Graph-Based Marginal Ranking for Update Summarization

Xuan Li*† Liang Du*† Yi-Dong Shen*

Abstract

Update summarization is to summarize a document collection B given that the users have already read another document collection A, which has time stamp prior to that of B. An important and challenging issue in update summarization is that contents in B already covered by A should be excluded from the update summary. In this paper, we propose a graph-based regularization framework MarginRank for update summarization. MarginRank extends the cost function of Zhou's Manifold Ranking with suppression terms, suppression of A on B, to fulfil the assumption that users have read A. MarginRank ranks sentences in B in a way that the top ranked sentences are most important and at the same time cover different contents from A. Experiments on the benchmark data sets TAC 2008 and 2009 show the effectiveness of the proposed method.

1 Introduction

Automatic document summarization is a long standing research topic, which dates back to the pioneering work of Luhn [12]. It has received great attention in recent years due to the explosive growth of documents on the World Wide Web. People are unable to go through the massive amount of documents, so it is desirable to design an engine that automatically summarizes the documents. Traditional summarization tasks can be mainly categorized as generic summarization [6, 5, 14, 7] and topic-focused summarization [13, 16, 21, 26]. Generic summarization is to automatically summarize a document or document collection using computers. Topic-focused summarization, receiving more attention in these years, is to generate summaries that are related to a given topic.

In 2007, a new summarization task[1] called update summarization is proposed to summarize evolutionary document collections by the Document Understanding Conferences[2] (DUC). The task can be generally described as, given an evolutionary document collection, generating an update summary that summarizes the contents of the collection at each time stamp. Update summarization has wide applications. For example, news reports about a typhoon disaster are continuously released, evolving with the development of the event from time to time. Readers may expect the summarization engine to provide them the summaries of the latest news reports. A consensus on the task is that when summarizing the documents, contents that are already read by the users in prior documents should not be included in the summary. There have been attempts [3, 10, 22] to employ different approaches to generate update summaries.

The state-of-the-art graph-ranking based method PNR^2 [10] extends the well-known traditional summarization method TextRank [14] for update summarization. PNR^2 performs positive reinforcements within the past and current document collections respectively and negative reinforcements between them. Positive reinforcements are to determine the importance of the sentences while negative reinforcements are to avoid content overlapping. However, the positive reinforcements within the current documents may suffer from the integration of the past documents. It can not reflect the actual reinforcements. Also, it is weird for the current documents to have negative reinforcements on the past ones, considering that the users read the documents in order of time. In this paper, we propose a novel regularization framework MarginRank (abbreviation of Marginal Ranking) that avoids these two problems. We call it marginal ranking because ranking aims to find sentences that have maximal marginal relevance to the topic, considering sentences covered by the past documents are of no relevance. The cost function of MarginRank mainly consists of three parts: manifold regularization terms and topic-relevance fitting terms to determine the importance and topic relevance of the sentences and suppression terms to avoid content overlapping. Through minimizing the cost function, MarginRank scores sentences in the current documents. The higher score a sentence has, the more marginal relevance

*State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing 100190, China {lixuan,duling,ydshen}@ios.ac.cn

†Graduated University, Chinese Academy of Sciences, Beijing 100049, China

[1] http://www-nlpir.nist.gov/projects/duc/duc2007/tasks.html

[2] http://duc.nist.gov/

the sentence has.

Another advantage of MarginRank is that it can easily incorporate the novelty requirement into the framework. Novelty is a common requirement in summarization. It requires that sentences in a summary should be distinct from each other in the sense that they describe different contents. With MarginRank, novelty of the summary can be achieved by imposing some inequality constraints.

Experiments on the benchmark data sets TAC 2008 and 2009 show that our algorithm significantly outperforms PNR^2 and several other state-of-the-art update summarization algorithms.

The paper is organized as follows. We discuss related work in section 2. We state the update summarization problem formally in section 3. MarginRank is described in section 4. The update summarization algorithm is given in section 5. Experimental results are shown in section 6. We conclude the paper in section 7.

2 Related Work

2.1 Document Summarization Update summarization is closely related to the traditional generic and topic-focused summarization. We first briefly mention some traditional summarization methods.

Some work [6, 19, 2, 27] views summarization as a sentence classification task, supervised or semisupervised. Sentences are classified to be summary-included or summary-excluded. Other work [23, 13, 7, 14, 15, 16, 21] views summarization as a sentence ranking task, supervised or unsupervised. Sentences are ranked according to some importance measure, and the most important sentences are extracted to form the summary. Among these methods, unsupervised graph-based ranking methods have gain much attention due to the lack of labeled data. Clustering methods are also utilized to cluster sentences in [18, 28, 26] so that the summary may cover different aspects of the document (collection).

Wan [22] proposes the TimedTextRank algorithm, adapted from TextRank [14], for evolutionary document collections. This method does not consider the removal of overlapping contents between current and previous documents but rather emphasizes the relative importance of current documents. Boudin et al. [3] extend the maximal marginal relevance method [4] for update summarization, where sentences are penalized based on their similarities to the previously read documents. The more similar a sentence to previous documents is, the more the sentence is penalized. PNR^2, proposed by Li et al. [10], performs both positive reinforcements to determine the importance of the sentences and negative reinforcements to avoid content overlapping for update summarization. To summarize a document collection "B" that evolves from "A", PNR^2 builds the adjacency matrix of $A \cup B$, in which the entries that connect A and B are negative while the rest are positive. It normalizes the columns of the adjacency matrix to form a pseudo transition matrix before applying the PageRank algorithm to perform reinforcements. However, incorporating A for the column normalization of B may cause some problems for the self-positive reinforcements of B. The positive reinforcements can not reflect the actual reinforcements. It is also weird for B to have negative reinforcements on A. Most of the participating systems in TAC 2008 and 2009 utilize traditional summarization methods for update summarization task by taking task-specific document preprocessing or postprocessing. For example, System 45 [29] and System 4 [9] both appeal to Manifold Ranking algorithm for TAC 2009. System 45 prunes severely overlapping sentences before ranking while System 4 reranks sentences after ranking.

There are also some summarization problem settings that are quite similar to update summarization. Wang et al. [24] propose to generate summaries of an evolutionary document collection that can discriminate the documents over different time periods. I.e., find the sentences that can represent the topic themes particular to the document collections over different time periods. They compute the mutual information between the sentences and the topic to select sentences. Comparative document summarization, proposed by Wang et al. [25], is also relevant to update summarization. Given a set of document collections, the problem is to generate a summary for each document collection that can discriminate the document collection from the others. It emphasizes the differences of the document collections while update summarization takes the differences into account for better summarization.

2.2 Novelty in Document Summarization Novelty (also called diversity) is an important requirement in document summarization. It requires that sentences in the summary should not refer to the same contents. Here we review some work that studies novelty in document summarization and the closely related question answering community.

Carbonell and Goldstein [4] first proposed Maximal Marginal Relevance for imposing novelty on automatic generated summaries. Sentences are ranked by combining their relevance to the query and similarities to the sentences selected already. Zhu et al. [31] propose Grasshopper that emphasizes novelty of summaries. Grasshopper is based on absorbing random walk. It computes the expected number of visits of an item (sentence) before absorbing and always selects the item that has the maximal visits. Li et al. [7] incorporate the novelty requirement into the summarizing process by employing structural SVM for document summarization. Both [31] and [7] study generic summarization, while we focus on topic-focused summarization in this paper. Li et al. [11] appeal to a probabilistic method that ensures novelty of the topic-focused summaries. Achananuparp et al. [1] proposed DiverseRank algorithm for diversifying the answers to a given question. However, since the method uses purely negative reinforcements among the answers, answers in a strongly connected clique may not be selected. Our method employs both positive and negative reinforcements.

3 Problem Definition

The update summarization task can be defined as follows.

Update summarization: given a topic T and a set of evolutionary document collection $< D_1, D_2, ... >$, where D_k $(k = 1, 2, ...)$ is a document collection and documents in D_i have time stamp prior to those in D_j for any $i < j$, generate a topic-focused summary for D_k $(k = 2, 3, ...)$ supposing the users have read $D_1, ..., D_{k-1}$.

The DUC 2007 update summarization pilot task have three document collections. Its successor TAC update summarization task reduces the number of document collections to two, A and B. A has time stamp prior to that of B. The topics in TAC are closely related to A and B. The following is a sample topic in TAC 2008. The title is general enough to cover document collections A and B, while narrative expresses the information need of the users.

```
<title> Hurricane Katrina </title>

<narrative>
  Describe the death toll, monetary aid,
  economic impact and humanitarian relief
  of Hurricane Katrina.
</narrative>
```

We take the update summarization task as a sentence ranking problem. Sentences should be ranked in a way that the more relevant to the topic and the more irrelevant to A a sentence is, the higher the sentence should be ranked. We call the ranking problem marginal ranking.

4 The Regularization Framework

In this section, we describe MarginRank, a graph-based ranking framework. We first introduce the cost function of MarginRank. Then we derive the algorithm for minimizing the cost function. We conclude this section by giving some interpretations to MarginRank.

4.1 The Cost Function Before we discuss the cost function, we first introduce some notations. Let $\mathbf{s} = [s_1, ..., s_n]^T$ be the sentences in the document collections A and B and $\mathbf{f} = [f_1, ..., f_n]^T$ the corresponding ranking scores. For simplicity, by $i \in A$ (or B) we denote sentence s_i belongs to A (or B). Similarity between two sentences s_i and s_j $(i \neq j)$ is denoted by w_{ij}. $w_{ii} = 0$ for all i. Let $D_i^A = \sum_{j\in A} w_{ij}$ $(i \in A)$, $D_i^B = \sum_{j\in B} w_{ij}$ $(i \in B)$, $D_i^{AB} = \sum_{j\in B} w_{ij}$ $(i \in A)$ and $D_i^{BA} = \sum_{j\in A} w_{ij}$ $(i \in B)$. Let $\mathbf{p} = [p_1, ..., p_n]^T$ denote the relevance of the sentences $s_1, ..., s_n$ to the topic respectively, where p_i $(1 \le i \le n)$ can be computed by the cosine similarity between s_i and the topic. We normalize $\mathbf{p}$ by dividing it by its L_2 norm $||\mathbf{p}||_2$.

We want to determine the ranking scores $\mathbf{f}$. We have several considerations for the cost function. According to the manifold ranking paradigm, similar sentences should have similar ranking scores in document collections A and B respectively. Another consideration is the update summarization requirement, i.e., contents in B that are already covered by A should not be repeated in the update summary. Thus such sentences should have low ranking scores. In other words, the ranking scores of the sentences in B that are similar to the sentences in A should be suppressed. A third consideration is the topic relevance, which can be taken as the prior ranking scores of the sentences. A possible way to integrate all these factors is to introduce the following cost function associated with $\mathbf{f}$:

$$\begin{aligned}\mathcal{L}(\mathbf{f}) =& \alpha_1 \sum_{i\in A}\sum_{j\in A} w_{ij}\left(\frac{f_i}{\sqrt{D_i^A}} - \frac{f_j}{\sqrt{D_j^A}}\right)^2 \\ &+ \alpha_2 \sum_{i\in B}\sum_{j\in B} w_{ij}\left(\frac{f_i}{\sqrt{D_i^B}} - \frac{f_j}{\sqrt{D_j^B}}\right)^2 \\ &- \alpha_3 \sum_{i\in A}\sum_{j\in B} w_{ij}\left(\frac{f_i}{\sqrt{D_i^{AB}}} - \frac{f_j}{\sqrt{D_j^{BA}}}\right)^2 \\ &+ \alpha_4 \sum_{i\in A\cup B} (f_i - p_i)^2\end{aligned}$$

where $\alpha_1, \alpha_2, \alpha_3$ and α_4 are tuning parameters.

The first and the second parts mean that similar sentences should have similar scores, with w_{ij} as the weighting parameter. Please note that the ranking scores are normalized by $\sqrt{D_i^A}$, $\sqrt{D_i^B}$, $\sqrt{D_i^{AB}}$ and $\sqrt{D_i^{BA}}$, the square root of the weighted degrees of the sentences. Such normalization can avoid sentences that have large degrees dominating the ranking results. The third part suppresses the scores of the sentences in B. The fourth part consists of the fitting terms, fitting the ranking scores to the prior topic relevance. However, there are two challenging problems remain in the cost function. The first is how to ensure that sentences in B are suppressed instead of being boosted. Note that boosting B and suppressing A may also lead to cost decline. The second is how to ensure that the cost function has lower bound.

For the first problem, we add the following constraints:

$$\frac{f_j}{\sqrt{D_j^{BA}}} \leq \frac{f_i}{\sqrt{D_i^{AB}}}, \text{ if } w_{ij} > 0 \text{ for all } j \in B, i \in A$$

The more similar a sentence to A is, the more the sentence will be suppressed, which is obvious by the weighting parameter w_{ij}. Also, the more important a sentence in A is (i.e., greater $f_i/\sqrt{D_i^{AB}}$), the more suppression it will have on the sentences in B. The following simple calculation confirms this effect.

(4.1)
$$[f_A-(f_B-\Delta f_B)]^2 = f_A^2+(f_B-\Delta f_B)^2-2f_Af_B+2f_A\Delta f_B$$

With greater f_A, a decline Δf_B of f_B will lead to greater decline $2f_A\Delta f_B$ of the cost function. In other words, minimizing the cost function would result in greater decline of f_B with greater f_A. For the second problem, we further add the following constraint:

$$\sum_{i \in A \cup B} f_i^2 = 1$$

I.e., we restrict the solution on a hypersphere with radius equal to 1.

Usually the inequality constraints are too hard to all be satisfied. We introduce slack variables to allow the inequality constraints to be violated. Denote the slack variables by $\boldsymbol{\xi} = [\xi_1, ..., \xi_m]^T$, with one inequality constraint associated with one slack variable. Now we formally state the optimization problems as follows:

$$\arg\min_{\mathbf{f},\boldsymbol{\xi}} \mathcal{L}(\mathbf{f}, \boldsymbol{\xi}) \tag{4.2}$$

where

$$\begin{aligned} \mathcal{L}(\mathbf{f}, \boldsymbol{\xi}) = & \frac{\alpha_1}{2} \sum_{i \in A} \sum_{j \in A} w_{ij} \left(\frac{f_i}{\sqrt{D_i^A}} - \frac{f_j}{\sqrt{D_j^A}} \right)^2 \\ & + \frac{\alpha_1}{2} \sum_{i \in B} \sum_{j \in B} w_{ij} \left(\frac{f_i}{\sqrt{D_i^B}} - \frac{f_j}{\sqrt{D_j^B}} \right)^2 \\ & - \alpha_2 \sum_{i \in A} \sum_{j \in B} w_{ij} \left(\frac{f_i}{\sqrt{D_i^{AB}}} - \frac{f_j}{\sqrt{D_j^{BA}}} \right)^2 \\ & + \alpha_3 \sum_{i \in A \cup B} (f_i - p_i)^2 + \beta \sum_{\substack{i \in A, j \in B \\ w_{ij} > 0}} w_{ij} \xi_{ij}^2 \end{aligned} \tag{4.3}$$

where α_1, α_2, α_3, $\beta > 0$, $\alpha_1 + \alpha_2 + \alpha_3 + \beta = 1$ and $\beta > \alpha_2$.

s.t.

(4.4)
$$\forall w_{ij} > 0 \ (i \in A, j \in B), -\frac{f_i}{\sqrt{D_i^{AB}}} + \frac{f_j}{\sqrt{D_j^{BA}}} - \xi_{ij} \leq 0$$

$$\sum_{i \in A \cup B} f_i^2 = 1 \tag{4.5}$$

Here we give some explanations to the above problem formulation. We simply set the first and the second parts of regularization terms to have the same tuning parameter, which may reduce the number of parameters. It works quite well in the experiments. We divide α_1 by 2 for more concise matrix representation of the problem, as we will see in the next subsection. β is multiplied by w_{ij} as the weight of the slack variable ξ_{ij}. By means of this parameter setting, the suppression of f_j ($j \in B$) can be one of the following two cases.

Case 1 : $\frac{f_i}{\sqrt{D_i^{AB}}} \geq \frac{f_j}{\sqrt{D_j^{BA}}}$ $(i \in A, j \in B)$. In this case, ξ_{ij} is zero. We suppress f_j by the term

$$-\alpha_2 w_{ij} \left(\frac{f_i}{\sqrt{D_i^{AB}}} - \frac{f_j}{\sqrt{D_j^{BA}}} \right)^2 \tag{4.6}$$

f_i and f_j are pulled apart.

Case 2 : $\frac{f_i}{\sqrt{D_i^{AB}}} < \frac{f_j}{\sqrt{D_j^{BA}}}$ $(i \in A, j \in B)$. In this case, we have $\xi_{ij} = \frac{f_j}{\sqrt{D_j^{BA}}} - \frac{f_i}{\sqrt{D_i^{AB}}}$ $(i \in A, j \in B)$. By substituting ξ_{ij} into $\mathcal{L}(\mathbf{f}, \boldsymbol{\xi})$, we can see that the actual suppression term is

$$(\beta - \alpha_2) w_{ij} \left(\frac{f_i}{\sqrt{D_i^{AB}}} - \frac{f_j}{\sqrt{D_j^{BA}}} \right)^2 \tag{4.7}$$

f_i and f_j are pulled together. Note that this is also a kind of suppression because f_j is already greater than f_i. This also explains why we add the requirement $\beta > \alpha_2$.

The cost function is a quadratically constrained quadratic programming problem. The optimization techniques used in Manifold Ranking algorithm can not be applied here. We discuss the algorithm for solving the problem in the next subsection.

4.2 Algorithm Derivation We present the algorithm for the optimization problem (4.2) in this subsection. We first rewrite the problem (Equation (4.2)) in matrix form.

Let $diag[x_i]_{i\in X}$ ($X = A$ or B) denote a diagonal matrix with the diagonal elements $x_1, ..., x_n$ ($1, ..., n \in X$). We define diagonal matrix

$$\mathbf{D}_{XY} = diag\left[\sum_{j\in Y} w_{ij}\right]_{i\in X} \tag{4.8}$$

where X, Y are either set A or B. Let $\mathbf{W} = [w_{ij}]_{i,j\in A\cup B}$ be the sentence similarity matrix of $A\cup B$, and $\mathbf{W}_{XY} = [w_{ij}]_{i\in X, j\in Y}$ the similarity matrix between X and Y, where $X, Y = A$ or B. Denote by $\mathbf{f}_A$ and $\mathbf{f}_B$ the ranking scores of the sentences in A and B respectively. With simple algebraic transformations, the first part on the right side of Equation (4.3) can be rewritten as follows:

$$\begin{aligned}
&\frac{\alpha_1}{2}\sum_{i\in A}\sum_{j\in A} w_{ij}\left(\frac{f_i}{\sqrt{D_i^A}} - \frac{f_j}{\sqrt{D_j^A}}\right)^2 \\
=&\alpha_1\left(\sum_{i\in A}(\sum_{j\in A} w_{ij})\frac{f_i^2}{D_i^A} - \sum_{i\in A}\sum_{j\in A} w_{ij}\frac{f_i}{\sqrt{D_i^A}}\frac{f_j}{\sqrt{D_j^A}}\right) \\
=&\alpha_1\mathbf{f}_A^T\mathbf{D}_{AA}^{-\frac{1}{2}}(\mathbf{I} - \mathbf{W}_{AA})\mathbf{D}_{AA}^{-\frac{1}{2}}\mathbf{f}_A
\end{aligned}$$

where $\mathbf{I}$ is the identity matrix. Similarly, the other terms can be rewritten as follows.

$$\begin{aligned}
&\frac{\alpha_1}{2}\sum_{i\in B}\sum_{j\in B} w_{ij}\left(\frac{f_i}{\sqrt{D_i^B}} - \frac{f_j}{\sqrt{D_j^B}}\right)^2 \\
=&\alpha_2\mathbf{f}_B^T\mathbf{D}_{BB}^{-\frac{1}{2}}(\mathbf{I} - \mathbf{W}_{BB})\mathbf{D}_{BB}^{-\frac{1}{2}}\mathbf{f}_B
\end{aligned}$$

$$\begin{aligned}
&\alpha_2\sum_{i\in A}\sum_{j\in B} w_{ij}\left(\frac{f_i}{\sqrt{D_i^{AB}}} - \frac{f_j}{\sqrt{D_j^{BA}}}\right)^2 \\
=&\alpha_2(\mathbf{f}_A^T\mathbf{f}_A + \mathbf{f}_B^T\mathbf{f}_B - 2\mathbf{f}_A^T\mathbf{D}_{AB}^{-\frac{1}{2}}\mathbf{W}_{AB}\mathbf{D}_{BA}^{-\frac{1}{2}}\mathbf{f}_B)
\end{aligned}$$

$$\alpha_3\sum_{i\in A\cup B}(f_i - p_i)^2 = \alpha_3(\mathbf{f}^T\mathbf{f} - 2\mathbf{p}^T\mathbf{f} + \mathbf{p}^T\mathbf{p})$$

Since $\mathbf{p}^T\mathbf{p}$ is a constant, it can be safely discarded.

$$\beta\sum_{\substack{i\in A, j\in B\\ w_{ij}>0}} w_{ij}\xi_{ij}^2 = \beta\boldsymbol{\xi}^T\mathbf{W}_\xi\boldsymbol{\xi}$$

where $\mathbf{W}_\xi$ is a diagonal matrix with diagonal elements w_{ij}.

If we define $\mathbf{S}_{XY} = \mathbf{D}_{XY}^{-\frac{1}{2}}\mathbf{W}_{XY}\mathbf{D}_{YX}^{-\frac{1}{2}}$, $\mathbf{g} = [\mathbf{f}^T, \boldsymbol{\xi}^T]^T$, $\mathbf{h} = [-2\alpha_3\mathbf{p}^T, \mathbf{0}^T]^T$ and

$$\mathbf{Q} = \begin{bmatrix} \gamma\mathbf{I} - \alpha_1\mathbf{S}_{AA} & \alpha_2\mathbf{S}_{AB} & \mathbf{0} \\ \alpha_2\mathbf{S}_{BA} & \gamma\mathbf{I} - \alpha_1\mathbf{S}_{BB} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \beta\mathbf{W}_\xi \end{bmatrix}$$

where $\mathbf{0}$ denotes a zero matrix or column vector with appropriate dimensions and $\gamma = \alpha_1 - \alpha_2 + \alpha_3$, $\mathcal{L}(\mathbf{f}, \boldsymbol{\xi})$ can be rewritten in the standard form

$$\mathcal{L}(\mathbf{g}) = \mathbf{g}^T\mathbf{Q}\mathbf{g} + \mathbf{h}^T\mathbf{g} \tag{4.9}$$

The inequality constraints (Equation (4.4)) can be reformulated in matrix form $\mathbf{A}_i^T\mathbf{g} \le 0$ ($1 \le i \le m$) by defining vectors $\mathbf{A}_i$, whose elements are all zeros except three elements $\frac{1}{\sqrt{D_i^B}}$, $-\frac{1}{\sqrt{D_j^A}}$ and -1 corresponding to the appropriate positions of $\mathbf{g}$. Similarly, the equality constraint (Equation (4.5)) can be reformulated as $\mathbf{g}^T\mathbf{B}\mathbf{g} - 1 = 0$ by defining

$$\mathbf{B} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

The augmented Lagrangian method[3] can solve the proposed optimization problem. It defines the following augmented Lagrangian function:

$$\begin{aligned}
\mathcal{L}_A^*(\mathbf{g}, \boldsymbol{\lambda}, \mu) =&\mathbf{g}^T\mathbf{Q}\mathbf{g} + \mathbf{h}^T\mathbf{g} \\
&+ \lambda_0(\mathbf{g}^T\mathbf{B}\mathbf{g} - 1) + \sum_{i=1}^m \lambda_i(\mathbf{A}_i^T\mathbf{g} + s_i^2) \\
&+ \mu\left[(\mathbf{g}^T\mathbf{B}\mathbf{g} - 1)^2 + \sum_{i=1}^m (\mathbf{A}_i^T\mathbf{g} + s_i^2)^2\right]
\end{aligned} \tag{4.10}$$

where $\boldsymbol{\lambda} = [\lambda_0, \lambda_1, ..., \lambda_m]^T$ and s_i ($1 \le i \le m$) are slack variables. The above Lagrangian function can be simplified by eliminating s_i, which leads to the following equation:

$$\begin{aligned}
\mathcal{L}_A(\mathbf{g}, \boldsymbol{\lambda}, \mu) =&\mathbf{g}^T\mathbf{Q}\mathbf{g} + \mathbf{h}^T\mathbf{g} + \lambda_0(\mathbf{g}^T\mathbf{B}\mathbf{g} - 1) \\
&+ \mu(\mathbf{g}^T\mathbf{B}\mathbf{g} - 1)^2 - \frac{1}{4\mu}\sum_{i=1}^m \lambda_i^2 \\
&+ \mu\sum_{i=1}^m\left(\max\left\{\mathbf{A}_i^T\mathbf{g} + \frac{\lambda_i}{2\mu}, 0\right\}\right)^2
\end{aligned} \tag{4.11}$$

It can be shown that minimizing the Lagrangian function (4.10) is equal to minimizing (4.11). Minimizing (4.11) is a unconstrained problem. We use gradient descent method to find a local minimum.

The algorithm AugmentedLagrangian shows how to solve the optimization problem (4.2) using augmented Lagrangian algorithm. The convergence of this method

[3] the interested readers may refer to [17] for more information about augmented Lagrangian method

AugmentedLagrangian Augmented Lagrangian method for solving the prosed regularization framework

Input:
$\mathbf{Q}, \mathbf{h}$: the quadratic cost function $\mathcal{L}(\mathbf{g}) = \mathbf{g}^T\mathbf{Q}\mathbf{g} + \mathbf{h}^T\mathbf{g}$.
$\mathbf{A}_i$: the inequality constraint $\mathbf{A}_i^T\mathbf{g} \leq 0$, $1 \leq i \leq m$.
$\mathbf{B}$: the quadratic equality constraint $\mathbf{g}^T\mathbf{B}\mathbf{g} = 1$.
t: parameter update step length.
Output:
$\mathbf{g}$: scores of the variables.

1. Initialize $\mathbf{g}^0$ by random assignment. Initialize $\boldsymbol{\lambda}$ to ones and μ to one.
2. **for** $k = 1, 2, ...$
3. Find a $\mathbf{g}^k$ that minimizes $\mathcal{L}_A(\mathbf{g}, \boldsymbol{\lambda}, \mu)$ starting from point $\mathbf{g}^{k-1}$.
4. **if** $\mathbf{g}^k$ converges, **return** $\mathbf{g}^k$.
5. Update λ, μ by
   $\lambda_0^k = \lambda_0^{k-1} + 2\mu^{k-1}(\mathbf{g}^T\mathbf{B}\mathbf{g} - 1)$,
   $\lambda_i^k = \max\{\lambda_i^{k-1} + 2\mu^{k-1}\mathbf{A}_i^T\mathbf{g}, 0\}, i = 1, ..., m$,
   $\mu^k = t\mu^{k-1}$.
6. **end for**

can be assured without increasing μ to a very large value. Note that the cost function may not be convex. A local minimum can be found.

The update summarization method for document collection B given collection A is to figure out $\mathbf{Q}, \mathbf{h}, \mathbf{A}_i$ $(1 \leq i \leq m)$ and $\mathbf{B}$, get the ranking scores of the sentences in B through algorithm AugmentedLagrangian, and take the first k sentences to form the summary.

4.3 Discussion If we set α_2 and β to zero in Equation (4.3), the cost function appears very similar to that of the Manifold Ranking algorithm [30] with the exception that sum of squares of the ranking scores is constrained to 1 (Equation (4.5)). In Manifold Ranking algorithm, such constraint is unnecessary for the convergence of the algorithm can be assured. We will show in section 6 that, when setting α_2 and β to zero, MarginRank gives almost the same ranking results as Manifold Ranking algorithm for A and B respectively.

If we view scoring the sentences as putting them in a one-dimensional space, MarginRank can be interpreted as the interaction between the points (sentences) labeled (or rather in) A and B. At first, all points are seated in their original positions (i.e., set to their prior relevance scores). Then, the interaction between the points drags them to new positions. On one hand, the points are always trying to pull neighboring points having the same label to their positions. On the other hand, the points labeled A also try to pull neighboring points labeled B in higher positions to their position and push neighboring points in B in lower positions away. The final positions (scores) are the results of such interaction.

5 Update Summarization through MarginRank

5.1 Novelty as Summary Update Novelty is an important requirement in the document summarization, document retrieval and question answering tasks. In document summarization, novelty requires that sentences in the summary should not repeat the same information. It can be easily incorporated into the MarginRank regularization framework.

Let B' denotes the set of sentences that are already selected to be included in the summary and $B^* = B - B'$ be the set of the remaining sentences in B. We view the novelty requirement as summary update, i.e, summarizing B^* given the users have already read B'. There are two strategies to impose novelty in the framework. The first one is to add B' into A. However, since A may have much more sentences than B' and sentences in B' may count for little in A, B' may be submerged by A. We recommend treating B' independently. We will show the experimental results of both strategies in section 6.

If we treat B' as independent of A, the regularization framework can be extended as:

(5.12)

$$\begin{aligned}\mathcal{L}_N(\mathbf{f}, \boldsymbol{\xi}) = \frac{\alpha_1}{2} \sum_{X \in \{A, B', B^*\}} \sum_{i \in X} \sum_{j \in X} w_{ij} \left(\frac{f_i}{\sqrt{D_i^X}} - \frac{f_j}{\sqrt{D_j^X}} \right)^2 \\ - \alpha_2 \sum_{X \in \{A, B'\}} \sum_{i \in X} \sum_{j \in B^*} w_{ij} \left(\frac{f_i}{\sqrt{D_i^{XB^*}}} - \frac{f_j}{\sqrt{D_j^{B^*X}}} \right)^2 \\ + \alpha_3 \sum_{i \in A \cup B' \cup B^*} (f_i - p_i)^2 + \beta \sum_{\substack{i \in A \cup B', j \in B^* \\ w_{ij} > 0}} w_{ij} \xi_{ij}^2\end{aligned}$$

Constraints are added analogously to Equations (4.4) and (4.5). By rewriting $\mathcal{L}_N(\mathbf{f}, \boldsymbol{\xi})$ in the matrix form $\mathcal{L}_N(\mathbf{h})$ as in the above section, we can apply the AugmentedLagrangian algorithm to the optimization problem.

5.2 Update Summarization We construct the adjacent matrix of A and B by the cosine similarities of the sentences. We set a threshold δ for all w_{ij} ($i \in A$ (B'), $j \in B^*$ or $i \in B^*$, $j \in A$ (B')). If $w_{ij} < \delta$, it is set to 0. It can avoid unnecessary suppresion of A (B') on B^*. It also reduces the number of inequality constraints, greatly speeding up the algorithm.

The update summarization algorithm considering novelty is given by the algorithm UpdateSum.

UpdateSum Update summarization algorithm using MarginRank

Input:
A, B: two evolutionary document collections. The time stamp of A is prior to that of B
T: the topic that expresses the information need of the users
$\alpha_1, \alpha_2, \alpha_3, \beta$: parameters in Equation (5.12)
t: parameter update step length in AugmentedLagrangian algorithm
Output:
S: the summary.

1. Set $B^* := B$, $B' := \phi$, $S := \phi$
2. Compute sentence similarity matrix $\mathbf{W}$, topic relevance $\mathbf{p}$ etc. using A, B, T.
3. Figure out the cost function $\mathcal{L}_N(\mathbf{h})$ (Equation (5.12)), constraints $\mathbf{A}_i$ $(0 \leq i \leq m)$ and $\mathbf{B}$.
4. **for** $k = 1, 2, ...$
5. Find a $\mathbf{h}$ that minimizes $\mathcal{L}_N(\mathbf{h})$ by calling AugmentedLagrangian
6. Add the sentence s_x that has the maximal score h_x in B^* to S. Move s_x from B^* to B'
7. **if** length of S reach the summary length limit, **return** S
8. Update $\mathcal{L}_N(\mathbf{h})$, $\mathbf{A}_i$ $(0 \leq i \leq m)$ and $\mathbf{B}$
9. **end for**

---

6 Experiments

6.1 Data Set, Evaluation Metrics and Preprocessing We use the popular update summarization benchmark data sets TAC 2008 and 2009[4] for our experiments. The data sets consist of a number of topics, with each topic associated with two document collections A and B. The automatic summarizer is expected to extract from each document collection B an update summary that does not exceed 100 words. Each document collection is associated with four summaries written by different experts as reference (ground-truth) summaries. Description of the data sets is given in Table 1.

We use two automatic evaluation measures: ROUGE and Basic Elements[5] for evaluation, which are officially adopted by TAC for evaluation of automatic generated summaries.

	TAC 2008	TAC 2009
Number of Topics	48	44
Number of documents per topic	A:10 B:10	A:10 B:10
Average time span between A and B	109 days	82 days
Summary length	100 words	100 words

Table 1: Description of the data sets

ROUGE measures the performance of a summarizer by counting the overlapping units between the automatic generated summaries and the reference summaries. Given a set of reference summaries $RefSum$ and an automatic generated summary $ExtractSum$, the n-gram recall and precision measure of ROUGE are computed as follows:

$$\text{ROUGE-N-R} = \frac{\sum_{S\in\{RefSum\}}\sum_{gram_n\in S} Count_{match}(gram_n)}{\sum_{S\in\{RefSum\}}\sum_{gram_n\in S} Count(gram_n)}$$

$$\text{ROUGE-N-P} = \frac{\sum_{S\in\{RefSum\}}\sum_{gram_n\in S} Count_{match}(gram_n)}{\sum_{S\in\{ExtractSum\}}\sum_{gram_n\in S} Count(gram_n)}$$

where $Count_{match}(gram_n)$ is the maximum number of n-grams $gram_n$ co-occurring in the reference summaries and the automatic generated summary. F-measure is calculated as usual:

$$\text{ROUGE-N-F} = \frac{2 \cdot \text{ROUGE-N-R} \cdot \text{ROUGE-N-P}}{\text{ROUGE-N-R} + \text{ROUGE-N-P}}$$

Parameter setting for ROUGE is the same as the official parameter setting[6] of TAC.

Basic Elements breaks down each reference sentence into minimal semantic units, called Basic Elements (BE). Each BE consists of a list of one to three words and their associated parts-of-speech or NER type. For example, all nouns, verbs, and adjectives found in the summary, subject+verb, verb+object, two head words connected via a preposition etc. To match nonidentical units that carry the same meaning, BE evaluation method applies rules to transform each unit into a number of different variants. The evaluation procedure is as follows. The method automatically creates BEs for human summaries and machine-generated summaries,

[4] http://www.nist.gov/tac/data/index.html

[5] ROUGE 1.5.5 package and BEwT-E package. Both can be downloaded at http://www.isi.edu/publications/licensed-sw/BE/index.html

[6] -n 4 -w 1.2 -m -2 4 -u -c 95 -r 1000 -f A -p 0.5 -t 0 -a -d

applies transformation rules to expand BE units into numerous variants, and performs matching of these units between the human summaries and machine-generated summaries.

For each document, we use the OpenNLP[7] tool to detect and tokenize sentences. A list of 707 words is used to filter stop words. The remaining words are stemmed by Snowball[8].

6.2 Experimental Results

6.2.1 Overall Performance As we have discussed in subsection 4.3, by setting α_2 and β in Equation (4.3) to zero, MarginRank differs from Manifold Ranking algorithm in the equation constraint (Equation (4.5)). Before we do experiments for the update summarization task, we first show that by setting α_2 and β to zero, MarginRank gives almost the same ranking results as the Manifold Ranking algorithm. The constraint parameter α_1 is set to 0.95 and the fitting parameter α_3 is set to 0.05 in both algorithms. We rank sentences using MarginRank and Manifold Ranking respectively and take the top ranked sentences as the summaries. We do experiments on the TAC 2009 A and B collections. Tables 2 and 3 shows the experimental results.

	ROUGE-1	ROUGE-2	ROUGE-SU
MarginRank $\alpha_2, \beta = 0$	0.36172	0.09120	0.12731
ManifoldRank	0.36530	0.09453	0.13156

Table 2: ROUGE scores on TAC 2009 collection A

	ROUGE-1	ROUGE-2	ROUGE-SU
MarginRank $\alpha_2, \beta = 0$	0.34448	0.07733	0.11989
ManifoldRank	0.3426	0.07804	0.11971

Table 3: ROUGE scores on TAC 2009 collection B

For the update summarization task, we compare MarginRank with the following algorithms.

- Human Avg: average performance of the human summarizers. The quality of summaries generated by one expert is measured by the ground-truth summaries generated by three other experts.
- Baseline 1: a summary comprised of all the leading sentences (up to 100 words) in the most recent document. Baseline 1 provides a lower bound on what can be achieved with a simple fully automatic extractive summarizer.
- Baseline 3: a summary consisting of sentences that have been manually selected from the document collections. Baseline 3 provides an approximate upper bound on what can be achieved with a purely extractive summarizer. It is provided by TAC and available for TAC 2009 datasets only.
- PNR[2]: state-of-the-art update summarization method proposed in [10].
- SMMR: extension of the MMR algorithm for update summarization proposed in [3].
- Relevance (N): an update summarization method that rank sentences according to their relevance (cosine similarity) to the topic in B. It first removes the set of sentences $S = \{s_i \in B | \max_{s_j \in A} sim(s_i, s_j) > N\}$ from B as a preprocessing step.
- ManifoldRank: Manifold Ranking algorithm [30] applied to document collection B without considering A. It serves as a baseline algorithm.
- Auto Avg: average performance of the participating systems in TAC 2008 and 2009.

We assign to each word equal weight – 1 in MarginRank and all the comparative algorithms for fair comparison. We set the threshold of the edge weight between two nodes that belong to A (or B') and B^* to be 0.3. We tune the parameters empirically on DUC 2007 update summarization datasets and apply the tuned parameters to TAC 2008 and 2009. In our experiments, we set α_2 to 0.015, β to 0.03 and α_3 to 0.05. As we have discussed in section 5, there are two cases when the suppression terms are in action. We simply set β to $2\alpha_2$ so that the weights of the suppression terms are the same in the two cases.

Tables 4 and 5 show the experimental results. For convenience, we refer to our summarization method UpdateSum as MarginRank. MarginRank-S considers B' separately while MarginRank-I incorporates B' into A. The numbers shown in brackets [], below the scores of MarginRank, give the corresponding confidence interval with confidence level 95%. We can see that the improvement of MarginRank over the comparative algorithms is statistical significant. Also, through Relevance, we can see that simple sentence pruning method dose not lead to much performance enhancement. It is hard to tune the parameter as well. From 0.5 to 0.3, the performance of Relevance degenerates drastically. Another observation is that ManifoldRank gives reasonable performance

[7] `http://opennlp.sourceforge.net/`
[8] `http://snowball.tartarus.org/index.php`

	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-SU	BEwT
Human Avg	0.39974	0.11690	0.36167	0.15034	0.24880
MarginRank-S	**0.36096*** [0.35250 - 0.36958]	**0.08569** [0.07879 - 0.09262]	**0.31633*** [0.30845 - 0.32430]	**0.12850*** [0.12276 - 0.13419]	**0.20143**
MarginRank-I	0.35525* [0.34731 - 0.36345]	0.08359 [0.07649 - 0.09012]	0.31145 [0.30330 - 0.31936]	0.12565 [0.11995 - 0.13139]	0.19930
PNR^2	0.31403	0.06802	0.28009	0.10578	0.15739
Smmr	0.31332	0.06797	0.28062	0.10526	0.15412
Relevance (>1)	0.31428	0.06677	0.27914	0.10499	0.15425
Relevance (0.5)	0.31409	0.06831	0.27900	0.10605	0.15964
Relevance (0.3)	0.28687	0.04590	0.25150	0.08729	0.11886
ManifoldRank	0.34618	0.07885	0.30394	0.12042	0.18887
Auto Avg	0.32117	0.06764	0.28309	0.10687	0.16542
Baseline 1	0.29924	0.06461	0.26493	0.10043	0.14384

Table 4: Performance comparison on the TAC 2008 data set

	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-SU	BEwT
Human Avg	0.39336	0.10647	0.21638	0.147780	0.23681
Baseline 3	0.37393	0.09889	0.20288	0.13735	0.21521
MarginRank-S	**0.35415*** [0.34696 - 0.36130]	**0.08357*** [0.07847 - 0.08860]	**0.22239** [0.21702 - 0.22793]	**0.12615*** [0.12166 - 0.13062]	**0.19255**
MarginRank-I	0.35294* [0.34518 - 0.36037]	0.08245 [0.07702 - 0.08764]	0.21889 [0.21303 - 0.22486]	0.12467* [0.11983 - 0.12935]	0.18872
PNR^2	0.33105	0.07045	0.21465	0.11303	0.15937
Smmr	0.33299	0.07146	0.21479	0.11366	0.16267
Relevance (>1)	0.33003	0.07106	0.21505	0.11251	0.16612
Relevance (0.5)	0.33306	0.07335	0.21768	0.11521	0.16756
Relevance (0.3)	0.30145	0.05348	0.19691	0.09473	0.13693
ManifoldRank	0.3426	0.07804	0.21353	0.11971	0.17539
Auto Avg	0.32958	0.07146	0.20076	0.11159	0.16642
Baseline 1	0.30142	0.05425	0.18251	0.09597	0.131124

Table 5: Performance comparison on the TAC 2009 data set

* indicates that the improvement of MarginRank over all the other comparative algorithms is statistical significant at a confidence level of 95%.

by considering B alone. However, if we are summarizing daily news reports instead of such manually created datasets, considering the current documents only is obviously inadequate.

To see more intuitively why MarginRank can outperform the comparative algorithms, we show the summaries of topic D0938 generated by the algorithms in table 6. Smmr and PNR^2 generate similar summaries, achieving similar performance. They tend to emphasize too much the difference between B and A instead of summarizing B, which may lead to selecting some unimportant sentences. On the other hand, the baseline algorithm ManifoldRank summarizes B without considering A. MarginRank algorithm selects more admissible sentences.

6.2.2 Parameter Tuning We vary the parameters to see how they affect the performance of the algorithm.

First, we do experiments to see the effects of the suppression terms using MarginRank-S. For this purpose, we fix α_1 to 0.95 and α_3 to 0.05 (both add up to 1). Recall that the suppression occurs in two cases (see to section 4.1). We clinch β to $2\alpha_2$ so that the suppression has the same weight in both cases. Then we variate the value of α_2 from 0 to 0.04. Figure 1 shows the experimental results.

topic	Discuss the planning, preparations, and funding issues for the memorial to the 9/11 World Trade Center disaster.
Summary of collection A by expert	Four US ex-presidents, a Hollywood legend, a retired fireman and a star TV journalist were among the 31 appointed in December 2004 to lead the project for a memorial to the victims of the September 11, 2001, attack on New York's World Trade Center. They were to be directors of the foundation that would construct, own, operate and maintain the memorial. Architect Michael Arad's design was chosen from 5,201 entries in an international competition. It featured two reflecting pools surrounded by gardens. Part of the $500 million in financing will come from New York's $816 million in federal grants.
Baseline 3, manually selected sentences	New York Governor George Pataki joined victims' family members, survivors, and rescue workers Thursday in laying out a timetable for two temporary Sept. 11 memorials. He will also use hundreds of millions of public dollars to jumpstart the fund-raising effort for the Sept. 11 memorial at the World Trade Center site, they said, and will push ahead the timetable for producing a redesign of the Freedom Tower office building. Nonetheless, some of the Sept. 11 families are unhappy with how the memorial is shaping up.
ManifoldRank	Beyond that, he said, the importance of the Trade Center memorial has been undiminished. Despite that withdrawal, critics of the cultural plan – including two members of the World Trade Center Memorial Foundation, which is charged with soliciting contributions – said the freedom center had to be removed before they would support the use of public money for the overall project. He will also use hundreds of millions of public dollars to jumpstart the fund-raising effort for the Sept. 11 memorial at the World Trade Center site, they said, and will push ahead the timetable for producing a redesign of
SMMR	In the ever fiercer fight over a year-old plan to build a home for the Drawing Center and the International Freedom Center alongside the World Trade Center memorial, some relatives of 9/11 victims called on Monday for a fund-raising boycott. He expects others will go there, too, instead of to the World Trade Center. We urge you to not donate to the World Trade Center memorial until the IFC and the Drawing Center are eliminated from the memorial plans, said An Open Letter to the American People. Beyond that, he said, the importance of the Trade Center memorial has been
PNR^2	He expects others will go there, too, instead of to the World Trade Center. In the ever fiercer fight over a year-old plan to build a home for the Drawing Center and the International Freedom Center alongside the World Trade Center memorial, some relatives of 9/11 victims called on Monday for a fund-raising boycott. We urge you to not donate to the World Trade Center memorial until the IFC and the Drawing Center are eliminated from the memorial plans, said An Open Letter to the American People. Beyond that, he said, the importance of the Trade Center memorial has been
MarginRank	Beyond that, he said, the importance of the Trade Center memorial has been undiminished. He will also use hundreds of millions of public dollars to jumpstart the fund-raising effort for the Sept. 11 memorial at the World Trade Center site, they said, and will push ahead the timetable for producing a redesign of the Freedom Tower office building. In the ever fiercer fight over a year-old plan to build a home for the Drawing Center and the International Freedom Center alongside the World Trade Center memorial, some relatives of 9/11 victims called on Monday for a fund-raising boycott. We urge

Table 6: Sample summaries generated by different algorithms

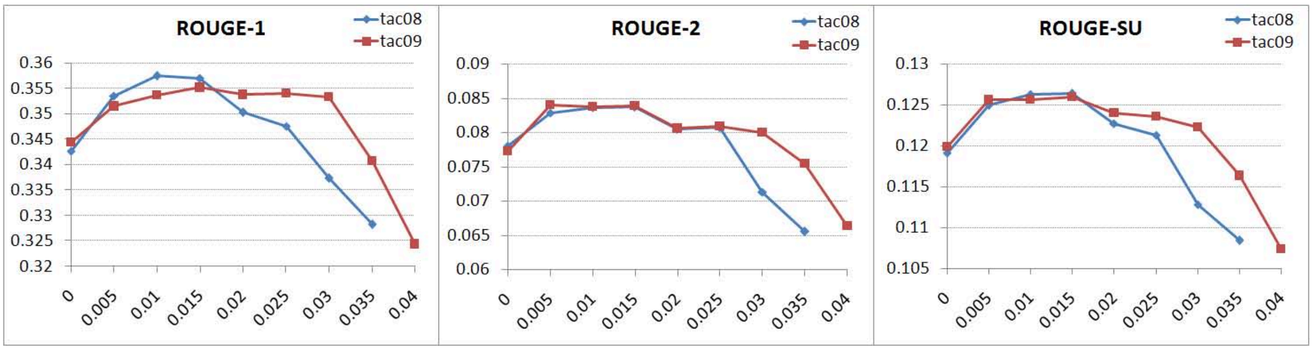

Figure 1: ROUGE scores vs. α_2 ($\beta = 2\alpha_2$). The X-axis is the value of α_2. The Y-axis is the ROUGE score.

We can see that generally MarginRank-S first improves and then degenerates. The improvement seems to be more robust to the parameters on TAC2009 than on TAC2008. It can be explained as below. Recall that the average time span between A and B of TAC2009 (Table 1) is much shorter than that of TAC2008. We could expect the content overlapping between A and B is more severe in TAC2009 than in TAC2008. Actually, we could see that from the performance of the baseline algorithm Relevance in Table 4 and 5. By pruning the sentences that are highly similar to those of A, Relevance (0.5) achieves approximate 1% performance enhancement on TAC2009, while the performance barely increases on TAC2008.

Second, we want to examine the effects of suppression in both cases separately. Again, we fix α_1 to 0.95 and α_3 to 0.05. Then by clinching β to α_2 and varying the value of α_2, we will see the effects of the suppression in the first case. Similarly, by setting α_2 to 0 and varying the value of β, we will see the effects of the suppression in the second case. Figures 2 and 3 show the ROUGE scores. We observe that the effects of the second kind of suppression are not so radical as the first one. As β goes to as high as 1, MarginRank-S still performs better than it performs when β is 0.

Figure 2: ROUGE-1 vs. α_2 ($\beta = \alpha_2$). Only the suppression in the first case is on action

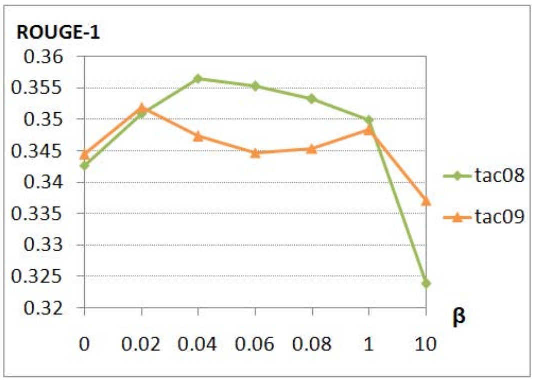

Figure 3: ROUGE-1 vs. β ($\alpha_2 = 0$). Only the suppression in the second case is on action

Third, the effects of α_3 are also examined. We set α_1, α_2 and β to 0.955, 0.015 and 0.03 respectively. α_3 are tuned from 0 to 0.45. Figure 4 shows the experimental results. When α_3 is set to a very large value, the performance should approximate that of Relevance (>1) in Tables 4 and 5.

6.2.3 Running Time The optimization problem of MarginRank is a quadratically constrained quadratic programming problem. Solving such problems are time consuming. Further, running the algorithm multiple times also increases the running time. We give the running time comparison between MarginRank and the related algorithms in Figure 5. The time is measured by second per document collection. The experiments are done on a desktop computer with an Intel Core2 Quad 2.66GHz CPU and 3.37GB memories. The operating system is Windows XP sp3. The programming language is Matlab.

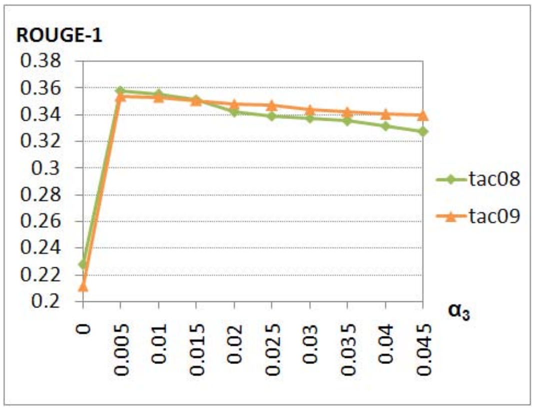

Figure 4: ROUGE-1 vs. α_3.

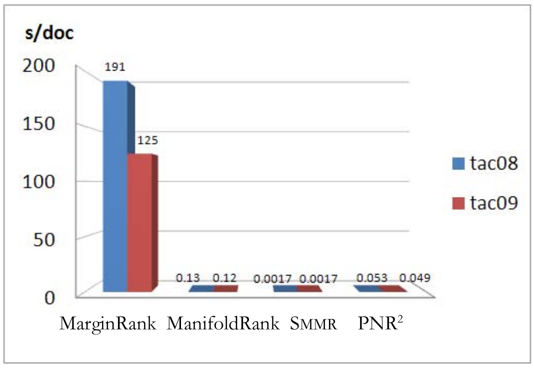

Figure 5: Running time of the algorithms

7 Conclusion

In this paper, we propose MarginRank, a graph-based regularization framework for update summarization. Through minimizing the cost function, MarginRank scores sentences in an evolutionary document collection in a way that the top ranked sentences summarize the document collection and at the same time cover different contents from previous document collections. The cost function of MarginRank is a quadratically constrained quadratic programming problem. It may not be convex. We employ the augmented Lagrangian algorithm for minimizing the cost function. The algorithm will find a local minimum. Since quadratically constrained quadratic programming problems are NP-hard, we will seek to reformulate the problem that can be solved more efficiently in the future. We also shown that novelty of the summary can be easily ensured using MarginRank. We do experiments on the TAC 2008 and 2009 data sets. Experimental results show the effectiveness of MarginRank.

8 Acknowledgments

This work is supported in part by NSFC grants 60970045 and 60833001, and by the National High-tech R&D Program (863 Program).

References

[1] Achananuparp, P., Yang, C. C., and Chen, X. Using negative voting to diversify answers in non-factoid question answering. In *Proc. of CIKM '09*, pages 1681–1684, 2009.

[2] M.-R. Amini and P. Gallinari. The use of unlabeled data to improve supervised learning for text summarization. In *Proc. of SIGIR '02*, pages 105–112, 2002.

[3] Florian Boudin, Marc El-Bèze, Juan-Manuel Torres-Moreno A Scalable MMR Approach to Sentence Scoring for Multi-Document Update Summarization In *Proc. of COLING '08*, pages 23–26, 2008.

[4] J. Carbonell and J. Goldstein. The use of mmr, diversity-based reranking for reordering documents and producing summaries. In *Proc. of SIGIR '98*, pages 335–336, 1998.

[5] G. Erkan and D. R. Radev. Lexpagerank: Prestige in multi-document text summarization. In *Proc. of EMNLP '04*, 2004.

[6] J. Kupiec, J. Pedersen, and F. Chen. A trainable document summarizer. In *Proc. of SIGIR '95*, pages 68–73, 1995.

[7] L. Li, K. Zhou, G.-R. Xue, H. Zha, and Y. Yu. Enhancing diversity, coverage and balance for summarization through structure learning. In *Proc. of WWW '09*, pages 71–80, 2009.

[8] C.-Y. Lin. Rouge: A package for automatic evaluation of summaries. In *Proc. of ACL workshop on Text Summarization Branches Out*, page 10, 2004.

[9] Li, S.-J., Wang, W., Zhang, Y.-W. TAC 2009 Update Summarization of ICL. In *Proceedings of TAC '09*, 2009.

[10] Wenjie Li, Furu Wei, Qin Lu, and Yanxiang He. PNR^2: ranking sentences with positive and negative reinforcement for query-oriented update summarization. In *Proc. of COLING '08*, pages 489–496, 2008

[11] Xuan Li, Yi-Dong Shen, Liang Du, Chen-Yan Xiong. Exploiting novelty, coverage and balance for topic-focused multi-document summarization. In *Proc. of CIKM '10*, pages 1765–1768, 2010

[12] H. P. Luhn. Automatic creation of literature abstracts. *IBM Journal Res. Develop.*, pages 159–165, 1958.

[13] D. Metzler and T. Kanungo. Machine learned sentence selection strategies for query-biased summarization. In *Proc. of SIGIR Learning to Rank Workshop*, 2008.

[14] R. Mihalcea and P. Tarau. TextRank: Bringing order into texts. In *Proc. of EMNLP '04*, 2004.

[15] R. Mihalcea. Language independent extractive summarization. In *Proc. of ACL '05*, pages 49–52, 2005.

[16] V. Nastase. Topic-driven multi-document summarization with encyclopedic knowledge and spreading activation. In *Proc. of EMNLP '08*, pages 763–772, 2008.

[17] Jorge Nocedal and Stephen J. Wright. Numerical Optimization. Springer-Verlag, New York, 1999.

[18] T. Nomoto and Y. Matsumoto. A new approach to unsupervised text summarization. In *Proc. of SIGIR '01*, pages 26–34, 2001.

[19] D. Shen, J.-T. Sun, H. Li, Q. Yang, and Z. Chen. Document summarization using conditional random fields. In *Proc. of IJCAI '07*, pages 2862–2867, 2007.

[20] Stephen Tratz and Eduard Hovy. Summarization Evaluation Using Transformed Basic Elements. In *Proc. of Text Analytics Conference '08*, 2008.

[21] X. Wan, J. Yang, and J. Xiao. Manifold-ranking based topic-focused multi-document summarization. In *Proc. of IJCAI '07*, pages 2903–2908, 2007.

[22] Xiaojun Wan. TimedTextRank: adding the temporal dimension to multi-document summarization. In *Proc. of SIGIR '07*, pages 867–868, 2007.

[23] C. Wang, F. Jing, L. Zhang, and H.-J. Zhang. Learning query-biased web page summarization. In *Proc. of CIKM '07*, pages 555–562, 2007.

[24] Wang, D., Zheng, L., Li, T., and Deng, Y. Evolutionary document summarization for disaster management. In *Proc. of SIGIR '09*, pages 680–681, 2009.

[25] Wang, D., Zhu, S., Li, T., and Gong, Y. 2009. Comparative document summarization via discriminative sentence selection. In *Proc. of CIKM '09*, pages 1963–1966, 2009.

[26] D. Wang, T. Li, S. Zhu, and C. Ding. Multi-document summarization via sentence-level semantic analysis and symmetric matrix factorization. In *Proc. of SIGIR '08*, pages 307–314, 2008.

[27] K.-F. Wong, M. Wu, and W. Li. Extractive summarization using supervised and semi-supervised learning. In *Proc. of COLING '08*, pages 985–992, 2008.

[28] H. Zha. Generic summarization and keyphrase extraction using mutual reinforcement principle and sentence clustering. In *Proc. of SIGIR '02*, pages 113–120, 2002.

[29] Zhang, J., Du, P., Xu, H.-B. and Cheng, X.-Q. ICT-Grasper at TAC2009: Temporal Preferred Update Summarization. In *Proceedings of TAC '09*, 2009.

[30] D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Schölkopf. Ranking on data manifolds. In *Proc. of NIPS '03*, pages 169–176, MIT Press, 2003.

[31] X. Zhu, A. Goldberg, J. V. Gael and D. Andrzejewski Improving Diversity in Ranking using Absorbing Random Walks. In *Proc. of HLT-NAACL '07*, pages 97–104, 2007.