Graph-Based Multi-Modality Learning for Topic-Focused Multi-Document Summarization

Xiaojun Wan and Jianguo Xiao
Institute of Computer Science and Technology
Peking University, Beijing 100871, China
{wanxiaojun, xiaojianguo}@icst.pku.edu.cn

Abstract

Graph-based manifold-ranking methods have been successfully applied to topic-focused multi-document summarization. This paper further proposes to use the multi-modality manifold-ranking algorithm for extracting topic-focused summary from multiple documents by considering the within-document sentence relationships and the cross-document sentence relationships as two separate modalities (graphs). Three different fusion schemes, namely linear form, sequential form and score combination form, are exploited in the algorithm. Experimental results on the DUC benchmark datasets demonstrate the effectiveness of the proposed multi-modality learning algorithms with all the three fusion schemes.

1 Introduction

Topic-focused (or query-based) multi-document summarization aims to create from a document set a summary which answers the need for information expressed in a given topic or query. Topic-focused summarization has drawn much attention in recent years and it has been one of the main tasks in recent Document Understanding Conferences (DUC). Topic-focused summary can be used to provide personalized news services for different users according to the users' unique information need. In a QA system, a question-focused summary is usually required to answer the information need in the issued question.

As compared with generic multi-document summarization, the challenge for topic-focused multi-document summarization is that a topic-focused summary is not only expected to deliver the important information contained in the whole document set as much as possible, but also is expected to guarantee that the information is biased to the given topic. Therefore, we need effective methods to take into account this topic-biased characteristic during the summarization process.

In recent years, a variety of graph-based methods have been proposed for topic-focused multi-document summarization [Wan et al., 2007; Wei et al., 2008]. The graph-based methods first construct a graph representing the sentence relationships at different granularities and then evaluate the topic-biased saliency of the sentences based on the graph. The manifold-ranking method is a typical graph-based summarization method [Wan et al., 2007] and it can naturally make uniform use of the sentence-to-sentence relationships and the sentence-to-topic relationships in a manifold-ranking process. The sentence relationships are treated as a single modality in the basic manifold-ranking method.

In this study, we classify the sentence relationships into within-document relationships and cross-document relationships, and consider each kind of relationships as a separate modality (graph). We believe that the two modalities have unique characteristics and it could be helpful to distinguish the two modalities in the sentence ranking process. We then propose to use the multi-modality learning algorithm for fusing the two modalities. The learning algorithm is an extension of the basic manifold-ranking algorithm. Three fusion schemes are proposed for the multi-modality scenario, i.e. the linear scheme, the sequential scheme and the score combination scheme.

Experiments have been performed on the DUC2005-2007 benchmark datasets, and the results demonstrate that the proposed multi-modality learning method can outperform the baseline manifold-ranking method. All the three fusion schemes are effective and the linear fusion scheme performs the best.

The rest of this paper is organized as follows: We briefly introduce the related work in Section 2. The basic manifold-ranking method is introduced in Section 3 and the proposed multi-modality learning method is described in Section 4. Empirical evaluation results are shown in Section 5. Lastly, we conclude this paper in Section 6.

2 Related Work

In this section, we focus on extraction-based methods. Extraction-based summarization usually involves assigning saliency scores to some units (e.g. sentences, paragraphs) of the documents and extracting the sentences with highest scores.

To date, various extraction-based methods have been proposed for generic multi-document summarization. A typical traditional method is the centroid-based method [Radev et al., 2004], which scores sentences based on such features as cluster centroids, position, TFIDF, and etc. The method computes a score based on each single feature and then linearly combines all the scores into an overall sentence score. New features such as topic signature are used to select important content in NeATS [Lin and Hovy, 2002]. Learning-based approaches have been proposed for combining various sentence features [Shen et al., 2007; Wong et al., 2008]. The MMR algorithm [Goldstein et al., 1999] is a popular way for removing redundancy between summary sentences. Themes (or topics, clusters) in documents have been discovered and used for sentence selection [Hardy et al., 2002; Harabagiu and Lacatusu, 2005; Wang et al., 2008].The influences of input difficulty on summarization performance have been investigated in [Nenkova and Louis, 2008]. Most recently, graph-based methods have been proposed to rank sentences. The methods first construct a graph model to reflect sentence relationships at different granularities, and then compute sentence scores using graph-based learning algorithms. For example, LexRank [Erkan and Radev, 2004] and TextRank [Mihalcea and Tarau, 2005] are such systems using algorithms similar to PageRank and HITS to compute sentence importance. Cluster-level information has been incorporated in the graph model to better evaluate sentences [Wan and Yang, 2008].

For topic-focused summarization, many methods are heuristic extensions of generic summarization methods by incorporating the information of the given topic or query into generic summarizers, and such related work can be found on DUC workshop publications. In recent years, a few novel methods have been proposed for topic-focused summarization. For example, Daumé and Marcu [2006] present a Bayesian model - BAYESUM for sentence extraction. Query expansion techniques have been used to overcome the mismatch between query and sentences [Nastase 2008]. Zhang et al. [2008] propose a novel adaptive model to mutually boost the summary and the topic representation. Learning-based methods have also been used for topic-focused summarization [Ouyang et al., 2007; Schilder and Kondadadi, 2008]. For graph-based methods, Wan et al. [2007] propose a manifold-ranking method to make uniform use of sentence-to-sentence and sentence-to-topic relationships. Wei et al. [2008] propose a query-sensitive mutual reinforcement chain for topic-focused summarization.

3 Basic Manifold-Ranking Algorithm

The manifold-ranking method [Zhou et al., 2003a; Zhou et al., 2003b] is a universal ranking algorithm and it is initially used to rank data points along their underlying manifold structure. The prior assumption of manifold-ranking is: (1) nearby points are likely to have the same ranking scores; (2) points on the same structure (typically referred to as a cluster or a manifold) are likely to have the same ranking scores. The manifold-ranking method has been used for topic-focused document summarization [Wan et al., 2007], where the data points refer to the topic description and all the sentences in the documents. The manifold-ranking process for the summarization task can be formalized as follows:

Given a set of data points $\chi=\{x_0,x_1,...,x_{\rm n}\}\subset R^{\rm m}$, the first point x_0 represents the topic description (query point) and the rest n points represent all the sentences in the documents (data points to be ranked). Note that the topic description is considered as a single query point, and it is processed in the same way as other sentences. Let $f:\chi\rightarrow R$ denote a ranking function which assigns to each point x_i ($0\le i\le n$) a ranking value f_i. We can view f as a vector $f=[f_0,\ldots,f_n]^{\rm T}$. We also define a prior vector $y=[y_0,\ldots,y_n]^{\rm T}$, in which $y_0=1$ because x_0 is the query object and $y_i=0$ ($1\le i\le n$) for all the remaining points that we want to rank.

An affinity graph is constructed by connecting any pair of different sentences. The affinity matrix is denoted as $W=(W_{ij})_{(n+1)\times(n+1)}$, and each element W_{ij} corresponds to the cosine similarity between data points x_i and x_j (we let W_{ii}=0 to void loops). W is then symmetrically normalized into S by $S=D^{-1/2}WD^{-1/2}$, where D is the diagonal matrix with (i,i)-element equal to the sum of the i-th row of W.

The cost function associated with f is defined to be

$$Q(f)=\alpha\cdot\sum_{i,j=0}^{n}W_{ij}\left|\frac{1}{\sqrt{D_{ii}}}f_i-\frac{1}{\sqrt{D_{jj}}}f_j\right|^2+\beta\cdot\sum_{i=0}^{n}\left|f_i-y_i\right|^2 \qquad (1)$$

where $\alpha\in[0,1)$, $\beta\in(0,1]$ are the regularization parameters and we have $\alpha+\beta=1$. The first term and second term of the right-hand side in the cost function are the *smoothness constraint*, and the *fitting constraint*, respectively.

The above equation can be re-written in a more concise form as follows:

$$Q(f)=\alpha\cdot f^T(I-S)f+(1-\alpha)\cdot(f-y)^T(f-y) \qquad (2)$$

Then the solution of the ranking process is:

$$f^*=\arg\min_f Q(f) \qquad (3)$$

According to [Zhou et al., 2003b], the ranking values can be obtained by iterating the following computation until convergence:

$$f^{(t+1)}=\alpha Sf^{(t)}+(1-\alpha)y \qquad (4)$$

The theorem in [Zhou et al., 2003b] guarantees that the sequence $\{f^{(t)}\}$ converges to

$$f^*=(1-\alpha)\cdot(I-\alpha S)^{-1}y \qquad (5)$$

Although f^* can be expressed in a closed form, for large scale problems, the iteration algorithm in Equation (4) is preferable due to computational efficiency.

The ranking value of a sentence indicates the topic-biased informativeness of the sentence. In order to remove redundancy between sentences, the sentences highly overlapping with other informative sentences are penalized by the same greedy algorithm in [Wan et al., 2007], which decreases the overall ranking score of less informative sentences by the part conveyed from the most informative one. The overall ranking score of each sentence reflects both the biased informativeness and the information novelty of the sentence. The sentences with high overall ranking scores are chosen into the summary.

4 Multi-Modality Learning Algorithm

The basic manifold-ranking algorithm makes uniform use of the sentence relationships in a single modality. However, the relationships between sentences in a document set can be classified as either within-document relationship or cross-document relationship: if two sentences come from the same document, the corresponding link is a within-document link; if two sentences come from different documents, the corresponding link is a cross-document link. Each kind of links can be considered as a separate modality. The two modalities reflect the local information channel and the global information channel between sentences, respectively, and they have unique and specific characteristics. Therefore, it would be more appropriate to distinguish the two modalities and apply the multi-modality manifold-ranking algorithm [Tong et al., 2005] for ranking sentences.

Based on each kind of modality, we can build an undirected graph to reflect each kind of sentence relationships. Let $W^a=[W^a_{ij}]_{(n+1)\times(n+1)}$ be the within-document affinity matrix containing only the within-document links for the n+1 data points, where W^a_{ij} is the cosine similarity value between x_i and x_j if x_i and x_j belong to the same document or one of x_i and x_j is x_0; Otherwise, W^a_{ij} is set to 0. Similarly, let $W^b=[W^b_{ij}]_{(n+1)\times(n+1)}$ be the cross-document affinity matrix containing the cross-document links, where W^b_{ij} is the cosine similarity value between x_i and x_j if x_i and x_j belong to different documents or one of x_i and x_j is x_0; Otherwise, W^b_{ij} is set to 0. Note that all the relationships between the topic x_0 and any document sentence x_i ($i\geq 1$) are included in both W^a and W^b. We then normalize W^a by $S^a=(D^a)^{-1/2}W^a(D^a)^{-1/2}$, where D^a is the diagonal matrix with (i,i)-element equal to the sum of the ith row of W^a. Similarly, W^b is normalized to S^b by $S^b=(D^b)^{-1/2}W^b(D^b)^{-1/2}$.

Then the multi-modality learning task for topic-focused summarization is to infer the ranking function f from W^a, W^b and y:

$$\{(W^a, D^a, S^a); (W^b, D^b, S^b); y\} \rightarrow f \qquad (6)$$

S^a, S^b and y can be considered as constraints in the learning task, where 1) if two data points (x_i and x_j) are measured as similar by S^a or S^b, they should receive similar ranking values in f (f_i and f_j) and vice versa; 2) if a data point x_i is within the initial query points, its ranking value f_i should be as consistent as possible with the initial value y_i.

In the following subsections, we will describe three different learning schemes for fusing the two modalities based on different optimization strategies: linear fusion scheme, sequential fusion scheme [Tong et al., 2005] and score combination scheme. The linear scheme and the sequential scheme fuse the constraints of the two modalities in the manifold-ranking process, while the score combination scheme directly fuses the ranking scores computed in separate modalities.

4.1 Linear Fusion

This scheme fuses the constraints from S^a, S^b and y simultaneously by a weighted sum. The cost function associated with f is defined to be:

$$\begin{aligned} Q(f) = \mu\cdot\sum_{i,j=0}^{n} W^a_{ij}\left|\frac{1}{\sqrt{D^a_{ii}}}f_i-\frac{1}{\sqrt{D^a_{jj}}}f_j\right|^2 + \\ \eta\cdot\sum_{i,j=0}^{n} W^b_{ij}\left|\frac{1}{\sqrt{D^b_{ii}}}f_i-\frac{1}{\sqrt{D^b_{jj}}}f_j\right|^2 + \theta\cdot\sum_{i=0}^{n}\left|f_i-y_i\right|^2 \end{aligned} \qquad (7)$$

where μ, η, θ capture the trade-off between the constrains, usually we have $0\leq\mu, \eta<1$, $0<\theta\leq 1$ and $\mu+\eta+\theta=1$. The first two terms of the right-hand side in the cost function are the *smoothness constraints* for the two modalities, and the last term is the *fitting constraint*, respectively.

The above equation can be written in a more concise form as follows

$$Q(f)=\mu\cdot f^T(I-S^a)f+\eta\cdot f^T(I-S^b)f+(1-\mu-\eta)\cdot(f-y)^T(f-y) \qquad (8)$$

And then the optimal ranking function f^* is achieved when $Q(f)$ is minimized:

$$f^*=\arg\min_f Q(f) \qquad (9)$$

According to [Tong et al., 2005], solving the above optimization problem leads to the following optimal ranking function f^*:

$$f^*=(1-\mu-\eta)\cdot(I-\mu\cdot S^a-\eta\cdot S^b)^{-1}y \qquad (10)$$

In practice, the following iterative form is more preferable than the above close form to obtain the ranking function f^*:

$$f^{(t+1)}=\mu\cdot S^a f^{(t)}+\eta\cdot S^b f^{(t)}+(1-\mu-\eta)\cdot y \qquad (11)$$

$$where\ f^{(0)}=y$$

And we have $f^*=\lim_{t\to\infty}f^{(t)}$ by similar analysis in [Zhou et al., 2003b].

4.2 Sequential Fusion

This scheme fuses the constraints from S^a, S^b and y sequentially. The optimization problem is formulated in a two-stage way:

$$Q_a(f)=\mu\cdot f^T(I-S^a)f+(1-\mu)\cdot(f-y)^T(f-y) \qquad (12)$$

$$f_a^*=\arg\min_f Q_a(f) \qquad (13)$$

$$Q_b(f)=\eta\cdot f^T(I-S^b)f+(1-\eta)\cdot(f-f_a^*)^T(f-f_a^*) \qquad (14)$$

$$f_b^*=\arg\min_f Q_b(f) \qquad (15)$$

where μ captures the trade-off between the constraints in $Q_a(f)$, and η captures the trade-off between the constraints in $Q_b(f)$. We have $0\leq\mu, \eta<1$.

The first stage defines an optimal f_a^* by considering the constraints from S^a and y, and the second stage defines an optimal f_b^* by considering the constraints from S^b and f_a^*. The final ranking score is decided by f_b^*, i.e., $f^* = f_b^*$.

According to [Tong et al., 2005], solving the above optimization problem leads to the following optimal ranking function f_b^*:

$$f_b^* = (1-\mu)(1-\eta)(I-\mu S^b)^{-1}(I-\mu S^b)^{-1} y \tag{16}$$

The iterative form for f_b^* is given as:

$$f_b^{(t+1)} = \mu \cdot S^a f_b^{(t)} + \eta \cdot S^b f_b^{(t)} - \mu\eta \cdot S^b S^a f_b^{(t)} + (1-\mu)(1-\eta) y, \quad where \ f_b^{(0)} = y \tag{17}$$

And we have $f_b^* = \lim_{t \to \infty} f_b(t)$ by similar analysis in [Zhou et al., 2003b].

4.3 Score Combination

This scheme first compute the ranking scores in each modality and then directly fuses the ranking scores:

$$Q_a(f) = \mu \cdot f^T (I - S^a) f + (1-\mu) \cdot (f-y)^T (f-y) \tag{18}$$

$$f_a^* = \arg\min_f Q_a(f) \tag{19}$$

$$Q_b(f) = \eta \cdot f^T (I - S^b) f + (1-\eta) \cdot (f-y)^T (f-y) \tag{20}$$

$$f_b^* = \arg\min_f Q_b(f) \tag{21}$$

where $\mu = \eta$ is defined the same as α in Equation (2). The two separate optimization problems are the same with the basic manifold ranking algorithm, and they can be solved by Equation (4).

And the final ranking function f^* is defined as follows:

$$f^* = \lambda \cdot f_a^* + (1-\lambda) \cdot f_b^* \tag{22}$$

where $\lambda \in [0,1]$ is the combination weight.

After we obtain the ranking values of the sentences by using any above fusion scheme, the same greedy algorithm in Section 3 is applied to remove redundancy between sentences and choose summary sentences.

5 Empirical Evaluation

5.1 Dataset and Evaluation Metrics

Topic-focused multi-document summarization has been the main task on DUC2005, DUC2006 and DUC2007, so we used the three DUC datasets for evaluation in this study. Table 1 gives a short summary of the three data sets. For each task, NIST assessors have developed topics/questions of interest to them[1] and they have chosen a set of documents relevant to each topic. These documents of newswire articles formed the document cluster for each topic. Reference summaries have been created for all the document clusters by NIST assessors. Given a DUC topic and relevant documents, the tasks aims to create from the documents a brief, well-organized, fluent summary which answers the need for information expressed in the topic.

	DUC2005	DUC 2006	DUC 2007
Task	Only task	Only task	Main task
Number of topics (clusters)	50	50	45
Document number per topic	32	25	25
Data source	TREC	AQUAINT	AQUAINT
Summary length	250 Words	250 words	250 words

Table 1: Summary of datasets

As a preprocessing step for similarity computation, the stop words in each sentence were removed and the remaining words were stemmed using the Porter's stemmer[2].

We used the ROUGE-1.5.5 toolkit[3] for evaluation, which was officially adopted by DUC for automatically summarization evaluation. The toolkit measures summary quality by counting overlapping units such as the n-gram, word sequences and word pairs between the candidate summary and the reference summary. ROUGE-N is an n-gram based measure and the recall oriented score, the precision oriented score and the F-measure score for ROUGE-N are computed as follows:

$$ROUGE-N_{\mathrm{Re}call} = \frac{\sum_{S \in \{\mathrm{Re}\,f\,erence\ Summaries\}} \sum_{n\text{-}gram \in S} Count_{match}(n-gram)}{\sum_{S \in \{\mathrm{Re}\,f\,erence\ Summaries\}} \sum_{n\text{-}gram \in S} Count(n-gram)} \tag{23}$$

$$ROUGE-N_{\mathrm{Pr}ecision} = \frac{\sum_{S \in \{\mathrm{Re}\,f\,erence\ Summaries\}} \sum_{n\text{-}gram \in S} Count_{match}(n-gram)}{\sum_{S \in \{Candidate\ Summary\}} \sum_{n\text{-}gram \in S} Count(n-gram)} \tag{24}$$

$$ROUGE-N_{F-Measure} = \frac{2 \times ROUGE-N_{\mathrm{Re}call} \times ROUGE-N_{\mathrm{Pr}ecision}}{ROUGE-N_{\mathrm{Re}call} + ROUGE-N_{\mathrm{Pr}ecision}} \tag{25}$$

where n stands for the length of the n-gram, and *Count*$_{match}$*(n-gram)* is the maximum number of n-grams co-occurring in a candidate summary and a set of reference summaries. *Count(n-gram)* is the number of n-grams in the reference summaries or candidate summary.

The ROUGE toolkit reports separate F-measure scores for 1, 2, 3 and 4-gram, and also for longest common subsequence co-occurrences. Among these different scores, unigram-based ROUGE score (ROUGE-1) has been shown to agree with human judgment most [Lin and Hovy, 2003]. In this study, we show three ROUGE F-measure scores in the experimental results: ROUGE-1 (unigram-based), ROUGE-2 (bigram-based), and ROUGE-W (based on weighted longest common subsequence, weight=1.2).

5.2 Evaluation Results

In the experiments, the proposed multi-modality manifold-ranking methods with the three fusion schemes introduced in Sections 4.1-4.3 are denoted as "MultiMR(LIN)", "MultiMR(SEQ)" and "MultiMR(COM)", respectively. Note that the sequential fusion scheme is a two-stage optimization process and it relies on the sequence of the two modalities. Here we use "MultiMR(SEQ1)" to denote the fusion scheme introduced in Section 4.2 and use "MultiMR(SEQ2)" to denote the other case. The proposed meth-

[1] Each topic consists of a title and a narrative text, and we concatenate the title and narrative text to represent the topic.

[2] http://www.tartarus.org/martin/PorterStemmer/

[3] http://haydn.isi.edu/ ROUGE/

ods are compared with the basic manifold-ranking method (i.e. "SingleMR") and the NIST baseline. The NIST baseline is the official baseline system established by NIST. We also list the average ROUGE scores of all the participating systems for each task (i.e. AverageDUC). Tables 2, 3, 4 show the comparison results on DUC 2005-2007, respectively.

In the experiments, the regularized parameter for the *fitting constraint* is fixed at 0.4, as in [Wan et al., 2007]. Therefore, we have α=0.6 for "SingleMR", $\mu+\eta$=0.6 for "MultiMR(LIN)", $\mu+\eta-\mu\eta$=0.6 for "MultiMR(SEQ1)" and "MultiMR(SEQ2)", and $\mu=\eta$=0.6 for "MultiMR(COM)". We heuristically let $\mu=\eta$ for "MultiMR(LIN)", i.e. $\mu=\eta$=0.3 for "MultiMR(LIN)". For the sake of simplicity, in "MultiMR(SEQ1)" and "MultiMR(SEQ2)" we use the same value of μ as in "MultiMR(LIN)". In "MultiMR(COM)", the combination weight λ is simply set to 0.5.

System	ROUGE-1	ROUGE-2	ROUGE-W
MultiMR(COM)	0.37183*	0.06761	0.12927*
MultiMR(SEQ1)	0.36978*	0.06786	0.12878*
MultiMR(LIN)	0.36909*	0.06836*	0.12877
MultiMR(SEQ2)	0.36712	0.06747	0.12807
SingleMR	0.36316	0.06603	0.12694
AverageDUC	0.33875	0.05851	0.11514
NIST Baseline	0.28760	0.04195	0.09874

Table 2: Comparison results (F-measure) on DUC 2005

System	ROUGE-1	ROUGE-2	ROUGE-W
MultiMR(LIN)	0.40306*	0.08508*	0.13997*
MultiMR(SEQ1)	0.40189*	0.08441	0.13963*
MultiMR(COM)	0.40068*	0.08529*	0.13944
MultiMR(SEQ2)	0.39987	0.08477	0.13906
SingleMR	0.39534	0.08335	0.13766
AverageDUC	0.37789	0.07483	0.12943
NIST Baseline	0.32095	0.05269	0.10993

Table 3: Comparison results (F-measure) on DUC 2006

System	ROUGE-1	ROUGE-2	ROUGE-W
MultiMR(LIN)	0.42041*	0.10302*	0.14595
MultiMR(COM)	0.41837*	0.10263	0.14530
MultiMR(SEQ1)	0.41803	0.10292	0.14511*
MultiMR(SEQ2)	0.41600	0.10095	0.14441
SingleMR	0.41303	0.10009	0.14203
AverageDUC	0.40059	0.09550	0.13726
NIST Baseline	0.33434	0.06479	0.11360

Table 4: Comparison results (F-measure) on DUC 2007

(* indicates that the improvement over the baseline "SingleMR" is statistically significant.)

Seen from the tables, the proposed multi-modality manifold-ranking methods outperform the basic manifold-ranking method. The proposed method with the linear fusion scheme (i.e. "MultiMR(LIN)") performs the best on the DUC2006 and DUC2007 datasets, and it can significantly outperform the baseline SingleMR method. Overall, the linear fusion scheme is better than the other two fusion schemes, which demonstrates that the linear scheme is more appropriate for fusing the within-document modality and the cross-document modality in the manifold-ranking process.

We also find that both the multi-modality manifold-ranking methods and the basic manifold-ranking method can much outperform the NIST baseline. They can also achieve much higher ROUGE scores than the average scores of all the participating systems. As compared with the participating systems on each DUC task, our proposed methods can achieve comparable ROUGE scores with the best performing system. For example, the highest ROUGE-1 F-measure score on DUC2005 is 0.37437, and the highest ROUGE-1 F-measure score on DUC2006 is 0.40997.

In order to further investigate the influences of the parameters in the proposed multi-modality manifold-ranking methods, the parameter value of μ in "MultiMR(LIN)" is varied from 0 to 0.6, and thus the parameter value of η ranges from 0.6 to 0. Figures 1 and 2 show the ROUGE-1 and ROUGE-W F-measure curves of "MultiMR(LIN)" on the three datasets, respectively. We can see from the figures that both modalities are beneficial to the overall summarization performance. We also vary the parameter value of λ from 0 to 1 in "MultiMR(COM)", and Figures 3 and 4 show the ROUGE-1 and ROUGE-W F-measure curves of "MultiMR(COM)", respectively. The curves also demonstrate that both the within-document modality and the cross-document modality are important for ranking sentences.

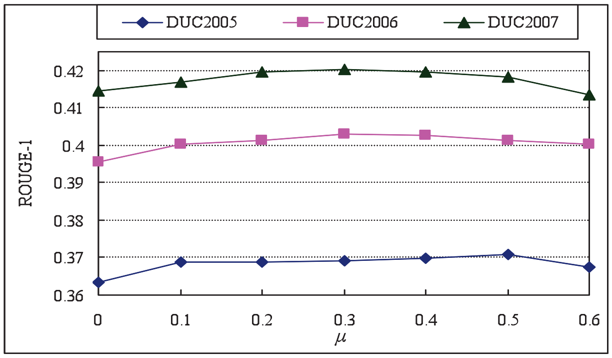

Figure 1: ROUGE-1 F-measure scores vs. μ for "MultiMR(LIN)"

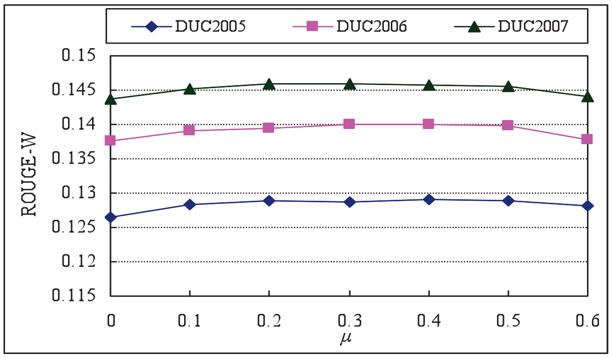

Figure 2: ROUGE-W F-measure scores vs. μ for "MultiMR(LIN)"

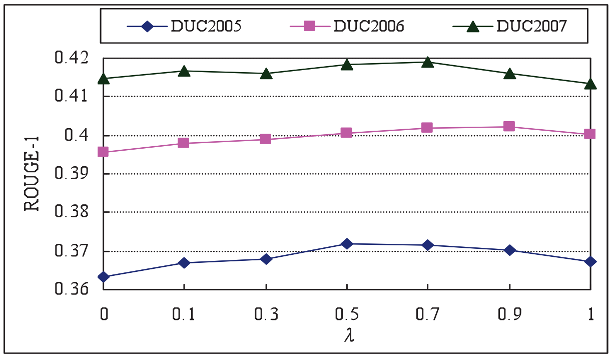

Figure 3: ROUGE-1 F-measure scores vs. λ for "MultiMR(COM)"

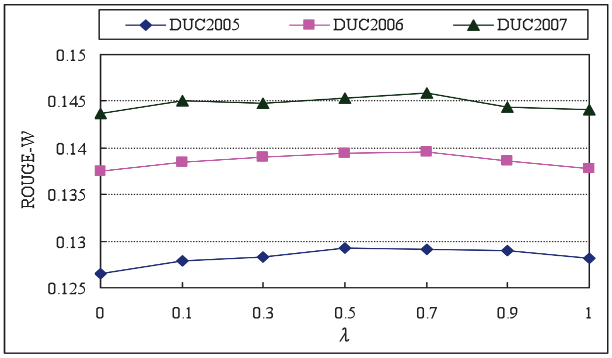

Figure 4: ROUGE-W F-measure scores vs. λ for "MultiMR(COM)"

6 Conclusion and Future Work

In this study we consider the within-document relationships and the cross-document relationships between sentences as two separate modalities, and propose to use the multi-modality manifold-ranking algorithm to fuse the two modalities. Experimental results demonstrate the effectiveness of the proposed methods.

In future work, we will analyze the DUC topic at a finer granularity by discovering relevant subtopics, and then consider the sentence relationships against each subtopic as a separate modality. The multi-modality manifold-ranking method can be exploited based on the constructed multiple modalities.

Acknowledgments

This work was supported by NSFC (60873155), RFDP (20070001059), Beijing Nova Program (2008B03), National High-tech R&D Program (2008AA01Z421) and NCET (08-0006).

References

[Daumé and Marcu, 2006] Hal Daumé and Daniel Marcu. Bayesian query-focused summarization. In *Proceedings of ACL-06*.

[Erkan and Radev, 2004] Gunes Erkan and Dragomir R. Radev. LexPageRank: prestige in multi-document text summarization. In *Proceedings of EMNLP-04*.

[Goldstein et al., 1999] J. Goldstein, M. Kantrowitz, V. Mittal and J. Carbonell. Summarizing Text Documents: Sentence Selection and Evaluation Metrics. *Proceedings of ACM SIGIR-99*.

[Harabagiu and Lacatusu, 2005] S. Harabagiu and F. Lacatusu. Topic themes for multi-document summarization. In *Proceedings of SIGIR-05*.

[Hardy et al., 2002] H. Hardy, N. Shimizu, T. Strzalkowski, L. Ting, G. B. Wise and X. Zhang. Cross-document summarization by concept classification. In *Proceedings of SIGIR-02.*

[Lin and Hovy, 2002] Chin.-Yew. Lin and Edword. H. Hovy. From Single to Multi-document Summarization: A Prototype System and its Evaluation. In *Proceedings of ACL-02*.

[Lin and Hovy, 2003] C.-Y. Lin and E.H. Hovy. Automatic Evaluation of Summaries Using N-gram Co-occurrence Statistics. In *Proceedings of HLT-NAACL -03*.

[Mihalcea and Tarau, 2005] R. Mihalcea and P. Tarau. A language independent algorithm for single and multiple document summarization. In *Proceedings of IJCNLP-05*.

[Nastase 2008] Vivi Nastase. Topic-driven multi-document summarization with encyclopedic knowledge and spreading activation. In *Proceedings of EMNLP-08*.

[Nenkova and Louis, 2008] Ani Nenkova and Annie Louis. Can you summarize this? Identifying correlates of input difficulty for generic multi-document summarization. In *Proceedings of ACL-08:HLT*.

[Ouyang et al., 2007] You Ouyang, Sujian Li, Wenjie Li. Developing learning strategies for topic-focused summarization. In *Proceedings of CIKM-07*.

[Radev et al., 2004] D. R. Radev, H. Y. Jing, M. Stys and D. Tam. Centroid-based summarization of multiple documents. *Information Processing and Management*, 40: 919-938, 2004.

[Schilder and Kondadadi, 2008] Frank Schilder and Ravikumar Kondadadi. FastSum: fast and accurate query-based multi-document summarization. In *Proceedings of ACL-08: HLT*.

[Shen et al., 2007] Dou Shen, Jian-Tao Sun, Hua Li, Qiang Yang, and Zheng Chen. Document summarization using conditional random fields. In Proceedings of IJCAI-07.

[Tong et al., 2005] Hanghang Tong, Jingrui He, Mingjing Li, Changshui Zhang and Wei-Ying Ma. Graph based multi-modality learning. In *Proceedings of ACM MM-05*.

[Wan et al., 2007] Xiaojun Wan, Jianwu Yang and Jianguo Xiao. Manifold-ranking based topic-focused multi-document summarization. In *Proceedings of IJCAI-07*.

[Wan and Yang, 2008] Xiaojun Wan and Jianwu Yang. Multi-document summarization using cluster-based link analysis. In *Proceedings of SIGIR-08*.

[Wang et al., 2008] Dingding Wang, Tao Li, Shenghuo Zhu, Chris Ding. Multi-document summarization via sentence-level semantic analysis and symmetric matrix factorization. In *Proceedings of SIGIR-08*.

[Wei et al., 2008] Furu Wei, Wenjie Li, Qin Lu and Yanxiang He. Query-sensitive mutual reinforcement chain and its application in query-oriented multi-document summarization. In *Proceedings of SIGIR-08*.

[Wong et al., 2008] Kam-Fai Wong, Mingli Wu and Wenjie Li. Extractive summarization using supervised and semi-supervised learning. In *Proceedings of COLING-08.*

[Zhang et al., 2008] Jin Zhang, Xueqi Cheng, Gaowei Wu, and Hongbo Xu. AdaSum: an adaptive model for summarization. In *Proceedings of CIKM-08*.

[Zhou et al., 2003a] D. Zhou, O. Bousquet, T. N. Lal, J. Weston and B. SchÖlkopf. Learning with local and global consistency. In *Proceedings of NIPS-03*.

[Zhou et al., 2003b] D. Zhou, J. Weston, A. Gretton, O. Bousquet and B. SchÖlkopf. Ranking on data manifolds. In *Proceedings of NIPS-03*.