EFFECTIVE PSEUDO-RELEVANCE FEEDBACK FOR LANGUAGE MODELING IN EXTRACTIVE SPEECH SUMMARIZATION

Shih-Hung Liu[*†], *Kuan-Yu Chen*[*†], *Yu-Lun Hsieh*[*], *Berlin Chen*[#],
Hsin-Min Wang[*], *Hsu-Chun Yen*[†], *Wen-Lian Hsu*[*]
[†]National Taiwan University, Taiwan
[#]National Taiwan Normal University, Taiwan
[*]Institute of Information Science, Academia Sinica, Taiwan
E-mail: [*]{journey, kychen, morphe,whm, hsu}@iis.sinica.edu.tw, [#] berlin@ntnu.edu.tw, [†]yen@cc.ee.ntu.edu.tw

ABSTRACT

Extractive speech summarization, aiming to automatically select an indicative set of sentences from a spoken document so as to concisely represent the most important aspects of the document, has become an active area for research and experimentation. An emerging stream of work is to employ the language modeling (LM) framework along with the Kullback-Leibler divergence measure for extractive speech summarization, which can perform important sentence selection in an unsupervised manner and has shown preliminary success. This paper presents a continuation of such a general line of research and its main contribution is two-fold. First, by virtue of pseudo-relevance feedback, we explore several effective sentence modeling formulations to enhance the sentence models involved in the LM-based summarization framework. Second, the utilities of our summarization methods and several widely-used methods are analyzed and compared extensively, which demonstrates the effectiveness of our methods.

Index Terms—Speech summarization, language modeling, Kullback-Leibler divergence, pseudo-relevance feedback

1. INTRODUCTION

Research on speech summarization has witnessed a booming interest in the speech processing community over the past decade [1, 2, 3, 4]. This is due in large part to the advances in automatic speech recognition and the popularity and ubiquity of multimedia associated with spoken documents [5, 6]. Extractive speech summarization selects indicative sentences from an original spoken document according to a target summarization ratio and concatenates them together to form a summary accordingly. By doing so, it can provide locations of important speech segments along with their corresponding transcripts for users to listen to and digest. The wide spectrum of extractive speech summarization methods that have been developed so far may roughly fall into three main categories [3, 4, 7]: 1) methods simply based on sentence structure or location 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 specific 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 specific domains or document structures. On the other hand, extractive spoken document based unsupervised statistical measures attempt to select salient sentences on top of some statistical features of spoken sentences, or of the words in the sentences, in an unsupervised manner. Statistical features derived for each sentence, can be word frequency, linguistic score, recognition confidence, similarity (proximity) measure and prosodic information, among others. The associated unsupervised summarization methods based on these features has garnered much research. Representative methods include, but are not limited to, vector space model (VSM) [9], latent semantic analysis (LSA) [9], maximum marginal relevance (MMR) method [10], Markov random walk (MRW) [11], LexRank [12] and the submodularity-based method [13]. 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) [14], the Bayesian classifier (BC) [15], the support vector machine (SVM) [16] and the conditional random fields (CRFs) [17], to name just a few. 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 expensive in terms of time and personnel. 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 [3, 4, 7, 18] for thorough and entertaining discussions of major methods that have been successfully developed and applied to a wide variety of text and speech summarization tasks.

Orthogonally to the foregoing summarization methods, a more recent line of research is to capitalize on the language modeling (LM) framework along with the Kullback-Leibler divergence measure for conducting important sentence selection in an unsupervised manner [19, 20, 21], which has emerged as a promising avenue to extractive speech summarization. Our work in this paper presents a continuation of this general line of research and its main contribution is two-fold. On one hand, we explore to leverage pseudo-relevance feedback in conjunction with several effective sentence modeling formulations to enhance the sentence models involved in the LM-based summarization framework. On the other hand, the utilities of our summarization methods and several widely-used methods are analyzed and compared extensively.

2. LM FREAMWORK FOR SUMMARIZATION

Extractive speech summarization aims at producing a concise summary by selecting salient sentences or paragraphs from the original spoken document according to a predefined target summarization ratio. Intuitively, this task could be framed as an ad hoc information retrieval (IR) problem [22], where the spoken document is treated as an information need and each sentence of

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

the document is regarded as a candidate information unit to be retrieved according to its relevance (or importance) to the information need. Therefore, the ultimate goal of extractive speech summarization could be stated as the selection of the most representative sentences that can succinctly describe the main theme of the spoken document. In the past several years, the language modeling framework has been introduced to a wide spectrum of IR tasks and demonstrated with good empirical success [23]; this modeling paradigm has been successfully adopted for speech summarization recently [19, 20, 21].

2.1. Document-likelihood Measure (DLM)

In the LM-based summarization framework, each sentence S of a spoken document D to be summarized is formulated as a probabilistic generative model for generating the document, and sentences are selected on the basis of their corresponding generative probability $P(D \mid S)$: the higher the probability $P(D \mid S)$, 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 $P(D \mid S)$ can be approximated by:

$$P(D \mid S) \approx \prod_{w \in D} P(w \mid S)^{c(w,D)}, \quad (1)$$

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 $P(D \mid S)$ if it occurs more frequently in D . The simplest way is to estimate the sentence model $P(w \mid S)$ on the basis of the frequency of words occurring in the sentence, with the maximum likelihood (ML) estimation [18]:

$$P(w \mid S) = \frac{c(w,S)}{|S|}, \quad (2)$$

where $c(w,S)$ is the number of times that word w occurs in S and $|S|$ is the length of S. In what follow, we will term (1) the document-likelihood measure (denoted by DLM for short).

2.2. KL-Divergence Measure (KLM)

Another basic formulation of LM for extractive speech summarization is the Kullback-Leibler divergence measure (denoted as KLM for short) [24], which determines the relationship between a document to be summarized and the sentences of the document from a more rigorous information-theoretic perspective. Two different language models are involved in KLM: one for the whole document and the other for each sentence of the document. KLM assumes that words in the document are simple random draws from a language distribution describing some aspects of interest and words in the sentences which belong to the summary should also be drawn from the same distribution. Therefore, we can use KLM to quantify how close the document D and one of its sentences S are: the closer the sentence model $P(w \mid S)$ to the document model $P(w \mid D)$, the more likely the sentence would be part of the summary. The divergence of the sentence model with respect to the document model is defined by [19, 23]:

$$\mathrm{KL}(D \,\|\, S) = \sum_{w \in V} P(w \mid D) \log \frac{P(w \mid D)}{P(w \mid S)}, \quad (3)$$

where w denotes a specific word in the vocabulary set V ; and a sentence S has a smaller value (or probability distance) in terms of $\mathrm{KL}(D \,\|\, S)$ is deemed to be more important. Then, the summary sentences of a given spoken document can be iteratively chosen (i.e., one at each iteration) from the spoken document in accordance with its corresponding divergence until the aggregated summary reaches a predefined target summarization ratio. The KLM not only can be thought as a natural generalization of DLM [23], but also has the additional merit of being able to accommodate extra information cues to improve the estimation of its component models (i.e., the sentence and document models) in a systematic way for better sentence ranking.

3. PSEUDO-RELEVANCE FEEDBACK WITH VARIOUS SENTENCE MODELING FORMULATIONS

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 $P(w \mid S)$ might not be appropriately estimated by the ML estimation. With the alleviation of this deficiency as motivation, in this paper we explore several effective sentence modeling formulations to enhance the sentence representation (or assign more accurate probability masses to words in the sentence) through leveraging the relevance cues gleaned from pseudo-relevance feedback (PRF) [22, 25]. A commonality among these formulations 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 $\mathbf{D}_S = \{D_1, \ldots, D_M\}$, counted as exemplars of pseudo-relevant documents, to be used for subsequent sentence modeling.

3.1. Relevance Model (RM)

In the context of speech summarization, each sentence S of a spoken document D to be summarized is assumed to be associated with an unknown relevance class R_S , and if words that are relevant to the semantic content expressed in S are samples drawn from R_S [26]. However, in reality, since there is no prior knowledge about R_S , we may use top-ranked contemporary text (or spoken) documents returned by the pseudo-relevance feedback procedure (denoted by $\mathbf{D}_S$) 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:

$$P_{\mathrm{RM}}(w \mid S) = \frac{\sum_{D_r \in \mathbf{D}_S} P(D_r) P(w \mid D_r) \prod_{w' \in S} P(w' \mid D_r)}{\sum_{D_r'' \in \mathbf{D}_S} P(D_r'') \prod_{w' \in S} P(w' \mid D_r'')}, \quad (4)$$

where the probability $P(D_r)$ can be simply kept uniform or determined in accordance with the relevance of D_r to S , while $P(w \mid D_r)$ and $P(w' \mid D_r'')$ are estimated on the grounds of the occurrence counts of w in D_r and D_r'' , respectively, with the ML estimation. The RM model assumes that words w that co-occur with the sentence S in the feedback documents will have higher probabilities. The resulting relevance model $P_{\mathrm{RM}}(w \mid S)$ can be linearly combined with or used to replace the original sentence model $P(w \mid S)$.

3.2. Simple Mixture Model (SMM)

In this paper, we also explore 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 returned by PRF are relevant and the resulting model $P_{\mathrm{SMM}}(w \mid S)$ estimated from these documents can potentially benefit sentence modeling. Specifically, SMM assumes that words in $\mathbf{D}_S$ are drawn from a two-component mixture model [27]: 1) One component is the SMM model $P_{\mathrm{SMM}}(w \mid S)$, and 2) the other is a background model $P(w \mid BG)$, which is set to be the baseline unigram language model in this study. The SMM model $P_{\mathrm{SMM}}(w \mid S)$ is estimated by maximizing

the log-likelihood of the set of feedback documents $\mathbf{D}_S$ expressed as follows, using the expectation-maximization (EM) algorithm [28]:

$$LL_{\mathbf{D}_S} = \sum_{D_r \in \mathbf{D}_S} \sum_{w \in V} c(w, D_r) \cdot \log[(1-\alpha) \cdot P_{\mathrm{SMM}}(w \mid S) + \alpha \cdot (w \mid BG)], \quad (5)$$

where α is the pre-defined mixing parameter used to control the degree of reliance between $P_{\mathrm{SMM}}(w \mid S)$ and $P(w \mid BG)$. The maximization of (5) can be conducted iteratively via the following EM update formulas:

E-step:

$$\tau_w^{(l)} = \frac{\alpha \cdot P_{\mathrm{SMM}}^{(l)}(w \mid S)}{\alpha \cdot P_{\mathrm{SMM}}^{(l)}(w \mid S) + (1-\alpha) \cdot P(w \mid BG)}, \quad (6)$$

M-step:

$$P_{\mathrm{SMM}}^{(l+1)}(w \mid S) = \frac{\sum_{D_r \in \mathbf{D}_S} c(w, D_r) \cdot \tau_w^{(l)}}{\sum_{w' \in V} \sum_{D_r'' \in \mathbf{D}_S} c(w', D_r'') \cdot \tau_w^{(l)}}, \quad (7)$$

where l denotes the l-th iteration of the EM algorithm. This estimation 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 $P_{\mathrm{SMM}}(w \mid S)$. Simply put, the SMM model $P_{\mathrm{SMM}}(w \mid S)$ 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. Accordingly, the SMM model $P_{\mathrm{SMM}}(w \mid S)$ can be combined with the original sentence language model through a simple linear interpolation.

3.3. Regularized Simple Mixture Model (RSMM)

Although SMM aims to extract extra word usage cues for enhanced sentence modeling, it may confront two intrinsic problems. One is the extraction of word usage cues from the set of feedback documents is not guided by the original sentence, as shown in (6) and (7). This would lead to a concern for SMM to be distracted from being able to appropriately model the sentence of interest, which is probably caused by some dominant distracting (or irrelevant) feedback documents. The other is that the mixing coefficient α is fixed across all feedback documents albeit that different (either relevant or irrelevant) documents would potentially contribute different amounts of word usage cues to the enhanced sentence model [29]. To mitigate these two problems, we may use the original sentence model $P(w \mid S)$ as a prior on the enhanced sentence model to be estimated, meanwhile introducing a trainable document-specific mixing coefficient α_{D_r} for each feedback document D_r. The resulting model is referred to hereafter as regularized simple mixture model (RSMM) and its corresponding EM update formulas are expressed as follows:

E-step:

$$\tau_{w,D_r}^{(l)} = \frac{\alpha_{D_r}^{(l)} \cdot P_{\mathrm{RSMM}}^{(l)}(w \mid S)}{\alpha_{D_r}^{(l)} \cdot P_{\mathrm{RSMM}}^{(l)}(w \mid S) + (1-\alpha_{D_r}^{(l)}) \cdot P(w \mid BG)}, \quad (8)$$

M-step:

$$\alpha_{D_r}^{(l+1)} = \frac{\sum_{w \in V} c(w, D_r) \cdot \tau_{w,D_r}^{(l)}}{\sum_{w' \in V} c(w', D_r)}, \quad (9)$$

$$P_{\mathrm{RSMM}}^{(l+1)}(w \mid S) = \frac{\mu \cdot P(w \mid S) + \sum_{D_r \in \mathbf{D}_S} c(w, D_r) \cdot \tau_{w,D_r}^{(l)}}{\mu + \sum_{w' \in V} \sum_{D_r'' \in \mathbf{D}_S} c(w', D_r'') \cdot \tau_{w',D_r''}^{(l)}}, \quad (10)$$

where μ is a weighting factor indicating the confidence on the prior information (viz. the original query model).

The notion of leveraging PRF for enhancing query modeling methods has recently attracted much attention and been applied with success to many IR tasks [22, 26, 27, 29]. 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 broadcast news (MATBN) corpus collected by the Academia Sinica and the Public Television Service Foundation of Taiwan between November 2001 and April 2003 [30], which has been segmented into separate stories and transcribed manually. 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. We chose 20 documents as the test set while the remaining 185 documents as the held-out development set. A subset of about 100,000 text news documents, compiled during the same period as the broadcast news documents to be summarized, was employed to estimate related models compared in this paper. A subset of 25-hour speech data compiled during November 2001 to December 2002 in MATBN was used to bootstrap the acoustic training with the minimum phone error rate (MPE) criterion and the training data selection scheme [31]. The vocabulary size is about 72 thousand words.

Three subjects were asked to create summaries of the 205 spoken documents for the summarization experiments as references (the gold standard) for evaluation. The reference summaries were generated by ranking the sentences in the manual transcript of a spoken document by importance without assigning a score to each sentence. For the assessment of summarization performance, we adopted the widely-used ROUGE metrics [32]. Three variants of the ROUGE metrics were used to quantify the utility of the proposed methods. They are, respectively, the ROUGE-1 (unigram) metric, the ROUGE-2 (bigram) metric and the ROUGE-L (longest common subsequence) metric. All the experimental results reported hereafter are obtained by calculating the F-scores [22] of these ROUGE metrics. 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

At the outset, we assess the performance level of the baseline KLM method for extractive speech summarization, by comparing it with several well-practiced unsupervised summarization methods, including LEAD, VSM, MRW, LexRank, MMR and the submodularity-based method. It should be noted that the LEAD-based method simply extracts the first few sentences from a document as the summary. VSM represents each sentence of a document and the whole document in a vector form, where each dimension specifies the weighted statistics, for example the product of the term frequency (TF) and inverse document frequency (IDF), associated with an index term (or word) in the sentence or document. Sentences with the highest relevance scores (usually calculated by the cosine similarity are included in the summary. MMR can be viewed as an extension of VSM, because it also represents each sentence of a document and the document itself in vector form and uses the cosine similarity for sentence selection. However, MMR performs sentence selection iteratively by simultaneously considering the criteria of topic relevance and redundancy. In addition, MRW conceptualizes the document to be summarized as a graph of sentences, where each node represents a sentence and the associated weight of each link represents the

lexical similarity relationship between a pair of nodes. Document summarization thus relies on the global structural information embedded in such conceptualized graph, rather than merely considering the local features of each node (sentence). Put simply, sentences more similar to others are deemed more salient to the main theme of the document. LexRank bears a close resemblance to MRW by selecting salient sentences based on the concept of eigen-centrality of the sentence graph. Lastly, the submodularity-based method (denoted by Submodularity hereafter) frames important sentence selection as a combinatorial optimization problem with objective functions defined on the sentence graph.

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. Several noteworthy observations can be drawn from Table 1. First, the various graph-based methods (viz. MRW, LexRank and Submodularity) are quite competitive with each other and perform better than LEAD and VSM for both the TD and SD cases. Second, MMR that presents an extension of VSM can work as well as the various graph-based methods for the TD case, and exhibit even better performance than the latter ones for the SD case. Third, it is evident that KLM shows comparable performance to the existing unsupervised methods, confirming the applicability of the language modeling framework for speech summarization. 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 [33, 34].

In the second set of experiments, we evaluate the utilities of leveraging pseudo-relevance feedback in conjunction with various modeling formulations, viz. RM, SMM and RSMM, to enhance the sentence models involved in the KLM method. The corresponding results of using these three formulations are shown in Table 2. Consulting Table 2 we notice two particularities. One is that all these three formulations can considerably improve the summarization performance of the KLM method, which corroborates the advantage of using PRF and the various formulations for enhanced sentence modeling. The other is that RSMM is the best-performing one among the three formulations for the TD case; however, such superiority tends to diminish for the SD case. One explanation is that imperfect speech recognition leads to inaccurate estimation of the prior constraint, viz. the original sentence model $P(w \mid S)$, of RSMM. Motivated by the above observation, we attempt to use the resulting model of RM (*cf.* (4)) as the prior constraint, in substitution to the original sentence model, for the estimation of RSMM (referred to as RSMM-RM for short). It seems intuitive that the resulting model of RM would be less susceptible to imperfect speech recognition in comparison to the original sentence model. As one can see from Table 2, RSMM-RM is able to deliver moderate improvements over RSMM for the KLM method in all evaluation conditions.

In the final set of experiments, we compare the KLM method with SVM; SVM is arguably one of the state-of-the-art supervised methods for extractive speech summarization [35, 36, 37]. In this paper, SVM was trained with the documents of the development set along with their summaries, where each sentence of a spoken document was characterized with a set of 19 commonly-used lexical and prosodic features [3, 18, 36, 37]. Comparing the results of SVM shown in Table 3 with that of the variants of the KLM method shown in Table 2, we notice that, although KLM and its variants are, in essence, unsupervised methods that merely use word occurrence or co-occurrence statistics for important sentence selection, they can perform on par with or even better than SVM that utilizes handcrafted summaries and a rich set of features for model training.

Table 1: Summarization results achieved by the baseline KLM method and several widely-used unsupervised methods.

		ROUGE-1	ROUGE-2	ROUGE-L
TD	KLM	41.1	29.8	37.1
	LEAD	31.0	19.4	27.6
	VSM	34.7	22.8	29.0
	MMR	40.7	29.4	35.8
	MRW	41.2	28.2	35.8
	LexRank	41.3	30.9	36.3
	Submodularity	41.4	28.6	36.3
SD	KLM	36.4	21.0	30.7
	LEAD	25.5	11.7	22.1
	VSM	34.2	18.9	28.7
	MMR	38.1	22.6	33.1
	MRW	33.2	19.1	29.1
	LexRank	30.5	14.6	25.4
	Submodularity	33.2	20.4	30.3

Table 2: Summarization results achieved by the KLM method integrated with various sentence modeling formulations.

		ROUGE-1	ROUGE-2	ROUGE-L
TD	RM	45.3	33.5	40.3
	SMM	43.9	32.0	38.8
	RSMM	47.2	36.5	42.3
	RSMM-RM	47.8	36.7	43.2
SD	RM	38.2	23.9	33.1
	SMM	38.3	22.9	32.7
	RSMM	38.1	23.5	32.9
	RSMM-RM	38.9	24.5	33.9

Table 3: Summarization results achieved by SVM and its integration with the KLM method.

		ROUGE-1	ROUGE-2	ROUGE-L
TD	SVM	45.4	35.3	40.7
SD	SVM	36.5	22.2	31.4

6. CONCLUSIONS

In this paper, we have explored several novel sentence modeling formulations that can work in concert with pseudo-relevance feedback to improve the performance of the KLM summarization method. Experimental evidence supports that the various methods instantiated from our modeling framework are quite comparable to a few existing state-of-the-art methods for extractive speech summarization. In future work, we plan to investigate jointly integrating proximity and other different kinds of relevance and lexical/semantic information cues into the process of feedback document selection so as to improve the empirical effectiveness of sentence modeling in KLM. We are also interested in investigating more robust indexing techniques to represent spoken documents. In addition, we intend to further adopt and formalize the proposed LM methods for 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 National Science Council, 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] S. Furui, T. Kikuchi, Y. Shinnaka and C. Hori, "Speech-to-text and speech-to-speech summarization of spontaneous speech," *IEEE Transactions on Speech and Audio Processing,* vol. 12, no. 4, pp. 401–408, 2004.

[2] K. McKeown, J. Hirschberg, M. Galley and S. Maskey, "From text to speech summarization," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 997–1000, 2005.

[3] Y. Liu and D. Hakkani-Tur, "Speech summarization," *Chapter 13* in *Spoken Language Understanding: Systems for Extracting Semantic Information from Speech*, G. Tur and R. D. Mori (Eds), New York: Wiley, 2011.

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

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

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

[7] I. Mani and M.T. Maybury (Eds.), *Advances in automatic text summarization*, Cambridge, MA: MIT Press, 1999.

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

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

[10] J. Carbonell and J. Goldstein, "The use of MMR, diversity based reranking for reordering documents and producing summaries," in *Proc. ACM SIGIR Conference on Research and Development in Information Retrieval*, 335–336, 1998.

[11] X. Wan and J. Yang, "Multi-document summarization using cluster-based link analysis," in *Proc. ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 299–306, 2008.

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

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

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

[15] J. Kupiec, J. Pedersen, and F. Chen, "A trainable document summarizer," in *Proc. ACM SIGIR Conf. on R&D in Information Retrieval*, pp. 68–73, 1995.

[16] A. Kolcz, V. Prabakarmurthi, and J. Kalita, "Summarization as feature selection for text categorization," in *Proc. ACM Conf. on Information and Knowledge Management*, pp. 365–370, 2001.

[17] M. Galley, "Skip-chain conditional random field for ranking meeting utterances by importance," in *Proc. Empirical Methods in Natural Language Processing*, pp. 364–372, 2006.

[18] G. Penn and X. Zhu, "A critical reassessment of evaluation baselines for speech summarization," in *Annual Meeting of the Association for Computational Linguistics,* pp. 470–478, 2008.

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

[20] A. Celikyilmaz and D. Hakkani-Tur, "A hybrid hierarchical model for multi-document summarization," in *Proc. Annual Meeting of the Association for Computational Linguistics*, pp. 815–824, 2010.

[21] B. Chen, H.-C. Chang, K.-Y. Chen, "Sentence modeling for extractive speech summarization," *IEEE International Conference on Multimedia & Expo*, pp. 1–6, 2013.

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

[23] C.X. Zhai, "Statistical language models for information retrieval: A critical review," *Foundations and Trends in Information Retrieval*, 2(3), pp. 137–213, 2008.

[24] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22(1), 79–86, 1951.

[25] J. Rocchio, "Relevance feedback in information retrieval," in G. Salton (Ed.), *The SMART Retrieval System: Experiments in Automatic Document Processing*, pp. 313–23, Prentice Hall, 1971.

[26] V. Lavrenko and W.B. Croft, "Relevance-based language models," in *Proc. ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 120–127, 2001.

[27] C. X. Zhai and J. Lafferty, "Model-based feedback in the language modeling approach to information retrieval," in *Proceedings of ACM SIGIR Conference on Information and knowledge management*, pp. 403–410, 2001.

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

[29] T. Tao and C. X. Zhai, "Regularized estimation of mixture models for robust pseudo-relevance feedback," in *Proceedings of ACM SIGIR Conference on Information and knowledge management*, pp. 162–169, 2006.

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

[31] S.-H. Liu, F.-H. Chu, S.-H. Lin, H.-S. Lee and B. Chen, "Training data selection for improving discriminative training of acoustic models," in *Proc. IEEE workshop on Automatic Speech Recognition and Understanding*, pp. 284–289, 2007.

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

[33] S. Xie and Y. Liu, "Using *N*-best lists and confusion networks for meeting summarization" *IEEE Transactions on Audio, Speech and Language Processing*, 19(5), pp. 1160–1169, 2011.

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

[35] S. Xie and Y. Liu, "Improving supervised learning for meeting summarization using sampling and regression," *Computer Speech & Language*, 24(3), pp. 495–514, 2010.

[36] J. Zhang and P. Fung, "Speech summarization without lexical features for Mandarin broadcast news," in *Proc. NAACL HLT, Companion Volume*, pp. 213–216, 2007.

[37] B. Chen, S.-H. Lin, Y.-M. Chang, J.-W. Liu, "Extractive speech summarization using evaluation metric-related training criteria," *Information Processing & Managemen*t, 49(1), pp. 1–12, 2013.