Enhanced Language Modeling for Extractive Speech Summarization with Sentence Relatedness Information

Shih-Hung Liu[1, 2], *Kuan-Yu Chen*[1, 2], *Yu-Lun Hsieh*[1], *Berlin Chen*[3], *Hsin-Min Wang*[1], *Hsu-Chun Yen*[2], *Wen-Lian Hsu*[1]

[1]Institute of Information Science, Academia Sinica, Taiwan
[2]National Taiwan University, Taiwan
[3]National Taiwan Normal University, Taiwan

E-mail: [1]{journey, kychen, morphe, whm, hsu}@iis.sinica.edu.tw, [2] yen@cc.ee.ntu.edu.tw, [3] berlin@ntnu.edu.tw

Abstract

Extractive summarization is intended to automatically select a set of representative sentences from a text or spoken document that can concisely express the most important topics of the document. Language modeling (LM) has been proven to be a promising framework for performing extractive summarization in an unsupervised manner. However, there remain two fundamental challenges facing existing LM-based methods. One is how to construct sentence models involved in the LM framework more accurately without resorting to external information sources. The other is how to additionally take into account the sentence-level structural relationships embedded in a document for important sentence selection. To address these two challenges, in this paper we explore a novel approach that generates overlapped clusters to extract sentence relatedness information from the document to be summarized, which can be used not only to enhance the estimation of various sentence models but also to allow for the sentence-level structural relationships for better summarization performance. Further, the utilities of our proposed methods and several state-of-the-art unsupervised methods are analyzed and compared extensively. A series of experiments conducted on a Mandarin broadcast news summarization task demonstrate the effectiveness and viability of our method.

Index Terms: speech summarization, language modeling, clustering, relevance, sentence relatedness

1. Introduction

Due in large part to the advances in automatic speech recognition and the popularity as well as ubiquity of multimedia associated with spoken documents [1, 2], research on speech summarization have attracted increasing interest in the speech processing community over the past decade [3-6]. Extractive speech summarization aims at producing a concise summary by selecting salient sentences or paragraphs from an original spoken document according to a predefined target summarization ratio. By doing so, it can indicate the important speech segments with their corresponding transcripts in the document for users to listen to and digest. The various extractive summarization methods that have been developed so far may roughly fall into three major categories [5-7]: 1) methods simply based on sentence structural or locational cues, 2) methods based on unsupervised statistical measures, and 3) methods based on supervised sentence classification.

For the first category, the important sentences can be selected from some salient parts of a spoken document [8]. As an illustration, sentences can be selected from the introductory and/or concluding parts. However, such methods can be only applied to some limited domains or document structures. On the other hand, extractive text or speech summarization using unsupervised statistical measures attempts to select salient sentences on top of some statistical features of sentences, or of the words in the sentences, in an unsupervised manner. Statistical features derived for each sentence, can be word frequency, linguistic score calculated from lexical and/or semantic cues, recognition confidence, similarity (proximity) measure and prosodic information, and so forth. The associated unsupervised summarization methods based on these features has garnered much research and may be further grouped into three subcategories: 1) the vector-based methods, which include, but are not limited to, the vector space model (VSM) [9], the latent semantic analysis (LSA) [9], and the maximum marginal relevance (MMR) method [10]; 2) the graph-based methods, which include, among others, the Markov random walk (MRW) [11], the LexRank [12], and the minimum dominating set algorithm [13]; 3) the combinatorial optimization-based methods, which include the submodularity-based method (Submodularity) [14] and the integer linear programming (ILP) method [15], to name just a few. Aside from that, a number of supervised classification-based methods using various kinds of indicative features also have been developed, such as the Gaussian mixture models (GMM) [16], the Bayesian classifier (BC) [17], the support vector machine (SVM) [18] and the conditional random fields (CRFs) [19]. In these methods, important sentence selection is usually formulated as a binary classification problem. A sentence can either be included in a summary or not. These classification-based methods need a set of training documents along with their corresponding handcrafted summaries (or labeled data) for training the classifiers (or summarizers). However, manual annotation is both time-consuming and labor-intensive. Even if the performance of unsupervised summarizers is not always comparable to that of supervised summarizers, their easy-to-implement and portable property still makes them attractive. Interested readers may also refer to [5-7, 20] for comprehensive and enjoyable discussions of major methods that have been successfully developed and applied to a wide variety of text and speech summarization tasks.

Orthogonal to the aforementioned summarization methods, a more recent line of research is devoted to capitalizing on the statistical language modeling (LM) framework in an unsupervised manner for important sentence selection [21-24], which has been applied to extractive speech summarization with preliminary success. However, there are still two fundamental challenges facing the existing LM-based methods. One is how to estimate sentence models involved in the LM framework more accurately without resorting to external information sources. The other is how to additionally take into account the sentence-level structural relationships embedded in a document to be summarized for important sentence selection. To address these two challenges, in this paper we explore a novel approach that generates overlapped clusters to

extract sentence relatedness information cues from the document to be summarized. Such information cues can be used not only to enhance the estimation of sentence models but also to allow for the sentence-level structural relationships for better summarization performance. In addition, the performance merits of our proposed methods and several widely-used unsupervised methods are analyzed and compared extensively.

2. LM Framework for Summarization

Intuitively, a speech summarization task can be framed as an ad hoc information retrieval problem [25], where the spoken document is treated as a query (information need) and each sentence of the document is regarded as a candidate information unit to be retrieved according to its degree of relevance (or importance) to the query. This way, the ultimate goal of extractive speech summarization can be stated as the selection of the most representative sentences that can succinctly describe the main theme of the spoken document. Over the years, the language modeling framework has been introduced to a wide spectrum of information retrieval tasks and demonstrated with good empirical success [25]; this modeling paradigm has been applied to speech summarization with some success recently [21-24].

2.1. Document-likelihood Measure (DLM)

When adopting the notion of language modeling for extractive speech summarization, a principal realization is to use a probabilistic generative paradigm for ranking each sentence S of a spoken document D to be summarized, which can be expressed by $P(S|D)$. Instead of calculating this probability directly, we can apply the Bayes' rule and rewrite it as follows:

$$P(S \mid D) = \frac{P(D \mid S)P(S)}{P(D)}, \quad (1)$$

where $P(D|S)$ is the sentence generative probability, i.e., the likelihood of D being generated by S, $P(S)$ is the prior probability of the sentence S being relevant, and $P(D)$ is the prior probability of the document D. $P(D)$ in Eq. (1) can be eliminated because it is identical for all sentences and will not affect the ranking of the sentences. Furthermore, $P(S)$ is a prior on sentences and is often assumed to be uniform without any additional prior knowledge about the sentences. As such, the sentences of a spoken document to be summarized can be ranked by means of the probability $P(D|S)$ instead of using the probability $P(S|D)$: the higher the probability $P(D|S)$, the more representative S is likely to be for D. If the document is treated as a sequence of words, where words are assumed to be conditionally independent given the sentence and their order is also assumed to be of no importance (i.e., the so-called "*bag-of-words*" assumption), then $P(D|S)$ can be approximated by:

$$P(D \mid S) \approx \prod_{w \in D} P(w \mid S)^{c(w,D)}, \quad (2)$$

where $c(w,D)$ is the occurrence count of a specific type of word (or term) w in D, reflecting that w will contribute more in the calculation of $P(D|S)$ if it occurs more frequently in D. The simplest way is to estimate the sentence model $P(w|S)$ on the basis of the frequency of words occurring in the sentence, with the maximum likelihood (ML) estimation [21]. The sentence model $P(w|S)$ can be further smoothed by a background unigram language model estimated from a large general collection to model the general properties of the language as well as to avoid the problem of zero probability. In what follow, we will term Eq. (2) the document-likelihood measure (denoted by DLM for short).

2.2. Enhanced Sentence Modeling

Due to that each sentence S of a spoken document D to be summarized usually consists of only a few words, the corresponding sentence model $P(w|S)$ might not be appropriately estimated by the ML estimation. With the alleviation of this deficiency as motivation, in this paper we adopt and formalize two effective query modeling methods [26, 27, 28, 29] that have been extensively studied in the information retrieval (IR) community to enhance sentence modeling. The commonality among these two methods is that each sentence S is regarded as a query and be posted to an IR system to retrieve a set of top ranked text or spoken documents from an external collection, counted as exemplars of pseudo-relevant documents, to be used for reformulating the sentence model (or assigning more accurate probability masses to words in the sentence) correspondingly.

2.2.1. Relevance Model (RM)

We may assume that each sentence S of a spoken document D to be summarized is associated with an unknown relevance class R_S and words that are relevant to the semantic content expressed in S are samples drawn from R_S [26]. However, since there is no prior knowledge about R_S in practice, a pseudo-relevance feedback (PRF) procedure may be performed, which takes each sentence S as a query and poses it to an IR system to retrieve a set of top-ranked documents $\mathbf{D}_S$ from an external collection to approximate the relevance class R_S. The corresponding relevance model (RM), on the grounds of a multinomial view of R_S, can be estimated using the following equation:

$$P_{\mathrm{RM}}(w \mid S) = \frac{\sum_{D_i \in \mathbf{D}_S} P(D_i) P(w \mid D_i) \prod_{w' \in S} P(w' \mid D_i)}{\sum_{D_j \in \mathbf{D}_S} P(D_j) \prod_{w' \in S} P(w' \mid D_j)}, \quad (3)$$

where the document probability $P(D_i)$ can be determined in accordance with the relevance of D_i to S (or simply kept uniform), while $P(w|D_i)$ is estimated based on the occurrence counts of w in D_i, with the ML estimation. The resulting relevance model $P_{\mathrm{RM}}(w|S)$ can be linearly combined with, or used to replace the original sentence model $P(w|S)$.

2.2.2. Simple Mixture Model (SMM)

In this paper, we also exploit an alternative formulation to extract relevance cues from PRF for sentence modeling in extractive speech summarization, which is referred to hereafter as the simple mixture model (SMM). The basic idea of SMM is to assume that the set of top-ranked documents $\mathbf{D}_S$ are relevant and the resulting unigram model $P_{\mathrm{SMM}}(w|S)$ estimated from these documents can potentially benefit sentence modeling. Specifically, SMM assumes that words in the set of pseudo-feedback documents are drawn from a two-component mixture model [27]: 1) One component is the SMM model $P_{\mathrm{SMM}}(w|S)$ and 2) the other is a background model $P(w|BG)$, which is set to be the unigram language model estimated from a large general collection (*cf.* Section 2.1). The SMM model $P_{\mathrm{SMM}}(w|S)$ is estimated by maximizing the log-likelihood of the set of top-ranked documents $\mathbf{D}_s$ expressed as follows:

$$LL_{\mathbf{D}_S} = \sum_{D_r \in \mathbf{D}_S} \sum_{w \in V} c(w, D_r) \cdot \log[(1-\alpha) \cdot P_{\mathrm{SMM}}(w \mid S) + \alpha \cdot P(w \mid BG)], \quad (4)$$

where α is a mixing parameter used to control the degree of reliance between $P_{\mathrm{SMM}}(w|S)$ and $P(w|BG)$, and V indicates the vocabulary size. The maximization of Eq. (4) can be conducted iteratively via the expectation-maximization (EM) update process [30].

This estimation of SMM will enable more specific words (i.e., words in $\mathbf{D}_S$ that are not well-explained by the background model) to receive more probability mass, thereby leading to a more discriminative sentence model $P_{\mathrm{SMM}}(w|S)$. Phrased another way, the SMM model $P_{\mathrm{SMM}}(w|S)$ is anticipated to extract useful word usage cues from $\mathbf{D}_S$, which are not only relevant to the sentence S but also external to those already captured by the background model.

3. Sentence Relatedness Information

As previously elaborated in Section 2, the various LM-based speech summarization methods focus exclusively on estimating the degree of relevance between a candidate summary sentence S and the spoken document D to be summarized. However, we believe that there remain some other useful cues that can aid in the process of important sentence selection. Furthermore, the success of RM and SMM methods comes at the expense of recourse to an external document collection for model estimation, which would be time-consuming and not always readily available. In light of these two deficiencies, in this paper we explore a novel clustering approach that generates overlapped sentence clusters to extract sentence relatedness information from the document to be summarized. Such extracted information can be directly used for determining the likelihood of each sentence being important (i.e., $P(S)$ in Eq. (1)), or/and used for resampling the most related sentences of each sentence in the document, which as a whole offer a good surrogate of the external document collection for constructing RM and SMM.

For the idea to work, the well-known k-nearest neighbors (k-NN) algorithm is employed to cluster all the sentences of a document to be summarized. Each sentence will play a central role in its own cluster with its k closest neighboring sentences in terms of the cosine similarity measure. Further, it is anticipated that a dominant (important) sentence that can cover the major topics of the document would have several neighboring sentences with high similarities and participates in several clusters as well. By doing so, the prior probability of each sentence S being important can be estimated (approximated) by

$$P(S) \approx \frac{|OverClu|_S \cdot P(D \mid Clu_S)}{\sum_{S'} |OverClu|_{S'} \cdot P(D \mid Clu_{S'})}, \quad (5)$$

where $|OverClu|_S$ is the number of distinct clusters that sentence S participates in, and $P(D|Clu_S)$ is the likelihood of the document D generated by the own cluster of sentence S. Note also that Clu_S is formed by concatenating those sentences that belong to it, and its corresponding cluster-based language model is obtained with the ML estimation (*cf.* Section 2.1). Once $P(S)$ has been properly estimated, the sentences of the spoken document to be summarized can be ranked by the product of $P(D|S)$ and $P(S)$, as previously illustrated in Eq. (1).

Alternatively, the derived sentence clusters can be used in place of the external document collection for the estimation of the RM- and SMM-based sentence models. The process proceeds as follows. For each sentence S of the document to be summarized, we rank all the derived clusters by calculating the likelihoods of their respective cluster-based language models to generate the sentence S. The sentences that participate in the top-ranked M clusters (except for S itself) are in turn selected as pseudo-relevant feedback sentences for use in constructing the RM (or SMM) model of S. It is worth pointing out that dominant sentences would be selected multiple times (or be emphasized) in this way. Since only the internal sentence relatedness (clustering) information of the document to be summarized is used here for estimating RM and SMM, we therefore term the resulting sentence models iRM and iSMM henceforth.

The notion of leveraging cluster-based information cues has recently attracted much attention and been integrated with success into many LM-based information retrieval models [31-33]. However, as far as we are aware, this notion has never been extensively explored for sentence modeling in extractive speech summarization.

4. Experimental Setup

The summarization dataset employed in this study is a Mandarin broadcast news (MATBN) corpus collected by the Academia Sinica and the Public Television Service Foundation of Taiwan between November 2001 and April 2003 [34]. Each story contains the speech of one studio anchor, as well as several field reporters and interviewees. A subset of 205 broadcast news documents compiled between November 2001 and August 2002 was reserved for the summarization experiments. Since broadcast news stories often follow a relatively regular structure as compared to other speech materials like conversations, the positional information would play an important role in extractive summarization of broadcast news stories; we, hence, chose 20 documents for which the generation of reference summaries is less correlated with the positional information (or the position of sentences) as the held-out test set to evaluate the general performance of the proposed summarization framework, and 100 documents as the development set. An external set of about 100,000 text news documents, compiled during the same period as the broadcast news documents to be summarized, was employed to estimate the related (component) models of the various summarization methods compared in this paper. Three subjects were asked to create summaries of the 205 spoken documents for the summarization experiments as references (the gold standard) for evaluation. For the assessment of summarization performance, we adopted the widely-used ROUGE metrics [35]. Three variants of the ROUGE metrics were used to quantify the utility of the proposed methods. They are, respectively, the ROUGE-1 (denoted by R-1: unigram) metric, the ROUGE-2 (denoted by R-2: bigram) metric and the ROUGE-L (denoted by R-L: longest common subsequence) metric. All the experimental results reported hereafter were obtained by calculating the F-scores of these three ROUGE metrics. In addition, the summarization ratio, defined as the ratio of the number of words in the automatic (or manual) summary to that in the reference transcript of a spoken document, was set to 10% in this research.

5. Experiments

To begin, we assess the performance level of the baseline DLM method for extractive speech summarization, by

comparing it with several well-practiced unsupervised summarization methods, including VSM, MRW, LexRank, Submodularity, ILP and MMR. The corresponding summarization results of these unsupervised methods are shown in Table 1, where TD denotes the results obtained based on the manual transcripts of spoken documents and SD denotes the results using the speech recognition transcripts that may contain speech recognition errors. The average Chinese character error rate (CER) obtained for the spoken documents was about 35%. Several noteworthy observations can be drawn from Table 1. First, DLM can match the performance of MRW, LexRank and Submodularity (all of the three latter methods belong to the graph-based methods) and exceed that of VSM for both the TD and SD cases, confirming the applicability of the LM-based summarization framework. Second, ILP appears to be the best-performing one among all the methods compared here. However, the superiority of ILP seems to diminish for the SD case, probably due to the effect of speech recognition errors. Third, integrating the minimum redundancy criterion of MMR into VSM and DLM (denoted by VSM-MMR and DLM-MMR, respectively) can further boost their summarization performance. Lastly, there is a sizable gap between the TD and SD cases, indicating room for further improvements. We may seek remedies, such as robust indexing techniques, to compensate for imperfect speech recognition [36, 37].

In the next set of experiments, we turn to evaluate the effect of using internal sentence relatedness (clustering) information, instead of relevance information gleaned from an external document collection, on the estimation of the RM- and SMM-based sentence models for improved summarization performance. It is evident from Table 2 that using internal sentence relatedness information to estimate the sentence models (viz. iRM and iSMM) works almost on par with that using relevance information gleaned from an external document collection (viz. RM and SMM) for the SD case, but obtains a lower performance level for the TD case. An obvious advantage of the former (viz. iRM and iSMM) is that the summarization process can be conducted more efficiently, without the perennial need of resorting to an IR system to retrieve relevant documents from an external collection. However, there is good reason to combine these two kinds of information sources (i.e., to linearly combine their models) for better summarization performance. As can be also seen from Table 2 (*cf.* RM+iRM and SMM+iSMM), these two kinds of information sources are indeed complementary to each other and their combination can bring slight but consistent gains.

In the last set of experiments, we continue to examine the merits of using the internal sentence relatedness information alternatively to estimate the sentence prior probability $P(S)$, which, in turn, can work in conjunction with the sentence generative probability $P(D|S)$ for important sentence selection (*cf.* Sections 2 and 3). It should be borne in mind that the sentence generative probability can be computed with the various sentence models mentioned above (viz. DLM, RM, iRM, RM+iRM, SMM, iSMM and SMM+iSMM). Consulting the results shown in Table 3, we find that with an appropriate use of the sentence prior probability, the summarization effectiveness of the various LM-based methods can indeed be significantly promoted for both the TD and SD cases, which corroborates the advantage of modeling the prior knowledge of sentence importance in the LM-based summarization framework. We thus argue that additional incorporation of more other prosodic or linguistic cues to infer the prior probability of a sentence being important, in tandem with the various sentence models, would further benefit extractive speech summarization [38]; this is left as future work.

Table 1: *Summarization results achieved by the baseline DLM and other several widely-used unsupervised methods.*

	TD			SD		
	R-1	R-2	R-L	R-1	R-2	R-L
VSM	34.7	22.8	29.0	34.2	18.9	28.7
DLM	41.1	29.8	36.1	36.4	21.0	30.7
MRW	41.2	28.2	35.8	33.2	19.1	29.1
LexRank	41.3	30.9	36.3	30.5	14.6	25.4
Submodularity	41.4	28.6	36.3	33.2	20.4	30.3
ILP	44.2	33.7	40.1	34.8	20.9	30.6
VSM-MMR	36.8	24.8	32.2	36.6	21.5	31.5
DLM-MMR	42.2	32.3	38.0	37.4	22.2	32.5

Table 2: *Summarization results achieved by the various sentence modeling formulations.*

	TD			SD		
	R-1	R-2	R-L	R-1	R-2	R-L
RM	45.3	33.5	40.3	38.2	23.9	33.1
iRM	42.9	30.6	37.4	37.5	22.7	32.3
RM+iRM	45.5	33.8	40.7	38.3	24.1	33.4
SMM	43.9	32.0	38.8	38.3	22.9	32.7
iSMM	42.4	32.9	38.0	37.7	23.2	32.6
SMM+iSMM	44.3	33.3	39.2	38.4	24.0	33.2

Table 3: *Summarization results achieved by additionally considering the sentence prior probability.*

	TD			SD		
	R-1	R-2	R-L	R-1	R-2	R-L
DLM+Prior	44.7	32.9	39.3	38.4	23.5	33.3
RM+Prior	46.5	35.2	41.9	39.2	23.4	33.5
iRM+Prior	45.2	34.6	39.9	38.7	23.1	33.3
RM+iRM+Prior	47.1	35.7	42.1	39.5	25.2	34.6
SMM+Prior	45.8	34.6	40.3	38.8	24.0	32.6
iSMM+Prior	45.3	34.6	39.8	38.5	23.5	32.2
SMM+iSMM+Prior	46.5	35.6	41.3	39.4	26.0	34.3

6. Conclusion and Outlook

In this paper, we have presented an enhanced language modeling (LM) framework for extractive speech summarization that can leverage additional information cues extracted from overlapped sentence clusters for sentence modeling. Experimental evidence supports that the various LM-based methods instantiated from our framework are quite comparable to a few existing state-of-the-art unsupervised summarization methods. In the future, we plan to adopt and formalize the proposed LM methods for other related tasks such as large vocabulary continuous speech recognition and spoken document retrieval.

7. Acknowledgement

This research is supported in part by the "Aim for the Top University Project" of National Taiwan Normal University (NTNU), sponsored by the Ministry of Education, Taiwan, and by the Ministry of Science and Technology, Taiwan, under Grants NSC 101-2221-E-003-024-MY3, NSC 102-2221-E-003-014-, NSC 101-2511-S-003-057-MY3, NSC 101-2511-S-003-047-MY3 and NSC 103-2911-I-003-301.

8. References

[1] Furui, S., Deng, L., Gales, M., Ney, H. and Tokuda, K., "Fundamental technologies in modern speech recognition," IEEE Signal Processing Magazine, 29(6): 16–17, 2012

[2] O'Shaughnessy, D., Deng, L. and Li, H., "Speech information processing: Theory and applications," Proceedings of the IEEE, 101(5): 1034–1037, 2013.

[3] Furui , S., Kikuchi, T., Shinnaka, Y. and Hori, C., "Speech-to-text and speech-to-speech summarization of spontaneous speech," IEEE Trans. Speech and Audio Proc., 12(4): 401–408, 2004.

[4] McKeown, K., Hirschberg, J., Galley, M. and Maskey, S., "From text to speech summarization," in ICASSP, pp. 997–1000, 2005.

[5] Liu, Y. and Hakkani-Tur, D., "Speech summarization," in G. Tur and R. D. Mori [Ed], Spoken Language Understanding: Systems for Extracting Semantic Information from Speech, Wiley, 2011.

[6] Nenkova, A. and McKeown, K., "Automatic summarization," Foundations and Trends in Information Retrieval, 5(2–3): 103–233, 2011.

[7] Mani, I. and Maybury, M.T. [Ed], Advances in automatic text summarization, Cambridge, MIT Press, 1999.

[8] Baxendale, P. B., "Machine-made index for technical literature-an experiment," IBM Journal, October 1958.

[9] Gong, Y. and Liu, X., "Generic text summarization using relevance measure and latent semantic analysis," in Proc. ACM SIGIR, pp. 19–25, 2001.

[10] Carbonell, J. and Goldstein, J., "The use of MMR, diversity based reranking for reordering documents and producing summaries," in Proc. SIGIR, 335–336, 1998.

[11] Wan, X. and Yang, J., "Multi-document summarization using cluster-based link analysis," in Proc. SIGIR, pp. 299–306, 2008.

[12] Erkan, G. and Radev, D.R., "LexRank: Graph-based lexical centrality as salience in text summarization", Journal of Artificial Intelligent Research, 22(1): 457–479, 2004.

[13] Shen, C. and Li, Tao, "Multi-document summarization via the minimum dominating set," in Proc. COLING, pp. 984-992, 2010.

[14] Lin, H. and Bilmes, J., "Multi-document summarization via budgeted maximization of submodular functions," in Proc. NAACL HLT, pp. 912–920, 2010.

[15] Riedhammer, K., Favre, B. and Hakkani-Tür, D. Z., "Long story short - Global unsupervised models for keyphrase based meeting summarization," Speech Communication, 52(10): 801–815, 2010.

[16] Fattah, M. A. and Ren, F., "GA, MR, FFNN, PNN and GMM based models for automatic text summarization," Computer Speech & Language, 23(1):126–144, 2009.

[17] Kupiec, J., Pedersen, J. and Chen, F., "A trainable document summarizer," in Proc. ACM SIGIR, pp. 68–73, 1995.

[18] Kolcz, A., Prabakarmurthi, V. and Kalita, J., "Summarization as feature selection for text categorization," in Proc. CIKM pp. 365–370, 2001.

[19] Galley, M., "Skip-chain conditional random field for ranking meeting utterances by importance," in Proc. EMNLP, pp. 364–372, 2006.

[20] Penn, G. and Zhu, X., "A critical reassessment of evaluation baselines for speech summarization," in Proc. ACL, pp. 470–478, 2008.

[21] Chen, Y. T., Chen, B. and Wang, H. M., "A Probabilistic Generative Framework for Extractive Broadcast News Speech Summarization," IEEE Trans. Audio, Speech and Language Proc., 17(1):95–106, 2009

[22] Lin, S. H., Yeh, Y. M. and Chen, B., "Leveraging Kullback-Leibler divergence measures and information-rich cues for speech summarization," IEEE Trans. Audio, Speech and Language Proc., 19(4): 871–882, 2011.

[23] Celikyilmaz, A. and Hakkani-Tur, D., "A hybrid hierarchical model for multi-document summarization," in Proc. ACL, pp. 815–824, 2010.

[24] Chen, B., Chang, H. C. and Chen, K. Y., "Sentence modeling for extractive speech summarization," in Proc. ICME, pp. 1–6, 2013.

[25] Baeza-Yates, R. and Ribeiro-Neto, B., Modern Information Retrieval: The Concepts and Technology behind Search, ACM Press, 2011.

[26] Lavrenko, V. and Croft, W. B., "Relevance-based language models," in Proc. SIGIR, pp. 120–127, 2001.

[27] Zhai, C. X. and Lafferty, J., "Model-based feedback in the language modeling approach to information retrieval," in Proc. CIKM, pp. 403–410, 2001.

[28] Chen, B., Chen, K. Y., Chen, P. N. and Chen, Y.W., "Spoken document retrieval with unsupervised query modeling techniques," IEEE Transactions on Audio, Speech and Language Processing, 20(9): 2602–2612, 2012.

[29] Chen, B., Chen, Y. W., Chen, K. Y., Wang, H. M. and Yu, K. T., "Enhancing query formulation for spoken document retrieval," Journal of Information Science and Engineering, 30(3): 553–569, 2014.

[30] Dempster, A. P., Laird, N. M. and Rubin, D. B., "Maximum likelihood from incomplete data via the EM algorithm," Journal of Royal Statistical Society B, 39(1):1–38, 1977.

[31] Liu, X. and Croft, W. B., "Cluster-based retrieval using language models," in Proc. SIGIR, pp. 186–193, 2004.

[32] Huang, Y., Sun, L., Nie, J. Y., "Smoothing document language model with local word graph," in Proc. CIKM, pp. 1943–1946, 2009.

[33] Lee, K. S. and Croft, W. B., "A deterministic resampling method using overlapping document clusters for pseudo-relevance feedback," Information Processing & Management, 49(4): 792–806, 2013.

[34] Wang, H. M., Chen, B., Kuo, J. W. and Cheng, S. S., "MATBN: A Mandarin Chinese broadcast news corpus," International Journal of Computational Linguistics and Chinese Language Processing, 10(2): 219–236, 2005.

[35] Lin, C. Y., "ROUGE: Recall-oriented Understudy for Gisting Evaluation," 2003. Available: http://haydn.isi.edu/ROUGE/.

[36] Xie, S. and Liu, Y., "Using N-best lists and confusion networks for meeting summarization" IEEE Trans. Audio, Speech and Language Proc., 19(5):1160–1169, 2011.

[37] Chelba, C., Silva, J. and Acero, A., "Soft indexing of speech content for search in spoken documents," Computer Speech & Language, 21(3): 458–478, 2007.

[38] Chen, B. and Lin, S. H., "A risk-aware modeling framework for speech summarization," IEEE Trans. Audio, Speech and Language Proc., 20(1): 211–222, 2012.