Manifold-Ranking Based Topic-Focused Multi-Document Summarization

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

Abstract

Topic-focused multi-document summarization aims to produce a summary biased to a given topic or user profile. This paper presents a novel extractive approach based on manifold-ranking of sentences to this summarization task. The manifold-ranking process can naturally make full use of both the relationships among all the sentences in the documents and the relationships between the given topic and the sentences. The ranking score is obtained for each sentence in the manifold-ranking process to denote the biased information richness of the sentence. Then the greedy algorithm is employed to impose diversity penalty on each sentence. The summary is produced by choosing the sentences with both high biased information richness and high information novelty. Experiments on DUC2003 and DUC2005 are performed and the ROUGE evaluation results show that the proposed approach can significantly outperform existing approaches of the top performing systems in DUC tasks and baseline approaches.

1 Introduction

Multi-document summarization aims to produce a summary delivering the majority of information content from a set of documents about an explicit or implicit main topic. Topic-focused multi-document summarization is a particular kind of multi-document summarization. Given a specified topic description (i.e. user profile, user query), topic-focused multi-document summarization (i.e. query-based multi-document summarization) is to create from the documents a summary which either answers the need for information expressed in the topic or explains the topic.

Automatic multi-document summarization has drawn much attention in recent years and it exhibits the practicability in document management and search systems. Multi-document summary can be used to concisely describe the information contained in a cluster of documents and facilitate the users to understand the document cluster. For example, a number of news services, such as *Google News*[1], *NewsBlaster*[2], have been developed to group news articles into news topics, and then produce a short summary for each news topic. The users can easily understand the topic they have interest in by taking a look at the short summary. Topic-focused summary can be used to provide personalized services for users after the user profiles are created manually or automatically. The above news services can be personalized by collecting users' interests, and both the retrieved related news articles and the news summary biased to the user profile are delivered to the specified user. Other examples include Question/Answer systems, where a question-focused summary is usually required to answer the information need in the issued question.

The challenges for topic-focused multi-document summarization are as follows: the first one is a common problem for general multi-document summarization, that the information stored in different documents inevitably overlaps with each other, and hence we need effective summarization methods to merge information stored in different documents, and if possible, contrast their differences; the second one is a particular challenge for topic-focused multi-document summarization that the information in the summary must be biased to the given topic, so we need effective summarization methods to take into account this topic-biased characteristic during the summarization process. In brief, a good topic-focused summary is expected to preserve the information contained in the documents as much as possible, and at the same time keep the information as novel as possible, and moreover, the information must be biased to the given topic. In recent years, a series of workshops and conferences on automatic text summarization (e.g. NTCIR[3], DUC[4]), special topic sessions in ACL, COLING, and SIGIR have advanced the technology and produced a couple of experimental online systems.

In this study, we propose a novel extractive approach based on manifold-ranking [Zhou et al., 2003a; Zhou et al., 2003b] of sentences to topic-focused multi-document summarization. The proposed approach

[1] http://news.google.com
[2] http://www1.cs.columbia.edu/nlp/newsblaster/
[3] http://research.nii.ac.jp/ntcir/index-en.html
[4] http://duc.nist.gov

first employs the manifold-ranking process to compute the manifold-raking score for each sentence that denotes the biased information richness of the sentence, and then uses the greedy algorithm to penalize the sentences highly overlapping with other informative sentences. The summary is produced by choosing the sentences with highest overall scores, which are deemed both informative and novel, and highly biased to the given topic. In the manifold-ranking algorithm, the intra-document and inter-document links between sentences are differentiated with different weights. Experimental results on two DUC tasks show that the proposed approach significantly outperforms the top performing approaches in DUC tasks and baseline approaches.

In the rest of this paper: Section 2 discusses previous work. The proposed summarization approach is proposed in Section 3. Section 4 describes the evaluation results. Section 5 presents our conclusion and future work.

2 Previous Work

A variety of multi-document summarization methods have been developed recently. Generally speaking, the methods can be either extractive summarization or abstractive summarization. Extractive summarization involves assigning salience scores to some units (e.g. sentences, paragraphs) of the documents and extracting the sentences with highest scores, while abstraction summarization (e.g. NewsBlaster) usually needs information fusion, sentence compression and reformulation. In this study, we focus on extractive summarization.

The centroid-based method [Radev et al., 2004] is one of the most popular extractive summarization methods. MEAD is an implementation of the centroid-based method that scores sentences based on such features, as cluster centroids, position, TF*IDF. NeATS [Lin and Hovy, 2002] uses sentence position, term frequency, topic signature and term clustering to select important content, and use MMR [Goldstein et al., 1999] to remove redundancy. XDoX [Hardy et al., 2002] first identifies the most salient themes within the document set by passage clustering and then composes an extraction summary, which reflects these main themes. Harabagiu and Lacatusu [2005] investigate five different topic representations and introduce a novel representation of topics based on topic themes. Recently, graph-based methods have been proposed to rank sentences or passages. Websumm [Mani and Bloedorn, 2000], LexPageRank [Erkan and Radev, 2004] and Mihalcea and Tarau [2005] are three such systems using algorithms similar to PageRank and HITS to compute sentence importance.

Most topic-focused document summarization methods incorporate the information of the given topic or query into generic summarizers and extracts sentences suiting the user's declared information need. In [Saggion et al., 2003], a simple query-based scorer by computing the similarity value between each sentence and the query is incorporated into a generic summarizer to produce the query-based summary. The query words and named entities in the topic description are investigated in [Ge et al., 2003] and CLASSY [Conroy and Schlesinger, 2005] for event-focused/query-based multi-document summarization. In [Hovy et al., 2005], the important sentences are selected based on the scores of basic elements (BE). CATS [Farzindar et al., 2005] is a topic-oriented multi-document summarizer which first performs a thematic analysis of the documents, and then matches these themes with the ones identified in the topic. More related work can be found on DUC 2003 and DUC 2005 publications.

To the best of our knowledge, the above systems are usually simple extensions of generic summarizers and do not uniformly fuse the information in the topic and the documents. While our approach can naturally and simultaneously take into account that information in the manifold-ranking process and select the sentences with both high biased information richness and information novelty.

3 The Manifold-Ranking Based Approach

3.1 Overview

The manifold-ranking based summarization approach consists of two steps: (1) the manifold-ranking score is computed for each sentence in the manifold-ranking process where the score denotes the biased information richness of a sentence; (2) based on the manifold-ranking scores, the diversity penalty is imposed on each sentence and the overall ranking score of each sentence is obtained to reflect both the biased information richness and the information novelty of the sentence. The sentences with high overall ranking scores are chosen for the summary. The definitions of biased information richness and information novelty are given as below:

Biased Information Richness: *Given a sentence collection* $\chi=\{x_i \mid 1\le i\le n\}$ *and a topic T, the biased information richness of sentence* x_i *is used to denote the information degree of the sentence* x_i *with respect to both the sentence collection and T, i.e. the richness of information contained in the sentence* x_i *biased towards T.*

Information Novelty: *Given a set of sentences in the summary* $R=\{x_i \mid 1\le i\le m\}$*, the information novelty of sentence* x_i *is used to measure the novelty degree of information contained in the sentence* x_i*, with respect to all other sentences in the set R..*

The underlying idea of the proposed approach is that a good summary is expected to include the sentences with both high biased information richness and high information novelty.

3.2 Manifold-Ranking Process

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. An intuitive description of manifold-ranking is as follows: A weighted network is formed on the data, and a positive rank score is assigned to each known relevant point and zero to the remaining points which are to be ranked. All points then spread their ranking score to their nearby neighbors via the weighted network. The spread process is repeated until a global stable state is achieved, and all points obtain their final ranking scores.

In our context, the data points are denoted by the topic description and all the sentences in the documents. The manifold-ranking process in our context can be formalized as follows:

Given a set of data points $\chi = \{x_0, x_1, ..., x_n\} \subset R^m$, the first point x_0 is the topic description and the rest n points are the sentences in the documents. Note that because the topic description is usually short in our experiments and we treat it as a pseudo-sentence[5], and then it can be processed in the same way as other sentences. Let $f : \chi \to R$ 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 $f=[f_0,\ldots,f_n]^T$. We also define a vector $y=[y_0,\ldots,y_n]^T$, in which $y_0=1$ because x_0 is the topic sentence and $y_i=0$ ($1 \le i \le n$) for all the sentences in the documents. The manifold ranking algorithm goes as follows:

1. Compute the pair-wise similarity values between sentences (points) using the standard Cosine measure. The weight associated with term t is calculated with the tf_t*isf_t formula, where tf_t is the frequency of term t in the sentence and isf_t is the inverse sentence frequency of term t, i.e. $1+\log(N/n_t)$, where N is the total number of sentences and n_t is the number of the sentences containing term t. Given two sentences (data points) x_i and x_j, the Cosine similarity is denoted as $sim(x_i,x_j)$, computed as the normalized inner product of the corresponding term vectors.
2. Connect any two points with an edge if their similarity value exceeds 0. We define the affinity matrix W by $W_{ij}=sim(x_i,x_j)$ if there is an edge linking x_i and x_j. Note that we let $W_{ii}=0$ to avoid loops in the graph built in next step.
3. Symmetrically normalize W by $S=D^{-1/2}WD^{-1/2}$ in which D is the diagonal matrix with (i,i)-element equal to the sum of the i-th row of W.
4. Iterate $f(t+1)= \alpha Sf(t)+(1-\alpha)y$. until convergence, where α is a parameter in (0,1).
5. Let f_i^* denote the limit of the sequence $\{f_i(t)\}$. Each sentences x_i ($1 \le i \le n$) gets its ranking score f_i^*.

Figure 1: The manifold-ranking algorithm.

[5] The topic can also be represented by more than one sentence, and in this case only the vector y needs to be modified to represent all the topic sentences in the manifold-ranking algorithm.

In the above iterative algorithm, the normalization in the third step is necessary to prove the algorithm's convergence. The fourth step is the key step of the algorithm, where all points spread their ranking score to their neighbors via the weighted network. The parameter of manifold-ranking weight α specifies the relative contributions to the ranking scores from neighbors and the initial ranking scores. Note that self-reinforcement is avoided since the diagonal elements of the affinity matrix are set to zero.

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

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

where $\beta=1-\alpha$. Although f^* can be expressed in a closed form, for large scale problems, the iteration algorithm is preferable due to computational efficiency. Usually the convergence of the iteration algorithm is achieved when the difference between the scores computed at two successive iterations for any point falls below a given threshold (0.0001 in this study).

Note that in our context, the links (edges) between sentences in the documents can be categorized into two classes: intra-document link and inter-document link. Given a link between a sentence pair of x_i and x_j, if x_i and x_j come from the same document, the link is an intra-document link; and if x_i and x_j come from different documents, the link is an inter-document link. The links between the topic sentence and any other sentences are all inter-document links. We believe that intra-document links and inter-document links have unequal contributions in the above iterative algorithm. In order to investigate this intuition, distinct weights are assigned to the intra-document links and the inter-document links respectively. In the second step of the above algorithm, the affinity matrix W can be decomposed as

$$W = W_{\text{intra}} + W_{\text{inter}} \qquad (2)$$

where W_{intra} is the affinity matrix containing only the intra-document links (the entries of inter-document links are set to 0) and W_{inter} is the affinity matrix containing only the inter-document links (the entries of intra-document links are set to 0).

We differentiate the intra-document links and inter-document links as follows:

$$\widetilde{W} = \lambda_1 W_{\text{intra}} + \lambda_2 W_{\text{inter}} \qquad (3)$$

We let $\lambda_1, \lambda_2 \in [0,1]$ in the experiments. If $\lambda_1<\lambda_2$, the inter-document links are more important than the intra-document links in the algorithm and vice versa. Note that if $\lambda_1=\lambda_2=1$, Equation (3) reduces to Equation (2). In the manifold-ranking algorithm, $\widetilde{W}$ is normalized into $\widetilde{S}$ in the third step and the fourth step uses the following iteration form: $f(t+1)= \alpha \widetilde{S} f(t)+(1-\alpha)y$.

3.3 Diversity Penalty Imposition

The original affinity matrix W is normalized by $\bar{S} = D^{-1}W$ to make the sum of each row equal to 1. Based on $\bar{S}$, the greedy algorithm similar to [Zhang et al., 2005] is applied to impose the diversity penalty and compute the final overall ranking scores, reflecting both the biased information richness and the information novelty of the sentences. The algorithm goes as follows:

1. Initialize two sets A=Ø, $B=\{x_i \mid i=1,2,\ldots,n\}$, and each sentence's overall ranking score is initialized to its manifold-ranking score, i.e. $RankScore(x_i) = f_i^*$, i=1,2,…n.
2. Sort the sentences in B by their current overall ranking scores in descending order.
3. Suppose x_i is the highest ranked sentence, i.e. the first sentence in the ranked list. Move sentence x_i from B to A, and then the diversity penalty is imposed to the overall ranking score of each sentence linked with x_i in B as follows:
   For each sentence $x_j \in B$,
   $RankScore(x_j) = RankScore(x_j) - \omega \cdot \overline{S}_{ji} \cdot f_i^*$
   where ω>0 is the penalty degree factor. The larger ω is, the greater penalty is imposed to the overall ranking score. If ω=0, no diversity penalty is imposed at all.
4. Go to step 2 and iterate until B= Ø or the iteration count reaches a predefined maximum number.

Figure 2: The algorithm of diversity penalty imposition.

In the above algorithm, the third step is crucial and its basic idea is to decrease the overall ranking score of less informative sentences by the part conveyed from the most informative one. After the overall ranking scores are obtained for all sentences, several sentences with highest ranking scores are chosen to produce the summary according to the summary length limit.

4 Experiments

4.1 Data Set

Topic-focused multi-document summarization has been evaluated on tasks 2 and 3 of DUC 2003 and the only task of DUC 2005, each task having a gold standard data set consisting of document clusters and reference summaries. In our experiments, task 2 of DUC 2003 was used for training and parameter tuning and the other two tasks were used for testing. Note that the topic representations of the three topic-focused summarization tasks are different: task 2 of DUC 2003 is to produce summaries focused by *events*; task 3 of DUC 2003 is to produce summaries focused by *viewpoints*; the task of DUC 2005 is to produce summaries focused by *DUC Topics*. In the experiments, the above topic representations were treated uniformly because they were deemed to have no substantial differences from each other. Table 1 gives a short summary of the three data sets.

As a preprocessing step, dialog sentences (sentences in quotation marks) were removed from each document. The stop words in each sentence were removed and the remaining words were stemmed using the Porter's stemmer[6].

	DUC 2003	DUC 2003	DUC 2005
Task	Task 2	Task 3	the only task
Number of clusters	30	30	50
Data source	TDT	TREC	TREC
Summary length	100 words	100 words	250 words

Table 1: Summary of data sets used in the experiments.

4.2 Evaluation Metric

We used the ROUGE [Lin and Hovy, 2003] toolkit[7] for evaluation, which was adopted by DUC for automatically summarization evaluation. It 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 recall measure computed as follows:

$$ROUGE-N = \frac{\sum_{S\in\{Re f Sum\}}\sum_{n\text{-}gram\in S} Count_{match}(n-gram)}{\sum_{S\in\{Re f Sum\}}\sum_{n\text{-}gram\in S} Count(n-gram)} \quad (4)$$

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

The ROUGE toolkit reports separate 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]. We show three of the ROUGE metrics in the experimental results, at a confidence level of 95%: ROUGE-1 (unigram-based), ROUGE-2 (bigram-based), and ROUGE-W (based on weighted longest common subsequence, weight=1.2).

In order to truncate summaries longer than length limit, we used the "-l" option in the ROUGE toolkit and we also used the "-m" option for word stemming.

4.3 Experimental Results

4.3.1 System Comparison

In the experiments, the proposed approach was compared with top three systems and four baseline systems on task 3 of DUC 2003 and the only task of DUC 2005 respectively. The top three systems are the systems with highest ROUGE scores, chosen from the performing systems on each task respectively. The *lead baseline* and *coverage baseline* are two baselines employed in the topic-focused multi-document summarization tasks of DUC 2003 and 2005. The *lead baseline* takes the first sentences one by one in the last document in the collection, where documents are assumed to be ordered chronologically. And

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

[7] We use ROUGEeval-1.4.2 downloaded from http://haydn.isi.edu/ ROUGE/

the *coverage baseline* takes the first sentence one by one from the first document to the last document.

In addition to the two standard baseline systems, we have implemented two other baseline systems, i.e. *Similarity-Ranking*1 and *Similarity-Ranking*2. The *Similarity-Ranking*1 first computes the similarity between the topic description and each sentence in the documents, and then the greedy algorithm proposed in Section 3.3 is employed to impose the diversity penalty on each sentence, with the normalized similarity value as the initial overall ranking score. The sentences with highest overall ranking scores are chosen to produce the summary. In essence, the *Similarity-Ranking*1 can be considered as a simplified version of the proposed manifold-ranking based system by ignoring the relationships between the sentences in the documents. And the *Similarity-Ranking*2 does not employ the diversity penalty imposition process and simply ranks the sentences by their similarity value with the topic description, which can be considered as a simplified version of *Similarity-Ranking*1 without the step of imposing diversity penalty.

Tables 2 and 3 show the system comparison results on the two tasks respectively. In the tables, S4-S17 are the system IDs of the top performing systems, whose details are described in DUC publications. The *Manifold-Ranking* adopts the proposed approach described in Section 3. The parameters of the *Manifold-Ranking* are set as follows: ω=8, λ_1=0.3 and λ_2=1, α=0.6. And the only parameter of the *Similarity-Ranking*1 is set as ω=8.

System	ROUGE-1	ROUGE-2	ROUGE-W
Manifold-Ranking	**0.37332**	**0.07677**	**0.11869**
Similarity-Ranking1	0.36088	0.07229	0.11540
S16	0.35001	0.07305	0.10969
Similarity-Ranking2	0.34542	0.07283	0.11155
S13	0.31986	0.05831	0.10016
S17	0.31809	0.04981	0.09887
Coverage Baseline	0.30290	0.05968	0.09678
Lead Baseline	0.28200	0.04468	0.09077

Table 2: System comparison on Task 3 of DUC 2003.

System	ROUGE-1	ROUGE-2	ROUGE-W
Manifold-Ranking	**0.38434**	**0.07317**	**0.10226**
S4	0.37396	0.06842	0.09867
S15	0.37383	0.07244	0.09842
Similarity-Ranking1	0.37356	0.06838	0.09949
S17	0.36901	0.07165	0.09751
Similarity-Ranking2	0.35752	0.06893	0.09596
Coverage Baseline	0.34568	0.05915	0.09103
Lead Baseline	0.30470	0.04764	0.08084

Table 3: System comparison on the task of DUC 2005.

Seen from Tables 2 and 3, the proposed system (i.e. *Manifold-Ranking*) outperforms the top performing systems and all baseline systems on all three tasks over all ROUGE scores[8]. The high performance achieved by the *Manifold-Ranking* benefits from the following factors:

1) **Manifold-ranking process**: The manifold-ranking process in the proposed approach makes full use of the inter-relationships between sentences by spreading the rank scores. In comparison with the *Similarity-Ranking*1, the ROUGE-1 scores of the proposed approach increase by 0.01244 and 0.01038 on the two tasks, respectively.

2) **Diversity penalty imposition**: If the proposed approach does not impose diversity penalty on sentences (i.e. ω=0), the ROUGE-1 scores will decrease by 0.02778 and 0.01952 on the two tasks, respectively. We can also see for the tables that the *Similarity-Ranking*1 much outperforms the *Similaity-Ranking*2 because of imposing diversity penalty on sentences.

3) **Intra-document/Inter-document link differentiation**: If the proposed approach does not differentiate the intra-document and inter-document links between sentences (i.e. λ_1=λ_2=1), the ROUGE-1 scores will slightly decrease by 0.00139 and 0.0007 on the two tasks, respectively.

In next sections we will mainly show ROUGE-1 performance due to page limit.

4.3.2 Parameter Tuning

Figure 3 demonstrates the influence of the penalty factor ω in the proposed approach (i.e. *Manifold-Ranking*) and the baseline approach (i.e. *Similarity-Ranking*1) when λ_1:λ_2=0.3:1 and α=0.6. We can see that when ω varies from 0 to 20, the performances of the *Manifold-Ranking* are always better than the corresponding performances of the *Similarity-Ranking*1 on the two tasks, respectively. This verifies that the use of the relationships between the sentences of the documents in the proposed approach can benefit the summarization task. It is also clear that no diversity penalty and too much diversity penalty will deteriorate the performances.

Figure 4 demonstrates the influence of the intra-document/inter-document link differentiating weight λ_1:λ_2 in the proposed approach when ω=8 and α=0.6. λ_1 and λ_2 range from 0 to 1 and λ_1: λ_2 denotes the real values λ_1 and λ_2 are set to. Different λ_1: λ_2 gives different contribution weights to the intra-document links and the inter-document links. It is observed that when more importance is attached to the intra-document links (i.e. λ_1 =1 and λ_2<0.9), the performances decrease evidently. It is the worst case when inter-document links are not taken into account (i.e. λ_1: λ_2=1:0), however, when intra-document links are not taken into account (i.e. λ_1: λ_2=0:1), the performances are still very well, which demonstrates that inter-document links are more important than intra-document links for the summarization task.

Figure 5 demonstrates the influence of the manifold weight α in the manifold-ranking algorithm of the proposed approach when ω=8, λ_1:λ_2=0.3:1.

[8] The improvement is significant over the ROUGE-1 score by comparing the 95% confidence intervals provided by the ROUGE package.

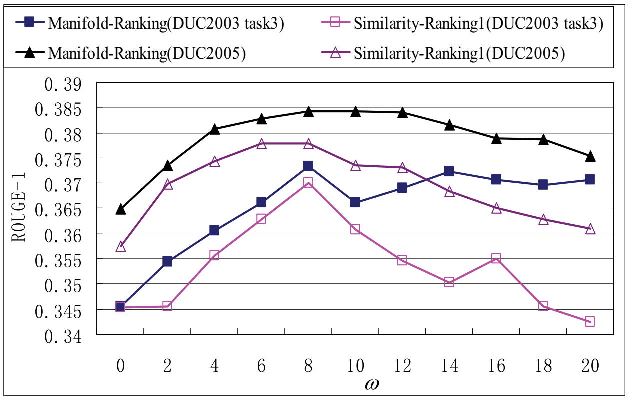

Figure 3: ROUGE-1 vs. ω.

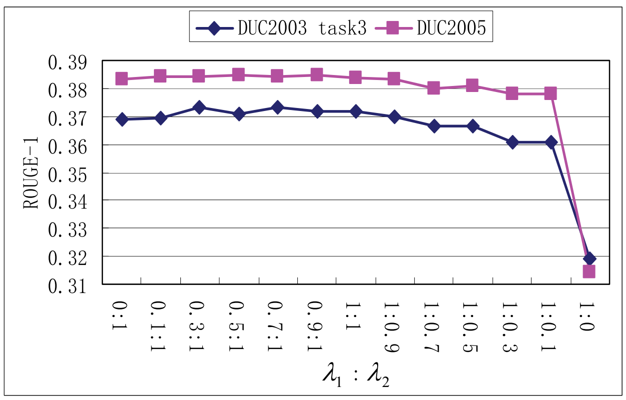

Figure 4: ROUGE-1 vs. λ_1: λ_2.

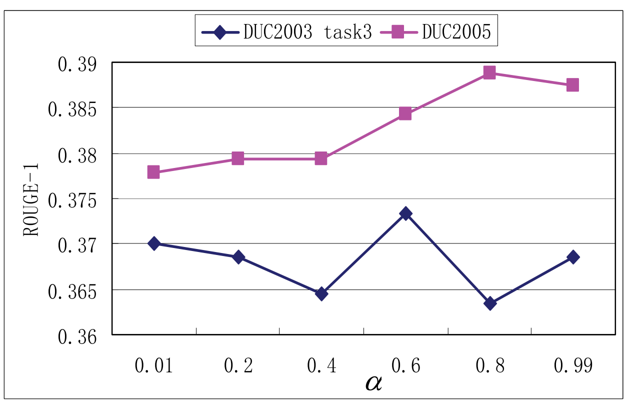

Figure 5: ROUGE-1 vs. α.

5 Conclusion and Future Work

In this paper we propose the manifold-ranking based approach to topic-focused multi-document summarization. The proposed approach employs the manifold-ranking process to make full use of the relationships among sentences and the relationships between the topic and the sentences.

In future work, we will employ machine learning approaches to automatically estimate the parameters in the proposed approach.

References

[Conroy and Schlesinger, 2005] J. M. Conroy and J. D. Schlesinger. 2005. CLASSY query-based multi-document summarization. In *Proceedings of DUC'2005*.

[Erkan and Radev, 2004] G. Erkan and D. Radev. LexPageRank: prestige in multi-document text summarization. In *Proceedings of EMNLP'2004*.

[Farzindar et al., 2005] A. Farzindar, F. Rozon and G. Lapalme. 2005. CATS a topic-oriented multi-document summarization system at DUC 2005. In *Proceedings of the 2005 Document Understanding Workshop (DUC2005)*.

[Ge et al., 2003] J. Ge, X. Huang and L. Wu. Approaches to event-focused summarization based on named entities and query words. In *Proceedings of the 2003 Document Understanding Workshop (DUC2003)*.

[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-1999*.

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

[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'2002*.

[Hovy et al., 2005] E. Hovy, C.-Y. Lin and L. Zhou. 2005. A BE-based multi-document summarizer with query interpretation. In *Proceedings of DUC2005*.

[Lin and Hovy, 2002] C.-Y. Lin and E. H. Hovy. From Single to Multi-document Summarization: A Prototype System and its Evaluation. In *Proceedings of ACL'2002*.

[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 2003*.

[Mani and Bloedorn, 2000] I. Mani and E. Bloedorn. Summarizing Similarities and Differences Among Related Documents. *Information Retrieval*, 1(1), 2000.

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

[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.

[Saggion et al., 2003] H. Saggion, K. Bontcheva and H. Cunningham. Robust generic and query-based summarization. In *Proceedings of EACL'2003*.

[Zhang et al., 2005] B. Zhang, H. Li, Y. Liu, L. Ji, W. Xi, W. Fan, Z. Chen, and W.-Y. Ma. Improving web search results using affinity graph. In *Proceedings of SIGIR'2005*.

[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'2003*.

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