UTTERANCE-LEVEL LATENT TOPIC TRANSITION MODELING FOR SPOKEN DOCUMENTS AND ITS APPLICATION IN AUTOMATIC SUMMARIZATION

Hung-yi Lee #1*, Yun-nung Chen* *2*, Lin-shan Lee* #3

Graduate Institute of Communication Engineering, National Taiwan University#
Graduate Institute of Computer Science and Information Engineering, National Taiwan University*
tlkagkb93901106@gmail.com1, vivian.ynchen@gmail.com2, lslee@gate.sinica.edu.tw3

ABSTRACT

In this paper, we propose to use an utterance-level latent topic transition model to estimate the latent topics behind the utterances, and test the performance of such model in extractive speech summarization. In this model, the latent topic weights behind an utterance are estimated, and these topic weights evolve from an utterance to the next in a spoken document based on a topic transition function represented by a matrix. We explore different ways of obtaining such topic transition matrices used in the model, and find using a set of matrices estimated with utterances clustered from a training spoken document set is very useful. This model was shown to be able to offer extra performance improvement when used with the popularly used Probability Latent Semantic Analysis (PLSA) in preliminary experiments on speech summarization.

Index Terms— Latent Topic Transition Modeling, Speech Summarization

1. INTRODUCTION

Latent topics have been widely used for analysing both text and spoken documents by discovering word clustering patterns in the documents and somehow projecting each document to a latent topic space constructed through such word clusters. Such latent topic information has been found very useful in many applications such as concept matching in information retrieval, document summarization and clustering, and key term extraction. In many cases, it is desired to have more precise latent topic information on sentence level (or utterance level) rather than for each document, and the latent topic information for each sentence (utterance) can be estimated by simply treating a sentence (utterance) as a short document and directly applying the latent topic analysis methods developed for documents. However, since a sentence (utterance) usually contains only several terms, very often the latent topics cannot be well estimated unless the context is considered. For example, assume a news story about golf is being analysed. An utterance with the term “Tiger Woods” but without the term “golf” may be viewed as about “animals” although it is very likely to be more related to “sports”. Therefore, utterance-level latent topic analysis may not be as straightforward as document-level. The problem is more difficult for spoken documents with serious recognition errors since it is possible that very limited terms are recognized correctly within an utterance. Moreover, utterance boundaries are not clear in a spoken document, so automatically segmented utterances may not be as well formed as in text. In reality, natural speech rarely consists of isolated, unrelated utterances but rather collocated, structured and coherent utterances. Hence, an utterance with the term “Tiger Woods” may be found to be about “sports” if the neighboring sentences are about “sports”, even if no terms related to “sports” actually appear in the utterance being considered.

Although unsupervised learning of linguistic structure have been widely studied[1], not too many topic models have ever attempted to model similar structural dependency among topics. Hidden Markov Models have been successfully used for capturing topic transition in summarization [2, 3], but the assumption that each sentence is generated from a single latent topic instead of a topic mixture may not be sufficient. The Hidden Topic Markov Model (HTMM)[4] was extended from Latent Dirichlet Allocation by modeling the topic dependency between adjacent sentences, but in HTMM the topic dependency is simply binary: the topics of a sentence are either independent of or exactly the same as the previous sentence. Structural Topic Model (strTM)[5] further improved HTMM by assuming that adjacent sentences follow a topic transition relation, but it is on text documents only rather than spoken documents.

In this paper, we propose to use an utterance-level latent topic transition model for spoken documents, in which a topic transition function is used to model the change of latent topic weights across adjacent utterances, so the topic weights of an utterance depend not only on the terms in the utterance being considered, but also on the topics of the preceding utterance. We then apply this model in extractive speech summarization, in which a number of indicative utterances was selected from the given spoken documents according to a target summarization ratio, and contatenated together to form the summary. In such speech summarization tasks, it is important to identify utterances that carry concepts closer to the document as a whole, and it has been verified that measuring the topic-based utterance-document similarity is more effective than the word-based alternative, because the former is vulnerable to problems like synonyms and recognition errors. By enhancing the topic estimation process by topic transition modelling, more accurate latent topic weights in each sentence can be used for sentence-document similarity measure.

2. SENTENCE-LEVEL LATENT TOPIC TRANSITION MODEL

The proposed utterance-level topic transition model for spoken documents is shown in Fig. 1. θ_t is a K-dimensional topic weight vector for the t-th utterance in the spoken document, and K is the number of topics. $\theta_t[k]$ is the k-th component of θ_t representing the weight of topic k for the t-th utterance. Here the topic weights of an utterance are not restricted to be non-negative. We assume that θ_t depends on θ_{t-1}, the topic weights of the preceding utterance based on the following relationship,

$$\theta_t = F_t(\theta_{t-1}) + \epsilon_t^\theta = M_t\theta_{t-1} + \epsilon_t^\theta. \quad (1)$$

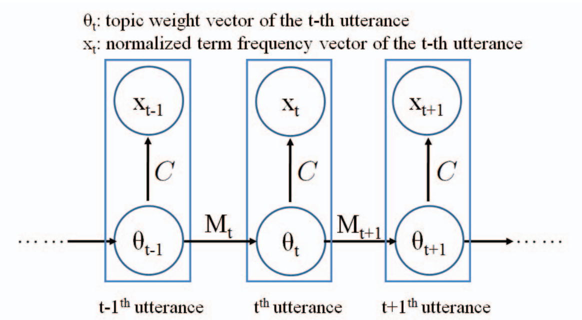
θt: topic weight vector of the t-th utterance
xt: normalized term frequency vector of the t-th utterance
xt-1
xt
xt+1
C
C
C
θt-1
θt
θt+1
Mt
Mt+1
t-1th utterance
tth utterance
t+1th utterance

Fig. 1: The utterance-level latent topic transition model. θ_t is a hidden K-dimensional topic weight vector of the t-th utterance in a spoken document, and K is the number of topics. The observation x_t is a V-dimensional vector representing the normalized term frequencies of the t-th utterance, and V is the lexicon size. θ_{t-1} evolves to θ_t based on a topic transition matrix M_t, while the observations x_t is generated from the hidden vector θ_t through a matrix C. M_t is a $K \times K$ matrix, and C is a $V \times K$ matrix.

$F_t(.)$ is the latent topic transition function modeling the relation between θ_{t-1} and θ_t. Although $F_t(.)$ can be of any form, here we assume the relation is linear and modeled by a $K \times K$ matrix M_t. ϵ_t^θ is a K-dimensional prediction error vector absorbing the difference between θ_t and $M_t\theta_{t-1}$. The transition matrix M_t can be defined based on some heuristic assumptions or trained with a document set, as will be clear in Section 3.

The relation between the latent topic weights and the normalized term frequencies for an utterance is formulated as

$$x_t = C\theta_t + \epsilon_t^x, \tag{2}$$

where x_t is a vector with dimension V which is the size of the lexicon, representing the normalized term frequencies for the t-th utterance. The i-th dimension of the vector x_t is the normalized term frequency of the i-th term w_i:

$$x_t[i] = \frac{n(s_t, w_i)}{\sum_{w_j \in s_t} n(s_t, w_j)}. \tag{3}$$

s_t is the transcription of the t-th utterance in the spoken document (either 1-best or in lattice form), and $n(s_t, w_i)$ is the frequency of term w_i in s_t. When s_t is in the lattice form, the occurrences of the term should be weighted by recognition scores such as the confidence measure. The matrix C in (2) is a $V \times K$ matrix that models the term-topic co-occurrence relationships. ϵ_t^x is a V-dimensional error vector absorbing the difference between $C\theta_t$ and x_t. Based on (2),

$$x_t[i] = \sum_{k=1}^{K} C[i,k]\theta_t[k] + \epsilon_t^x[i], \tag{4}$$

where $C[i,k]$ is the $[i,k]$ element of the matrix C, and $\epsilon_t^x[i]$ is the i-th element of ϵ_t^x. Based on (4), the value of $x_t[i]$, the normalized term frequency for term w_i, is the multiplication of $C[i,k]$ and $\theta_t[k]$ summed over the K latent topics with error $\epsilon_t^x[i]$. $C[i,k]$ can be understood as the normalized term frequency of term w_i in an utterance given a unit weight of latent topic k.

The matrix C above can be obtained in different ways. In the experiments reported below, C is obtained based on utterance-level Probability Latent Semantic Analysis (PLSA) [6], which uses a set of latent topic variables $\{T_k, k = 1, 2, ..., K\}$ to characterize the "term-topic" co-occurrence relationships. Given a training utterance set, PLSA training yields $\{P(w_i|T_k), i = 1, 2, \ldots, V, k = 1, 2, \ldots, K\}$, the probability of observing the term w_i in an utterance given the topic T_k, and $\{P(T_k|s), k = 1, 2, \ldots, K\}$, the mixture weight of topic T_k for all the utterance transcriptions s in the training set. This is accomplished by the EM algorithm for maximizing a likelihood function. Here $C[i,k]$ used in (2) and (4) is simply set to the probabilities $P(w_i|T_k)$ obtained from PLSA.

The problem now is to find the hidden sequence of topic weight vectors $\{\theta_1, \theta_2, \ldots, \theta_T\}$ for a spoken document of T utterances given the observation sequence of normalized term frequency vectors $\{x_1, x_2, \ldots, x_T\}$, whereas the latent topic weights θ_t for each utterance and the error vectors ϵ_t^θ and ϵ_t^x are all hidden. To make the problem tractable, we assume ϵ_t^θ and ϵ_t^x are both sample vectors generated from zero-mean normal distributions respectively with dimensions K and V, and covariance matrices Σ_θ and Σ_x. In other words, although we never know the true latent topic weights generating the utterances, based on the assumption of the probability distributions for the error vectors, we can estimate the most possible topic weight sequence $\{\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_T\}$ given the observations $\{x_1, x_2, \ldots, x_T\}$. With (1), (2) and the assumptions about the distributions generating ϵ_t^θ and ϵ_t^x, we have

$$\begin{aligned} &P(\theta_t|\theta_{t-1}) \\ &\propto exp\Big\{-\frac{1}{2}[\theta_t - M_t\theta_{t-1}]^T\Sigma_\theta^{-1}[\theta_t - M_t\theta_{t-1}]\Big\}(2\pi)^{-K/2}|\Sigma_\theta|^{-1/2}, \end{aligned} \tag{5}$$

$$P(x_t|\theta_t) \propto exp\Big\{-\frac{1}{2}[x_t - C\theta_t]^T\Sigma_x^{-1}[x_t - C\theta_t]\Big\}(2\pi)^{-V/2}|\Sigma_x|^{-1/2}, \tag{6}$$

and the joint probability for a sequence of normalized term frequencies $\{x_1, x_2, \ldots, x_t\}$ and a sequence of topic weight vectors $\{\theta_1, \theta_2, \ldots, \theta_t\}$,

$$\begin{aligned} &P(\{x_1, x_2, \ldots, x_t\}, \{\theta_1, \theta_2, \ldots, \theta_t\}) \\ &= \prod_{n=1}^{t} P(\theta_n|\theta_{n-1}) \prod_{n=1}^{t} P(x_n|\theta_n). \end{aligned} \tag{7}$$

When the topic transition matrix M_t and the term-topic co-occurrence relationship matrix C are given, it is possible to find the topic weight vector $\hat{\theta}_t$ for the t-th utterance given the observations $\{x_1, x_2, \ldots, x_t\}$ such that

$$\hat{\theta}_t = \arg\max_{\theta_t} P(\theta_t|x_1, x_2, ..., x_t), \tag{8}$$

where the value of $P(\theta_t|x_1, x_2, ..., x_t)$ is obtainable based on (7). $\hat{\theta}_t$ is the best estimate for the topic weight vector θ_t of the t-th utterance with maximum posterior probability given observations x_1 to x_t. Equation (8) here is actually solved by the algorithm of Kalman filtering [7].

3. DIFFERENT TOPIC TRANSITION MATRIX

Here we assume two forms of the topic transition matrix M_t as presented below.

3.1. Identity Transition Matrix

Since language is intrinsically cohesive and coherent, the neighboring sentences are usually about very similar topics; hence it is reasonable to assume that the topic weight vectors of adjacent sentences are nearly the same. Therefore, we may set the topic transition matrix M_t equal to the identity matrix for all t ($M_t = I$), and the variations in the topic weight vectors of adjacent sentences are thus all absorbed by the error vectors ϵ_t^θ in (1). The topic weight vector for the t-th utterance obtained by (8) with the assumption that the transition matrix is identity is denoted as $\hat{\theta}_t^0$, which can be used as the initial topic weight vector for estimation in the next subsection.

3.2. Estimated Transition Matrices

If we have a training spoken document set $\mathcal{D}$ [1] and initial topic weight vectors of all utterances in this set $\mathcal{D}$, we can estimate the transition matrices iteratively using these data. We first cluster all the utterances in the document set $\mathcal{D}$ in an unsupervised way. In the experiments below, we used k-means over the PLSA topic weight vectors for clustering. We assume that the utterances in each cluster thus obtained share the same transition matrix, that is, if there are R clusters for the document set $\mathcal{D}$, we should have R different transition matrices, one for each cluster. At iteration i, a set of transition matrices $\{M_1^i, \dots, M_r^i, \dots, M_R^i\}$ is estimated such that M_r^i for cluster r minimizes the prediction error over all utterances belonging to the cluster r in $\mathcal{D}$ when (1) is applied:

$$M_r^i = \arg\min_M \sum_{d \in \mathcal{D}} \sum_{\substack{s_{t-1} \in d, \\ C(s_{t-1})=r}} (\hat{\theta}_t^{i-1} - M\hat{\theta}_{t-1}^{i-1})^T (\hat{\theta}_t^{i-1} - M\hat{\theta}_{t-1}^{i-1}), \tag{9}$$

where $C(s_{t-1})$ represents the cluster ID for the transcription of the $t-1$-th utterance, or the cluster which s_{t-1} belongs to, the superscript ^{T} stands for matrix transpose, $\hat{\theta}_t^{i-1}$ is the topic weight vector of s_t obtained at the iteration $i-1$, and the initial topic weight vector $\hat{\theta}_t^0$ is already obtained in Section 3.1. The second summation in (9) is over all utterances in the document d belonging to the cluster r, and the first summation is over all documents d in the set $\mathcal{D}$. After obtaining the R transition matrices, the topic transition in (1) is modified into

$$\theta_t = M_{C(s_{t-1})}^i \theta_{t-1} + \epsilon_t^\theta. \tag{10}$$

That is, the transition matrix from θ_{t-1} to θ_t is $M_{C(s_{t-1})}^i$, and $C(s_{t-1})$ is the cluster s_{t-1} belongs to. Then Kalman filtering is used to estimate the topic weight vectors for all utterances as in (8), which can then be used as the topic weight vectors $\hat{\theta}_t^i$ for estimating the transition matrices in the next iteration. After N training iterations, the topic weight vectors $\hat{\theta}_t^N$ are further used for summarization in the next section.

4. SPEECH SUMMARIZATION WITH SENTENCE-LEVEL LATENT TOPICS

In the initial experiments on speech summarization, we used the Maximum Marginal Relevance (MMR) score [8]. This approach selects in each iteration one utterance from the document to be added to the summary, which is the utterance with the highest similarity to the whole document, while adding minimum redundancy to the summary. This is achieved at each iteration by evaluating a MMR score for each utterance s_a which has not been added to the summary, and then the utterance with the highest MMR score is selected,

$$MMR(s_a) = \lambda S(s_a, d) - (1-\lambda) S(s_a, d_{sum}), \tag{11}$$

where $S(s_a, d)$ is the similarity measure between s_a and the whole document d, d_{sum} is the summary obtained in the current iteration, and

$$S(s_a, d) = \frac{1}{|d|} \sum_{s_b \in d} SIM(s_a, s_b), \tag{12}$$

where $SIM(s_a, s_b)$ is the similarity measure between s_a and s_b, and $|d|$ is the number of utterances in d. Therefore, the first term on the right hand side of (11) is to be maximized, while the second term is to be minimized, and the parameters λ is to properly weight these two goals.

In the experiments to be reported below, three different approaches were used to estimate the similarity $SIM(s_a, s_b)$ in (12) all based on the cosine similarities between the vector representations v_a and v_b for s_a and s_b. These three approaches are respectively referred to as word-based, PLSA-based, and transition-based here. For word-based similarity $SIM_{word}(s_a, s_b)$, each component of v_i corresponds to a word in the lexicon, whose value is the term frequency weighted by the latent topic entropy for the term [9]. For PLSA-based similarity $SIM_{plsa}(s_a, s_b)$, the dimension of v_a is the number of latent topics K, and the value of each component is simply $P(T_k|s)$ from PLSA. For transition-based similarity $SIM_{tran}(s_a, s_b)$, v_a can be either $\hat{\theta}_a^0$ obtained above in Section 3.1 or $\hat{\theta}_a^N$ obtained in Section 3.2 with different cluster numbers R. Different similarity measures can be further integrated.

5. EXPERIMENTS

5.1. Experimental Setup

The corpus used in this research is the lectures for a course offered at National Taiwan University. The lectures were given in the host language of Mandarin Chinese but with many technical terms uttered in the guest language of English. The lectures covered a total of 17 chapters, with a total length of 45.2 hours. The corpus was segmented into 193 documents based on the slides used. The average document length was about 17.5 minutes. One-best ASR transcriptions were used for testing. For ASR, the acoustic models were trained on the ASTMIC corpus for Mandarin and on the Sinica Taiwan English corpus for English (both included hundreds of speakers), and then adapted using a 25-minute bilingual corpus from the target speaker (the course instructor). The language model was trained with a set of corpus in Chinese, and then adapted with two other courses offered by the same instructor and the course slides. The accuracies for the ASR transcriptions were 78.15% for Mandarin characters, 53.44% for English words, and 76.26% overall. The reference summaries of 40 documents were provided by graduate students who had taken this course. Only these 40 documents were used for testing. The utterances in all documents were used in PLSA and transition matrix training. The transition matrix training was performed with 5 iterations ($N = 5$). The ROUGE-1, 2, 3 and L F-measures with summarization ratios 5%, 10%, and 15% obtained from the package ROUGE [10] were used to evaluate the summarization results. In the experiments, both Σ_θ and Σ_x for ϵ_t^θ and ϵ_t^x in (1) and (2) were set to $\sigma^2 I$ with σ set to 0.01. The number of latent topics K was set to 16. λ for MMR in (11) was 0.9.

5.2. Experimental Results

Table 1 lists the ROUGE-1, 2, 3 and L results with 5%, 10% and 15% summarization ratios using different sentence similarity measures $SIM(s_a, s_b)$ for the MMR scores in (12). Column (A) shows the results for word-based similarity, or $S_{word}(s_a, s_b)$ was used

[1] In the experiments below, document set $\mathcal{D}$ is the documents for summarization.

Table 1: The ROUGE-1, 2, 3 and L results with 5%, 10% and 15% summarization ratios using different utterance similarity measures $SIM(s_a, s_b)$ in (12). Pairwised t-test with significance level at 0.05 was used to test the significance. The superscripts $^\alpha$, $^\beta$, $^\gamma$ and $^\delta$ respectively represent significantly better than the results in columns (A), (B), (C1) and (C2) on the same row. The ROUGE-1 of the "random", "head" and "longest" baselines with 5% summarization ratio are 0.261, 0.244 and 0.228 respectively.

		(A)	**(B)**	**(C)**		
		word	word	word+PLSA+Topic Transition		
	ratio		+PLSA	**(C1)**	Estimated	
				Iden.	**(C2)**	**(C3)**
				Matrix	1 cluster	2 clusters
ROUGE-1	5%	0.292	0.289	0.296	0.295	0.298
	10%	0.384	0.395	0.400	0.404	0.406^{α}
	15%	0.425	0.443^{α}	0.440	0.446^{α}	$0.458^{\alpha\gamma}$
ROUGE-2	5%	0.121	0.115	0.129	0.123	0.131
	10%	0.162	0.173	0.179	0.180	0.189^{α}
	15%	0.197	0.211	0.209	0.217	$0.240^{\alpha\beta\gamma\delta}$
ROUGE-3	5%	0.085	0.078	0.092	0.084	0.097
	10%	0.112	0.125	0.131	0.131	0.140^{α}
	15%	0.146	0.156	0.157	0.163	$0.187^{\alpha\beta\gamma\delta}$
ROUGE-L	5%	0.285	0.281	0.291	0.288	0.295
	10%	0.377	0.388	0.393	0.397	0.399^{α}
	15%	0.417	0.435^{α}	0.432	0.438^{α}	$0.451^{\alpha\gamma}$

for $SIM(s_a, s_b)$ in (12). Column (B) is the results when PLSA-based similarity was used in addition, or $S_{word}(s_a, s_b)S_{plsa}(s_a, s_b)$ was used for $SIM(s_a, s_b)$ in (12). Column (C) is the results further integrating the information considering topic transition, or $S_{word}(s_a, s_b)S_{plsa}(s_a, s_b)S_{trans}(s_a, s_b)$ was used for $SIM(s_a, s_b)$ in (12). Columns (C1), (C2), and (C3) under column (C) are the results using different topic transition matrices. Pairwised t-test with significance level at 0.05 was used to test the significance of the improvements obtained. The superscripts $^\alpha$, $^\beta$, $^\gamma$ and $^\delta$ respectively represent significantly better than the results in columns (A), (B), (C1) and (C2) on the same row. The ROUGE-1 values of the "random", "head"(selecting the sentences based on their positions in the documents) and "longest"(selecting the sentences based on their lengths) baselines with 5% summarization ratios are 0.261, 0.244 and 0.228 respectively, and all the experimental results reported in Table 1 are much better than the naive baselines.

Comparing the results in columns (A) and (B), we find that although the integration with PLSA topic distributions $P(T_k|s)$ offered improvements in all evaluation measures for 10% and 15% summairzation ratios, it was useless for 5% summarization ratio. A possible reason may be the difficulties of utterance-level latent topic estimation as previously mentioned. An utterance has only a few terms, so the PLSA latent topic distributions for the utterance did not correctly reveal the real latent topics for the utterance, thus the similarity measure obtained in this way is rough. When only very few utterances are to be selected (5% summarization ratio), the PLSA topic distributions did not help select the correct ones.

Next consider columns (C1), (C2) and (C3) using the latent topic transition model. Comparing column (C1) using the identity matrix with columns (A) and (B), we note that latent topic transition model was always better in all the evaluation metrics for 5% and 10% summarization ratios, even if only the identity matrix was used. Next, columns (C2) and (C3) are the results of using estimated transition matrices. Only the results of using 1 and 2 clusters, or using 1 and 2 latent topic transition matrices are shown here for space limitation. We notice that there is no significance difference between using a single estimated matrix (column (C2)) and the identity matrix (column (C1)), probably because both of them used only a single topic transition matrix, and when only one transition matrix can be estimated, this transition matrix estimated from the training data set turned out to be somewhat similar to an identity matrix, which is reasonable as mentioned above. Moreover, the single estimated matrix (column (C2)) outperformed PLSA (column (B)) in all cases although the improvements were not significant. When we further extended the cluster number to 2, we see with 2 transition matrices to model the topic transitions in column (C3), the results outperformed the identity matrix (column (C1)) and the single matrix (column (C2)) in all cases. Also it can be found that column (C3) was significantly better than columns (A), (B), (C1) and (C2) in many cases (indicated by superscriptions $^\alpha$, $^\beta$, $^\gamma$ and $^\delta$). For example, (C3) was significantly better than column (C2) for ROUGE-2 and ROUGE-3 with 15% summarization ratio. This verified that the proposed latent topic transition model offered more accurate latent topic weights behind each utterance.

6. CONCLUSION

In this paper, we propose to model the latent topic transition between adjacent utterances in a spoken document via a topic transition matrix. We show that this latent topic transition modeling offered improvements in speech summarization. Also, a set of transition matrices estimated for clustered utterances in a training document set turned out to yield very good results.

7. REFERENCES

[1] S. Goldwater and T. L. Griffiths, "A fully bayesian approach to unsupervised part-of-speech tagging," in *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, 2007.

[2] P. Fung, G. Ngai, and C.-S. Cheung, "Combining optimal clustering and hidden Markov models for extractive summarization," in *Proceedings of the ACL 2003 workshop on Multilingual summarization and question answering*, 2003.

[3] R. Barzilay and L. Lee, "Catching the drift: Probabilistic content models, with applications to generation and summarization," in *HLT-NAACL*, 2004.

[4] A. Gruber, M. Rosen-zvi, and Y. Weiss, "Hidden topic Markov models," in *In Proceedings of Artificial Intelligence and Statistics*, 2007.

[5] HongningWang, D. Zhang, and C. Zhai, "Structural topic model for latent topical structure analysis," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, 2011.

[6] T. Hofmann, "Probabilistic latent semantic analysis," in *In Proc. of Uncertainty in Artificial Intelligence, UAI99*, 1999.

[7] B. M. Yu, K. V. Shenoy, and M. Sahani, "Derivation of Kalman filtering and smoothing equations," Stanford University, Tech. Rep., 2004.

[8] S. Xie and Y. Liu, "Using corpus and knowledge-based similarity measure in maximum marginal relevance for meeting summarization," in *ICASSP*, 2008.

[9] S.-Y. Kong and L.-S. Lee, "Semantic analysis and organization of spoken documents based on parameters derived from latent topics," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, pp. 1875 –1889, 2011.

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