Exploiting Graph Regularized Nonnegative Matrix Factorization for Extractive Speech Summarization

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

***Abstract*— Extractive summarization systems attempt to automatically pick out representative sentences from a source text or spoken document and concatenate them into a concise summary so as to help people grasp salient information effectively and efficiently. Recent advances in applying nonnegative matrix factorization (NMF) on various tasks including summarization motivate us to extend this line of research and provide the following contributions. First, we propose to employ graph-regularized nonnegative matrix factorization (GNMF), in which an affinity graph with its similarity measure tailored to the evaluation metric of summarization is constructed and in turn serves as a neighborhood preserving constraint of NMF, so as to better represent the semantic space of sentences in the document to be summarized. Second, we further consider sparsity and orthogonality constraints on NMF and GNMF for better selection of representative sentences to form a summary. Extensive experiments conducted on a Mandarin broadcast news speech dataset demonstrate the effectiveness of the proposed unsupervised summarization models, in relation to several widely-used state-of-the-art methods compared in the paper.**

I. INTRODUCTION

The recent decade has witnessed a booming interest in the research on speech summarization from the speech processing community [1]-[4]. It is largely attributed to continued rapid progress in automatic speech recognition (ASR) as well as the popularity and ubiquity of multimedia associated with spoken documents [5], [6]. As one predominant branch of this line of research, extractive speech summarization targets to select important sentences from an original spoken document according to a predefined summarization ratio, and subsequently concatenate them to form a compact summary that can represent the major theme of the original document. As a result, it provides all locations of important speech segments along with their corresponding transcripts, based on which users can efficiently access and assimilate the semantic content of the document.

One important aspect of extractive summarization is to measure the relevance between the document to be summarized and its constituent sentences, and then select representative sentences based on the corresponding relevance or ranking scores. Among existing unsupervised methods, latent semantic analysis (LSA) and nonnegative matrix factorization (NMF) [7] are typical approaches to extract latent semantic vectors and can be viewed as a kind of representation learning. Essentially, NMF is a parts-based learning for latent semantics, which means that the input sentence only allows linear combination of semantic vectors with positive weights. This is similar to human cognition process for objects and text documents, which has been proven in cognitive studies [8], [9]. Previous work also pointed out that LSA is inferior to NMF, since LSA produces positive and negative weights in linear combination of latent semantic vectors for a given sentence, which is a less meaningful representation of the input sentence in comparison with NMF [10]-[12]. Several studies applied NMF in text summarization; for instance, Lee et al. [12] directly applied NMF to generic summarization and examine its effectiveness, and Wang et al. [11] additionally encoded external knowledge into NMF for multi-document summarization.

With the aforementioned in mind, this paper presents a continuation of this line of research and its contributions are two-fold. First, we propose to exploit graph-regularized nonnegative matrix factorization (GNMF), where an affinity graph with its similarity measure directly linked to the extractive speech summarization task is constructed and in turn serves as a neighborhood preserving constraint of NMF, in order to better capture the semantic structure of sentences in the document to be summarized. Second, we further consider the sparsity and orthogonality constraints on NMF and GNMF for better selection of representative sentences to form a summary. The idea of exploring extensions of NMF has recently drawn much attention and been applied with success to various tasks [10]-[14]. However, to our best knowledge, it has never been extensively explored for representing latent semantics of spoken documents and their constituent sentences in the realm of extractive speech summarization.

The remainder of this paper is organized as follows. We first briefly review related work on extractive summarization in Section II. Section III introduces the notion of NMF, as

well as its extensions that are imposed with sparsity and orthogonal constraints, for use in a summarization task. Afterwards, Section IV sheds light on the novel use of a GNMF-based summarization method for better sentence and document representations. Finally, experiments and conclusions are presented in Sections V and VI, respectively.

II. Related Work

The wide spectrum of extractive speech summarization methods developed so far may roughly fall into three main categories [3], [4], [15]: 1) methods solely based on sentence structure or location cues, 2) methods based on unsupervised statistical analysis without the need of human-annotated ground truth while constructing the summarizers, and 3) methods based on supervised sentence classification.

For the first category, important sentences are selected from some specific parts of a spoken document, e.g., the introductory and/or concluding parts [17]. Such methods can only be applied to limited domains or well-structured documents. These unsupervised methods attempt to extract salient sentences on the basis of prior knowledge about the summarization process conducted by human. Information such as acoustic, phonetic, and prosodic features of spoken words in an automatic transcript, or statistics obtained from the transcript such as word frequency, linguistic features, and recognition confidence, are derived for measuring the importance of sentence and/or the similarity among all sentences, in the spoken document. Methods based on these features have attracted much attention. Representative methods include, but are not limited to, vector space model (VSM) [18], latent semantic analysis (LSA) [18], maximum marginal relevance (MMR) method [19], data reconstruction (DSDR) [20], Markov random walk (MRW) [21], LexRank [22], submodularity-based method [23], integer linear programming (ILP) method [24] and language modeling-based methods [25], [26].

On the other hand, a number of supervised methods using various indicative features and explicit objectives for classification also have been developed, such as Gaussian mixture models (GMM) [27], Bayesian classifiers (BC) [28], support vector machines (SVM) [29] and conditional random fields (CRFs) [30], to name just a few. In these models, the selection of representative sentences is usually cast as a binary classification problem, i.e., to determine whether a given sentence should be included into the summary or not. However, these supervised methods require a set of training documents accompanied with their corresponding hand-crafted summaries (or labeled data) for training the classifiers (or summarizers). In practice, manual annotation is expensive in terms of time and labor. Therefore, even though the performance of unsupervised summarizers is not always comparable to that of supervised ones, the fact that they are easy-to-implement and portable still appeals to both academic research and practical applications. Interested readers may also refer to [3], [4], [15], [16] for a thorough and insightful discussions of major methods that have been successfully developed and applied to a wide variety of text and speech summarization tasks.

III. Nonnegative Matrix Factorization and its Extensions

A. Nonnegative Matrix Factorization (NMF)

NMF is a matrix factorization algorithm that concentrates on the analysis of an input data matrix with nonnegative elements. In formal terms, given a data matrix $X=[x_1, \ldots, x_N] \in R^{V\times N}$, where each column of X is a sample vector, NMF attempts to find two nonnegative matrices $B \in R^{V\times K}$ and $H \in R^{N\times K}$ whose product can well approximate the original matrix X:

$$X \approx BH^T. \tag{1}$$

One of the most commonly-used cost functions that quantify the quality of the approximation is the square of the Euclidean distance between two matrices (viz. the square of the Frobenius norm of the difference between two matrices). Hence, it minimizes the following object function [7]:

$$\min_{B\geq 0, H\geq 0} \| X - BH^T \|_F^2, \tag{2}$$

where $\|.\|_F$ means the matrix Frobenius norm.

Albeit that the objective function expressed in Eq. (2) is convex in B only or H only, it cannot be convex as considering both sets of variables at the same. Thus, it is unfeasible to seek an algorithm to find the global minimum. Lee and Seung [7] presented an iterative algorithm (a.k.a. multiplicative update); the corresponding update formulas that minimize the objective function defined in Eq. (2) are

$$\hat{B}_{ik} = B_{ik} \frac{(XH)_{ik}}{(BH^TH)_{ik}}, \tag{3}$$

and

$$\hat{H}_{jk} = H_{jk} \frac{(X^TB)_{jk}}{(HB^TB)_{jk}}. \tag{4}$$

The above update formulas have been proven to find a local-minimum solution [7].

In practice, the dimension of K is far less than V and N. Thus, the essence of NMF is to seek a low-rank approximation of the original data matrix. We can treat this approximation of the input data matrix in a column-by-column manner as

$$x_j \approx \sum_{k=1}^{K} b_k h_{jk}, \tag{5}$$

where b_k is the k-th column vector of B. Thus, each data vector x_j is approximated by a linear combination of the columns of B, weighted by the corresponding row entries of H. Following the convention, B can be regarded as containing a basis, which is optimized for the linear approximation of the data present in X. On the other hand, H can be regarded as the new representation, and each sample vector can be

represented as a column of H^T with respect to the new basis B. Since relatively few basis vectors are used to represent many data vectors, a good low-rank approximation can only be obtainable if the basis vectors discover the latent structure inherent in the data.

In the context of extractive summarization, the data matrix is the word-by-sentence matrix $S \in R^{V \times N}$ for a document to be summarized, where V is the vocabulary size and N is the number of sentences in the document. By using NMF to decompose the word-by-sentence matrix S into the product of B and H^T, we can obtain the latent semantic representation of a sentence in H. We thus refer to B and H as the basis matrix and the representation matrix, respectively. Note that in this study we append the document vector in the last column of the word-by-sentence matrix S. By doing so, we can simultaneously produce the latent semantic vectors of a document to be summarized and its constituent sentences. Once we obtain the latent semantic representations of the document and its constituent sentences, we can simply apply the cosine similarity to measure the degree of relevance between any sentence of the document to be summarized and the document itself in the hidden semantic space, and subsequently select sentences that have highest relevance scores to form a summary.

B. Sparsity and Orthogonality Constraints on NMF

Despite that NMF has been successfully applied in several applications, it does not always result in parts-based representations [31], [32]. An intuitive way to force NMF to perform parts-based learning is to explicitly control the sparsity property of the entries in the representation matrix. Numerous sparsity constraints have been proposed and used in the literature, and among them the l_1-norm regularization is the most successful constraint of achieving sparsity in representation (viz. to enforce the entries in the representation matrix towards zero). The objective function of NMF to be minimized with the sparsity constraint is thus defined as follows [32]:

$$\min_{B \geq 0, H \geq 0} \| X - BH^T \|_F^2 + \beta \cdot \sum_{j=1}^{N} \| H_j \|_1, \qquad (6)$$

where $\|.\|_1$ is used to designate the 1-norm and N is the number of sentences in a document. The parameter β is to control the trade-off between the reconstruction error of approximation and the sparsity of H. We denote such NMF with the sparsity property by SNMF hereafter. The update formula for the basis matrix B is same as in Eq. (3) and that for the representation matrix H is as follows [33]:

$$\hat{H}_{jk} = H_{jk} \frac{(X^T B)_{jk}}{(HB^T B)_{jk} + \beta}. \qquad (7)$$

On the other hand, previous empirical studies have revealed that imposing the orthogonality constraint on the derivation of NMF will lead to better empirical performance [34]. However, the current decomposition of NMF will not guarantee each column of the basis matrix or the representation matrix to be independent of each other. In order to impose the orthogonality constraint on the derivation of NMF, one can incorporate extra orthogonality constraint into the NMF objective function to ensure that the updated basis and/or representation matrices will preserve orthogonality. One can impose such orthogonal constraint on one-side, either basis matrix B or representation matrix H, or both. However, it is more reasonable to permit use of orthogonal constraint on basis matrix B, i.e. $B^TB=I$. The objective function of NMF to be minimized with an orthogonality constraint is defined as follows:

$$\min_{B \geq 0, H \geq 0, B^T B = I} \| X - BH^T \|_F^2 . \qquad (8)$$

The update formula of representation matrix H is unchanged as in Eq. (4) and that of basis matrix B goes as follows [36]:

$$\hat{B}_{ik} = B_{ik} \frac{(XH)_{ik}}{(BB^T XH)_{ik}}. \qquad (9)$$

The more detailed descriptions and proofs of the update formula in Eq. (8) can be found in [35], [36]. We will refer to such NMF with the orthogonality property as ONMF hereafter.

Likewise, in the context of extractive summarization we can obtain the sentence and document representations at the same time by appending document vector in the last column of word-by-sentence matrix S. With the additional extra constraint (sparsity or orthogonality) on the NMF formulation, we can more precisely capture the latent semantic representation of sentences and document in expectation of resulting in better summarization performance. To the best of our knowledge, such extensions of NMF have yet to be fully explored in the field of speech summarization research.

IV. GRAPH-REGULARIZED NONNEGATIVE MATRIX FACTORIZATION (GNMF)

When NMF is adopted for representation learning of data, a major deficiency is that it fails to take the geometrical structure of the data space in account. To tackle this limitation, Cai et al. [14] proposed an efficient algorithm to produce a new representation matrix which is more amenable to intrinsic geometric structure of the data space. More precisely, they incorporated NMF with a geometric-based regularizer, which was constructed by a graph of nearest neighbors in the data. This idea is feasible due to the local invariance assumption [37]; that is, if two data points are close in the intrinsic geometric structure of the data space, then the new representations of these two data points with respect to a new basis are close to each other as well. Formally, the objective function of graph-regularized NMF is defined as follows:

$$\min_{B \geq 0, H \geq 0} \| X - BH^T \|_F^2 + \lambda \cdot Tr\left(H^T (D - W) H\right), \qquad (10)$$

where $Tr(.)$ denotes the trace of a matrix; the regularization parameter λ control the trade-off between accuracy of approximation and smoothness of the new representation; W is the symmetric weight matrix (or graph) which is built from

the nearest neighbor graph of the original data; and D is a diagonal matrix whose entries are column (or row) sums of W. This kind of graph-regularized NMF is designated as GNMF. The update formulas for Eq. (10) are expressed by [14]

$$\hat{B}_{ik} = B_{ik} \frac{(XH)_{ik}}{(BH^T H)_{ik}}, \tag{11}$$

and

$$\hat{H}_{jk} = H_{jk} \frac{(X^T B + \lambda WH)_{jk}}{(HB^T B + \lambda DH)_{jk}}. \tag{12}$$

It is easy to validate that when $\lambda = 0$ the update formulas shown in Eqs. (11) and (12) are identical to that shown in Eqs. (3) and (4), respectively. There are many choices for designing the nearest neighborhood relationship in the weight matrix W. In the context of extractive summarization, one intuitive way to compute the weights between any pair of sentences in W is the cosine similarity, where the sentences are represented in vector form and each element in vector is the product of term frequency and inverse document frequency (TF-IDF). However, a more sensible choice is to use the metric for summarization evaluation (such as ROUGE [40]) to represent the closeness of each pair of sentences. Thus, by virtue of GNMF, we not only obtain the parts-based representations (or semantic vectors) for spoken sentences and document but also retain the intrinsic geometric information in the corresponding latent semantic space.

GNMF has been applied with success in several tasks like face recognition and document clustering [14]. However, as far as we know, this work is the first attempt to leverage GNMF for extractive speech summarization.

V. EXPERIMENTS

A. Experimental Setup

The summarization dataset was compiled from a publicly available broadcast news corpus (MATBN) collected by the Academia Sinica and the Public Television Service Foundation of Taiwan between November 2001 and April 2003 [39]. It has been segmented into separate stories and transcribed manually. Each story contains the speech of one news anchorperson, as well as several field reporters and interviewees. A subset of 205 broadcast news stories was selected for the summarization experiments, with 20 documents being the test set and the remaining 185 documents as the held-out development set. A subset of 25-hour speech data in MATBN was used to bootstrap the acoustic model training with the minimum phone error rate (MPE) criterion and the training data selection scheme. The vocabulary size is about 72K words.

Three subjects were asked to create summaries of the 205 spoken documents for the summarization experiments as the reference (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 adopt the widely-used ROUGE metrics [40], including ROUGE-1 (unigram), ROUGE-2 (bigram), and ROUGE-L (the longest common subsequence). All the experimental results reported hereafter are obtained by calculating the F-scores [38] 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.

Each news story consists of two kinds of transcripts, viz. TD and SD, where TD denotes the results obtained based on the manual transcripts of spoken documents and SD denotes the results using the speech recognition transcripts that may contain speech recognition errors. The parameters of all unsupervised methods compared in the paper were optimized based on the development set.

B. Experimental Results

In the first set of experiments, we evaluate the performance of several popular unsupervised methods for extractive speech summarization, including 1) the position-based method, i.e. LEAD [17], 2) the vector-based methods, i.e. VSM, LSA[18], MMR [19], DSDR [20] (consisting of linear and non-linear versions which are denoted by DSDR-lin and DSDR-non respectively), 3) the graph-based methods, viz. MRW [21], LexRank [22], the submodularity-based method (denoted by Submodularity hereafter) [23], 4) the optimization-based method, i.e., the ILP method [24], and 5) the language model-based method [25].

The corresponding summarization results of these unsupervised methods are depicted in Table I. Several noteworthy observations can be drawn. First, among the various vector-based methods (viz. VSM, LSA, NMF, DSDR-lin and DSDR-non), NMF performs better than VSM and LSA and is on par with DSDR-non in the TD case. However, this situation is reversed in the SD case, presumably due to the influence of imperfect speech recognition. DSDR-non is the best-performing method in the TD case, while LSA performs slightly better than all other vector-based methods in the SD case. Second, the graph-based methods (viz. MRW, LexRank, and Submodularity) are quite competitive to each other and perform better than the vector-based methods in both the TD and SD cases. Third, MMR, an extension of VSM that performs removal of redundant information as an additional criterion, can work as well as the various graph-based methods in the TD case, delivering even better performance than the latter ones for the SD case. Fourth, it is evident that ULM yields a performance comparable to other unsupervised methods, confirming the applicability of the language modeling approach for speech summarization. Fifth, the ILP method turns out to be the best-performing one among all unsupervised summarization methods compared here for the TD case, but it only offers mediocre performance for the SD case. Lastly, there is a sizable gap between the TD and SD cases, indicating a room for further improvements. We may seek remedies, such as robust indexing techniques, to compensate for imperfect speech recognition [41], [42].

TABLE I
SUMMARIZATION RESULTS OF THE BASELINE NMF METHOD AND SEVERAL WIDELY-USED UNSUPERVISED METHODS

		ROUGE-1	ROUGE-2	ROUGE-L
TD	NMF	0.370	0.233	0.289
	LEAD	0.310	0.194	0.276
	VSM	0.347	0.228	0.290
	LSA	0.362	0.233	0.316
	MMR	0.368	0.248	0.322
	ULM	0.411	0.298	0.371
	DSDR-lin	0.353	0.225	0.301
	DSDR-non	0.386	0.235	0.310
	MRW	0.412	0.282	0.358
	LexRank	0.413	0.309	0.363
	Submodularity	0.414	0.286	0.363
	ILP	**0.442**	**0.337**	**0.401**
SD	NMF	0.326	0.175	0.266
	LEAD	0.255	0.117	0.221
	VSM	0.342	0.189	0.287
	LSA	0.345	0.201	0.301
	MMR	**0.366**	**0.215**	**0.315**
	ULM	0.364	0.210	0.307
	DSDR-lin	0.247	0.121	0.196
	DSDR-non	0.342	0.183	0.261
	MRW	0.332	0.191	0.291
	LexRank	0.305	0.146	0.254
	Submodularity	0.332	0.204	0.303
	ILP	0.348	0.209	0.306

TABLE II
SUMMARIZATION RESULTS OF THE NMF METHODS AND ITS EXTENSIONS

		ROUGE-1	ROUGE-2	ROUGE-L
TD	NMF	0.370	0.233	0.289
	SNMF	0.417	0.286	0.335
	ONMF	0.435	0.318	0.360
	GNMF-TFIDF	0.451	0.336	0.374
	GNMF-ROUGE2	**0.459**	**0.360**	**0.377**
SD	NMF	0.326	0.175	0.266
	SNMF	0.354	0.210	0.279
	ONMF	0.364	0.223	0.291
	GNMF-TFIDF	0.366	0.222	0.293
	GNMF-ROUGE2	**0.370**	**0.237**	**0.295**

TABLE III
SUMMARIZATION RESULTS OF THE GNMF METHODS WITH CONSIDERING ADDITIONAL SPARSITY AND ORTHOGONALITY CONSTRAINTS.

		ROUGE-1	ROUGE-2	ROUGE-L
TD	GNMF	0.459	0.360	0.377
	SGNMF	0.461	0.362	0.385
	OGNMF	0.463	0.365	0.390
	SOGNMF	**0.468**	**0.370**	**0.398**
SD	GNMF	0.370	0.237	0.295
	SGNMF	0.372	0.242	0.309
	OGNMF	0.371	0.245	0.311
	SOGNMF	**0.377**	**0.246**	**0.315**

In the second set of experiments, we assess the effectiveness of various extensions of NMF using extra constraints, including sparsity (SNMF), orthogonality (ONMF) and graph-regularization (GNMF), to enhance the latent semantic representations of sentences and documents. From the results shown in Table II, it is evident that both the sparsity and orthogonality constraints imposed on the NMF formulation can considerably improve the summarization performance, which validates the proposition of using the sparsity and orthogonality constraints to improve the hidden semantic representation of sentences. Interestingly, we also observe that imposing the orthogonality constraint on NMF can result in better summarization performance than the sparsity constraint for both the TD and SD cases. On the other hand, we have experimented with two different weight matrices (viz. nearest-neighbor graphs) in GNMF: one is TFIDF (denoted by GNMF-TFIDF) and the other is to leverage ROUGE-2 metric (denoted by GNMF-ROUGE2). When consulting the results shown in Table II, we find that both GNMF-TFIDF and GNMF-ROUGE2 outperform all other variants of NMF for both the TD and SD cases. Particularly, GNMF-ROUGE2 brings about marked improvements in terms of the ROUGE-2 evaluation metric for both the TD and SD cases. It implies that that such a metric is more suitable to be a criterion when building the nearest-neighbor graph in the application domain of summarization. These empirical findings confirm the claim that the intrinsic geometric structure does exist in the conventional NMF-based sentence space, and GNMF can successfully capture this information to improve the latent semantic representations of the document to be summarized and its constituent sentences.

In the last set of experiments, it is interesting to see what happen when GNMF method (here we only consider the ROUGE-2 based affinity graph) with additionally considering sparsity and orthogonality constraints. Thus, we step by step to impose the extra constraints on GNMF method, i.e. GNMF with sparsity alone (denoted as SGNMF), orthogonality alone (OGNMF) and both constraints (denoted as SOGNMF). The corresponding results of those particular considerations are listed in Table III. As in expectation, we observe that all those extra constraints on GNMF will give help to improve the summarization performance whether in TD and SD cases. Especially, when considering the sparsity and orthogonality constraints on GNMF simultaneously (viz. SOGNMF), we can achieve the best summarization performance in comparison with all other approaches no matter in TD and SD cases. These empirical observations corroborate that somehow the additional sparsity and orthogonality considerations will aid to capture more meaningful latent representation so as to improve the summarization performance.

VI. CONCLUSIONS

In this paper, we have explored several novel extensions of NMF for use in extractive speech summarization. These extensions involve imposing extra constraints on the derivation of NMF, including sparsity, orthogonality and graph-regularization. Experimental evidence validates that the various methods instantiated from our summarization framework outperform several existing state-of-the-art unsupervised methods for extractive speech summarization. In particular, graph-regularized NMF (GNMF) can simultaneously encode the hidden semantic representations and the intrinsic geometric information cues within the original document space, thereby boosting the summarization performance. In particular, the ROUGE-2 based weight

matrix leads to the best performance improvement. There are three potential directions for future work. First, we plan to simultaneously integrate different kinds of constraints into the derivation of NMF (for example, the sparsity constraint in conjunction with the orthogonality constraint) so as to further improve the empirical effectiveness of NMF. Second, we would like to explore the possibility of combining NMF and existing start-of-the-art methods (such as ILP) for better summarization performance. Lastly, we are also interested in investigating more robust indexing techniques for representing spoken documents in order to bridge the performance gap between the TD and SD cases.

VII. Acknowledgements

This research is supported in part by the "Aim for the Top University Project" of National Taiwan Normal University (NTNU), sponsored by the Ministry of Education, Taiwan, and by the Ministry of Science and Technology, Taiwan, under Grants MOST 103-2221-E-003-016-MY2, MOST 104-2221-E-003-018-MY3, MOST 104-2911-I-003-301.

References

[1] S. Furui, *et al.*, "Speech-to-text and speech-to-speech summarization of spontaneous speech," *IEEE Transactions on Speech and Audio Processing,* 12(4), pp. 401–408, 2004.

[2] K. McKeown, *et al.*, "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, 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, *et al.*, "Fundamental technologies in modern speech recognition," *IEEE Signal Processing Magazine*, 29(6), pp. 16–17, 2012

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

[7] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, 401, pp. 788-791, 1999.

[8] E. Wachsmuth, M. W. Oram, and D. I. Perrett, "Recognition of objects and their component parts: responses of single units in the temporal cortex of the macaque," *Cerebral Cortex*, 4, pp. 509-522, 1994.

[9] S. E. Palmer, "Hierarchical structure in perceptual representation," *Cognitive Psychology*, 9, pp. 441-474, 1977.

[10] W. Xu, X. Liu and Y. Gong, "Document clustering based on non-negative matrix factorization," in *Proc. of The Annual International ACM SIGIR Conference*, pp. 267–273, 2003.

[11] D. D. Wang, *et al.*, "Multi-document summarization via sentence-level semantic analysis and symmetric matrix factorization," in *Proc. of The Annual International ACM SIGIR Conference*, pp. 307-314, 2008.

[12] J. H. Lee, *et al.*, "Automatic generic document summarization based on non-negative matrix factorization," *Information Processing & Management*, 45(1), pp. 20-34, 2009.

[13] W. Y. Chu, J. W. Hung and Berlin Chen, "Modulation spectrum factorization for robust speech recognition,", in *Proc. of APSIPA Annual Summit and Conference*, Xian, China, October 18-21, 2011.

[14] D. Cai, X. He, J. Han and T. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8), pp. 1548-1560, Dec. 2010.

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

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

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

[18] Y. Gong and X. Liu, "Generic text summarization using relevance measure and latent semantic analysis," in *Proc. of The Annual International ACM SIGIR Conference*, pp. 19–25, 2001.

[19] J. Carbonell and J. Goldstein, "The use of MMR, diversity based reranking for reordering documents and producing summaries," in *Proc. of The Annual International ACM SIGIR Conference*, pp. 335–336, 1998.

[20] Z. Y. He, *et al.*, "Document summarization based on data reconstruction," in *Proc. of AAAI Conference on Artificial Intelligence*, pp. 620-626, 2012.

[21] X. Wan and J. Yang, "Multi-document summarization using cluster-based link analysis," in *Proc. of The Annual International ACM SIGIR Conference*, pp. 299–306, 2008.

[22] 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.

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

[24] R. McDonald, "A study of global inference algorithms in multi-document summarization," in *Proc. European conference on IR research*, pp. 557–564, 2007.

[25] S.-H. Liu, *et al.*, "Combining relevance language modeling and clarity measure for extractive speech summarization," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23(6), pp. 957-969, 2015

[26] 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.

[27] 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.

[28] J. Kupiec, *et al.*, "A trainable document summarizer," in *Proc. of The Annual International ACM SIGIR Conference*, pp. 68–73, 1995.

[29] A. Kolcz, *et al.*, "Summarization as feature selection for text categorization," in *Proc. ACM Conference on Information and Knowledge Management*, pp. 365–370, 2001.

[30] 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.

[31] J. Kim and H. Park, "Sparse nonnegative matrix factorization for clustering," *CSE Technical Reports*, Georgia Institute of Technology, 2008.

[32] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *Journal of Machine Learning Research*, 5, pp. 1457-1469, 2004.

[33] J. Eggert and E. Korner, “Sparse coding and NMF,” in *Proc. of IEEE International Joint Conference on Neural Network*, pp. 2529-2533, 2004.
[34] H. L. Li, *et al.*, “Non-negative matrix factorization with orthogonality constraintsand its application to raman spectroscopy,” *Journal of VLSI Signal Processing*, 48(1), pp. 83-97, 2007.
[35] S. Choi, “Algorithms for orthogonal nonnegative matrix factorization,” in *Proc. of IEEE International Joint Conference on Neural Networks*, pp. 1828-1832, 2008.
[36] C. Ding, *et al.*, “Orthogonal nonnegative matrix t-factorizations for clustering,” in *Proc. of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 126-135, 2006.
[37] M. Belkin and P. Niyogi, “Laplacian eigenmaps and spectral techniques for embedding and clustering,” *Advances in Neural Information Processing Systems*, pp. 585-591, MIT Press, 2001.
[38] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval: The Concepts and Technology behind Search*, ACM Press, 2011.
[39] H.-M. Wang, *et al.*, “MATBN: A Mandarin Chinese broadcast news corpus,” *International Journal of Computational Linguistics and Chinese Language Processing*, 10(2), pp. 219–236, 2005.
[40] C.-Y. Lin, “ROUGE: Recall-oriented Understudy for Gisting Evaluation,” 2003. Available: http://haydn.isi.edu/ROUGE/.
[41] 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.
[42] C. Chelba, *et al.*, “Soft indexing of speech content for search in spoken documents,” *Computer Speech & Language*, 21(3), pp. 458–478, 2007.