Multi-Layer Mutually Reinforced Random Walk with Hidden Parameters for Improved Multi-Party Meeting Summarization

Yun-Nung Chen and Florian Metze

School of Computer Science, Carnegie Mellon University
5000 Forbes Ave., Pittsburgh, PA 15213, USA

{yvchen, fmetze}@cs.cmu.edu

Abstract

This paper proposes an improved approach of summarization for spoken multi-party interaction, in which a multi-layer graph with hidden parameters is constructed. The graph includes utterance-to-utterance relation, utterance-to-parameter weight, and speaker-to-parameter weight. Each utterance and each speaker are represented as a node in the utterance-layer and speaker-layer of the graph respectively. We use terms/topics as hidden parameters for estimating utterance-to-parameter and speaker-to-parameter weight, and topical similarity between utterances as the utterance-to-utterance relation. By within- and between-layer propagation in the graph, the scores from different layers can be mutually reinforced so that utterances can automatically share the scores with the utterances from the speakers with similar terms/topics and similar utterances. For both ASR output and manual transcripts, experiments confirmed the efficacy of including hidden parameters and involving speaker information in the multi-layer graph for summarization. We find that choosing latent topics as hidden parameters can significantly reduce the running time and does not hurt the performance.

Index Terms: summarization, multi-party meeting, mutual reinforcement, random walk

1. Introduction

Speech summarization is important for spoken or even multimedia documents, which are more difficult to browse than text, and has therefore been investigated in the past [1]. Recent work has been increasingly directed towards conversational speech such as telephone conversation and multi-party meeting [2, 3, 4, 5, 6]. In this work, we perform extractive summarization on the output of automatic speech recognition (ASR) and corresponding manual transcripts of multi-party academic meeting recordings [7].

A general approach has been found be very successful [8], in which each utterance in the document can be represented as a sequence of terms, and the importance score of the utterance can integrate the score from the grammatical structure of the utterance, some statistical measure (such as TF-IDF), some linguistic measure (e.g., POS tags), a confidence score, and an n-gram score for each term in the utterance. For each document, the utterances to be used in the summary are then selected based on this score.

Many approaches to text summarization focus on graph-based methods to compute lexical centrality of each utterance, in order to extract summaries [9, 10]. Speech summarization carries intrinsic difficulties due to the presence of recognition errors, spontaneous speech effects, and lack of segmentation.

In recent work, we proposed a graphical structure to rescore the importance scores of utterances, which can model the topical coherence between utterances using a random walk process within documents [4, 11, 12, 13]. Unlike lecture and news, meeting recordings contain spoken multi-party interactions, so that the "speaker importance" scores can be added to the estimation of the importance of individual utterance [14]. However, the utterance-to-speaker relation is not easy to model [12], so this paper additionally includes a middle layer to provide the common parameters between utterances and speakers. Then the proposed multi-layer mutually reinforced random walk can compute the importance of hidden parameters and then increase the scores of utterances similar to other utterances based on hidden parameter modeling. It models intra- and inter-speaker topics together in the graph by automatically propagating scores from the utterance- and speaker-layer to hidden-parameter-layer for improving meeting summarization [10, 12, 15].

Section 2 describes the construction of the multi-layer graph and the algorithms about computing the importance of utterances with integration of within- and between-layer propagation through hidden parameters. Section 3 shows the results of applying proposed approaches, evaluates the effectiveness of hidden parameters, and discusses the difference parameter types and relation types result in for both ASR and manual transcripts. Section 4 concludes the achievements.

2. Proposed Approach

We first preprocess the utterances in all meetings by applying word stemming, stop word removal, and noise utterance filtering [16]. For extractive summarization, we set a cut-off ratio to retain only the most important utterances to form the summary of each document based on the "importance" of utterances. Thus, we formulate the utterance selection problem as computing the importance of each utterance. Then we construct a multi-layer graph to compute the importance for all utterances, speakers, and hidden parameters in utterance-layer, speaker-layer, and hidden-parameter-layer respectively. In the multi-layer directed graph, each utterance is represented by a node in the utterance layer, and the edges between these are weighted by topical or lexical similarity described in Section 2.4. Each speaker in the meeting is a node in speaker layer. The hidden parameters represent terms or latent topics, and the edges between different layers are weighted by the relation between the two nodes as described in Section 2.3.

The basic idea is that an utterance similar to more important utterances should be more important [11], so the importance of each utterance considers the scores propagated from other utterances weighted by the similarity between them. In this ap-
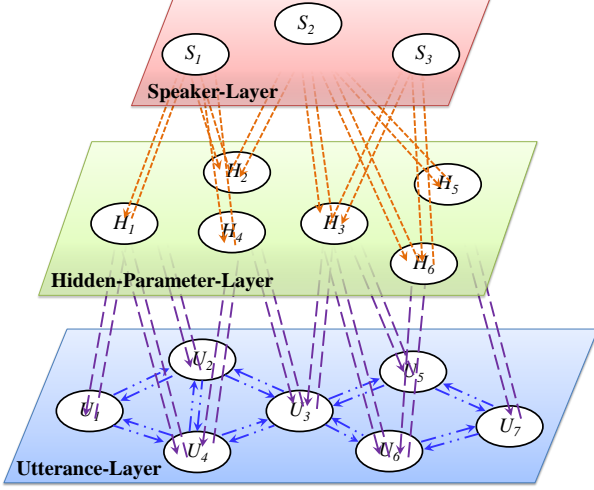
Figure 1: *A simplified example of the multi-layer graph with hidden parameters, where a speaker S_i is represented as a speaker node, an utterance U_j is represented as an utterance node, and a hidden parameter H_l is represented as a parameter node of the graph. There are three different types of edges corresponding to different relations (utterance-to-utterance, utterance-to-parameter, and speaker-to-parameter).*

proach, the propagated scores additionally consider speaker information, which is automatically modeled via hidden parameters in the graph. Figure 1 shows a simplified example for such a multi-layer graph with hidden parameters, in which there are a speaker-layer, an utterance-layer, and a hidden-parameter-layer.

2.1. Parameters from Topic Model

Topic models such as probabilistic semantic analysis (PLSA) and latent Dirichlet allocation (LDA) have been widely used to analyze the semantics of documents based on a set of latent topics [17, 19]. Given a set of documents $\{d_j, j = 1, 2, ..., J\}$ and all terms $\{t_i, i = 1, 2, ..., M\}$ they include, PLSA uses a set of latent topic variables, $\{T_k, k = 1, 2, ..., K\}$, to characterize the "term-document" co-occurrence relationships. The PLSA model can be optimized using the EM algorithm, by maximizing a likelihood function [17]. We utilize two parameters, latent topic significance (LTS) and latent topic entropy (LTE), for proposed approach [18].

Latent topic significance (LTS) for a given term t_i with respect to a topic T_k can be defined as

$$\text{LTS}_{t_i}(T_k) = \frac{\sum_{d_j \in D} n(t_i, d_j) P(T_k \mid d_j)}{\sum_{d_j \in D} n(t_i, d_j)[1 - P(T_k \mid d_j)]}, \quad (1)$$

where $n(t_i, d_j)$ is the occurrence count of term t_i in a document d_j. Thus, a higher $\text{LTS}_{t_i}(T_k)$ indicates that the term t_i is more significant for the latent topic T_k.

Latent topic entropy (LTE) for a given term t_i can be calculated from the topic distribution $P(T_k \mid t_i)$,

$$\text{LTE}(t_i) = -\sum_{k=1}^{K} P(T_k \mid t_i) \log P(T_k \mid t_i), \quad (2)$$

where the topic distribution $P(T_k \mid t_i)$ can be estimated from the topic model. $\text{LTE}(t_i)$ is a measure of how the term t_i is focused on a few topics, so a lower latent topic entropy implies the term carries more topical information.

2.2. Statistical Measures of a Term

The statistical measure of a term t_i, $s(t_i, d)$ measures the importance of t_i such as TF-IDF. In this work, it can be defined based on $\text{LTE}(t_i)$ as $s(t_i, d) = \gamma \cdot n(t_i, d) / \text{LTE}(t_i)$, where γ is a scaling factor such that $s(t_i, d)$ lies within the interval $[0, 1]$, so the score $s(t_i, d)$ is inversely proportion to the latent topic entropy $\text{LTE}(t_i)$. This measure outperformed the very successful "significance score" [18, 8] in speech summarization, so we use the LTE-based statistical measure as our baseline.

2.3. Between-Layer Relation via Hidden Parameters

Given a set of utterances $\{U_i, i = 1, 2, ..., |U|\}$ and a set of speakers $\{S_j, j = 1, 2, |S|\}$, where a speaker node in the graph is represented by combining all utterances from the speaker, for the middle layer of the graph, we choose two different types of hidden parameters $\{H_l, l = 1, 2, ..., L\}$ – terms in the vocabulary and latent topics from the topic model, which are the common parameters shared with other two layers. We derive the weights of between-layer relation as the significance of the hidden parameters given the utterance or the speaker.

2.3.1. Term Layer

We use terms from the vocabulary to represent hidden parameters, and the weight is computed as follows.

$$\text{Lex}(U_i, H_l) = \frac{1}{|U_i|} \sum_{t \in U_i} idf_t, \quad (3)$$

where $|H_l|$ is the size of vocabulary and idf_t is the inverse document frequency (IDF) of term t. Hence $\text{Lex}(U_i, H_l)$ is average TF-IDF of term t in the utterance U_i. $\text{Lex}(S_j, H_l)$ can be derived in the similar way.

2.3.2. Topic Layer

We use the trained topic model to represent each latent topic as a node in the layer of hidden parameters. Based on latent topics, which are the common parameters shared by utterances and speakers, we can weight between-layer edges below.

$$\text{Topic}(U_i, H_l) = \frac{1}{|U_i|} \sum_{t \in U_i} \text{LTS}_t(T_l), \quad (4)$$

where $|H_l|$ is set to be K, which is the number of latent topics. $\text{Topic}(S_j, H_l)$ can be computed in the similar way.

2.4. Within-Layer Relation via Similarity

We compute two different types of similarity between utterances based on topical and lexical distribution.

Within a document d, we can first compute the probability that the topic T_k is addressed by an utterance U_i,

$$P(T_k \mid U_i) = \frac{\sum_{t \in U_i} n(t, U_i) P(T_k \mid t)}{\sum_{t \in U_i} n(t, U_i)}. \quad (5)$$

Then an asymmetric topical similarity $\text{Sim}(U_i, U_j)$ for utterances U_i to U_j (with direction $U_i \rightarrow U_j$) can be defined by accumulating $\text{LTS}_t(T_k)$ in (1) weighted by $P(T_k \mid U_i)$ for all terms t in U_j over all latent topics,

$$\text{Sim}(U_i, U_j) = \sum_{t \in U_j} \sum_{k=1}^{K} \text{LTS}_t(T_k) P(T_k \mid U_i), \quad (6)$$

where the idea is similar to generative probability in information retrieval. We call this generative significance of U_i given U_j.

2.5. Multi-Layer Mutually Reinforced Random Walk

For each document d, we construct a directed multi-layer graph G containing utterance set, speaker set, and hidden parameter set to compute the importance of each utterance. $G = \langle V_U, V_S, V_H, E_{UU}, E_{UH}, E_{SH} \rangle$, where $V_U = \{U_i \in d\}$, $V_S = \{S_i \in d\}$, $V_H = \{H_i\}$, $E_{UU} = \{e_{ij} \mid U_i, U_j \in V_U\}$, $E_{UH} = \{e_{ij} \mid U_i \in V_U, H_j \in V_H\}$, and $E_{SH} = \{e_{ij} \mid S_i \in S, H_j \in V_H\}$. E_{UU}, E_{UH}, and E_{SH} correspond the utterance-to-utterance, utterance-to-parameter, and speaker-to-parameter relation respectively [10].

We compute $W_{UU} = [w_{U_i,U_j}]_{|V_U| \times |V_U|}$, where w_{U_i,U_j} is from $\text{Sim}(U_i, U_j)$. Word overlap between utterances may be sparse due to recognition errors, so it's possible that topical similarity via topic models can be more informative than lexical similarity. $W_{UH} = [w_{U_i,H_j}]_{|V_U| \times |V_H|}$, where w_{U_i,H_j} is the either from $\text{Topic}(U_i, H_j)$ or $\text{Lex}(U_i, H_j)$. Similarly, $W_{SH} = [w_{S_i,H_j}]_{|V_S| \times |V_H|}$, where w_{S_i,H_j} is the either from $\text{Topic}(S_i, H_j)$ or $\text{Lex}(S_i, H_j)$. Row-normalization and column-normalization are applied to obtain L_{UU}, L_{UH}, and L_{SH} as normalized affinity matrices [20].

Traditional random walk integrates the original scores and the scores propagated from other utterance nodes [11, 14, 21]. Here the proposed approach additionally considers the speaker information and integrates importance propagated from speaker nodes via hidden parameters to model intra- and inter-speaker relation automatically. The algorithm is detailed as follows.

Here we use mutually reinforced random walk to propagate the scores based on external mutual reinforcement between different layers and internal importance propagation within the layer.

$$
\begin{cases}
F_H^{(t+1)} = (1-2\alpha)F_H^{(0)} + \alpha \cdot L_{UH}^T F_U^{(t)} + \alpha \cdot L_{SH}^T F_S^{(t)} \\
F_U^{(t+1)} = (1-2\alpha)F_U^{(0)} + 2\alpha \cdot L_{UU}^T L_{UH} F_H^{(t)} \\
F_S^{(t+1)} = (1-2\alpha)F_S^{(0)} + 2\alpha \cdot L_{SH} F_H^{(t)}
\end{cases},
$$
(7)

where $F_H^{(t)}$, $F_U^{(t)}$, and $F_S^{(t)}$ denote the importance scores of hidden parameter set V_H, utterance set V_U, and speaker set V_S in t-th iteration respectively. In the algorithm, they are the interpolations of the initial importance and the scores propagated from another layer, where $F_H^{(t)}$ integrates the scores propagated from both utterance-layer and speaker-layer to measure the importance of each hidden parameter.

For utterance set, $L_{UU}^T L_{UH} F_H^{(t)}$ is the score propagated from hidden parameter set according to utterance-to-parameter relation and then weighted by utterance-to-utterance similarity L_{UU}. Similarly, nodes in the speaker-layer also include the scores propagated from hidden-parameter-layer but without within-layer propagation, because the speaker-to-speaker relation cannot be estimated accurately, and if we set the uniform distribution for the speaker-to-speaker relation, the results may not be influenced a lot. Then $F_H^{(t+1)}$, $F_U^{(t+1)}$, and $F_S^{(t+1)}$ can be mutually updated by the latter parts in (7) iteratively. The algorithm will converge and then (8) can be satisfied [10].

$$
\begin{cases}
F_H^* = (1-2\alpha)F_H^{(0)} + \alpha \cdot L_{UH}^T F_U^* + \alpha \cdot L_{SH}^T F_S^* \\
F_U^* = (1-2\alpha)F_U^{(0)} + 2\alpha \cdot L_{UU}^T L_{UH} F_H^* \\
F_S^* = (1-2\alpha)F_S^{(0)} + 2\alpha \cdot L_{SH} F_H^*
\end{cases}
$$
(8)

We can solve F_H^* as below.

$$
\begin{aligned}
F_H^* =\ & (1-2\alpha)F_H^{(0)} && (9)\\
& + \alpha \cdot L_{UH}^T \left((1-2\alpha)F_U^{(0)} + 2\alpha \cdot L_{UU}^T L_{UH} F_H^* \right) \\
& + \alpha \cdot L_{SH}^T \left((1-2\alpha)F_S^{(0)} + 2\alpha \cdot L_{SH} F_H^* \right) \\
=\ & (1-2\alpha)F_H^{(0)} \\
& + \alpha(1-2\alpha) \cdot L_{UH}^T F_U^{(0)} + \alpha(1-2\alpha) \cdot L_{SH}^T F_S^{(0)} \\
& + 2\alpha^2 \cdot L_{UH}^T L_{UU}^T L_{UH} F_H^* + 2\alpha^2 \cdot L_{SH}^T L_{SH} F_H^* \\
=\ & \Big((1-2\alpha)F_H^{(0)} e^T \\
& + \alpha(1-2\alpha) \cdot (L_{UH}^T F_U^{(0)} e^T + L_{SH}^T F_S^{(0)} e^T) \\
& + 2\alpha^2 \cdot (L_{UH}^T L_{UU}^T L_{UH} F_H^* + L_{SH}^T L_{SH} F_H^*) \Big) F_H^* \\
=\ & M F_H^*,
\end{aligned}
$$

where the $e = [1, 1, ..., 1]^T$. It has been shown that the closed-form solution F_H^* of (9) is the dominant eigenvector of M [22], which is the eigenvector corresponding to the largest absolute eigenvalue of M. Then we can compute the solution of F_U^* using (8), which denotes the updated importance scores for all utterances. Similar to the PageRank [23], the solution can also be obtained by iteratively updating $F_H^{(t)}$, $F_U^{(t)}$, and $F_S^{(t)}$.

We set $F_U^{(0)}$ to be the baseline score after normalization such that the sum of them is equal to 1, $F_H^{(0)} = e^T/|V_H|$, and $F_S^{(0)} = e^T/|V_S|$, which means we assume all hidden parameters and speakers in the document have equal importance in the beginning.

3. Experiments

3.1. Corpus

The corpus used here is a sequence of academic meetings, which features largely overlapping participant sets and topics of discussion. For each meeting, SmartNotes [3] was used to record both the audio from each participant, as well as the notes. The meetings were transcribed both manually and using a speech recognizer; the word error rate is around 44%. In this paper we use 10 meetings held from April to June of 2006. On average, each meeting had about 28 minutes of speech. Across these 10 meetings, there were 6 unique participants; each meeting featured between 2 and 4 of these participants (average: 3.7). Total number of utterances is 9837 across 10 meetings. In this paper, we empirically set $\alpha = 0.45$ for the unsupervised experiments because $(1 - 2 \times 0.45)$ is a proper damping factor [23, 21]. Note that for previous approaches, α is set to be 0.9 such that damping factor is $(1 - 0.9)$. We use PLSA as our topic model and set the number of topics to be 32.

The reference summaries are given by the set of "noteworthy utterances": two annotators manually labelled the degree (three levels) of "noteworthiness" for each utterance, and we extract the utterances with the highest level of "noteworthiness" to form the summary of each meeting. Note that this experiment does not consider redundancy of information but focuses on the importance of utterances. After performing the algorithm, iteratively selecting utterances based on redundancy can achieve final summarized results. In the following experiments, for each meeting, we extract about 10% and 20% of the number of terms as the shorter summary considering reasonable ratios of meeting data, which are different from previous experiments [12], where the ratio is set to be 30%.

Table 1: *The results of all proposed approaches and maximum relative improvement with respect to the baseline (%).*

F-measure		10% Summary				20% Summary			
		ASR		Manual		ASR		Manual	
		R-1	R-L	R-1	R-L	R-1	R-L	R-1	R-L
(a)	Baseline: LTE	44.27	43.32	43.10	41.99	44.73	44.11	42.30	41.68
(b)	Two-Layer MRRW-WBP (LexSim)	45.82	44.82	44.89	44.00	45.64	44.78	43.92	43.26
(c)	Two-Layer MRRW-WBP (TopicSim)	46.53	45.77	44.46	43.57	45.18	44.34	43.95	43.20
(d)	Multi-Layer MRRW-Term	**50.36**	**49.62**	**49.36**	**48.42**	48.02	47.35	45.69	44.95
(e)	Multi-Layer MRRW-Topic	50.00	49.16	48.68	47.82	**48.35**	**47.78**	**46.69**	**45.99**
Max Relative Improvement		+13.76	+14.54	+14.52	+15.31	+8.09	+8.32	+10.38	+10.34

3.2. Evaluation Metrics

Our automated evaluation utilizes the standard DUC (Document Understanding Conference) evaluation metric, ROUGE [24], which represents recall over various n-grams statistics from a system-generated summary against a set of human generated summaries. F-measures for ROUGE-1 (unigram; R-1) and ROUGE-L (longest common subsequence; R-L) can be evaluated in exactly the same way.

3.3. Results

Table 1 shows the performance achieved by all proposed approaches. Row (a) is the baseline, which uses an LTE-based statistical measure to compute the importance of utterances. Row (b) is the result after applying two-layer mutually reinforced random walk (MRRW) using within- and between-layer propagation (WBP), which uses lexical similarity to measure utterance-to-utterance relation [12]. Row (c) is the same as row (c) except it uses topical similarity for utterance-to-utterance relation. Row (d) and (e) are the results of proposed multi-layer MRRW with terms (Term) and latent topics (Topic) as hidden parameters respectively.

3.3.1. Effectiveness of Hidden Parameters

We can see all performance of the multi-layer graph with hidden parameters (row (d)-(e)) significantly outperforms two-layer approaches (row (b)-(c)) for both ASR and manual transcripts. The largest improvement is from multi-layer MRRW-Term (row (d)) for 10% summary and from multi-layer MRRW-Topic (row (e)) for 20% summary. The results confirm that hidden parameters carry more information and help summarization.

3.3.2. Comparing Types of Hidden Parameters

For shorter summary, there's no obvious difference between using lexical and topical similarity in two-layer approaches (row (b)-(c)). Therefore, we analyze the difference between choosing terms (similar to lexical information) and latent topics (similar to topical information) as hidden parameters.

When using terms as hidden parameters, we have about 2000 nodes in the middle layer. The selected terms correspond to words in vocabulary, which are collected from the corpus after word stemming and stop word removal. After performing the algorithm, each term in the hidden-parameter-layer has the score indicating the term significance in terms of utterances and speakers. Then the importance of each utterance integrates the scores propagated from other utterances and from the terms the utterance contains, where the latter carries speaker information. Using terms as parameters allows scores to be computed more

accurately, and more accurate scores are better for extracting shorter summary so that MRRW-Term (row (d)) performs better than MRRW-Topic (row (e)) for 10% summary.

In the case of latent topics as hidden parameters, the size of middle layer is equal to the number of topics, which is 32. We experiment using different numbers of topics, 16, 32, 64, and 128, but the different settings do not influence the results a lot, which means that topic models can still capture the most important topics for all settings. With the proposed algorithm, the score of each latent topic considers utterance-to-parameter and speaker-to-parameter relation to indicate the topic significance in the multi-layer graph. With latent topics as parameters, scores can be computed more generally. Topic models capture not only lexically but conceptually similar information so that using latent topics as parameters is more suitable for extracting a 20% summary. Hence the results show that MRRW-Topic (row (e)) performs better than MRRW-Term (row (d)).

3.3.3. Running Time Reduction

MRRW-Term includes larger size of hidden-parameter-layer and the proposed algorithm needs to compute the eigenvector of $M_{2000 \times 2000}$ in (9), which requires longer running time. MRRW-Topic can use smaller size of hidden-parameter-layer by modeling similar terms in the same latent topic so that the running time can be significantly reduced. Comparing the results between them, reducing the running time does not hurt the performance a lot for 10% summary and even results in further improvement for 20% summary. It shows that our proposed algorithm is effective and can be applied in a practical way.

4. Conclusions and Future Work

Extensive experiments are performed and evaluation shows that multi-layer mutually reinforced random walk with hidden parameters can model importance of utterances and speakers through hidden parameters in the multi-layer graph. The speaker information can be automatically included in importance of utterances by between-layer propagation, achieving about 13% and 8% relative improvement compared to the LTE baseline for shorter summary of ASR and manual transcripts respectively. Reducing the running time by using latent topics as hidden parameters does not hurt the performance, showing the practicality and effectiveness of the proposed algorithm. In the future, we plan to model additional parameters such as prosodic features and to integrate different types of features in a single multi-layer graph.

5. References

[1] L.-S. Lee and B. Chen, "Spoken document understanding and organization," *IEEE Signal Processing Magazine*, vol. 22, no. 5, pp. 42–60, 2005.

[2] D. Harwath and T. J. Hazen, "Topic identification based extrinsic evaluation of summarization techniques applied to conversational speech," in *Proceedings of IEEE ICASSP*, 2012.

[3] S. Banerjee and A. I. Rudnicky, "An extractive-summarizaion baseline for the automatic detection of noteworthy utterances in multi-party human-human dialog," in *Proceedings of IEEE SLT*, 2008.

[4] Y.-N. Chen and F. Metze, "Intra-speaker topic modeling for improved multi-party meeting summarization with integrated random walk," in *Proceedings of NAACL-HLT*, 2012, number 382–385.

[5] F. Liu and Y. Liu, "Using spoken utterance compression for meeting summarization: A pilot study," in *Proceedings of IEEE SLT*, 2010.

[6] P.-Y. Hsueh and J. D. Moore, "Improving meeting summarization by focusing on user needs: a task-oriented evaluation," in *Proceedings of Intelligent User Interfaces*, 2009.

[7] Y. Liu, S. Xie, and F. Liu, "Using N-best recognition output for extractive summarization and keyword extraction in meeting speech," in *Proceedings of IEEE ICASSP*, 2010.

[8] S. Furui, T. Kikuchi, Y. Shinnaka, and C. Hori, "Speech-to-text and speech-to-speech summarization of spontaneous speech," *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 4, pp. 401–408, 2004.

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

[10] X. Cai and W. Li, "Mutually reinforced manifold-ranking based relevance propagation model for query-focused multi-document summarization," *IEEE Transactions on Acoustics, Speech and Language Processing*, vol. 20, pp. 1597–1607, 2012.

[11] Y.-N. Chen, Y. Huang, H.-Y. Lee, and L.-S. Lee, "Spoken lecture summarization by random walk over a graph constructed with automatically extracted key terms," in *Proceedings of InterSpeech*, 2011.

[12] Y.-N. Chen and F. Metze, "Two-layer mutually reinforced random walk for improved multi-party meeting summarization," in *Proceedings of IEEE SLT*, 2012.

[13] F. Lin, *Scalable Methods for Graph-Based Unsupervised and Semi-Supervised Learning*, Ph.D. thesis, Carnegie Mellon University, 2012.

[14] Y.-N. Chen and F. Metze, "Integrating intra-speaker topic modeling and temporal-based inter-speaker topic modeling in random walk for improved multi-party meeting summarization," in *Proceedings of InterSpeech*, 2012.

[15] N. Garg, B. Favre, K. Reidhammer, and D. Hakkani-Tür, "Clusterrank: A graph based method for meeting summarization," in *Proceedings of InterSpeech*, 2009.

[16] M. F. Porter and et al., "An algorithm for suffix stripping," Program, 1980, http://tartarus.org/~martin/PorterStemmer/.

[17] T. Hofmann, "Probabilistic latent semantic indexing," in *Proceedings of SIGIR*, 1999.

[18] S.-Y. Kong and L.-S. Lee, "Semantic analysis and organization of spoken documents based on parameters derived from latent topics," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, no. 7, pp. 1875–1889, 2011.

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

[20] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[21] W. Hsu and L. Kennedy, "Video search reranking through random walk over document-level context graph," in *Proceedings of MM*, 2007.

[22] A. Langville and C. Meyer, "A survey of eigenvector methods for web information retrieval," *SIAM Review*, 2005.

[23] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," in *Proceedings of WWW*, 1998.

[24] C. Lin, "Rouge: A package for automatic evaluation of summaries," in *Workshop on Text Summarization Branches Out*, 2004.