I-VECTOR BASED LANGUAGE MODELING FOR QUERY REPRESENTATION

Kuan-Yu Chen[+], *Hsin-Min Wang*[+], *Berlin Chen*[*], *Hsin-His Chen*[#]

[+]Institute of Information Science, Academia Sinica, Taipei, Taiwan
[*]National Taiwan Normal University, Taipei, Taiwan
[#]National Taiwan University, Taipei, Taiwan

[+]{kychen, whm}@iis.sinica.edu.tw, [*]berlin@ntnu.edu.tw, [#]hhchen@ntu.edu.tw

ABSTRACT

Since more and more multimedia data associated with spoken documents have been made available to the public, spoken document retrieval (SDR) has become an important research subject in the past two decades. Following the research tendency, many efforts have been devoted towards developing indexing and modeling techniques for representing spoken documents, but only few have been made on improving query formulation for better representing users' information needs. The i-vector based language modeling (IVLM) framework, stemming from the state-of-the-art i-vector framework for language identification and speaker recognition, has been proposed and formulated to represent documents in SDR with good promise recently. However, a major challenge of using IVLM for query modeling is that a query usually consists of only a few words; thus, it is hard to learn a reliable representation accordingly. In this paper, we focus our attention on query reformulation and propose three novel methods on top of IVLM to more accurately represent users' information needs. In addition, we also explore the use of multi-levels of index features, including word- and subword-level units, to work in concert with the proposed methods. A series of empirical SDR experiments conducted on the TDT-2 (Topic Detection and Tracking) collection demonstrate the good effectiveness of our proposed methods as compared to existing state-of-the-art methods.

Index Terms— Spoken document retrieval, i-vector, language modeling, query representation

1. INTRODUCTION

Over the past two decades, spoken content analysis [1, 2] has garnered considerable interest in the speech processing community. A significant amount of research efforts has been devoted towards developing robust indexing (or representation) techniques [3-6] so as to extract probable spoken terms or phrases inherent in a spoken document that could match the query words or phrases literally. On the other hand, spoken document retrieval (SDR) that revolves more around the notion of relevance of a spoken document in response to a query has also been a prominent subject of much recent research. It is generally agreed that a document is relevant to a query if it can address the stated information need of the query, but not because it happens to contain all the words in the query [7].

In the past, several retrieval models, such as the vector space model (VSM) [7, 8], the latent semantic analysis (LSA) [6, 7, 9], and the Okapi BM25 model [7, 10], have been proposed and used in many information retrieval (IR) applications, including SDR. Their efficient and effective abilities have been proved by many researchers and practitioners for a wide variety of IR-related tasks. Recently, an emerging stream of thought is to employ a statistical language model (LM) [10-14], which has become an attractive choice due to its simplicity and clear probabilistic meaning, as well as state-of-the-art performance. In practice, each document is interpreted as a generative model composed of a mixture of multinomial (or *n*-gram) distributions for observing a query, while the query is regarded as observations, expressed as a sequence of words. Accordingly, documents can be ranked according to their likelihoods of generating the query, viz. the query-likelihood measure (QLM) [11, 12]. Another basic formulation of LM for SDR is the Kullback-Leibler divergence measure (KLM) [7, 14]. On the basis of KLM, both query and document are usually modeled by a unigram language model, respectively. The relevance degree between the pair of query and document is recast to calculate the divergence distance between the two unigram models. It is easy to show that KLM covers QLM. They are equivalent when we merely use the empirical query word distribution derived by the maximum likelihood (ML) estimator to infer the query language model. Distinctively, KLM has the merit of being able to accommodate extra information cues to improve the estimation of its component models (e.g., the query model) for better document ranking in a theoretical way.

Recently, the i-vector technique has stood out as one of the state-of-the-art paradigms for language identification (LID) [15-18] and speaker recognition (SR) [19-21]. One challenge of these tasks is the need to process and analyze a high-dimensional vector, which is typically constructed from the variable-length series of acoustic feature vectors of a given input utterance based on some reference models. As such, for the LID and SR tasks, the major contribution of the i-vector technique is to concisely and effectively represent the consecutive sequence of acoustic feature vectors of an speech utterance as a single low-dimensional vector while retaining the most representative (e.g., language-specific for LID or speaker-specific for SR) information encapsulated in the original utterance, and subsequently a range of well-developed post-processing techniques (such as probabilistic linear discriminative analysis, PLDA) can be readily applied. However, when a given speech utterance consists of only a few acoustic feature vectors, the representation of the utterance learned by the i-vector technique becomes understandably problematic, and the recognition performance may degrade severely [21-26].

Stimulated by the concept of the i-vector technique, a novel i-vector based language modeling (IVLM) [27] framework has been recently proposed and introduced to SDR. Since a document is composed of a variable-length series of words, the idea of the core is to apply the i-vector technique to represent a document by a low-dimensional vector, which retains the most representative information of the document. Preliminary SDR experiments have demonstrated the performance merits of IVLM integrated with

QLM (i.e., the document is regarded as an IVLM model while the query as an observation sequence) over several well-practiced retrieval models. In this paper, we try to make a step forward to better represent the users' information needs with the IVLM modeling framework, and to couple such a framework with KLM for improving document ranking (i.e., each document is represented by a conventional unigram language model derived by the ML estimator, while the query is represented by an IVLM model). However, the fact that a query usually consists of only a few words inevitably causes deficiency in deriving a low-dimensional representation for the query. To mitigate the problem, in this paper three novel reformulation methods are proposed for use in SDR. It is also expected that the conventional LID and SR applications can benefit from our methods. In addition, we also investigate representing queries and documents with different granularities (i.e., word- and subword-levels) of index features to work in conjunction with KLM.

2. LANGUAGE MODELING FOR SPOKEN DOCUMENT RETRIEVAL

2.1. Query Likelihood Measure (QLM)

The fundamental formulation of the language modeling (LM) approach to SDR is to compute the conditional probability $P(Q|d)$, which is the likelihood of a query Q generated by a spoken document d (i.e., the so-called query-likelihood measure, QLM) [10-14]. A spoken document d is deemed to be relevant 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 $P(Q|d)$ can be further decomposed as a product of the probabilities of the query words generated by the document [7, 14]:

$$P(Q \mid d) = \Pi_{l=1}^{L} P(q_l|d), \tag{1}$$

where $P(q_l|d)$ is the likelihood of generating q_l by document d, which is typically estimated based on the occurrence frequency of q_l in d by the empirical ML estimator. To capture the general properties of a language as well as to avoid the problem of zero probability, each document model is usually smoothed by a background language model [11-14].

2.2. Kullback-Leibler Divergence Measure (KLM)

Another promising realization of the LM approach to SDR is the Kullback-Leibler divergence measure (KLM) [7, 14], which determines the degree of relevance between a spoken document and a query from a more rigorous information-theoretic perspective. Two different language models are involved in KLM: one for the document and the other for the query. KLM assumes that words in the query are random draws from a language distribution that describes the information need of a user, and words in the relevant spoken documents should also be drawn from the same distribution. Accordingly, we can use KLM to quantify how close a spoken document d and a given query Q are: the closer the document model $P(w|d)$ to the query model $P(w|Q)$, the more likely the document would be a relevant document. The documents are ranked according to the divergence of the document language model with respect to the query language model [14]:

$$-KL(Q \| d) = -\sum_{w \in V} P(w|Q) \log \frac{P(w|Q)}{P(w|d)} \propto \sum_{w \in V} P(w|Q) \log P(w|d), \tag{2}$$

where V denotes the vocabulary inventory in the language and the minus operator for KL-divergence is used to convert the divergence distance to a relevance measure. A spoken document d that has a smaller value (or probability distance) in terms of $KL(Q||d)$ is considered more relevant to Q. The retrieval effectiveness of the KLM method depends primarily on the accurate estimation of the query language model $P(w|Q)$ and the document language model $P(w|d)$. Furthermore, it is known that KLM will give the same ranking as QLM (*cf.* Eq. (1)), when the query language model is simply derived with an empirical ML estimator. As such, KLM not only can be thought as a generalization of QLM, 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) in a theoretical and systematic way.

3. I-VECTOR BASED LANGUAGE MODELING & QUERY REPRESENTATION

3.1. I-Vector Technique

The i-vector technique [15-21] is a simplified variant of the joint factor analysis approach [28, 29], and both are well-known paradigms for LID and SR. Their major contribution is providing an elegant way to convert the cepstral coefficient vector sequence of a variable-length utterance into a fixed-size low-dimensional vector representation. To do so, a Gaussian mixture model is first used to collect the Baum-Welch statistics from the utterance. Then, the first-order statistics from each mixture component are concatenated to form a high-dimensional "supervector" S, which is assumed to obey an affine linear model [27-29]:

$$S = \mathbf{m} + \mathbf{T} \cdot \varphi_S, \tag{3}$$

where $\mathbf{T}$ is a total variability matrix, φ_S is an utterance specific latent variable, and $\mathbf{m}$ denotes a global statistics vector. More specifically, the column vectors of $\mathbf{T}$ form a set of basis vectors spanning a subspace covering the important variability, e.g., the language-specific clues for LID or the speaker-specific clues for SR, and the utterance specific variable φ_S indicates the combination of the variability of the utterance. In this way, a variable-length utterance is represented by a low-dimensional vector φ. Finally, the low-dimensional vector is applied to some well-developed post-processing techniques, such as PLDA, for LID and SR. Since the component models of the i-vector technique can be trained in an unsupervised manner while those of the joint factor analysis approach must be trained along with manual annotation information, the former has become a more attractive paradigm for LID and SR recently. A thorough and entertaining discussion of the i-vector technique and its continued practice can be found in [28, 29].

3.2. I-Vector Based Language Modeling

In our recent work [27], the i-vector technique has been investigated in the context of language modeling for SDR. More concretely, each document d is first represented by a high-dimensional feature vector $\nu_d \in \mathbb{R}^{\beta}$. All of the representative (e.g., lexical-, semantic-, and structure-specific) statistics are encoded in the β-dimensional vector ν_d, which obeys an affine linear model:

$$\nu_d = \mathbf{m} + \mathbf{T} \cdot \varphi_d, \tag{4}$$

where $\mathbf{T} \in \mathbb{R}^{\beta \times \gamma}$ is a total variability matrix, γ is a desired value ($\gamma \ll \beta$), and $\mathbf{m} \in \mathbb{R}^{\beta}$ denotes a global statistics vector. Similarly, the column vectors of $\mathbf{T}$ span a subspace covering the important characteristics for documents. Moreover, each document has a document specific variable $\varphi_d \in \mathbb{R}^{\gamma}$, which encodes the combination of the fundamental variability of the document. Based

on the methodology, a special version is to characterize the representative information of a document only by words. In this respect, each element of the β-dimensional vector corresponds to a distinct word, and the probability of a word w occurring in a document d can be inferred through a log-linear function:

$$P(w \mid d, \mathbf{T}, \mathbf{m}, \varphi_d) = \frac{\exp(\mathbf{T}_w \varphi_d + \mathbf{m}_w)}{\sum_{w' \in V} \exp(\mathbf{T}_{w'} \varphi_d + \mathbf{m}_{w'})}, \quad (5)$$

where $\mathbf{T}_w$ denotes the row vector of $\mathbf{T}$ corresponding to word w and $\mathbf{m}_w$ denotes the statistics value of $\mathbf{m}$ corresponding to word w. The resulting model is termed the i-vector based language model (IVLM). Based on Eqs. (4) and (5), the parameters (i.e., $\mathbf{T}$, φ_d and $\mathbf{m}$) of IVLM can be estimated by maximizing the total likelihood over all training documents:

$$L = \prod_d \prod_{w \in d} \left(\frac{\exp(\mathbf{T}_w \varphi_d + \mathbf{m}_w)}{\sum_{w' \in V} \exp(\mathbf{T}_{w'} \varphi_d + \mathbf{m}_{w'})} \right)^{c(w,d)}, \quad (6)$$

where $c(w,d)$ denotes the number of times word w appearing in document d. Since estimating all the parameters jointly is intractable, we estimate them through an iterative process; i.e., we first estimate $\mathbf{T}$ and $\mathbf{m}$ with fixed φ_d, and then estimate φ_d with fixed $\mathbf{T}$ and $\mathbf{m}$ [27, 30]. More derivational detail and illustration of the IVLM for SDR can be found in [27].

3.3. Query Representation

An obvious deficiency inherent in the i-vector technique for both LID and SR is that, when a given speech utterance consists of only a few acoustic (cepstral coefficient) feature vectors, the low-dimensional representation learned by the i-vector technique is understandably problematic and the performance may degrade dramatically [21-26]. In the context of SDR, a similar deficiency occurs when we interpret a user's information need by a low-dimensional representation, since a query usually composes of only a few words and the representative (e.g., lexical-, semantic-, and structure-specific) statistics would be scarce and vague. With the alleviation of the scarcity problem as motivation, an intuitive idea for deriving a reliable representation for the query is to select a set of references that are "close" to the query to form a conglomerate. As such, an immediate challenge is how to determine the closeness between a candidate reference and the query. Without loss of generality, the closeness score can be one of or the combination of the degrees of acoustic, topical, semantic, syntactic, and/or literal similarities. To conjugate with the special case of the proposed IVLM model (*c.f.* Section 3.2), the closeness is measured by considering only the literal similarity score (e.g., the KLM score (*c.f.* Section 2.2)). Similar to the scenario of applying pseudo-relevance feedback for query expansion and document re-ranking in information retrieval [31-34], the references are selected from the target spoken document collection. In the following, we shed light on three novel methods we propose to derive the new query representation with a set of selected references, $\mathbf{R}=\{r_1,\ldots,r_{|\mathbf{R}|}\}$.

3.3.1 Sample Pooling

A straightforward way to crystallize the idea is to gather a set of selected references to form a conglomerate. Rich statistics can be mined from the conglomerate and rendered by a new β-dimensional vector $\nu_{\hat{Q}}$. To do so, we pool every β-dimensional vector ν_{r_i}, $r_i \in \mathbf{R}$, with its closeness score to distinguish highly correlated references from less correlated references to yield a new representation, $\nu_{\hat{Q}}$, for a given query:

$$\nu_{\hat{Q}} = \alpha \cdot \nu_Q + (1-\alpha) \cdot \left(\sum_{i=1}^{|\mathbf{R}|} s(Q, r_i) \cdot \nu_{r_i} \right), \quad (7)$$

where $s(Q,r_i)$ is the normalized closeness score for r_i. Finally, the query representation, $\varphi_{\hat{Q}}$, can be derived by performing a fold-in process with $\nu_{\hat{Q}}$, $\mathbf{T}$ and $\mathbf{m}$. As such, each query Q has its own IVLM model, including the query specific variable $\varphi_{\hat{Q}}$ and common $\mathbf{T}$ and $\mathbf{m}$. We name this pooling function as the "sample pooling" method.

3.3.2 I-Vector Pooling

Due to the fact that the ultimate goal of the framework is to obtain a new query representation in a low-dimensional feature vector, one reasonable type of manipulation is to craft the representation at the feature level directly. We can first interpret each reference r_i by its own representation φ_{r_i}, which is derived by performing the fold-in process with ν_{r_i}, $\mathbf{T}$ and $\mathbf{m}$. Then, the query representation can be obtained by pooling together all φ_{r_i} weighted by their normalized closeness scores:

$$\varphi_{\hat{Q}} = \alpha \cdot \varphi_Q + (1-\alpha) \cdot \left(\sum_{i=1}^{|\mathbf{R}|} s(Q, r_i) \cdot \varphi_{r_i} \right). \quad (8)$$

We term this pooling function as the "i-vector pooling" method. Comparing the sample pooling method and the i-vector pooling method, it is evident that the former follows the original idea to enrich the statistics, based on which the new query representation is derived, while the latter composes the new query representation at the post stage directly.

3.3.3 Model Pooling

In addition to the above two pooling methods, we also propose a model-level pooling method (hereafter named "model pooling") to derive a distributed representation for a given query:

$$P(w \mid \hat{Q}) = \sum_{i=1}^{|\mathbf{R}|} s(Q, r_i) \cdot P(w \mid r_i, \mathbf{T}, \mathbf{m}, \varphi_{r_i}), \quad (9)$$

where $P(w|r_i, \mathbf{T}, \mathbf{m}, \varphi_{r_i})$ designates the corresponding IVLM model of reference r_i.

3.3.4 Retrieval Model

In the retrieval phase, each query Q will have its own enhanced IVLM-based query model, which can be linearly combined with or used to replace the original query model $P(w|Q)$ in the KL-divergence measure (*c.f.* Eq. (2)) to distinguish relevant documents from irrelevant ones [14, 27].

4. EXPERIMENTAL SETUP

We used the Topic Detection and Tracking collection (TDT-2) [34] in the experiments. 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 obtained for the spoken documents is about 35%. The titles of the Chinese news stories from Xinhua News Agency were used as our test queries in the experiments. Table 1 shows some statistics of the TDT-2 collection. It is known that the way to systemically determine the optimal number of latent variables is still an open issue and needs further investigation. In this paper, the variable γ, that is the number of basis vectors (*c.f.* Section 3.2), is set to 8. The retrieval performance is evaluated in terms of non-interpolated mean average precision (MAP) following the TREC evaluation [35].

In this paper, we also integrate subword-level information cues into the various retrieval models compared in this paper [2, 9, 13, 27]. To do this, syllable pairs are taken as the basic units for indexing in addition to words. The recognition transcript of each spoken document, in form of a word stream, was automatically

Table 1. *Statistics of the TDT-2 collection.*

	TDT-2 (Development Set) 1998, 02~06			
# 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	4,841	153	287.1
Query length (in characters)	8	27	13	14
# Relevant documents per query	2	95	13	29.3

Table 2. *Retrieval results (in MAP) of different retrieval models with word- and subword-level index features.*

	VSM	LSA	SCI	QLM	LDA	IVLM
Word	0.273	0.296	0.270	0.321	0.328	0.336
Subword	0.257	0.384	0.270	0.329	0.377	0.360

Table 3. *Retrieval results (in MAP) of different pooling methods with word- and subword-level index features with respect to the number of references (|**R**|).*

|**R**|	Word			Subword		
	Sample Pooling	I-vector Pooling	Model Pooling	Sample Pooling	I-vector Pooling	Model Pooling
1	0.359	0.360	0.357	0.397	0.398	0.380
3	0.365	0.368	0.364	**0.451**	**0.464**	**0.446**
5	**0.372**	0.373	0.375	0.448	0.459	0.440
10	0.372	**0.374**	**0.379**	0.450	0.460	0.440
15	0.371	0.372	0.377	0.447	0.456	0.440

converted into a stream of overlapping syllable pairs. Then, all the distinct syllable pairs occurring in the spoken document collection were identified to form a vocabulary of syllable pairs for indexing. Accordingly, we can simply use syllable pairs, in replace of words, to represent the spoken documents and construct the associated language models.

5. EXPERIMENTAL RESULTS

In the first set of experiments, we compare several retrieval models, including the vector space model (VSM) [8], the latent semantic analysis (LSA) [6], the semantic context inference (SCI) [3], and the basic LM-based method (i.e., QLM) [14]. The results when using word- and subword-level index features are shown in Table 2. At first glance, QLM in general outperforms the other three methods in most cases, validating the applicability of the LM framework for SDR. Next, we compare two extensions of QLM, namely the latent Dirichlet allocation (LDA) [36] and the IVLM method [27], with QLM. The experimental results are also shown in Table 2. As expected, both LDA and IVLM outperform QLM, and they are almost on par with each other. The results also reveal that LDA and IVLM can give more accurate estimates of the document language models than the empirical ML estimator used in QLM, and thus improve the retrieval effectiveness.

In the next set of experiments, we evaluate the capability of IVLM to enhance query representation in SDR. The results when using different pooling methods (i.e., the sample pooling, i-vector pooling, and model pooling) and different levels of index units, as well as different numbers of references, are shown in Table 3. It is worth noting that, KLM is equivalent to QLM when the query model is simply estimated by an empirical ML estimator. Thus, the baseline performance here is equivalent to that of QLM shown in Table 2. Several observations can be drawn from Table 3. First, it

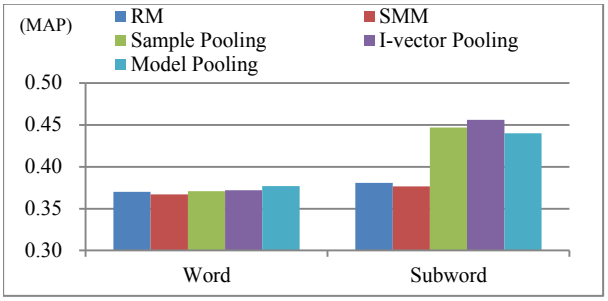

Figure 1. *Retrieval results (in MAP) of i-vector based query representation techniques, relevance model (RM), and simple mixture model (SMM) with word- and subword-level index features.*

is clear that the proposed framework outperforms the baseline KLM model (*c.f.* QLM in Table 2) in all cases. This indicates that IVLM is able to improve the estimation of the query model for better document ranking in SDR. Second, all the proposed pooling methods have comparable performance, and they outperform all of the retrieval models compared in Table 2. Third, the experimental results indicate that the best setting of the number of references is around 5~10 for the word-level index features and 3 for the subword-level index features. Comparing the results in Table 3 with that of IVLM in Table 2, it can be seen that accurate query modeling seems to be more crucial to the retrieval performance than enhanced document modeling. A reason might be that a document is usually long enough for building a reliable representation while an accurate query representation is usually much harder to be inferred from a short query. Moreover, it can also be seen that most retrieval models seem to benefit from the use of subword-level index features, probably because the subword-level index units can shadow the impact of imperfect speech recognition results to some extent.

In the last set of experiments, we further compare the proposed framework with two representative LM-based methods for query reformulation [37], namely the relevance model (RM) and the simple mixture model (SMM), which have been well-practiced and proved their capability in various text IR tasks. The number of the pseudo-relevant documents for RM and SMM (and the references respectively for the proposed three IVLM-based methods) is set to 15. The corresponding retrieval results with different levels of index units are depicted in Figure 1. The results indicate that all of these models deliver comparable performance when using word-level index features, while the proposed three IVLM-based query models outperform the two representative query models by a big margin when using subword-level index units. The reason might be that the model parameters are more accurately estimated, since the observations will increase when fine-grained index units are used to index queries and documents. In sum, the marked results have confirmed that IVLM indeed is efficient and effective for representing queries and documents in SDR.

6. CONCLUSIONS & FUTURE WORK

This paper presents a novel extension of the i-vector based language modeling (IVLM) framework for spoken document retrieval. We have advanced the IVLM framework by proposing several effective models in query representation. The utility of the proposed models have been validated by extensive comparisons with several existing retrieval models. Our future work includes the development of supervised training, incorporation of various representative information or knowledge for larger-scale SDR, and extending the IVLM framework to speech recognition and document summarization.

7. REFERENCES

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

[2] L. S. Lee and B. Chen, "Spoken document understanding and organization," *IEEE Signal Processing Magazine*, 22(5), pp. 42-60, 2005.

[3] C. L. Huang, B. Ma, H. Li, and C. H. Wu, "Speech indexing using semantic context inference," in *Proc. INTERSPEECH*, pp. 717-720, 2011.

[4] W. Naptali, M. Tsuchiya, and S. Nakagawa, "Word co-occurrence matrix and context dependent class in LSA based language model for speech recognition", *International Journal of Computers*, pp. 85-95, 2009.

[5] D. F. Harwath, T. J. Hazen, and J. R. Glass, "Zero resource spoken audio corpus analysis", in *Proc. ICASSP*, pp. 8555-8559, 2013.

[6] S. Deerwester, S. Dumais, G. Furnas, T. Landauer, and R. Harshman, "Indexing by latent semantic analysis", *Journal of the American Society of Information Science*, 41(6), pp. 391-407, 1990.

[7] C. D. Manning, P. Raghavan and H. Schtze, *Introduction to Information Retrieval*, New York: Cambridge University Press, 2008.

[8] G. Salton , A. Wong , and C. S. Yang, "A vector space model for automatic indexing," *Communications of the ACM*, 18(11), pp. 613-620, Nov. 1975

[9] K. Y. Chen, H. M. Wang, B. Chen, and H. H. Chen, "Weighted matrix factorization for spoken document retrieval," in *Proc. ICASSP*, pp. 8530-8534, 2013.

[10] K. S. Jones, S. Walker, and S. E. Robertson. "A probabilistic model of information retrieval: development and comparative experiments (parts 1 and 2)," *Information Processing and Management*, 36(6), pp. 779-840, 2000.

[11] J. M. Ponte and W. B. Croft, "A language modeling approach to information retrieval," in *Proc. SIGIR*, pp. 275-281, 1998.

[12] W. B. Croft and J. Lafferty (eds.), "Language modeling for information retrieval," Kluwer International Series on Information Retrieval, Volume 13, Kluwer Academic Publishers, 2003.

[13] K. Y. Chen, H. M. Wang, and B. Chen, "Spoken document retrieval leveraging unsupervised and supervised topic modeling techniques," *IEICE Transactions on Information and Systems*, pp. 1195-1205, 2012.

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

[15] D. Martinez, O. Plchot, L. Burget, O. Glembek, and P. Matejka, "Language recognition in ivector space," in *Proc. INTERSPEECH*, pp. 861-864, 2011.

[16] M. Soufifar, S. Cumani, L. Burget, and J. Cernocky, "Discriminative classifiers for phonotactic language recognition with ivectors," in *Proc. ICASSP*, pp. 4853-4856, 2012.

[17] L. F. D'Haro, O. Glembek, O. Plchot, P. Matejka, M. Soufifar, R. Cordoba, and J. Cernocky, "Phonotactic language recognition using i-vectors and phoneme posteriogram counts," in *Proc. INTERSPEECH*, pp. 42-45, 2012.

[18] M. Soufifar, M. Kockmann, L. Burget, and O. Plchot, O. Glembek, and T. Svendsen, "Ivector approach to phonotactic language recognition," in *Proc. INTERSPEECH*, pp. 2913-2916, 2011.

[19] O. Glembek, L. Burget, P. Matejka, M. Karafiat, and P. Kenny, "Simplification and optimization of i-vector extraction," in *Proc. ICASSP*, pp. 4516-4519, 2011.

[20] D. Garcia-Romero, and C. Y. Espy-Wilson, "Analysis of i-vector length normalization in speaker recognition systems," in *Proc. INTERSPEECH*, pp. 249-252, 2011.

[21] A. Kanagasundaram, R. Vogt, D. Dean, S. Sridharan, and M. Mason, "I-vector based speaker recognition on short utterances," in *Proc. INTERSPEECH*, pp. 2341-2344, 2011.

[22] A. Kanagasundaram, D. Dean, J. Gonzalez-Dominguez, S. Sridharan, D. Ramos, and J. Gonzalez-Rodriguez, "Improving short utterance based i-vector speaker recognition using source and utterance-duration normalization techniques," in *Proc. INTERSPEECH*, pp. 2465-2469, 2013.

[23] A. K. Sarkar, D. Matrouf, P. M. Bousquet, and J. F. Bonastre, "Study of the effect of i-vector modeling on short and mismatch utterance duration for speaker verification," in *Proc. INTERSPEECH*, pp. 2662-2665, 2012.

[24] P. Kenny, T. Stafylakis, P. Ouellet, Md. J. Alam, and P. Dumouchel, "PLDA for speaker verification with utterances of arbitrary duration," in *Proc. ICASSP*, pp. 7649-7653, 2013.

[25] T. Hasan, R. Saeidi, J. H. L. Hansen, and D. A. van Leeuwen, "Duration mismatch compensation for i-vector based speaker recognition systems," in *Proc. ICASSP*, pp. 7663-7667, 2013.

[26] V. Hautamaki, Y. C. Cheng, P. Rajan, and C. H. Lee, "Minimax i-vector extractor for short duration speaker verification," in *Proc. INTERSPEECH*, pp. 3708-3712, 2013.

[27] K. Y. Chen, H. S. Lee, H. M. Wang, B. Chen, and H. H. Chen, "I-vector based language modeling for spoken document retrieval," in *Proc. ICASSP*, pp. 7083-7087, 2014.

[28] P. Kenny, G. Boulianne, P. Ouellet, and P. Dumouchel, "Joint factor analysis versus eigenchannels in speaker recognition," *IEEE Transactions on Audio, Speech, and Language Processing*, 15(4), pp. 1435-1447, 2007.

[29] P. Kenny, G. Boulianne, P. Ouellet, and P. Dumouchel, "Speaker and session variability in GMM-based speaker verification," *IEEE Transactions on Audio, Speech, and Language Processing*, 15(4), pp. 1448-1460, 2007.

[30] A. L. Maas, and A. Y. Ng, "A probabilistic model for semantic word vectors," in *Proc. NIPS Workshop*, 2010.

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

[32] T. Tao, and C. Zhai, "Regularized estimation of mixture models for robust pseudo-relevance feedback," in *Proc. SIGIR*, pp. 162-169, 2006.

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

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

[35] J. Garofolo, G. Auzanne, and E. Voorhees, "The TREC spoken document retrieval track: A success story," in *Proc. TREC*, pp. 107-129, 2000.

[36] X. Wei and W. B. Croft, "LDA-based document models for ad-hoc retrieval," in *Proc. SIGIR*, pp. 178-185, 2006.

[37] K. Y. Chen, S. H. Liu, B. Chen, E. E. Jan, H. M. Wang, W. L. Hsu, and H. H. Chen, "Leveraging effective query modeling techniques for speech recognition and summarization," in *Proc. EMNLP*, 2014.