $S_{i,g}$
$S_{i,g}$'s RM Model
$S_{i,g}$'s HMM Model
Document Likelihood
$P_{HMM}(D_i \mid S_{i,g})$

Fig. 2. A diagram of spoken document summarization using the HMM and RM models

$$P\left(D_l \mid S_{i,g}\right) \approx \frac{P\left(D_l\right)P\left(S_{i,g} \mid D_l\right)}{\sum_{D_u \in \{\mathbf{D}\}_{\text{Top}\,K}} P\left(D_u\right)P\left(S_{i,g} \mid D_u\right)}. \tag{5}$$

A uniform prior probability $P(D_l)$ can be further assumed for the top K retrieved documents, and the sentence likelihood $P(S_{i,g} \mid D_l)$ can be calculated using an equation similar to Eq. (1) if the IR system is implemented with the HMM retrieval model. Consequently, the relevance model $P(w_j \mid RM_{i,g})$ is combined linearly with the original sentence model $P(w_j \mid S_{i,g})$ to form a more accurate sentence model:

$$\hat{P}\left(w_j \mid S_{i,g}\right) = \alpha \cdot P\left(w_j \mid S_{i,g}\right) + \left(1-\alpha\right) \cdot P\left(w_j \mid RM_{i,g}\right), \tag{6}$$

where α is a weighting parameter. The final sentence HMM is thus expressed as:

$$\hat{P}_{HMM}\left(D_i \mid S_{i,g}\right) = \prod_{w_j \in D_i} \left[\lambda \cdot \hat{P}\left(w_j \mid S_{i,g}\right) + \left(1-\lambda\right)P\left(w_j \mid C\right)\right]^{c\left(w_j, D_i\right)}. \tag{7}$$

Fig. 2 shows a diagram of spoken document summarization using the HMM and RM models.

3 Experiment Setup

3.1 Speech and Text Corpora

The speech data set was comprised of approximately 176 hours of radio and TV broadcast news documents collected from several radio and TV stations in Taipei between 1998 and 2004 [12]. From them, a set of 200 documents (1.6 hours) collected in August 2001, was reserved for the summarization experiments [1]. The remainder of the speech data was used to train an acoustic model for speech recognition, of which about 4.0 hours of data with corresponding orthographic transcripts was used to bootstrap the acoustic model training, while 104.3 hours of the remaining un-transcribed speech data was reserved for unsupervised acoustic model training [13]. The acoustic models were further optimized by the minimum phone error (MPE) training algorithm. A large number of text news documents collected from the Central News Agency (CNA) between 1991 and 2002 (the Chinese Gigaword Corpus released by LDC) was also used. The text news documents collected in 2000 and 2001 were used to train *n*-gram language models for speech recognition; and a subset of about 14,000 text news documents collected in the same period as that of the broadcast news documents to be summarized (August 2001) was used to construct the HMM and RM models.

3.2 Broadcast News Transcription

Front-end processing was performed with the HLDA-based (Heteroscedastic Linear Discriminant Analysis) data-driven Mel-frequency feature extraction approach and further processed by MLLT (Maximum Likelihood Linear Transformation) transformation for feature de-correlation. The speech recognizer was implemented with a left-to-right frame-synchronous Viterbi tree search as well as a lexical prefix tree organization of the lexicon. The recognition hypotheses were organized into a word graph for further language model rescoring. We used a word bigram language model in the tree search procedure and a trigram language model in the word graph rescoring procedure. The Chinese character error rate (CER) for the 200 broadcast news documents reserved for summarization was 14.17%.

3.3 Evaluation Metric

Three subjects were asked to summarize the 200 broadcast news documents, which were to be used as references for evaluation, in two ways:1) to rank the importance of the sentences in the reference transcript of the broadcast news document from the top to the middle; and 2) to write an abstract of the document with a length roughly equal to 25% of the original broadcast news document. Several summarization ratios of the summary length to the total document length [1] were tested. In addition, the ROUGE measure [14, 15] was used to evaluate the performance levels of the proposed models and the conventional models. The measure evaluates the quality of the summarization by counting the number of overlapping units, such as *n*-grams and word sequences, between the automatic summary and a set of reference (or manual) summaries. ROUGE-N is an *n*-gram recall measure defined as follows:

$$ROUGE-N=\frac{\sum\limits_{S\in\mathbf{S}_R}\sum\limits_{gram_n\in S}Count_{match}(gram_n)}{\sum\limits_{S\in\mathbf{S}_R}\sum\limits_{gram_n\in S}Count(gram_n)}, \tag{8}$$

Table 1. The results achieved by the HMM and other summarization models under different summarization ratios.

Summarization Ratio	HMM-1	HMM-2	VSM	LSA-1	LSA-2	SenSig	Random
10%	0.2989	0.2945	0.2845	0.2755	0.2498	0.2760	0.1122
20%	0.3295	0.3052	0.3110	0.2911	0.2917	0.3190	0.1263
30%	0.3670	0.3334	0.3435	0.3081	0.3378	0.3491	0.1834
50%	0.4743	0.4755	0.4565	0.4070	0.4666	0.4804	0.3096

where *N* denotes the length of the *n*-gram; *S* is an individual reference (or manual) summary; $\mathbf{S}_R$ is a set of reference summaries; $Count_{match}(gram_n)$ is the maximum number of *n*-grams co-occurring in the automatic summary and the reference summary; and $Count(gram_n)$ is the number of *n*-grams in the reference summary. In this paper, we adapted the ROUGE-2 measure, which uses word bigrams as matching units.

4 Experiment Results

4.1 Comparison of HMM and Other Summarization Models

The summarization results obtained by the HMM summarization model using word indexing terms (HMM-1) are shown in the second column of Table 1; and the corresponding ROUGE-2 recall rates are approximately 0.30, 0.33, 0.37, and 0.47 for the summarization ratios 10%, 20%, 30%, and 50%, respectively. The summarization results of the HMM summarization model using syllable indexing terms (HMM-2) are shown in the third column of the table; and it is obvious that the results are comparable to that of the HMM summarization model using word indexing terms. In the following experiments, unless specified otherwise, the HMM model corresponds to the HMM summarization model using word indexing terms. In addition, all the other summarization models discussed in this subsection also use word indexing terms.

We compared the HMM model with the conventional VSM [4] and LSA models. Two variants of LSA, namely, the model mentioned in Section 1 [4] (LSA-1) and the model in [6] (LSA-2), were evaluated. For a spoken document, LSA-2 simply evaluates the score of each sentence based on the norm of its vector representation in the lower *L*-dimensional latent semantic space. A fixed number of sentences with relatively large scores are therefore selected to form the summary. In the experiments, we set the value of *L* at 5, the same as that in [6]. The two LSA models were implemented with the MIT SVD Toolkit [16]. We also tried to select indicative sentences from the spoken document based on the sentence significance score (denoted as the SenSig model) [5]. For example, given a sentence

Table 2. The results of combining the HMM and RM models under different summarization ratios; RM was constructed with the IR system using word indexing terms.

Summarization Ratio	M_{doc}=5	M_{doc}=10	M_{doc}=15	M_{doc}=20
10%	0.3074	0.3078	0.3078	0.3078
20%	0.3265	0.3284	0.3260	0.3260
30%	0.3667	0.3650	0.3661	0.3676
50%	0.4759	0.4764	0.4762	0.4768

Table 3. The results of combining the HMM and RM models under different summarization ratios; RM was constructed with the IR system using syllable indexing terms.

Summarization Ratio	M_{doc}=5	M_{doc}=10	M_{doc}=15	M_{doc}=20
10%	0.3057	0.3111	0.3152	0.3152
20%	0.3254	0.3344	0.3341	0.3332
30%	0.3673	0.3659	0.3659	0.3659
50%	0.4782	0.4770	0.4768	0.4759

$S_{i,g} = \{w_1, w_2, ..., w_r, ..., w_{N_{i,g}}\}$ of length $N_{i,g}$, the sentence significance score is expressed by the following formula:

$$Sig\left(S_{i,g}\right) = \sum_{r=1}^{N_{i,g}} \left[\beta_1 \cdot I\left(w_r\right) + \beta_2 \cdot L\left(w_r\right)\right], \quad (9)$$

where $I(w_r)$ is the product of the term frequency (TF) and the inverse document frequency (IDF) of term w_r [17]; $L(w_r)$ is the logarithm of the bigram probability of w_r given its predecessor term w_{r-1} in $S_{i,g}$, which is estimated from a large contemporary text collection; and β_1 and β_2 are tunable weighting parameters. The results for the above models are shown in columns 4 to 7 of Table 1; the results obtained by random selection (Random) are also listed for comparison. We observe that HMM outperforms the VSM, LSA, and SenSig models, which demonstrates that the HMM-based probabilistic ranking model is indeed a good candidate for the extractive spoken document summarization task addressed by this study.

4.2 Combination of HMM and RM

As mentioned in Section 2.2, when the HMM is used for summarization, the sentence model $P(w_j \mid S_{i,g})$ might not be accurately estimated, since each sentence of a spoken document consists of only a few words and the portions of words present in the sentence are not necessarily the same as the probabilities of those words in the true model. Therefore, we combine the RM model $P(w_j \mid RM_{i,g})$ with the sentence

Table 4. The results of combining the HMM and RM models, using syllable indexing terms; the RM model was constructed with the IR system using syllable indexing terms.

Summarization Ratio	M_{doc}=5	M_{doc}=10	M_{doc}=15	M_{doc}=20
10%	0.3190	0.3276	0.3285	0.3285
20%	0.3327	0.3414	0.3439	0.3439
30%	0.3473	0.3544	0.3542	0.3542
50%	0.4735	0.4750	0.4724	0.4724

Table 5. The results of combining the HMM and RM models, using both word and syllable indexing terms; the RM model was constructed with the IR system using syllable indexing terms.

Summarization Ratio	M_{doc}=5	M_{doc}=10	M_{doc}=15	M_{doc}=20
10%	0.3305	0.3285	0.3335	0.3352
20%	0.3411	0.3391	0.3442	0.3468
30%	0.3641	0.3641	0.3612	0.3645
50%	0.4809	0.4816	0.4781	0.4782

model $P(w_j \mid S_{i,g})$ to produce a better estimated sentence model, as expressed in Eq. (6). To construct the RM model, each sentence of the spoken document to be summarized is taken as a query and posed to the IR system to obtain a set of M relevant documents from the contemporary text news collection. We implement the IR system with the HMM retrieval model using either words or syllables as the indexing terms. The results of combining the HMM and RM models are shown in Tables 2 and 3. In Table 2, the IR system uses words as the indexing terms to construct the RM model, while, in Table 3, syllables are adopted as the indexing terms for the IR system. Each column in the tables indicates the number of relevant documents (M_{doc}) returned by the IR system for construction of the RM model.

A number of conclusions can be drawn from the results. First, the combination of HMM and RM boosts the summarization performance when the summarization ratios are low (e.g., 10%), while the gains are almost negligible at higher summarization ratios. Second, the RM model constructed based on the IR system using syllables as indexing terms is better than that based on the IR system using words as indexing terms. One possible reason is that the automatic transcript of a sentence in a broadcast news document often contains speech recognition errors and, in Chinese, syllable accuracy is always higher than word accuracy. Therefore, the IR system that uses syllables as indexing terms might retrieve a set of more relevant documents than the system using single words. Finally, the summarization performance seems to become saturated when the IR system returns 15 relevant documents for construction of the RM model.

4.3 Information Fusion using Word- and Syllable-level Indexing Terms

The summarization results reported in Sections 4.1 and 4.2 were obtained in such a way that the summarization models were only implemented with words as the indexing terms, although the IR system used to construct the RM model can use either words or syllables as indexing terms. Hence, we also implemented the models by using syllables as indexing terms. The summarization results of combining the HMM and RM models and using syllable indexing terms, are shown in Table 4. In this case, the RM model was also constructed with the IR system using syllable indexing terms. Compared with the results in Table 3, the summarization model implemented with syllable indexing terms is considerably better than the one implemented with word indexing terms, especially at lower summarization ratios. Finally, the results derived by combining the HMM and RM models, as well as by using both word and syllable indexing terms, are shown in Table 5. Compared with the results in Table 4, the fusion of these two kinds of indexing information clearly yields additional performance gains. This is because word-level indexing terms contain more semantic information, while syllable-level indexing terms are more robust against errors in speech recognition. Thus, combining these two kinds of indexing terms for the Chinese spoken document summarization task is effective.

5 Conclusions

We have presented an HMM-based probabilistic model for extractive Chinese spoken document summarization. The model's capabilities were verified by comparison with other summarization models. Moreover, the RM model of each sentence of a spoken document to be summarized was integrated with the sentence HMM model for better model estimation. The experiment results are very promising. In our current implementation, the relevant model trained on relevant documents retrieved for a sentence from a contemporary text collection is integrated with the sentence HMM. These relevant documents can be used to train the sentence HMM directly. We believe this is a more effective way to utilize relevant documents.

Acknowledgments. This work was supported in part by the National Science Council, Taiwan, under Grants: NSC95-2221-E-003-014-MY3 and NSC95-2422-H-001-008.

References

1. L.S. Lee and B. Chen, "Spoken Document Understanding and Organization," *IEEE Signal Processing Magazine*, Vol. 22, No. 5 (2005) 42-60
2. S. Furui, T. Kikuchi, Y. Shinnaka and C. Hori, "Speech-to-text and speech-to-speech summarization of spontaneous speech," *IEEE Trans. on Speech and Audio Processing*, Vol. 12, No. 4 (2004) 401-408
3. I. Mani and M. T. Maybury, Eds. Advances in Automatic Text Summarization. Cambridge. MA: MIT Press (1999)

4. Y. Gong and X. Liu, “Generic text summarization using relevance measure and latent semantic analysis,” in *Proc. ACM SIGIR Conference on R&D in Information Retrieval*, (2001) 19-25
5. J. Goldstein et al., “Summarizing text documents: sentence selection and evaluation metrics,” in *Proc. ACM SIGIR Conference on R&D in Information Retrieval* (1999) 121-128
6. M. Hirohata et al., “Sentence Extraction-based Presentation Summarization Techniques and Evaluation Metrics,” in Proc. IEEE International Conference on Acoustics, Speech, and Signal processing (2005) 1065-1068
7. K. Koumpis and S. Renals, “Automatic Summarization of Voicemail Messages Using Lexical and Prosodic Features,” *ACM Transaction. on Speech and Language Processing*, Vol. 2, No.1 (2005) 1-24
8. S. Maskey and J. Hirschberg, “Comparing Lexical, Acoustic/Prosodic, Structural and Discourse Features for Speech Summarization,” in *Proc. European Conference on Speech Communication and Technology* (2005) 621-624
9. B. Chen, H.M. Wang, L.S. Lee, “A Discriminative HMM/N-Gram-Based Retrieval Approach for Mandarin Spoken Documents,” *ACM Transactions on Asian Language Information Processing*, Vol. 3, No. 2 (2004) 128-145
10. Croft, W.B., Lafferty, J. (Eds.). Language Modeling for Information Retrieval. Kluwer-Academic Publishers (2003)
11. M. D. Smucker et al., “Dirichlet Mixtures for Query Estimation in Information Retrieval,” CIIR Technical Report, Center for Intelligent Information Retrieval, University of Massachusetts (2005)
12. B. Chen et al., “Speech Retrieval of Mandarin Broadcast News via Mobile Devices,” in *Proc. European Conference on Speech Communication and Technology* (2005) 109-112
13. B. Chen et al., “Lightly Supervised and Data-Driven Approaches to Mandarin Broadcast News Transcription,” in *Proc. IEEE International Conference on Acoustics, Speech, and Signal processing* (2004) 777-780
14. C.Y. Lin, “ROUGE: Recall-Oriented Understudy for Gisting Evaluation,” (2003) Available from: <http://www.isi.edu/~cyl/ROUGE/>
15. C.-Y. Lin, “Looking for a few good metrics: ROUGE and its evaluation,” Working Notes of NTCIR-4 (Vol. Supl. 2) (2004) 1-8
16. D. Rohde. Doug Rohde's SVD C Library, Version 1.34 (2005) Available from: <http://tedlab.mit.edu:16080/~dr/SVDLIBC/>
17. R. Baeza-Yates and B. Ribeiro-Neto. Modern Information Retrieval. Addison-Wesley (1999)