LEVERAGING MANIFOLD LEARNING FOR EXTRACTIVE BROADCAST NEWS SUMMARIZATION

Shih-Hung Liu, Kuan-Yu Chen, Berlin Chen, Hsin-Min Wang, Wen-Lian Hsu*

*National Taiwan Normal University, Taiwan
Institute of Information Science, Academia Sinica, Taiwan

ABSTRACT

Extractive speech summarization is intended to produce a condensed version of the original spoken document by selecting a few salient sentences from the document and concatenate them together to form a summary. In this paper, we study a novel use of manifold learning techniques for extractive speech summarization. Manifold learning has experienced a surge of research interest in various domains concerned with dimensionality reduction and data representation recently, but has so far been largely under-explored in extractive text or speech summarization. Our contributions in this paper are at least twofold. First, we explore the use of several manifold learning algorithms to capture the latent semantic information of sentences for enhanced extractive speech summarization, including isometric feature mapping (ISOMAP), locally linear embedding (LLE) and Laplacian eigenmap. Second, the merits of our proposed summarization methods and several widely-used methods are extensively analyzed and compared. The empirical results demonstrate the effectiveness of our unsupervised summarization methods, in relation to several state-of-the-art methods. In particular, a synergy of the manifold learning based methods and state-of-the-art methods, such as the integer linear programming (ILP) method, contributes to further gains in summarization performance.

Index Terms— Manifold learning, nonlinear dimension reduction, local invariance, extractive summarization

1. INTRODUCTION

In the recent past, research on speech summarization has witnessed a booming interest from the speech processing community [1]-[4]. This is largely attributed to the dramatic advances in automatic speech recognition (ASR) as well as the popularity and ubiquity of multimedia associated with spoken documents [5], [6]. Extractive speech summarization, one predominant branch of this research area, aims to choose salient sentences from an original spoken document according to a predefined summarization ratio and concatenate them together to form a compact summary representing the major theme of the original document. As a result, it facilitates to provide all locations of important speech segments along with their corresponding transcripts for users to access and assimilate the most important semantic content, potentially saving a great deal of time.

Generally, the most common perspective of extractive summarization is to measure the relevance degree between each sentence in a given document to be summarized and the document itself. Based on the estimated relevance or ranking scores, a set of salient sentences can thereby be selected. The relevance score is computed by using either shallow lexical features or latent semantic features; however, the latter usually shows superior to the former. Latent semantic analysis (LSA) [13] and nonnegative matrix factorization (NMF) [8] are two representative classic methods, which can be exploited to extract latent semantic features and also be regarded as instantiations of representation learning.

The motivation of this work is to assume that a low-dimensional manifold structure is embedded in the original high-dimensional textual vector space (represented using the term frequency and inverse document frequency (TF-IDF) mechanism). Traditional dimension reduction approaches, such as LSA and NMF, do not target at discovering the corresponding low-dimensional manifold structure, and thus might lead to an inaccurate measure of the relevance degree between a pair of sentence and document, such as that using the cosine similarity. It seems that a feasible remedy is to leverage manifold learning to explore the inherent manifold structures of high-dimensional textual vectors. Furthermore, the deduced low-dimensional representations can be readily used to determine the relevance degree between the sentences and the document to be summarized for better selecting salient sentences to be included in the summary.

To date, several well-developed manifold learning algorithms have been developed and applied to various domains involving dimensionality reduction (or representation learning), showing encouraging results. This study follows such a general line of research and the main contributions are at least two-fold. First, we empirically explore several manifold learning algorithms, including isometric feature mapping (ISOMAP) [30], locally linear embedding (LLE) [29] and Laplacian eigenmap [31], to reduce the dimensionality of original high-dimensional sentence vectors, as well as to capture the latent semantic information and retain the intrinsic geometric structure pertaining to the representations of sentences in a spoken document for use in extractive speech summarization. Second, the utilities of our proposed summarization methods and several well-established methods are extensively analyzed and compared. When integrating the manifold learning approach based method with state-of-the-art methods, such as the integer linear programming (ILP) method, we can even achieve the best performance. Although numerous latent semantic or matrix factorization based methods have been extensively investigated in various multimedia-related tasks, as far as we are aware, there appears to be a dearth of research that considers capturing the information of intrinsic manifold structures with respect to the representations of sentences in a document for extractive summarization.

2. RELATED WORK

The methodologies of extractive speech summarization may be broadly clustered in two groups: ASR-based and non-ASR-based methods [7]. The former conducts summarization with automatic (imperfect) transcripts generated by an ASR system, and thereby can harness and extend the power of a broad range of well-practiced text summarization methods to the context of speech summarization. On the contrary, the latter endeavors to estimate the importance of a spoken sentence and/or its relevance to the spoken document to be summarized based on the acoustic or prosodic features derived from the raw speech signal. That is, this school of methods does not resort to any ASR system for generating the corresponding transcripts. In general, the empirical performance of the latter is usually inferior to that of the former. Nevertheless, how to systematically and effectively combine the strengths of both schools of methods for better performance in speech summarization still awaits further exploration. In this study, we will focus exclusively on the design and development of ASR-based methods.

The wide spectrum of ASR-based extractive speech summarization methods developed so far may roughly fall into three main categories [3], [4], [10]: 1) methods simply based on sentence structure or location cues, 2) methods based on unsupervised statistical mechanisms without the need of human-annotated ground truth while constructing the associated 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 [12]. Such methods can only be applied to some limited domains or document structures known in advance. Next, the unsupervised methods attempt to extract salient sentences simply on the basis of various acoustic, phonetic and prosodic features of spoken words in the automatic transcript, or statistical information residing in the transcript, such as word frequency, linguistic score and recognition confidence, among others. Representative methods include, but are not limited to, the vector space model (VSM) [13], latent semantic analysis (LSA) [13], maximum marginal relevance (MMR) method [14], data reconstruction (DSDR) [15], nonnegative matrix factorization (NMF) [16], Markov random walk (MRW) [17], LexRank [18], submodularity-based method [19], integer linear programming (ILP) method [20] and language modeling methods [21][22]. Third, the selection of important sentences performed by supervised methods is usually casted as a binary classification problem by incorporating various kinds of indicative features and explicit objectives, i.e., to determine whether a given sentence should be included into the summary or not. Gaussian mixture models (GMM) [23], Bayesian classifiers (BC) [24], support vector machines (SVM) [25] and conditional random fields (CRFs) [26], to name just a few, are good representatives. Interested readers may also refer to [3], [4], [10], [11] for comprehensive reviews and new insights into the mainstream methods that have been successfully developed and applied to a wide range of text/speech summarization tasks.

3. MANIFOLD LEARNING TECHNIQUES

Manifold learning generally is a process of dimensionality reduction, which usually assumes that the data instances are intrinsically low-dimensional ones but instead are lying in a very high-dimensional space. Classic dimensionality reduction algorithms, such as principle component analysis (PCA) [27] and multidimensional scaling (MDS) [28], seek to infer a low-dimensional embedding for each data instance and meanwhile preserve the variance of data instances in the original input space. However, most of the classic algorithms fail to capture the geometric structure in the original high-dimensional space. In order to detect the underlying hidden structure of data instances embedding in the original high-dimensional space, many manifold learning algorithms have been proposed, such as Locally Linear Embedding (LLE) [29], ISOMAP [30] and Laplacian eigenmap [31], among others. All of these algorithms follow the so-called local invariance perspective [31], i.e., the nearby instances in the originally high-dimensional space are likely to have similar embeddings in the transformed low-dimensional space. In the following, we introduce three manifold learning algorithms that we will formulate and crystallize for use in extractive speech summarization.

3.1. Isometric feature mapping

The isometric feature mapping (ISOMAP) algorithm builds on the classic multidimensional scaling (MDS) and further seeks to preserve the intrinsic geometry of data instances, as those captured by the geodesic manifold distances between all pairs of data instances. The major challenge is how to appropriately estimate the geodesic distance between a pair of far-away data instances when only their input-space distance is given. Accordingly, for neighboring data instances, input space distance provides a good approximation to geodesic distance; for a pair of far-away instances, their geodesic distance can be approximated by adding up a sequence of "short hops" between neighboring instances. These approximations are computed efficiently by finding shortest paths in a graph with edges connecting neighboring data instances.

To be concrete, the complete isometric feature mapping process essentially proceeds with three main steps. The first step is to decide the K-nearest neighbors of any pair of data instances in the input space, where K is any arbitrary number. These neighborhood relations are represented as a weighted graph G over the data instances, with edges of weight between neighboring instances. In the second step, ISOMAP estimates the geodesic distances between all pairs of instances by computing their shortest path distances in the graph G. The final step applies classic MDS to the matrix of graph distances, constructing embeddings of all data instances in a d-dimensional Euclidean space that best preserves the manifold's estimated intrinsic geometry [30]. Through the three-step process, ISOMAP can produce a globally optimal low-dimensional representation for high-dimensional data instances.

3.2. Locally linear embedding

Different from ISOMAP, the locally linear embedding (LLE) algorithm eliminates the need to estimate pairwise distances between widely separated data instances. LLE attempts to discover nonlinear structure inherent in data by exploiting the local symmetries of linear reconstructions. Basically, the LLE algorithm is based on simple geometric intuitions and employs a three-step process. The first one is to assign K-neighbors to each data point x_i by K-nearest neighbors (KNN, K is a predefined parameter). The second step is to compute a set of weights that can be used to linearly reconstruct the original data instance as faithfully as possible from its neighbors [29]:

$$c(W)=\sum_{i=1}^{N}\Big| x_i-\sum_{x_j\in neighbor(x_i)} w_{ij}\cdot x_j\Big|^2, \tag{1}$$

where N is the total number of data instances and x_j is the neighbor of data instances x_i. The weight w_{ij} is an entry of the matrix W and denotes the contribution of the j-th data instance to the reconstruction of the i-th one. The optimal weight w_{ij} can be found by solving a least-square problem [29]. The third step of LLE is to derive the corresponding low-dimensional vectors y_i, which in turn can be reconstructed by referencing to the weight w_{ij} that can be obtained by minimizing the following cost function:

$$c(Y)=\sum_{i=1}^{N}\Big| y_i-\sum_{y_j\in neighbor(y_i)} w_{ij}\cdot y_j\Big|^2, \tag{2}$$

where the notions are the same as those used in Eq. (1). The embedding cost in Eq. (2) defines a quadratic form in the vectors y_i

and can be minimized by a sparse N-by-N eigenvalue problem [29]. Consequently, the LLE algorithm constructs low-dimensional representations of data instances with neighborhood-preserving mapping, while the resulting representations can be subsequently used in downstream tasks.

3.3. Laplacian eigenmap

Essentially, the Laplacian eigenmap algorithm is a geometrically motivated dimensionality reduction approach. Similar to ISOMAP and LLE, Laplacian eigenmap is also performed through a process that contains three steps. The first and second steps are to choose a desired edge measure and in turn to construct the neighborhood-preserving graph (or matrix). After these two steps, a weight matrix W is constructed. Then, the objective of the third step is to compute eigenvalues and eigenvectors for the following generalized eigen-decomposition system [31]:

$$Ly = \lambda Dy, \tag{3}$$

where L=D-W is the Laplacian matrix and D is a diagonal weight matrix whose entries are column (or row, since W is symmetric) sums of W. Generally, Laplacian eigenmap is related to a family of spectral clustering and widely-applied in various tasks [32]. More detailed analysis of the Laplacian eigenmap algorithm and its theoretical justifications can be found in [31].

3.4. Analytic comparison of manifold learning techniques

Although ISOMAP, LLE and Laplacian eigenmap share the commonality of exploiting a three-step process for dimensionality reduction so as to unveil the inherent manifold structures of spoken documents and sentences, they actually have been developed under disparate philosophies. First, the process of ISOMAP resembles that of MDS except that the distance between each pair of data instances (such as spoken documents or sentences) is measured by the shortest path algorithm that is used to achieve a geodesic approximation and turns out to be nonlinear dimensionality reduction, instead of linear dimensionality reduction. Second, since the global calculation of all pairwise shortest paths is time-consuming, LLE thus attempts to select only a few nearby instances pertaining to each instance of interest to alleviate such a problem. The basic assumption behind LLE is that if each instance can be linear reconstructed by its nearby instances in the original high-dimensional space, then it can also be linear reconstructed by these nearby samples in the deduced low-dimensional space with the same set of combination weights. Finally, Laplacian eigenmap suggests that if two instances are close to each other in the original high-dimension space, then they should also be close to each other in a low-dimensional space. Hence, the objective function of Laplacian eigenmap is defined by the summation of the distance between any pair of nearest low-dimensional embeddings weighted by their corresponding distance in the original high-dimensional space. Consequently, the derivation of the associated embeddings of data instances boils down to a kind of generalized eigen-decomposition of the Laplacian matrix [31].

3.5. Leveraging manifold learning for summarization

To recap, given a set of high-dimensional input data instances, the aforementioned manifold learning algorithms aim at discovering low-dimensional representations of the data instances, retaining the intrinsic manifold structure originally embedding in the high-dimension data space. In the context of extractive speech summarization, sentences in a given document to be summarized are first represented as high-dimensional vectors (i.e., data instances). Each element of the sentence vector is associated with the term frequency multiplied by the inverse document frequency (TF-IDF) of a word in the vocabulary. Then, the corresponding locality-preserving weight matrix can be produced based on the cosine similarity measure between any pair of sentences. Each pair of sentences is connected if the corresponding similarity measure is greater than a predefined threshold. After that, by virtue of a manifold learning algorithm (e.g., ISOMAP, LLE, and Laplacian eigenmap), we can obtain the corresponding low-dimensional representations for each sentence and the document itself. Finally, based on measuring the cosine similarity of each of the sentences within the document to be summarized and the document itself, with respect to their low-dimensional embeddings, we can thus select salient sentences according to their similarity (relevance) scores to the document so as to form a concise summary.

4. EXPERIMENTS

4.1. Experimental setup

The summarization dataset is 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 [34], which has been segmented into separate stories and transcribed manually. Three subjects were asked to create summaries of the 205 spoken documents for the summarization experiments, which were used 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 [35], 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 [33] of these ROUGE metrics, respectively. 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, i.e., 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.

4.2. Experiment results

At the outset of this subsection, we evaluate the performance levels of several existing widely-applied unsupervised methods for extractive speech summarization, including 1) the vector-based methods, i.e., VSM, LSA[13], MMR [14], NMF [36], CBOW [37], and SG [37], 2) the graph-based methods, i.e., MRW [17] and LexRank [18], 3) the optimization-based method, i.e., the submodularity-based method (denoted by Submodularity hereafter) [19] and the ILP method [20], and 4) the unigram language model-based method (ULM) [21]. The corresponding summarization results of these unsupervised methods are illustrated in Table 1, where the best result within each column (corresponding to a specific evaluation metric) for either the TD case or the SD case is type-set boldface. We may draw attention to several observations here. First, among the various vector-based methods (i.e., VSM, LSA, NMF, CBOW and SG), NMF performs better than VSM and LSA, while on par with CBOW and SG in the TD case. However, this situation is reversed in the SD case probably due to the negative influence of imperfect speech recognition. CBOW is the top-performing method in the TD case, while SG performs slightly better than all the other vector-based methods in the SD case.

Table 1. Summarization results achieved by several widely-used unsupervised methods.

		ROUGE-1	ROUGE-2	ROUGE-L
TD	VSM	0.347	0.228	0.290
	LSA	0.362	0.233	0.316
	NMF	0.370	0.233	0.289
	MMR	0.368	0.248	0.322
	ULM	0.411	0.298	0.371
	CBOW	0.382	0.249	0.322
	SG	0.371	0.239	0.311
	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	VSM	0.342	0.189	0.287
	LSA	0.345	0.201	0.301
	NMF	0.326	0.175	0.266
	MMR	**0.366**	**0.215**	**0.315**
	ULM	0.364	0.210	0.307
	CBOW	0.362	0.214	0.314
	SG	0.364	0.215	0.311
	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 2. Summarization results achieved by several manifold learning based methods.

		ROUGE-1	ROUGE-2	ROUGE-L
TD	NMF	0.370	0.233	0.289
	MDS	0.402	0.289	0.329
	ISOMAP	0.424	0.323	0.350
	Laplacian	0.422	0.319	**0.364**
	LLE	**0.437**	**0.337**	0.360
SD	NMF	0.326	0.175	0.266
	MDS	0.347	0.209	0.286
	ISOMAP	0.369	0.231	0.313
	Laplacian	0.367	0.217	0.289
	LLE	**0.375**	**0.240**	**0.299**

Table 3. Summarization results achieved by the synergy of LLE and ILP.

		ROUGE-1	ROUGE-2	ROUGE-L
TD	LLE	0.437	0.337	0.360
	ILP	0.442	0.337	**0.401**
	ILP-LLE	**0.465**	**0.369**	0.380
SD	LLE	0.375	0.240	0.299
	ILP	0.348	0.209	0.306
	ILP-LLE	**0.377**	**0.245**	**0.314**

Second, the various graph-based methods (i.e., MRW and LexRank) are quite competitive to one another and perform better than the vector-based methods in both the TD and SD cases. Third, MMR, which presents an extension of VSM by taking the 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 performance comparable to other unsupervised methods, confirming the applicability of the language modeling framework for speech summarization. Fifth, the ILP method stands out in performance among all the unsupervised summarization methods compared here in the TD case, but it only offers mediocre performance in the SD case. Lastly, there is a sizable gap between the TD and SD cases, indicating room for further improvements. We may seek remedies, such as robust indexing techniques, to compensate for imperfect speech recognition [38], [39].

In the second set of experiments, we assess the utilities of several manifold learning approaches, including the isometric feature mapping (ISOMAP) based method, the locally linear embedding (LLE) based method and Laplacian eigenmap (Laplacian) based method, in comparison with two matrix factorization-based strong baselines (viz. NMF and MDS). When looking into Table 2, the empirical results reveal that all of these manifold learning based methods achieve better performance than NMF and MDS in both the TD and SD cases. The results suggest that performing dimensionality reduction by additionally considering the intrinsic geometric structure of data instances (i.e., sentences or documents) can capture more precisely latent semantic information and thus boost the summarization performance. Among these three manifold learning based summarization methods, LLE is the top-performing one in both the TD and SD cases, while the performance levels of ISOMAP and Laplacian are almost the same. Further, compared to the results shown in Tables 1 and 2, LLE is quite competitive to the start-of-the-art ILP method in TD case and performs the best in the SD case. Those results corroborate our assumption that the manifold learning techniques, to some extent, can unveil the low-dimensional manifold structures inherent in the original high-dimensional vector space, and thus lead to a more accurate measure (with the cosine similarity metric) of the relevance degree between a pair of sentence and document.

Furthermore, it is worthwhile to pursuit the combination of the manifold learning based methods and the global optimization (i.e. ILP) method for better summarization performance. Since LLE is the top-performing method among these manifold learning based methods, we thus choose it to be integrated with the ILP method, determining the similarity (relevance) measures among sentences and between sentences and the document to be summarized. The corresponding results are depicted in Table 3. We find that the integration of LLE and ILP will boost performance compared to that using LLE or ILP in isolation, achieving the best results in both the TD and SD cases. This indicates that LLE enables to capture more precise latent semantic representation and such information indeed can aid ILP to perform global selection with redundancy consideration in the latent semantic space, so as to further improve summarization performance.

5. CONCLUSION AND OUTLOOK

In this paper, we have empirically explored several well-developed manifold learning algorithms for use in extractive speech summarization, including isometric feature mapping (ISOMAP), locally linear embedding (LLE) and Laplacian eigenmap. Experimental evidence supports that the various methods instantiated from our summarization framework outperform several existing widely-used unsupervised methods and quite comparable with some start-of-the-art methods. As for future work, we plan to explore more sophisticated manifold learning algorithms and their synergies with start-of-the-art summarization methods. We are also interested in investigating robust indexing techniques for representing spoken documents in order to bridge the performance gap between the TD and SD cases.

6. ACKNOWLEDGEMENT

This research is supported in part by the "Aim for the Top University Project" of National Taiwan Normal University (NTNU), sponsored by the Ministry of Education, Taiwan, and by the Ministry of Science and Technology, Taiwan, under Grants MOST 104-2221-E-003-018-MY3, MOST 104-2911-I-003-301 and MOST 105-2221-E-003-018-MY3.

7. REFERENCES

[1] S. Furui et al., “Speech-to-text and speech-to-speech summarization of spontaneous speech,” *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 4, pp. 401–408, 2004.

[2] K. McKeown 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*, vol. 5, no. 2–3, pp. 103–233, 2011.

[5] S. Furui et al., “Fundamental technologies in modern speech recognition,” *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 16–17, 2012.

[6] D. O'Shaughnessy et al., “Speech information processing: Theory and applications,” *Proceedings of the IEEE*, vol. 101, no. 5, pp. 1034–1037, 2013.

[7] X. Zhu et al., “Summarizing multiple spoken documents: finding evidence from untranscribed audio,” in *Proc. of Joint Conference of ACL and IJCNLP*, pp. 549–557, 2009.

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

[9] W. Xu et al., “Document clustering based on non-negative matrix factorization,” in *Proc. of the Annual International ACM SIGIR Conference*, pp. 267–273, 2003.

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

[11] 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.

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

[13] 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.

[14] 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.

[15] Z. Y. He, et al., “Document summarization based on data reconstruction,” in *Proc. of AAAI Conference on Artificial Intelligence*, pp. 620–626, 2012.

[16] J. H. Lee et al., “Automatic generic document summarization based on non-negative matrix factorization,” *Information Processing & Management*, vol. 45, no. 1, pp. 20–34, 2009.

[17] 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.

[18] G. Erkan and D. R. Radev, “LexRank: Graph-based lexical centrality as salience in text summarization,” *Journal of Artificial Intelligent Research*, vol. 22, no. 1, pp. 457–479, 2004.

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

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

[21] 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*, vol. 23, no. 6, pp. 957–969, 2015

[22] 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.

[23] M. A. Fattah and F. Ren, “GA, MR, FFNN, PNN and GMM based models for automatic text summarization,” *Computer Speech & Language*, vol. 23, no. 1, pp. 126–144, 2009.

[24] J. Kupiec et al., “A trainable document summarizer,” in *Proc. of the Annual International ACM SIGIR Conference*, pp. 68–73, 1995.

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

[26] 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.

[27] I. T. Jolliffe, *Principal Component Analysis*, Springer-Verlag, New York, 1989

[28] T. F. Cox, and M. A. A. Cox, *Multidimensional scaling*, Chapman and Hall, 2001.

[29] S. Roweis and L. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[30] J. Tenenbaum et al., “A global geometric framework for nonlinear dimensionality reduction,” *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[31] M. Belkin and P. Niyogi, “Laplacian eigenmaps and spectral techniques for embedding and clustering,” *Advances in Neural Information Processing Systems 14*, pp. 585-591, MIT Press, 2001.

[32] U. V. Luxburg, A tutorial on spectral clustering, *Statistics and Computing*, vol. 17, no.5, pp.395-416, 2007.

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

[34] H.-M. Wang, et al., “MATBN: A Mandarin Chinese broadcast news corpus,” *International Journal of Computational Linguistics and Chinese Language Processing*, vol. 10, no. 2, pp. 219–236, 2005.

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

[36] J. H. Lee et al., “Automatic generic document summarization based on non-negative matrix factorization,” *Information Processing & Management*, vol. 45, no. 1, pp. 20-34, 2009.

[37] T. Mikolov et al., “Efficient estimation of word representations in vector space,” in *Proc. of ICLR*, pp. 1–12, 2013.

[38] S. Xie and Y. Liu, “Using N-best lists and confusion networks for meeting summarization” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, no. 5, pp. 1160–1169, 2011.

[39] C. Chelba, et al., “Soft indexing of speech content for search in spoken documents,” *Computer Speech & Language*, vol. 21, no. 3, pp. 458–478, 2007.