UNSUPERVISED DOMAIN ADAPTATION FOR SPOKEN DOCUMENT SUMMARIZATION WITH STRUCTURED SUPPORT VECTOR MACHINE

Hung-yi Lee #1*, Yu-yu Chou* *2*, Yow-Bang Wang* †3 *and Lin-shan Lee* *†4

Research Center for Information Technology Innovation, Academia Sinica #
Graduate Institute of Communication Engineering, National Taiwan University *
Graduate Institute of Electrical Engineering, National Taiwan University †
tlkagkb93901106@gmail.com[1], yoyuchou@gmail.com[2], piscesfantasy@gmail.com[3], lslee@gate.sinica.edu.tw[4]

ABSTRACT

Supervised approaches can learn a spoken document summarizer generating high-quality summaries using a set of training examples matched to the domain of target documents. However, preparing a sufficient number of in-domain training examples is expensive. In this paper we propose an approach for unsupervised domain adaptation for spoken document summarization, so no in-domain training examples are needed. A summarizer is first learned from a set of out-of-domain training examples by a supervised summarization approach based on structured support vector machine, and this summarizer is used to generate a set of initial summaries for the target spoken documents. The target documents and their initial machine-generated summaries then serve as extra training examples for learning a new summarizer, which further updates the summaries of the target spoken documents. This process is continued iteratively to incrementally improve the summarizer for the target spoken documents. Moreover, extra approaches transforming the feature representations based on the data distribution in the target domain and augmenting the representations with an extra set of domain-specific features are also proposed. Encouraging results were obtained in summarizing Mandarin-English code-switching course lectures using training examples from Mandarin broadcast news.

***Index Terms*—** Speech Summarization, Unsupervised Domain Adaptation, Structured Support Vector Machine

1. INTRODUCTION

This paper focuses on extractive summarization of spoken documents, or automatically selecting a subset of utterances in the spoken document as the summary [1]. Although unsupervised approaches such as Maximum Marginal Relevance (MMR) methods [2, 3] and graph-based approaches [4, 5, 6, 7] have been successful, supervised learning has been widely used [8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20]. For the latter, the task is usually treated as a binary classification problem determining whether to include an utterance in the summary. With a set of training spoken documents with reference summaries, the binary classifier can be very well trained with positive and negative examples. Some approaches were also developed to directly select a whole utterance subset instead of selecting them individually, so the relationships among the utterances can be considered [21, 22].

If the target spoken documents and the training spoken documents are in the same domain, it is possible to learn a high-quality summarizer for the target domain. However, preparing a sufficient number of training examples in the target domain may be very expensive. For example, considering spoken lectures, because the content of the course lectures is usually for some specialized area, it is very hard to collect enough number of related training spoken documents, not to mention hiring experts understanding the content to produce reference summaries. Semi-supervised learning and supervised domain adaptation have been investigated for spoken documents to address this problem [23, 22], but some in-domain training examples are still needed for these approaches. In this paper, we consider unsupervised domain adaptation for spoken document summarization, that is, to leverage the out-of-domain training examples at hand to summarize target spoken documents without in-domain training examples.

Self-labeling [24], widely used in speaker adaptation [25, 26, 27], is used here. A summarizer is first learned from a set of out-of-domain training examples. It is then used to summarize the target spoken documents. The target documents and their initial machine-generated summaries then serve as extra training examples for learning a new summarizer, which is in turn used to update the summaries of the target spoken documents. This process is continued iteratively to incrementally improve the summarizer for the target spoken documents. Self-labeling has been applied on e-mail summarization based on binary classification [28], but here more powerful summarization technique based on structured SVM [21] is used including considering prosodic and other types of features, as well as feature transformation across domains. Encouraging results were obtained in the preliminary experiments of summarizing Mandarin-English code-switching spoken lectures using training examples in Mandarin broadcast news.

2. SUPERVISED SPOKEN DOCUMENT SUMMARIZATION BASED ON STRUCTURED SVM

Here we first describe the supervised summarization algorithm used in this paper. Given a spoken document d, the task is to select a subset of utterances s_d from d to form a summary, which can be formulated as selecting s_d maximizing the following objective function $F(s_d)$ [29, 30].

$$F(s_d) = \sum_{x_i \in s_d} R(x_i) - \lambda \sum_{x_i, x_j \in s_d} Sim(x_i, x_j) \quad (1)$$
$$s.t. \sum_{x_i \in s_d} L(x_i) \leq \bar{L},$$

where $R(x_i)$ is the importance score of the utterance x_i in the subset s_d, $Sim(x_i, x_j)$ is the similarity between any two utterances x_i and x_j, $L(x_i)$ is the length of utterance x_i, and $\bar{L}$ is the length constraint for the summary. The first term in the right hand side of $F(s_d)$ rewards the subset s_d including more important utterances, while the second term penalizes the subset s_d including utterances similar to each other. The parameter λ in (1) is to properly weight these two goals. While obtaining the exact solution for (1) is computationally hard, some approximation algorithms exist [29].

We proposed a supervised spoken document summarization algorithm using structured SVM [21]. $R(x_i)$ in (1) is expressed as a linear function with a weight vector w to be learned,

$$R(x_i) = w \cdot h(x_i), \tag{2}$$

where $h(x_i)$ is a feature vector representing the utterance x_i, which will be further described in Section 3.1. With the availability of a training set $\{(d_n, r_{d_n})\}_{n=1}^N$, where N is the number of training examples, d_n is the n-th training document, and r_{d_n} is the reference summary for d_n (which is an utterance subset of d_n), we jointly learn the weight vector w and the parameter λ which ensure that for every training document d_n the reference summary r_{d_n} is the subset of d_n which gives the highest objective function in (1), or $F(r_{d_n}) > F(s_{d_n})$ for any other subset s_{d_n} of d_n.

The above goal of jointly learning w and λ is accomplished by solving the optimization problem below using structured SVM [31, 21]:

$$\min_{w,\lambda} \frac{1}{2}(\|w\|_2 + \lambda^2) + \frac{C}{N}\sum_{n=1}^{N} \epsilon_n, \tag{3}$$

$$\begin{aligned} s.t.\ & \forall n, \forall s_{d_n}, s_{d_n} \neq r_{d_n} : \\ & F(r_{d_n}) - F(s_{d_n}) \geq 1 - \epsilon_n,\ \ \epsilon_n \geq 0. \end{aligned}$$

The constraints in (3) require that for each training document d_n the differences between the objective function of the reference summary r_{d_n} and any other possible subset s_{d_n} are larger than a margin, which makes the algorithm more robust against various unknown disturbances. Each constraint is padded with a per-document slack variable ϵ_n whose sum over the training set is minimized. The norm of the parameters to be learned and the scale of the slack variables are traded off with a parameter C. Since $F(s_d)$ is a linear function of the parameters w and λ, the optimization problem in (3) is a quadratic programming problem with global optimal solution obtainable just as the ordinary SVM. By solving (3), we directly learn w and λ which can be used in extracting summaries for all new documents d based on (1) and (2). In this way, all utterances in the subset s_d are considered jointly rather than individually, and the weight parameter λ in (1) can be automatically learned as well [21]. The optimization problem in (3) has a huge number of constraints, but an approximate solution can be found in reasonable time with the cutting plane algorithm by selecting a set of active constraints from all the constraints [31].

3. UNSUPERVISED DOMAIN ADAPTATION FOR SPOKEN DOCUMENT SUMMARIZATION

In the scenario of unsupervised domain adaptation for spoken document summarization, there are a set of testing spoken documents $\{d_u\}_{u=1}^U$ to be summarized, and a training set $\{(d_l, r_{d_l})\}_{l=1}^L$, where d_l is the l-th training spoken document, and r_{d_l} is the human-generated reference summary for d_l.[1] The training and testing documents are from different domains.

[1] L and l indicate *labeled*, while U and u indicate *unlabeled*.

3.1. Feature Transformation Across Domains and Domain-specific Features

For each utterance x, a D-dimensional feature vector $f(x)$ including such information as utterance's length, position, similarity to the whole document, prosodic features, and so on is extracted. The feature components in $f(x)$ should be the general characteristics of utterances which can be extracted from documents in different domains. Although $f(x)$ can be directly taken as $h(x)$ in (2) to learn the parameters w and λ in (3) based on $\{(d_l, r_{d_l})\}_{l=1}^L$, this set of parameters may not be suitable for summarizing the testing documents because the distributions of $f(x)$ in training and testing domains may be very different. To handle this problem, a $D' \times D$ dimension reduction transformation matrix W ($D' < D$) is obtained by principle component analysis (PCA) based on the feature vectors $f(x)$ for all the utterances x in the *testing documents*, and W is used to reduce $f(x)$ into a D'-dimensional feature vector $\bar{f}(x)$, where $\bar{f}(x) = Wf(x)$. Therefore, the feature components in $f(x)$ not representative (or with small variances) in testing documents are excluded in $\bar{f}(x)$, and thus not involved in training when taking $\bar{f}(x)$ as $h(x)$.

Some utterance characteristics useful for summarization in the testing documents may be undefined or without correspondents in the training documents. For example, speaker role information is useful to determine the importance of an utterance in document genre like meetings [1], but such features can not be extracted from some domains like course lectures offered by a single speaker. To include this kind of features, $f(x)$ or $\bar{f}(x)$ are expanded into $f'(x)$ and $\bar{f}'(x)$:

$$f'(x) = \begin{bmatrix} f(x) \\ g(x) \end{bmatrix}, \quad \bar{f}'(x) = \begin{bmatrix} \bar{f}(x) \\ g(x) \end{bmatrix},$$

where $g(x)$ is a vector of features specific for the testing domain, and set to zero vectors for utterances x in training documents d_l. Either $f(x)$, $\bar{f}(x)$, $f'(x)$ and $\bar{f}'(x)$ can be taken as $h(x)$ in (2).

3.2. Self-labeling

With $h(x)$ defined above, the parameters w^0 and λ^0 are learned from $\{(d_l, r_{d_l})\}_{l=1}^L$ based on (3), and then used in (1) to generate an initial summary $r_{d_u}^0$ for each testing document d_u. It is possible to use an unsupervised summarization approach to generate $r_{d_u}^0$ [20], but the supervised method learns from the training set and generates better $r_{d_u}^0$ in general. Both $\{(d_l, r_{d_l})\}_{l=1}^L$ and $\{(d_u, r_{d_u}^0)\}_{u=1}^U$ are then taken as training examples to learn a new set of parameters w^1 and λ^1, which are further used in (1) to generate summaries $r_{d_u}^1$ for each d_u. Then $\{(d_l, r_{d_l})\}_{l=1}^L$ and $\{(d_u, r_{d_u}^1)\}_{u=1}^U$ are again used to learn w^2 and λ^2. This process is continued iteratively. After I iterations, the summary $r_{d_u}^I$ is taken as the output summary for each testing document d_u.

Self-labeling is useful because of two reasons. First, although the quality of extra training examples $\{(d_u, r_{d_u}^i)\}_{u=1}^U$ cannot be as good as $\{(d_l, r_{d_l})\}_{l=1}^L$ generated manually, $\{(d_u, r_{d_u}^i)\}_{u=1}^U$ can be helpful because they are exactly matched to the testing documents. Second, when $f'(x)$ or $\bar{f}'(x)$ are taken as $h(x)$ in (2), the components in w corresponding to the domain-specific features $g(x)$ cannot be directly learned from $\{(d_l, r_{d_l})\}_{l=1}^L$ because $g(x)$ are zero vectors for all utterances x in d_l. The weights for $g(x)$ can be learned only when $\{(d_u, r_{d_u}^i)\}_{u=1}^U$ obtained for testing documents are taken as training examples.

4. EXPERIMENTS

4.1. Experimental Setup

4.1.1. Testing Spoken Documents - Course Lectures

The testing spoken documents used in the preliminary experiments were from the course lectures offered at National Taiwan University by a single speaker with a total length of 45 hours. The lectures were Mandarin-English code-switching with Mandarin as the host language but many technical terms uttered in the guest language of English embedded in the Mandarin utterances. The corpus was segmented into 193 documents based on the slides used, and the average document length was about 17.5 minutes. One-best ASR transcriptions were used for summarization, and the accuracies for the ASR transcriptions were 88.0%.

Only 40 documents in the testing corpus were used for the experiments here. The reference summaries for them were generated by graduate students who had taken the course. Each document has three reference summaries. The length constraint (in number of Chinese characters plus English words in the manual transcriptions) is 10% of the documents. These reference summaries were used for evaluating the machine-generated summaries only, not for training.

4.1.2. Training Spoken Documents - Broadcast News

The training set included 200 Mandarin broadcast news stories. The average story length was about 29 seconds. One-best ASR transcriptions were used for summarization as well, with character accuracy of 81.7%. Each training spoken document has 3 reference summaries generated by graduate students of National Taiwan University. A training document with 3 reference summaries was regarded as 3 training examples.

4.2. General Features

Here we report the feature components of $f(x)$ general for all domains in Section 3.1.

4.2.1. Similarity with the whole document

The similarity measure between an utterance x and the whole document d was defined as $S(x,d) = \frac{1}{|d|}\sum_{x' \in d} Sim(x,x')$, where $Sim(x,x')$ was the similarity measure between utterances x and x', and $|d|$ the number of utterances in d. $Sim(x,x')$ was the cosine similarity between the vector representations v and v' for x and x', where the vector representations could be either lexical-based or topic-based.

For lexical-based similarity, each component of v corresponded to a term in the lexicon, whose value was the term frequency in the utterance weighted by the inverse document frequency for the term. For topic-based similarity, we used Probability Latent Semantic Analysis (PLSA) [32] with a set of latent topic variables $\{T_k, k = 1,2,...,K\}$ to characterize the "term-utterance" co-occurrence relationships. PLSA training based on utterances yielded $P(w|T_k)$, the probability of observing the term w in an utterance given the topic T_k, and $P(T_k|x)$, the mixture weight of topic T_k given the transcription of an utterance x. We trained two separate PLSA models individually with the utterances in testing and training documents rather than training a common model jointly because the training and testing documents did not share any common topics. Now the dimension of v was the number of latent topics K and the values of the components in v were simply $\{P(T_k|x), k = 1,2,...,K\}$. $K = 16, 32, 64, 128$ in the experiments below. Therefore, there were a total of 5 feature components regarding the similarities with the whole document.

To consider the context of an utterance, the same 5 components for the previous and next utterances were taken as another 10 components in $f(x)$.

4.2.2. Prosodic features

It was well known that prosody is very helpful to spoken document summarization [11, 33, 34]. We used 60 prosodic features for each utterance, 27 features related to pause and syllable duration, 13 to energy and 20 to pitch. The details are left out for space limitation.

4.2.3. Other features

There were other 4 feature components:

- Utterance length: Number of Chinese characters plus English words in the utterance's transcription.
- Normalized utterance position: i/N for the i-th utterance in a document with N utterances.
- Significance scores:
  - The sum of the significance scores $I(w)$ for all terms w in the utterance's transcription. $I(w) = tf(w) \times idf(w)$, where $tf(w)$ is the number of w in the whole document, and $idf(w)$ is the inverse document frequency of w [35].
  - The sum of another significance scores $I'(w)$ for all terms w in the utterance's transcription. $I'(w) = tf(w)/LTE(w)$, where $LTE(w)$ is the latent topic entropy for w based on PLSA with $K = 32$ [36].

The above 4 feature components for the previous and next utterances were also included in $f(x)$.

4.3. Features Specific for Testing Documents

Here we present feature components of $g(x)$ in Section 3.1 specific for testing documents.

4.3.1. Key Term related Features

There were two sets of key terms specific for the course lectures considered here:

- All English words were considered as key terms because most of them were terminologies. This key term definition is specific for the code-switching lectures here.
- 205 key terms automatically extracted by the key term extraction approach proposed previously for course lectures [37].

Based on each key term definition, there were 3 feature components for an utterance:

- The number of key terms in an utterance.
- Assume a key term occurring the first time in a document gave more new information than the same key term appearing latter. Hence, in an utterance the number of key terms occurring first time in the document among the same key terms was taken as a feature component.

- Assume the position of a key term in an utterance also brought some information, so the average positions of the key terms occurring first time in the spoken document was also taken as a feature component.

Since there were 2 key term definitions, there was a total of 6 features related to key terms.

4.3.2. Latent Topic Distribution

The K probabilities $\{P(T_k|x), k = 1, 2, \ldots, K\}$ obtained by PLSA trained from testing spoken documents are taken as K components for $g(x)$, and $K = 32$ in the experiments reported below. There were another PLSA model with the same number of latent topics learned from the training documents, but the two PLSA models were learned individually with completely independent topics. Therefore, the features mentioned above were considered as domain specific.

4.4. Experimental Results

In the following experiments, ROUGE F-measures [38] were used to evaluate the summarization results. The greedy algorithm [29] was always used to optimize (1) when w in (2) and λ were given. The lexical-based similarity in Section 4.2.1 was used for the similarity $Sim(x_i, x_j)$ in (1).

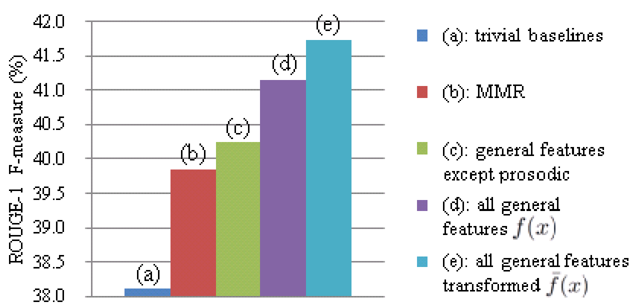

Fig. 1. *ROUGE-1 F-measures for the initial summaries $\{r^0_{d_u}\}_{u=1}^U$ before self-labeling for the testing spoken documents using different feature sets in Section 3.1.*

Fig. 1 is the ROUGE-1 F-measures for the initial summaries $\{r^0_{d_u}\}_{u=1}^U$ obtained before self-labeling with different feature sets described in Sections 3.1 and 4.2. Bar (a) is for a trivial baseline, in which the longest utterances were selected for the length constraint $\bar{L}$, while bar (b) is the MMR unsupervised approach. Bar (c) is for the feature vector $f(x)$ including the general feature components described in Section 4.2 except the prosodic features, and bar (d) is for all general feature components in Section 4.2 including prosodic features. Bar (e) is for $\bar{f}(x)$ transformed from $f(x)$ used for bar (d), and $D' = 0.6D$ in Section 3.1. The results for $f'(x)$ and $\bar{f}'(x)$ are not shown in Fig. 1. Since $g(x)$ in the out-of-domain training examples were zero vectors, without self-labeling the results of $f'(x)$ and $\bar{f}'(x)$ were exactly the same as $f(x)$ and $\bar{f}(x)$ respectively. We see from Fig. 1 that even though the training documents were out-of-domain, the supervised learning methods (bars (c), (d) and (e)) still yielded better results than the trivial baseline and MMR (bars (a) and (b)). The prosodic features were helpful ((d) vs (c)), and the feature transformation further improved the performance ((e) vs (d)).

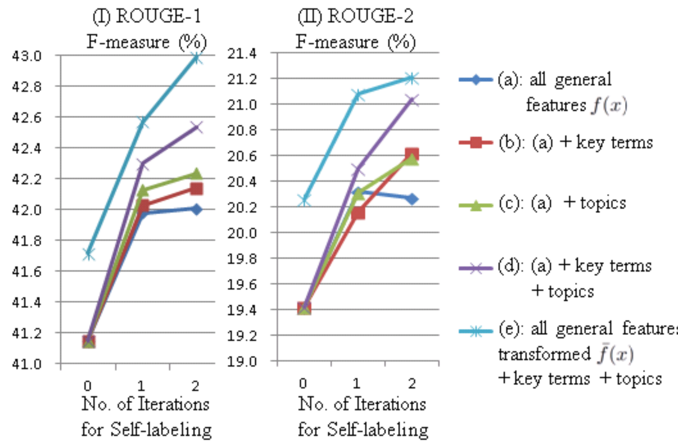

Fig. 2. *The results yielded by self-labeling with 0,1 and 2 training iterations and different feature sets for (I) ROUGE-1 and (II) ROUGE-2.*

Fig. 2 are the ROUGE-1 and -2 F-measures yielded by self-labeling in Section 3.2 with different sets of features with 0, 1 and 2 iterations. Zero iteration means without self-labeling. Curve (a) is the results yielded by $f(x)$ in Section 3.1 with all general features in Section 4.2. Curves (b), (c), and (d) are respectively for $f'(x)$ with $g(x)$ including the features related to key terms, latent topic distribution, and both. Curve (e) is for transformed $\bar{f}'(x)$ for all general features plus $g(x)$ including all domain-specific features in Section 4.3. Therefore, the two points for zero iterations for ROUGE-1 in Fig 2 (a) correspond to bars (d) and (e) in Fig. 1. We see that some improvements were always achievable with self-labeling regardless of the features used (iterations 1,2 vs 0) [2]. Also, both the feature sets related to key terms and topic distribution were helpful ((b), (c) vs (a)), and can be integrated ((d) vs (b),(c)), while feature transformation yielded further improvements ((e) vs (d)).

5. CONCLUSION

In this paper, we investigate unsupervised domain adaptation for spoken lecture summarization with the supervised summarization method based on structured SVM. Encouraging results were obtained with self-labeling, feature transformation and domain-specific features when summarizing course lectures using training examples in broadcast news. More different document domains and genres will be explored in the future.

6. REFERENCES

[1] Yang Liu and Dilek Hakkani-Tur, *Spoken Language Understanding - Systems for Extracting Semantic Informaiont from speech*, chapter 13, pp. 357 – 396, Wiley, 2011.

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

[3] Jaime Carbonell and Jade Goldstein, "The use of mmr, diversity-based reranking for reordering documents and producing summaries," in *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, 1998.

[2]The performance was saturated with more iterations in most cases.

[4] Nikhil Garg, Benoit Favre, Korbinian Reidhammer, and Dilek Hakkani-Tur, "ClusterRank: A graph based method for meeting summarization," in *Interspeech*, 2009.

[5] Hui Lin, J. Bilmes, and Shasha Xie, "Graph-based submodular selection for extractive summarization," in *ASRU*, 2009.

[6] Yun-Nung Chen, Yu Huang, Ching-Feng Yeh, and Lin-Shan Lee, "Spoken lecture summarization by random walk over a graph constructed with automatically extracted key terms," in *Interspeech*, 2011.

[7] Yun-Nung Chen and Florian Metze, "Integrating intra-speaker topic modeling and temporal-based inter-speaker topic modeling in randomwalk for improved multi-party meeting summarization," in *Interspeech*, 2012.

[8] Jian Zhang, Ho Yin Chan, Pascale Fung, and Lu Cao, "A comparative study on speech summarization of broadcast news and lecture speech," in *Interspeech*, 2007.

[9] Shih-Hsiang Lin, Berlin Chen, and Hsin-Min Wang, "A comparative study of probabilistic ranking models for chinese spoken document summarization," *ACM Transactions on Asian Language Information Processing (TALIP)*, vol. 8, pp. 3:1–3:23, 2009.

[10] J.J. Zhang, R.H.Y. Chan, and P. Fung, "Extractive speech summarization using shallow rhetorical structure modeling," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, pp. 1147 –1157, 2010.

[11] Shasha Xie, D. Hakkani-Tur, B. Favre, and Yang Liu, "Integrating prosodic features in extractive meeting summarization," in *ASRU*, 2009.

[12] Anne Hendrik Buist, Wessel Kraaij, and Stephan Raaijmakers, "Automatic summarization of meeting data: A feasibility study," in *in Proc. Meeting of Computational Linguistics in the Netherlands (CLIN)*, 2004.

[13] Sameer Maskey and Julia Hirschberg, "Comparing lexical, acoustic/prosodic, structural and discourse features for speech summarization," in *Interspeech*, 2005.

[14] Jian Zhang and Pascale Fung, "Speech summarization without lexical features for mandarin broadcast news," in *Proceedings of the Human Language Technology Conference of the NAACL*, 2007, pp. 213–216.

[15] Shih-Hsiang Lin, Berlin Chen, and Hsin-Min Wang, "A comparative study of probabilistic ranking models for chinese spoken document summarization," *ACM Transactions on Asian Language Information Processing (TALIP)*, vol. 8, pp. 3:1–3:23, 2009.

[16] J.J. Zhang, R.H.Y. Chan, and P. Fung, "Extractive speech summarization by active learning," in *ASRU*, 2009.

[17] Michel Galley, "A skip-chain conditional random field for ranking meeting utterances by importance," in *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, 2006.

[18] Shasha Xie and Yang Liu, "Improving supervised learning for meeting summarization using sampling and regression," *Computer Speech & Language*, vol. 24, pp. 495 – 514, 2010.

[19] Shih-Hsiang Lin, Yu-Mei Chang, Jia-Wen Liu, and B. Chen, "Leveraging evaluation metric-related training criteria for speech summarization," in *ICASSP*, 2010.

[20] Shih-Hsiang Lin, Yueng-Tien Lo, Yao-Ming Yeh, and Berlin Chen, "Hybrids of supervised and unsupervised models for extractive speech summarization," in *Interspeech*, 2009.

[21] Hung-Yi Lee, Yu-Yu Chou, Yow-Bang Wang, and Lin-Shan Lee, "Supervised spoken document summarization jointly considering utterance importance and redundancy by structured support vector machine," in *Interspeech*, 2012.

[22] Hitoshi Nishikawa, Toshiro Makino, and Yoshihiro Matsuo, "Domain adaptation with augmented space method for multi-domain contact center dialogue summarization," in *MLSLP*, 2012.

[23] Shasha Xie, Hui Lin, and Yang Liu, "Semi-supervised extractive speech summarization via co-training algorithm," in *Interspeech*, 2010.

[24] Anna Margolis, *A Literature Review of Domain Adaptation with Unlabeled Data*, 2011.

[25] Mukund Padmanabhan, George Saon, and Geoffrey Zweig, "Lattice-based unsupervised mllr for speaker adaptation," in *ASR*, 2000.

[26] R. Wallace, K. Thambiratnam, and F. Seide, "Unsupervised speaker adaptation for telephone call transcription," in *ICASSP*, 2009.

[27] Langzhou Chen, M.J.F. Gales, and K.K. Chin, "Constrained discriminative mapping transforms for unsupervised speaker adaptation," in *ICASSP*, 2011.

[28] Oana Sandu, Giuseppe Carenini, Gabriel Murray, and Raymond Ng, "Domain adaptation to summarize human conversations," in *Proceedings of the 2010 Workshop on Domain Adaptation for Natural Language Processing*, 2010.

[29] Ryan McDonald, "A study of global inference algorithms in multi-document summarization," in *Proceedings of the 29th European conference on IR research*, 2007.

[30] D. Gillick, K. Riedhammer, B. Favre, and D. Hakkani-Tur, "A global optimization framework for meeting summarization," in *ICASSP*, 2009.

[31] Ioannis Tsochantaridis, Thomas Hofmann, Thorsten Joachims, and Yasemin Altun, "Support vector machine learning for interdependent and structured output spaces," in *ICML*, 2004.

[32] Thomas Hofmann, "Probabilistic latent semantic analysis," in *UAI*, 1999.

[33] Sameer Maskey and Julia Hirschberg, "Comparing lexical, acoustic/prosodic, structural and discourse features for speech summarization," in *Interspeech*, 2005.

[34] Sameer Maskey and Julia Hirschberg, "Summarizing speech without text using hidden markov models," in *Proceedings of the Human Language Technology Conference of the NAACL*, 2006.

[35] Sadaoki Furui, Tomonori Kikuchi, Yousuke Shinnaka, and Chiori Hori, "Speech-to-text and speech-to-speech summarization of spontaneous speech," *IEEE Trans. on Speech and Audio Processing*, vol. 12, no. 4, pp. 401–408, 2004.

[36] Sheng-Yi Kong and Lin-Shan 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.

[37] Yun-Nung Chen, Yu Huang, Hung-Yi Lee, and Lin-Shan Lee, "Unsupervised two-stage keyword extraction from spoken documents by topic coherence and support vector machine," in *ICASSP*, 2012.

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