EFFECTIVE PSEUDO-RELEVANCE FEEDBACK FOR SPOKEN DOCUMENT RETRIEVAL

Yi-Wen Chen[*], *Kuan-Yu Chen*[†], *Hsin-Min Wang*[†], *Berlin Chen*[*]

[*]National Taiwan Normal University, Taipei, Taiwan
[†]Institute of Information Science, Academia Sinica, Taipei, Taiwan
E-mail: [*]{699470462, berlin}@ntnu.edu.tw, [†]{kychen, whm}@iis.sinica.edu.tw

ABSTRACT

With the exponential proliferation of multimedia associated with spoken documents, research on spoken document retrieval (SDR) has emerged and attracted much attention in the past two decades. Apart from much effort devoted to developing robust indexing and modeling techniques for representing spoken documents, a recent line of thought targets at the improvement of query modeling for better reflecting the user's information need. Pseudo-relevance feedback is by far the most commonly-used paradigm for query reformulation, which assumes that a small amount of top-ranked feedback documents obtained from the initial round of retrieval are relevant and can be utilized for this purpose. Nevertheless, simply taking all of the top-ranked feedback documents obtained from the initial retrieval for query modeling (reformulation) does not always work well, especially when the top-ranked documents contain much redundant or non-relevant information. In the view of this, we explore in this paper an interesting problem of how to effectively glean useful cues from the top-ranked documents so as to achieve more accurate query modeling. To do this, different kinds of information cues are considered and integrated into the process of feedback document selection so as to improve query effectiveness. Experiments conducted on the TDT (Topic Detection and Tracking) task show the advantages of our retrieval methods for SDR.

Index Terms—Spoken document retrieval, pseudo-relevance feedback, query modeling, Kullback-Leibler (KL)-divergence

1. INTRODUCTION

In the recent past, spoken document retrieval (SDR) has received a growing amount of interest and activity in the speech processing community. This is due in large part to the advances in automatic speech recognition (ASR) and the ever-increasing volumes of multimedia associated with spoken documents made available to the public [1, 2, 3]. Unlike research on spoken term detection (STD) [3] that usually embraces the goal of extracting probable spoken terms or phrases inherent in a spoken document that could match the query words or phrases literally, research on SDR revolves more around the notion of relevance of a spoken document in response to a query [4].

There are at least two fundamental problems facing SDR. On one hand, the imperfect speech recognition transcript carries wrong information and thus would deviate somewhat from representing the true theme of a spoken document. On the other hand, a query is often only a vague expression of an underlying information need, and there probably would be word usage mismatch between a query and a spoken document even if they are topically related to each other. A large body of SDR work has been placed on the exploration of robust indexing or modeling techniques to represent spoken documents [3, 5, 6, 7, 8], but very limited research has been conducted to look at the other side of the coin, namely, the improvement of query formulation for better reflecting the underlying information need of a user [9]. As for the latter problem, pseudo-relevance feedback [4] is by far the most commonly-used paradigm, which assumes that a small amount of top-ranked documents obtained from the initial round of retrieval are relevant and can be utilized for query reformulation. Subsequently, the system performs a second round of retrieval with the enhanced query representation to search for more relevant documents.

We had recently introduced a new perspective on query modeling [9], saying that it can be approached with pseudo-relevance feedback and the language modeling (LM) retrieval approach [10] leveraging the notion of relevance [11], which seems to show preliminary promise for query reformulation. The success of such query modeling depends largely on the assumption that the set of top-ranked feedback documents obtained from the initial round of retrieval are relevant and can be used to estimate a more accurate query model. Nevertheless, simply taking all of the top-ranked feedback documents obtained from the initial round of retrieval does not always work well for query modeling (or reformulation), especially when the top-ranked documents contain much redundant or non-relevant information.

With the above background, in this paper we turn our attention to a more challenging problem of how to additionally glean useful cues from the top-ranked feedback documents to achieve more accurate query modeling. Towards this end, several kinds of information cues are considered and integrated to select representative feedback documents for better retrieval performance. The rest of this paper is organized as follows. We briefly review the basic mathematical formulations of the LM-based retrieval models for SDR, as well as the idea of pseudo-relevance feedback, in Section 2. In Section 3, we describe and explain several cues we explore to select representative feedback documents during pseudo-relevance feedback. After that, the experimental settings and a series of retrieval experiments are presented in Sections 4 and 5, respectively. Finally, Section 6 concludes the paper with directions for future research.

2. RETRIEVAL MODELS

2.1. Unigram Language Model (ULM)

A recent trend in building SDR systems is to use the language modeling (LM) approach [7, 8, 9]. This is due to the fact that the LM approach has sound theoretical underpinnings and excellent empirical performance. The fundamental formulation of the LM approach to SDR is to compute the conditional probability $P(Q \mid D)$, i.e., the likelihood of a query Q generated by each spoken document D (the so-called query-likelihood measure) [10]. A spoken document D is deemed to be relevant with respect to the query Q if the corresponding document model is more likely to generate the query. If the query Q is treated as a sequence of words, $Q = q_1, q_2, \cdots, q_L$, where the query words are assumed to be conditionally independent given the document D and their order is also assumed to be of no importance (i.e., the so-called "*bag-of-words*" assumption), the similarity measure $P(Q \mid D)$ can be further decomposed as a product of the probabilities of the query words generated by the document:

$$P(Q|D) = \prod_{l=1}^{L} P(q_l|D), \tag{1}$$

where $P(q_l \mid D)$ is the likelihood of generating q_l by document D (a.k.a. the document model), which is estimated based on the word occurrence frequencies in a document by the maximum-likelihood (ML) estimator and can be further smoothed with a background unigram model $P(w \mid BG)$ to model the general properties of the language as well as to avoid the problem of zero probability [10].

2.2. Kullback-Leibler (KL)-Divergence Measure

Another basic formulation of LM for SDR is the Kullback-Leibler (KL)-divergence measure [10, 12]:

$$\begin{aligned} -KL(Q\|D) &= -\sum_{w \in V} P(w|Q) \log \frac{P(w|Q)}{P(w|D)} \\ &\overset{\text{rank}}{=} \sum_{w \in V} P(w|Q) \log P(w|D), \end{aligned} \tag{2}$$

where the query and the document are, respectively, framed as a (unigram) language model (i.e., $P(w|Q)$ and $P(w|D)$), $\overset{\text{rank}}{=}$ means equivalent in terms of being used for the purpose of ranking documents, and V denotes the vocabulary. A document D has a smaller value (or probability distance) in terms of $KL(Q \| D)$ is deemed to be more relevant with respect to Q. The retrieval effectiveness of the KL-divergence measure depends primarily on the accurate estimation of the query modeling $P(w|Q)$ and the document modeling $P(w|D)$. Furthermore, it is easy to show that the KL-divergence measure will give the same ranking as the ULM model (*cf.* (1)) when the query language model is simply derived with the ML estimator [11]. Accordingly, the KL-divergence measure not only can be thought as a generalization of the query-likelihood measure, but also has the additional merit of being able to accommodate extra information cues to improve the estimation of its component models (especially, the query model) for better document ranking in a systematic way [9].

2.3. Pseudo-Relevance Feedback

Due to the fact that a query often consists of only a few words, the query model that is meant to represent the user's information need might not be appropriately estimated by the ML estimator. Furthermore, merely matching words between a query and documents might not be an effective approach, as the word overlaps alone could not capture the semantic intent of the query. To cater for this, an LM-based SDR system with the KL-divergence measure can adopt the idea of pseudo-relevance feedback and perform two rounds of retrieval to search for more relevant documents. In the first round of retrieval, an initial query is input into the SDR system to retrieve a number of top-ranked feedback documents. Subsequently, on top of these top-ranked feedback documents, a refined query model is constructed and a second round of retrieval is conducted with this new query model and the KL-divergence measure depicted in (2). It is usually anticipated that the SDR system can thus retrieve more documents relevant to the query.

However, an LM-based SDR system with the pseudo-relevance feedback process may confront two intrinsic challenges. One is how to purify the top-ranked feedback documents obtained from the first round of retrieval so as to remove redundant and non-relevant information. The other is how to effectively utilize the selected set of representative feedback documents for estimating a more accurate query model. For the latter, there are a number of studies proposing various query modeling techniques directly exploiting the top-ranked feedback text (or spoken) documents, such as the simple mixture model (SMM) [13], the relevance model (RM) [11] and their extensions [9] , among others. However, for the former, there is relatively little work done on selecting useful and representative feedback documents from the top-ranked ones for SDR, as far as we are aware. Recently, the so-called "Gapped Top *K*" and "Cluster Centroid" selection methods [14] have been proposed for text information retrieval (IR). "Gapped Top *K*" selects top *K* documents with a ranking gap *L* in between any two top-ranked documents, while "Cluster Centroid" groups the top-ranked documents into *K* clusters and selects one representative document from each cluster to obtain diversified feedback documents. Another more attractive and sophisticated method proposed for text IR is "Active-RDD" [15], which takes into account the relevance, diversity and density cues of the top-ranked documents for feedback document selection. The above three methods have not been extensively studied for SDR.

In this paper, we go a step further by additionally exploring the non-relevance cue during feedback document selection, apart from the relevance, diversity and density cues. As we will see later, the additional use of the non-relevance cue can further boost the SDR performance.

3. LEVERAGING RELEVANCE, NON-RELEVANCE, DIVERSITY AND DENSITY MEASURES FOR PSEUDO-RELEVANCE FEEDBACK

Our SDR system first takes the initial query and employs the ULM retrieval model to obtain a number of top-ranked documents $\mathbf{D}_{\text{Top}} = \{D_1, D_2, \ldots, D_N\}$. Then in the pseudo-relevance feedback process, the system iteratively selects documents from $\mathbf{D}_{\text{Top}}$ to form a representative set of feedback documents by simultaneously considering the relevance, non-relevance, diversity and density cues. More specifically, each candidate feedback document *D* is associated with a score that is a linear combination of measures of these cues, expressed as follows:

Table 1. Statistics for TDT-2 Collection.

# Spoken documents	2,265 stories 46.03 hours of audio			
# Distinct test queries	16 Xinhua text stories (Topics 20001~20096)			
	Min.	Max	Med.	Mean
Document length (in characters)	23	4841	153	287
Length of query (in characters)	8	27	13	14
# Relevant documents per test query	2	95	13	29

$$D^* = \arg\max_{D \in \mathbf{D}_{\text{Top}} - \mathbf{D}_{\text{P}}} \Big[(1-\alpha-\beta-\gamma) \cdot M_{Rel}(Q,D) + \alpha \cdot M_{NR}(Q,D) + \beta \cdot M_{Diversity}(D) + \gamma \cdot M_{Density}(D) \Big], \quad (3)$$

where $\mathbf{D}$p is the set of already selected feedback documents; $M_{Rel}(Q,D)$, $M_{NR}(Q,D)$, $M_{Diversity}(D)$ and $M_{Density}(D)$ are measures of relevance, non-relevance, diversity and density for each document D in $\mathbf{D}_{\text{Top}}$, respectively; α , β and γ are weighting coefficients. The selection process illustrated in (3) will be executed iteratively until $\mathbf{D}_{\text{P}}$ contains a pre-defined number of feedback documents. It is worth mentioning that the method described in (3) bears a close resemblance in spirit to the maximal marginal relevance (MMR) ranking algorithm [16, 17] which was originally proposed for extractive document summarization. Note also that $M_{Rel}(Q,D)$ is just the similarity (query-likelihood) measure of the ULM retrieval model depicted in (1). In the following, we will describe how to model the other information cues we explore for a given candidate feedback document.

3.1. Non-Relevance Measure

For a given query Q, we can estimate a non-relevance model $P(w|NR_Q)$ of it based on the low-ranked documents obtained from the initial round of retrieval, and the non-relevance measure of a candidate feedback document D is thus defined by

$$M_{NR}(D) = KL(NR_Q \| D). \quad (4)$$

The additional incorporation of $M_{NR}(D)$ for feedback document selection will prefer those documents that have only a small probability distance to the original query model but also a larger probability distance to the non-relevance model. Since the number of relevant documents with respect to a given query is usually very small compared to that of non-relevant ones in practice, we may assume that the entire spoken document collection (more specifically, the background language model $P(w|BG)$) could offer an alternative estimate of the non-relevance model.

3.2. Diversity Measure

Recently, diversification of retrieval results has gained popularity in the text IR community, since it can be used to complement the conventional document ranking criteria which only consider relevance information and often suffer from returning too many redundant documents. By analogy, in the context of pseudo-relevance feedback, if we use the top-ranked documents that contain too much redundant information to estimate the query model, then the second round of retrieval is prone to return too many "redundant" documents to the user. In order to diversify the selected feedback documents for better query reformulation, we compute the diversity measure of a candidate feedback document with respect to the set $\mathbf{D}_{\text{P}}$ of already selected feedback documents, which is expressed as follows:

$$M_{Diversity}(D) = \min_{D_j \in \mathbf{D}_{\text{P}}} \frac{1}{2} \cdot \big[KL(D_j \| D) + KL(D \| D_j) \big]. \quad (5)$$

3.3. Density Measure

Intuitively, the structural information among the top-ranked documents can be taken into account as well during feedback document selection. For this idea to work, we can compute the average negative, symmetric probability distance between a document D and all the other documents D_h in $\mathbf{D}_{\text{Top}}$, which is expressed as follows:

$$M_{Density}(D) = \frac{-1}{|\mathbf{D}_{\text{Top}}| - 1} \cdot \sum_{\substack{D_h \in \mathbf{D}_{\text{Top}} \\ D_h \neq D}} \big[KL(D_h \| D) + KL(D \| D_h) \big], \quad (6)$$

where $|\mathbf{D}_{\text{Top}}|$ is the number of documents in $\mathbf{D}_{\text{Top}}$. A document D having a higher value of $M_{Density}(D)$ is deemed to be closer to the other documents in $\mathbf{D}_{\text{Top}}$ and thus to be more representative (and less likely to be an outlier).

4. EXPERIMENTAL SETUP

4.1. Spoken Document Collection

We used the Topic Detection and Tracking collection (TDT-2) [9, 18] for this work. The Mandarin news stories from Voice of America news broadcasts were used as the spoken documents. All news stories were exhaustively tagged with event-based topic labels, which served as the relevance judgments for performance evaluation. The average word error rate (WER) of the spoken documents is about 35% [19]. The retrieval results, assuming that manual transcripts for the spoken documents to be retrieved (denoted TD, text documents, in the tables below) are known, are also shown for reference, compared to the results when only the erroneous transcripts by speech recognition are available (denoted SD, spoken documents, in the tables below). The retrieval results are expressed in terms of non-interpolated mean average precision (mAP) following the TREC evaluation [4]. Table 1 shows some basic statistics about the TDT-2 collection. In order to evaluate the performance of the various feedback document selection methods studied in this paper, the number of the top-ranked documents obtained from the first round of retrieval is set to 25 (i.e., $|\mathbf{D}_{\text{Top}}|$ =25) and the target number of selected feedback document is set to 5 (i.e., $|\mathbf{D}_{\text{P}}|$ =5). Albeit that, it is known that the way to systemically determine the values of the free parameters that the feedback document selection methods, as well as the retrieval models, incorporate is still an open issue and needs further investigation and proper experimentation.

4.2. Query Modeling

In this paper, we employ RM and SMM for query reformulation in concert with the various feedback document selection methods studied in this paper. The refined query model based on RM [11] is formulated by

$$P_{\mathrm{RM}}(w|Q)=\frac{\sum_{D_j\in\mathbf{D}_\mathrm{P}}P(D_j)P(w\,|\,D_j)\prod_{l=1}^{L}P(q_l\,|\,D_j)}{\sum_{D_j\in\mathbf{D}_\mathrm{P}}P(D_j)\prod_{l=1}^{L}P(q_l\,|\,D_j)}, \qquad (7)$$

where the probability $P(D_j)$ can be simply kept uniform or determined in accordance with the relevance of D_j to Q, while $P(w\,|\,D_j)$ and $P(q_l\,|\,D_j)$ are estimated on the grounds of the word occurrence counts in D_j with the ML estimator. The RM model assumes that words w that co-occur with the query Q in the feedback documents will have higher probabilities. We had recently presented an improved version of the RM model by further incorporating a set of latent topics into the modeling of $P(w\,|\,D_j)$ and $P(q_l\,|\,D_j)$, which is referred to as the topic-based relevance model (TRM) [9] hereafter.

On the other hand, SMM [13] assumes words in the set of feedback documents $\mathbf{D}_\mathrm{P}$ are drawn from two models: 1) the feedback model $P(w\,|\,FB)$ and 2) the background model $P(w\,|\,BG)$. The feedback model $P(w\,|\,FB)$ is estimated by maximizing the log-likelihood of the set of feedback documents $\mathbf{D}_\mathrm{P}$ using the expectation-maximization (EM) algorithm [20]. The resulting feedback model $P(w\,|\,FB)$ can be linearly combined with or used to replace the original query model $P(w|Q)$.

5. EXPERIMENTAL RESULTS

In the first set of experiments, we compare the performance of RM, TRM and SMM when the top-ranked (i.e., top 25) documents obtained from the initial round of retrieval is used for constructing the refined query models. The corresponding results are shown in Table 2, where the results of ULM and LDA (latent Dirichlet allocation) [21] are also listed for reference. LDA is a state-of-the-art (more sophisticated) LM-based retrieval model that incorporates a set of latent topics for representing (spoken) documents. It is worth mentioning that both ULM and LDA perform retrieval only with the initial query. Inspection of Table 2 reveals two noteworthy points. First, the performance gap between the retrieval using manual transcripts (denoted by TD) and the recognition transcripts (denoted by SD) is about 0.05 in terms of mAP, such degradation is apparently less pronounced as compared to the WER of spoken documents [9]. Second, RM and SMM tend to perform on par with each other, and they deliver substantial improvements over ULM (and perform comparably to LDA), while TRM exhibits superior performance over RM and SMM, confirming the merits of leveraging topical information for query modeling.

In the second set of experiments, we evaluate the utility of the various feedback document selection methods investigated in this paper, including "Gapped Top *K*" (denoted by "Gapped" for short), "Cluster Centroid" (denoted by "Cluster" for short), "Active-RDD" and our purposed method (*cf.* Sections 2 and 3), in concert with some of the above retrieval (query) models (the number of selected feedback documents is set to 5). The corresponding results are shown in Table 3, whereas the results of simply using the top 5 documents obtained from the initial round of retrieval to construct the refined query models are listed in Table 4 for comparison. There are three noteworthy points to these results. First, using either "Active-RDD" or our proposed method to select feedback documents seems to outperform that simply using the top 5 documents (*cf.* Table 4) or the top 25 documents (*cf.* Table 2) obtained from the initial round of retrieval as the feedback documents by a big margin, indicating that appropriate feedback document selection is critical to the success of query reformulation. Second, our proposed method delivers better performance gains over "Active-RDD" for all cases, which exhibits the advantage of using the non-relevance cue for feedback document selection. Third, "Gapped Top *K*" and "Cluster Centroid" both result in performance that appears to be much inferior to that of "Active-RDD" and our proposed method.

Table 2. Retrieval results (in mAP) achieved by various retrieval models.

	ULM	LDA	RM	TRM	SMM
TD	0.371	0.401	0.421	0.456	0.415
SD	0.323	0.341	0.369	0.397	0.361

Table 3. Retrieval results (in mAP) achieved by various combinations of retrieval models and feedback document selection methods.

		RM	TRM	SMM
TD	Gapped	0.414	0.452	0.406
	Cluster	0.396	0.441	0.380
	Active-RDD	0.471	0.492	0.457
	Our Method	0.491	0.507	0.490
SD	Gapped	0.357	0.391	0.333
	Cluster	0.378	0.395	0.325
	Active-RDD	0.437	0.461	0.403
	Our Method	0.448	0.475	0.424

Table 4. Retrieval results (in mAP) achieved when simply using the top 5 documents obtained from the initial round of retrieval for constructing various query models.

	RM	TRM	SMM
TD	0.405	0.440	0.438
SD	0.369	0.396	0.399

6. CONCLUSIONS

We have proposed a language modeling (LM) framework to combine several kinds of information cues into the process of feedback document selection for enhanced query formulation in SDR. The utility of the retrieval methods deduced from such a framework has also been validated by extensively comparisons with several existing methods. The experimental results seem to reveal the superiority of our LM framework for SDR. As to future work, we would like to adopt this LM framework for speech recognition and summarization [22, 23].

7. ACKNOWLEDGEMENT

This work was sponsored in part by "Aim for the Top University Plan" of National Taiwan Normal University and Ministry of Education, Taiwan, and the National Science Council, Taiwan, under Grants NSC 101-2221-E-003-024-MY3, NSC 101-2511-S-003-057-MY3, NSC 101-2511-S-003-047-MY3, NSC 99-2221-E-003-017-MY3, and NSC 98-2221-E-003-011-MY3.

8. REFERENCES

[1] L. S. Lee and B. Chen, "Spoken document understanding and organization," *IEEE Signal Processing Magazin*e, 22(5), 42–60, 2005.

[2] M. Ostendorf, "Speech technology and information access," *IEEE Signal Processing Magazine*, 25(3), 150–152, 2008.

[3] C. Chelba, T. J. Hazen and M. Saraclar, "Retrieval and browsing of spoken content," *IEEE Signal Processing Magazine*, 25(3), 39–49, 2008.

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

[5] V. T. Turunen and M. Kurimo, "Indexing confusion networks for morph-based spoken document retrieval," in *Proc. ACM SIGIR Conference on Research and Development in Information Retrieval*, 631–638, 2007.

[6] S. Parlak and M. Saraçlar, "Performance analysis and improvement of Turkish broadcast news retrieval," *IEEE Transactions on Audio, Speech and Language Processing*, 20(3), 731–743, 2012.

[7] B. Chen, "Word topic models for spoken document retrieval and transcription," *ACM Transactions on Asian Language Information Processing*, 8(1), 2:1–2:27, 2009.

[8] T. K. Chia, K. C. Sim, H. Li, and H. T. Ng, "Statistical lattice-based spoken document retrieval," *ACM Transactions on Information Systems*, 28 (1), 2:1–2:30, 2010.

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

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

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

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

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

[14] X. Shen and C. Zhai, "Active feedback in ad hoc information retrieval," in *Proc. ACM SIGIR Conference on Research and Development in Information Retrieval*, 55–66, 2005.

[15] 24. Z. Xu, R. Akella and Y. Zhang, "Incorporating diversity and density in active learning for relevance feedback," in *Proc. European conference on IR research*, 245–257, 2007.

[16] 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.

[17] B. Chen and S.-H. Lin, "A risk-aware modeling framework for speech summarization," *IEEE Transactions on Audio, Speech and Language Processing*, 20(1), 199–210, 2012.

[18] LDC, "Project topic detection and tracking," *Linguistic Data Consortium,* 2000.

[19] H. Meng, B. Chen, S. Khudanpur, G. A. Levow, W. K. Lo, D. Oard, P. Schone, K. Tang, H. M. Wang, and J. Wang, "Mandarin–English information (MEI): investigating translingual speech retrieval," *Computer Speech and Language*, 18(2), 163–179, 2004.

[20] 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), 1–38, 1977.

[21] D. M. Blei, A. Y. Ng and M. I. Jordan, "Latent Dirichlet allocation," *Journal of Machine Learning Research*, 3, 993–1022, 2003

[22] B. Chen and K.-Y. Chen, "Leveraging relevance cues for language modeling in speech recognition," *Information Processing & Management*, 49(4), pp. 807-816, 2013.

[23] B. Chen, H.-C. Chang and K.-Y. Chen, "Sentence modeling for extractive speech summarization," in *Proc. IEEE International Conference on Multimedia & Expo*, 2013.