Towards a Unified Approach to Simultaneous Single-Document and Multi-Document Summarizations

Xiaojun Wan

Institute of Compute Science and Technology
The MOE Key Laboratory of Computational Linguistics
Peking University
`wanxiaojun@icst.pku.edu.cn`

Abstract

Single-document summarization and multi-document summarization are very closely related tasks and they have been widely investigated independently. This paper examines the mutual influences between the two tasks and proposes a novel unified approach to simultaneous single-document and multi-document summarizations. The mutual influences between the two tasks are incorporated into a graph model and the ranking scores of a sentence for the two tasks can be obtained in a unified ranking process. Experimental results on the benchmark DUC datasets demonstrate the effectiveness of the proposed approach for both single-document and multi-document summarizations.

1 Introduction

Single-document summarization aims to produce a concise and fluent summary for a single document, and multi-document summarization aims to produce a concise and fluent summary for a document set consisting of multiple related documents. The two tasks are very closely related in both task definition and solution method. Moreover, both of them are very important in many information systems and applications. For example, given a cluster of news articles, a multi-document summary can be used to help users to understand the whole cluster, and a single summary for each article can be used to help users to know the content of the specified article.

To date, single-document and multi-document summarizations have been investigated extensively and independently in the NLP and IR fields. A series of special conferences or workshops on automatic text summarization (e.g. SUMMAC, DUC, NTCIR and TAC) have advanced the technology and produced a couple of experimental online systems. However, the two summarization tasks have not yet been simultaneously investigated in a unified framework.

Inspired by the fact that the two tasks are very closely related and they can be used simultaneously in many applications, we believe that the two tasks may have mutual influences on each other. In this study, we propose a unified approach to simultaneous single-document and multi-document summarizations. The mutual influences between the two tasks are incorporated into a graph-based model. The ranking scores of sentences for single-document summarization and the ranking scores of sentences for multi-document summarization can boost each other, and they can be obtained simultaneously in a unified graph-based ranking process. To the best of our knowledge, this study is the first attempt for simultaneously addressing the two summarization tasks in a unified graph-based framework. Moreover, the proposed approach can be easily adapted for topic-focused summarizations.

Experiments have been performed on both the single-document and multi-document summarization tasks of DUC2001 and DUC2002. The results demonstrate that the proposed approach can outperform baseline independent methods for both the two summarization tasks. The two tasks are validated to have mutual influences on each other.

The rest of this paper is organized as follows: Section 2 introduces related work. The details of the proposed approach are described in Section 3. Section 4 presents and discusses the evaluation results. Lastly we conclude our paper in Section 5.

2 Related Work

Document summarization methods can be either extraction-based or abstraction-based. In this section, we focus on extraction-based methods.

Extraction-based methods for single-document summarization usually assign a saliency score to each sentence in a document and then rank and select the sentences. The score is usually computed based on a combination of statistical and linguistic features, such as term frequency, sentence position, cue words and stigma words (Luhn, 1969; Edmundson, 1969; Hovy and Lin, 1997). Machine learning techniques have also been used for sentence extraction (Kupiec et al., 1995; Conroy and O'Leary, 2001; Shen et al., 2007; Li et al., 2009). The mutual reinforcement principle has been exploited to iteratively extract key phrases and sentences from a document (Zha, 2002; Wan et al, 2007a). Wan et al. (2007b) propose the CollabSum algorithm to use additional knowledge in a cluster of documents to improve single document summarization in the cluster.

In recent years, graph-based ranking methods have been investigated for document summarization, such as TextRank (Mihalcea and Tarau, 2004; Mihalcea and Tarau, 2005) and LexPageRank (ErKan and Radev, 2004). Similar to PageRank (Page et al., 1998), these methods first build a graph based on the similarity relationships between the sentences in a document and then the saliency of a sentence is determined by making use of the global information on the graph recursively. The basic idea underlying the graph-based ranking algorithm is that of "voting" or "recommendation" between sentences.

Similar methods have been used for generic multi-document summarization. A typical method is the centroid-based method (Radev et al., 2004). For each sentence, the method computes a score based on each single feature (e.g. cluster centroids, position and TFIDF) and then linearly combines all the scores into an overall sentence score. Topic signature is used as a novel feature for selecting important content in NeATS (Lin and Hovy, 2002). Various sentence features have been combined by using machine learning techniques (Wong et al., 2008). A popular way for removing redundancy between summary sentences is the MMR algorithm (Carbonell and Goldstein, 1998). Themes (or topics,

clusters) in documents have been discovered and used for sentence selection (Harabagiu and Lacatusu, 2005). Hachey (2009) investigates the effect of various source document representations on the accuracy of the sentence extraction phase of a multi-document summarization task. Graph-based methods have also been used to rank sentences in a document set. The methods first construct a graph to reflect sentence relationships at different granularities, and then compute sentence scores based on graph-based learning algorithms. For example, Wan (2008) proposes to use only cross-document relationships for graph building and sentence ranking. Cluster-level information has been incorporated in the graph model to better evaluate sentences (Wan and Yang, 2008).

For topic-focused multi-document summarization, many methods are extensions of generic summarization methods by incorporating the information of the given topic or query into generic summarizers. In recent years, a few novel methods have been proposed for topic-focused summarization (Daumé and Marcu, 2006; Wan et al., 2007c; Nastase 2008; Li et al., 2008; Schilder and Kondadadi, 2008; Wei et al., 2008).

The above previous graph-based summarization methods aim to address either single-document summarization or multi-document summarization, and the two summarization tasks have not yet been addressed in a unified graph-based framework.

3 The Unified Summarization Approach

3.1 Overview

Given a document set, in which the whole document set and each single document in the set are required to be summarized, we use *local saliency* to indicate the importance of a sentence in a particular document, and use *global saliency* to indicate the importance of a sentence in the whole document set.

In previous work, the following two assumptions are widely made for graph-based summarization models:

Assumption 1: A sentence is locally important in a particular document if it is heavily linked with many locally important sentences in the same document.

Assumption 2: A sentence is globally important in the document set if it is heavily linked with many globally important sentences in the document set.

The above assumptions are the basis for PageRank-like algorithms for single document summarization and multi-document summarization, respectively. In addition to the above two assumptions, we make the following two assumptions to consider the mutual influences between the two summarization tasks:

Assumption 3: A sentence is locally important in a particular document, if it is heavily linked with many globally important sentences in the document set.

The above assumption is reasonable because the documents in the set are relevant and the globally important information in the document set will be expressed in many single documents. Therefore, if a sentence is salient in the whole document set, the sentence may be salient in a particular document in the set.

Assumption 4: A sentence is globally important in the document set, if it is heavily linked with many locally important sentences.

The above assumption is reasonable because the documents in the set are relevant and the globally important information in the whole set is the aggregation of the locally important information in each single document. Therefore, if a sentence is salient in a particular document, the sentence has the potential to be salient in the whole document set.

In brief, the local saliency and global saliency of a sentence can mutually influence and boost each other: high local saliency will lead to high global saliency, and high global saliency will lead to high local saliency.

Based on the above assumptions, our proposed approach first builds affinity graphs (each graph is represented by an affinity matrix) to reflect the different kinds of relationships between sentences, respectively, and then iteratively computes the local saliency scores and the global saliency scores of the sentences based on the graphs. Finally, the algorithm converges and the local saliency score and global saliency score of each sentence are obtained. The sentences with high local saliency scores in a particular document are chosen into the summary of the single document, and the sentences with

high global saliency scores in the set are chosen into the summary of the document set.

Note that for both summarization tasks, after the saliency scores of sentences have been obtained, the greedy algorithm used in (Wan et al., 2007c) is applied to remove redundancy and finally choose both informative and novel sentences into the summary.

3.2 Algorithm Details

Formally, the given document set is denoted as $D=\{d_i|1\leq i\leq m\}$, and the whole sentence set is denoted as $S=\{s_i|1\leq i\leq n\}$. We let $Info_{single}(s_i)$ denote the local saliency score of sentence s_i in a particular document $d(s_i)\in D$, and it is used to select summary sentences for the single document $d(s_i)$. And we let $Info_{multi}(s_i)$ denote the global saliency score of sentence s_i in the whole document set D, and it is used to select summary sentences for the document set D.

The four assumptions in Section 3.1 can be rendered as follows:

$$Info_{single}(s_i) \propto \sum_j (W_A)_{ji} Info_{single}(s_j) \qquad (1)$$

$$Info_{multi}(s_i) \propto \sum_j (W_B)_{ji} Info_{multi}(s_j) \qquad (2)$$

$$Info_{single}(s_i) \propto \sum_j (W_C)_{ji} Info_{multi}(s_j) \qquad (3)$$

$$Info_{multi}(s_i) \propto \sum_j (W_D)_{ji} Info_{single}(s_j) \qquad (4)$$

where W_A, W_B, W_C, W_D are $n \times n$ affinity matrices reflecting the different kinds of relationships between sentences in the document set, where n is the number of all sentences in the document set. The detailed derivation of the matrices will be presented later.

After fusing the above equations, we can obtain the following unified forms:

$$Info_{single}(s_i) = \mu \sum_j (W_A)_{ji} Info_{single}(s_j) \\ + (1-\mu)\sum_j (W_C)_{ji} Info_{multi}(s_j) \qquad (5)$$

$$Info_{multi}(s_i) = \mu \sum_j (W_B)_{ji} Info_{multi}(s_j) \\ + (1-\mu)\sum_j (W_D)_{ji} Info_{single}(s_j) \qquad (6)$$

However, the above summarization method ignores the feature of sentence position, which has been validated to be very important for document summarizations. In order to incorporate this important feature, we add one prior score to each computation as follows:

$$Info_{single}(s_i) = \alpha \sum_j (W_A)_{ji} Info_{single}(s_j) \\ + \beta \sum_j (W_C)_{ji} Info_{multi}(s_j) + \gamma \cdot prior_{single}(s_i) \qquad (7)$$

$$Info_{multi}(s_i) = \alpha \sum_j (W_B)_{ji} Info_{multi}(s_j)$$
$$+ \beta \sum_j (W_D)_{ji} Info_{single}(s_j) + \gamma \cdot prior_{multi}(s_i) \qquad (8)$$

where α, β, $\gamma \in [0,1]$ specify the relative contributions to the final saliency scores from the different factors, and we have $\alpha+\beta+\gamma=1$. $prior_{single}(s_i)$ is the prior score for the local saliency of sentence s_i, and here $prior_{single}(s_i)$ is computed based on sentence position of s_i in the particular document $d(s_i)$. $prior_{multi}(s_i)$ is the prior score for the global saliency of sentence s_i, and we also compute $prior_{multi}(s_i)$ based on sentence position of s_i.

We use two column vectors $\vec{u} = [Info_{single}(s_i)]_{n\times1}$ and $\vec{v} = [Info_{multi}(s_i)]_{n\times1}$ to denote the local and global saliency scores of all the sentences in the set, respectively. And the matrix forms of the above equations are as follows:

$$\vec{u} = \alpha W_A^T \vec{u} + \beta W_C^T \vec{v} + \gamma \vec{p}_{single} \qquad (9)$$

$$\vec{v} = \alpha W_B^T \vec{v} + \beta W_D^T \vec{u} + \gamma \vec{p}_{multi} \qquad (10)$$

where $\vec{p}_{single} = [prior_{single}(s_i)]_{n\times1}$ and $\vec{p}_{multi} = [prior_{multi}(s_i)]_{n\times1}$ are the prior column vectors.

The above matrices and prior vectors are constructed as follows, respectively:

W_A: This affinity matrix aims to reflect the local relationships between sentences in each single document, which is defined as follows:

$$(W_A)_{ij} = \begin{cases} sim_{cosine}(s_i, s_j), & \text{if } d(s_i) = d(s_j) \\ & \text{and } i \neq j \\ 0, & \text{Otherwise} \end{cases} \qquad (11)$$

where $d(s_i)$ refers to the document containing sentence s_i. $sim_{cosine}(s_i, s_j)$ is the cosine similarity between sentences s_i and s_j.

$$sim_{cosine}(s_i, s_j) = \frac{\vec{s}_i \cdot \vec{s}_j}{|\vec{s}_i| \times |\vec{s}_j|} \qquad (12)$$

where $\vec{s}_i$ and $\vec{s}_j$ are the corresponding term vectors of s_i and s_j. Note that we have $(W_A)_{ij} = (W_A)_{ji}$, and we have $(W_A)_{ii} = 0$ to avoid self loops.

We can see that the matrix contains only the within-document relationships between sentences.

W_B: This affinity matrix aims to reflect the global relationships between sentences in the document set, which is defined as follows:

$$(W_B)_{ij} = \begin{cases} sim_{cosine}(s_i, s_j), & \text{if } d(s_i) \neq d(s_j) \\ 0, & \text{Otherwise} \end{cases} \qquad (13)$$

We can see that the matrix contains only the cross-document relationships between sentences. We do not include the within-document sentence relationships in the matrix because it has been shown that the cross-document relationships are more appropriate to reflect the global mutual influences between sentences than the within-document relationships in (Wan, 2008).

W_C: This affinity matrix aims to reflect the cross-document relationships between sentences in the document set. However, the relationships in this matrix are used for carrying the influences of the sentences in other documents on the local saliency of the sentences in a particular document. If we directly use Equation (13) to compute the matrix, the mutual influences would be overly used. Because other documents might not be sampled from the same generative model as the specified document, we probably do not want to trust them so much as the specified document. Thus a confidence value is used to reflect out belief that the document is sampled from the same underlying model as the specified document. Heuristically, we use the cosine similarity between documents as the confidence value. And we use the confidence value as the decay factor in the matrix computation as follows:

$$(W_c)_{ij} = \begin{cases} sim_{cosine}(s_i, s_j) \times sim_{cosine}(d(s_i), d(s_j)), \\ \qquad\qquad\qquad \text{if } d(s_i) \neq d(s_j) \\ 0, \quad \text{Otherwise} \end{cases} \qquad (14)$$

W_D: This affinity matrix aims to reflect the within-document relationships between sentences. Thus we have $W_D = W_A$, which means that the global saliency score of a sentence is influenced only by the local saliency scores of the sentences in the same document, without considering the sentences in other documents.

Note that the above four matrices are symmetric and we can replace W_A^T, W_B^T, W_C^T and W_D^T by W_A, W_B, W_C and W_D in Equations (9) and (10), respectively.

$prior_{single}(s_i)$: It is computed under the assumption that the first sentences in a document are usually more important than other sentences.

$$prior_{single}(s_i) = 0.5 + \frac{1}{position(s_i)+1} \qquad (15)$$

where $position(s_i)$ returns the position number of sentence s_i in its document $d(s_i)$. For example, if

s_i is the first sentence in its document, *position*(s_i) is 1.

The prior weight is then normalized by:

$$prior_{\sin gle}(s_i) = \frac{prior_{\sin gle}(s_i)}{\sum_i prior_{\sin gle}(s_i)} \quad (16)$$

prior_{multi}(s_i): We also let the prior weight reflect the influence of sentence position.

$$prior_{multi}(s_i) = prior_{\sin gle}(s_i) \quad (17)$$

And then the prior weight is normalized in the same way.

The above definitions are for generic document summarizations and the above algorithm can be easily adapted for topic-focused summarizations. Given a topic *q*, the only change for the above computation is *prior_{multi}(s_i)*. The topic relevance is incorporated into the prior weight as follows:

$$prior_{multi}(s_i) = sim_{\cos ine}(s_i, q) \quad (18)$$

$$prior_{multi}(s_i) = \frac{prior_{multi}(s_i)}{\sum_i prior_{multi}(s_i)} \quad (19)$$

In order to solve the iterative problem defined in Equations (9) and (10), we let $\vec{r} = [\vec{u}^T \ \vec{v}^T]^T$, $\vec{p} = [\lambda \vec{p}_{single}^T \ \lambda \vec{p}_{multi}^T]^T$, $W = \begin{bmatrix} \alpha W_A^T & \beta W_C^T \\ \beta W_D^T & \alpha W_B^T \end{bmatrix}$, and then the iterative equations correspond to the following linear system:

$$\vec{r} = W\vec{r} + \vec{p} \quad (20)$$

$$(I - W)\vec{r} = \vec{p} \quad (21)$$

To guarantee the solution of the above linear system, *W* is normalized by columns. If all the elements of a column are zero, we replace the elements with $1/(2n)$, where $2n$ equals to the element number of the column. We then multiply *W* by a decay factor θ ($0<\theta<1$) to scale down each element in *W*, but remain the meaning of *W*. Here, θ is empirically set to 0.6[1]. Finally, Equation (21) is rewritten as follows:

$$(I - \theta \cdot W)\vec{r} = \vec{p} \quad (22)$$

Thus, the matrix ($I-\theta W$) is a strictly diagonally dominant matrix and the solution of the linear system exists and we can apply the Gauss-Seidel method used in (Li et al., 2008) to solve the linear system. The GS method is a well-know method for numeric computation in

mathematics and the details of the method is omitted here.

4 Empirical Evaluation

4.1 Dataset and Evaluation Metric

Generic single-document and multi-document summarizations have been the fundamental tasks in DUC 2001 and DUC 2002 (i.e. tasks 1 and 2 in DUC 2001 and tasks 1 and 2 in DUC 2002), and we used the two datasets for evaluation. DUC2001 provided 309 articles, which were grouped into 30 document sets. Generic summary of each article was required to be created for task 1, and generic summary of each document set was required to be created for task 2. The summary length was 100 words or less. DUC 2002 provided 59 document sets consisting of 567 articles (D088 is excluded from the original 60 document sets by NIST) and generic summaries for each article and each document set with a length of approximately 100 words were required to be created. The sentences in each article have been separated and the sentence information has been stored into files. The summary of the two datasets are shown in Table 1.

	DUC 2001	DUC 2002
Task	Tasks 1, 2	Tasks 1, 2
Number of documents	309	567
Number of clusters	30	59
Data source	TREC-9	TREC-9
summary length	100 words	100 words

Table 1. Summary of datasets

We used the ROUGE toolkit[2] (Lin and Hovy, 2003) for evaluation, which has been widely adopted by DUC for automatic summarization evaluation. It measured summary quality by counting overlapping units such as the n-gram, word sequences and word pairs between the candidate summary and the reference summary.

The ROUGE toolkit reported separate recall-oriented scores for 1, 2, 3 and 4-gram, and also for longest common subsequence co-occurrences. We showed three of the ROUGE metrics in the experimental results: 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 the length limit,

---

[1] In our pilot study, we can observe good performance when θ is in a wide range of [0.4, 0.8].

[2] We used ROUGEeval-1.4.2 in this study.

we used the "-l 100" option in ROUGE toolkit. We also used the "-m" option for word stemming.

4.2 Evaluation Results

4.2.1 System Comparison

In the experiments, the combination weight γ for the prior score is fixed at 0.15, as in the PageRank algorithm. Therefore, we have $\alpha+\beta=0.85$. Here, we use $\alpha/(\alpha+\beta)$ to indicate the relative contributions of the first two parts in Equations (9) and (10). We empirically set $\alpha/(\alpha+\beta)=0.4$ in the experiments. The proposed unified approach (i.e. UnifiedRank) is compared with a few baseline approaches and the top three participating systems.

The graph-based baselines for single-document summarization are described as follows:

BasicRank: This baseline approach adopts the basic PageRank algorithm to rank sentences based on all sentence relationships in a single document, similar to previous work (Mihalcea and Tarau, 2004).

PositionRank: This baseline approach improves the basic PageRank algorithm by using the position weight of a sentence as the prior score for the sentence. The position weight of a sentence is computed by using Equation (15).

CollabRank1: This baseline approach is the "UniformLink(Gold)" approach proposed in (Wan et al. 2007b). It uses a cluster of multiple documents to improve single document summarization by constructing a global affinity graph.

CollabRank2: This baseline approach is the "UnionLink(Gold)" approach proposed in (Wan et al. 2007b).

The graph-based baselines for multi-document summarization are described as follows:

BasicRank: This baseline approach adopts the basic PageRank algorithm to rank sentences based on all sentence relationships in document set. Both within-document and cross-document sentence relationships are used for constructing the affinity graph.

PositionRank: Similarly, this baseline approach improves the basic PageRank algorithm by using the position weight of a sentence as the prior score for the sentence.

TwoStageRank: This baseline approach leverages the results of single document summarization for multi-document summarization. It first computes the score of each sentence within each single document by using the PositionRank method, and then computes the final score of each sentence within the document set by considering the document-level sentence score as the prior score in the improved PageRank algorithm.

The top three systems are the systems with highest ROUGE scores, chosen from the participating systems on each task, respectively. Tables 2 and 3 show the comparison results for single-document summarization on DUC2001 and DUC2002, respectively. Tables 4 and 5 show the comparison results for multi-document summarization on DUC2001 and DUC2002, respectively. In the tables, SystemX (e.g. System28, SystemN) represents one of the top performing systems. The systems are sorted by decreasing order of the ROUGE-1 scores.

For single-document summarization, the proposed UnifiedRank approach always outperforms the four graph-based baselines over all three metrics on both two datasets. The performance differences are all statistically significant by using t-test (p-value<0.05). The ROUGE-1 score of UnifiedRank is higher than that of the best participating systems and the ROUGE-2 and ROUGE-W scores of UnifiedRank are comparable to that of the best participating systems.

For multi-document summarization, the proposed UnifiedRank approach outperforms all the three graph-based baselines over all three metrics on the DUC2001 dataset, and it outperforms the three baselines over ROUGE-1 and ROUGE-W on the DUC2002 dataset. In particular, UnifiedRank can significantly outperform BasicRank and TwoStageRank over all three metrics on the DUC2001 dataset (t-test, p-value<0.05). Moreover, the ROUGE-1 and ROUGE-W scores of UnifiedRank are higher than that of the best participating systems and the ROUGE-2 score of UnifiedRank is comparable to that of the best participating systems.

The results demonstrate that the single-document and multi-document summarizations can benefit each other by making use of the mutual influences between the local saliency and

global saliency of the sentences. Overall, the proposed unified graph-based approach is effective for both single document summarization and multi-document summarization. However, the performance improvement for single-document summarization is more significant than that for multi-document summarization, which shows that the global information in a document set is very beneficial to summarization of each single document in the document set.

System	ROUGE-1	ROUGE-2	ROUGE-W
UnifiedRank	**0.45377**	**0.17649**	**0.14328**
CollabRank2	0.44038	0.16229	0.13678
CollabRank1	0.43890	0.16213	0.13676
PositionRank	0.43596	0.15936	0.13684
BasicRank	0.43407	0.15696	0.13629

Table 2. Comparison results for single-document summarization on DUC2001[3]

System	ROUGE-1	ROUGE-2	ROUGE-W
UnifiedRank	**0.48478**	0.21462	0.16877
System28	0.48049	**0.22832**	**0.17073**
System21	0.47754	0.22273	0.16814
CollabRank1	0.47187	0.20102	0.16318
CollabRank2	0.47028	0.20046	0.16260
PositionRank	0.46618	0.19853	0.16180
System31	0.46506	0.20392	0.16162
BasicRank	0.46261	0.19457	0.16018

Table 3. Comparison results for single-document summarization on DUC2002

System	ROUGE-1	ROUGE-2	ROUGE-W
UnifiedRank	**0.36360**	0.06496	**0.10950**
PositionRank	0.35733	0.06092	0.10798
BasicRank	0.35527	0.05608	0.10641
TwoStageRank	0.35221	0.05500	0.10515
SystemN	0.33910	0.06853	0.10240
SystemP	0.33332	0.06651	0.10068
SystemT	0.33029	**0.07862**	0.10215

Table 4. Comparison results for multi-document summarization on DUC2001

System	ROUGE-1	ROUGE-2	ROUGE-W
UnifiedRank	**0.38343**	0.07855	**0.12341**
PositionRank	0.38056	0.08238	0.12292
TwoStageRank	0.37972	0.08166	0.12261
BasicRank	0.37595	**0.08304**	0.12173
System26	0.35151	0.07642	0.11448
System19	0.34504	0.07936	0.11332
System28	0.34355	0.07521	0.10956

Table 5. Comparison results for multi-document summarization on DUC2002

4.2.2 *Influences of Combination Weight*

In the above experiments, the relative contributions from the first two parts in Equations (9) and (10) are empirically set as $\alpha/(\alpha+\beta)=0.4$. In this section, we investigate how the relative contributions influence the summarization performance by varying $\alpha/(\alpha+\beta)$ from 0 to 1. A small value of $\alpha/(\alpha+\beta)$ indicates that the contribution from the same kind of saliency scores of the sentences is less important than the contribution from the different kind of saliency scores of the sentences, and vice versa. Figures 1-8 show the ROUGE-1 and ROUGE-W curves for single-document summarization and multi-document summarization on DUC2001 and DUC2002, respectively.

For single document summarization, very small value or very large value for $\alpha/(\alpha+\beta)$ will lower the summarization performance values on the two datasets. The results demonstrate that both the two kinds of contributions are important to the final performance of single document summarization.

For multi-document summarization, a relatively large value (≥ 0.4) for $\alpha/(\alpha+\beta)$ will lead to relatively high performance values on the DUC2001 dataset, but a very large value for $\alpha/(\alpha+\beta)$ will decrease the performance values. On the DUC2002 dataset, a relatively small value (≤ 0.4) will lead to relatively high performance values, but a very small value for $\alpha/(\alpha+\beta)$ will decrease the performance values. Though the trends of the curves on the DUC2001 and DUC2002 datasets are not very consistent with each other, the results show that both the two kinds of contributions are beneficial to the final performance of multi-document summarization.

5 Conclusion and Future Work

In this study, we propose a novel unified approach to simultaneous single-document and multi-document summarization by making using of the mutual influences between the two tasks. Experimental results on the benchmark DUC datasets show the effectiveness of the proposed approach.

In future work, we will perform comprehensive experiments for topic-focused document

---

[3] The summarization results for participating systems on DUC2001 are incomplete.

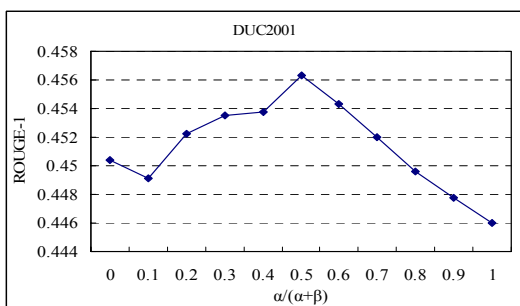summarizations to show the robustness of the proposed approach.

Figure 1. ROUGE-1 vs. combination weight for single-document summarization on DUC2001
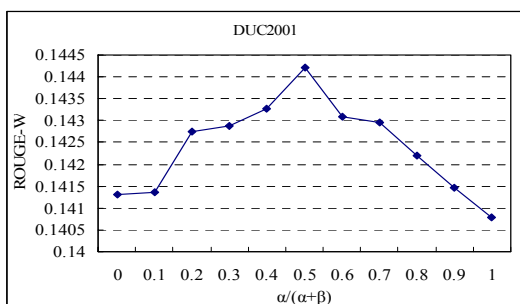

Figure 2. ROUGE-W vs. combination weight for single-document summarization on DUC2001
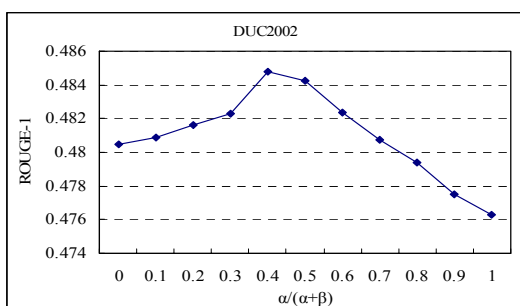

Figure 3. ROUGE-1 vs. combination weight for single-document summarization on DUC2002
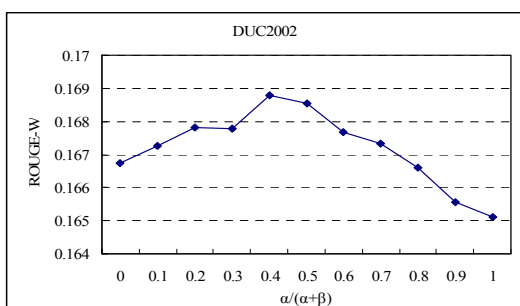

Figure 4. ROUGE-W vs. combination weight for single-document summarization on DUC2002
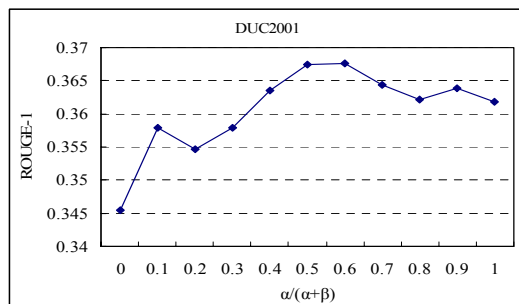

Figure 5. ROUGE-1 vs. combination weight for multi-document summarization on DUC2001
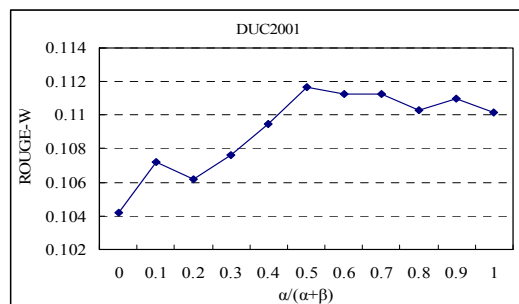

Figure 6. ROUGE-W vs. combination weight for multi-document summarization on DUC2001
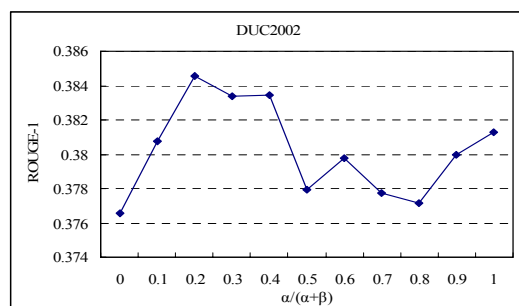

Figure 7. ROUGE-1 vs. combination weight for multi-document summarization on DUC2002
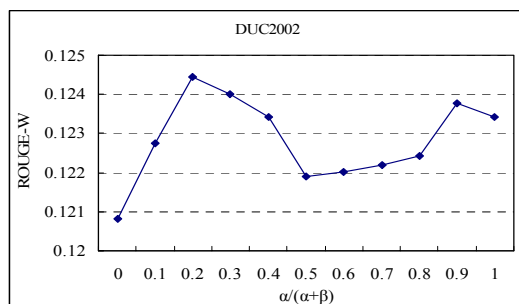

Figure 8. ROUGE-W vs. combination weight for multi-document summarization on DUC2002

1144

References

J. Carbonell, J. Goldstein. 1998. The Use of MMR, Diversity-based Reranking for Reordering Documents and Producing Summaries. In *Proceedings of SIGIR1998*, 335-336.

J. M. Conroy, D. P. O'Leary. 2001. Text Summarization via Hidden Markov Models. In *Proceedings of SIGIR2001*, 406-407.

H. Daumé and D. Marcu. 2006. Bayesian query-focused summarization. In *Proceedings of ACL-06*.

H. P. Edmundson. 1969. New Methods in Automatic Abstracting. *Journal of the Association for computing Machinery*, 16(2): 264-285.

G. ErKan, D. R. Radev. 2004. LexPageRank: Prestige in Multi-Document Text Summarization. In *Proceedings of EMNLP2004*.

B. Hachey. 2009. Multi-document summarisation using generic relation extraction. *In Proceedings of EMNLP2009*.

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

E. Hovy, C. Y. Lin. 1997. Automated Text Summarization in SUMMARIST. In *Proceeding of ACL'1997/EACL'1997 Worshop on Intelligent Scalable Text Summarization*.

J. Kupiec, J. Pedersen, F. Chen. 1995. A.Trainable Document Summarizer. In *Proceedings of SIGIR1995*, 68-73.

W. Li, F. Wei, Q. Lu and Y. He. 2008. PNR2: ranking sentences with positive and negative reinforcement for query-oriented update summarization. In *Proceedings of COLING-08*.

L. Li, K. Zhou, G.-R. Xue, H. Zha, Y. Yu. 2009. Enhancing diversity, coverage and balance for summarization through structure learning. In *Proceedings of WWW-09*.

C..-Y. Lin and E.. H. Hovy. 2002. From Single to Multi-document Summarization: A Prototype System and its Evaluation. In *Proceedings of ACL-02*.

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

H. P. Luhn. 1969. The Automatic Creation of literature Abstracts. *IBM Journal of Research and Development*, 2(2).

R. Mihalcea, P. Tarau. 2004. TextRank: Bringing Order into Texts. In *Proceedings of EMNLP2004*.

R. Mihalcea and P. Tarau. 2005. A language independent algorithm for single and multiple docu-ment summarization. In *Proceedings of IJCNLP-05*.

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

L. Page, S. Brin, R. Motwani, and T. Winograd. 1998. The pagerank citation ranking: Bringing order to the web. *Technical report*, Stanford Digital Libraries.

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

F. Schilder and R. Kondadadi. 2008. FastSum: fast and accurate query-based multi-document summarization. In *Proceedings of ACL-08: HLT*.

D. Shen, J.-T. Sun, H. Li, Q. Yang, and Z. Chen. 2007. Document Summarization using Conditional Random Fields. In *Proceedings of IJCAI2007*.

X. Wan. 2008. Using Only Cross-Document Relationships for Both Generic and Topic-Focused Multi-Document Summarizations. *Information Retrieval*, 11(1): 25-49.

X. Wan and J. Yang. 2008. Multi-document summarization using cluster-based link analysis. In *Proceedings of SIGIR-08*.

X. Wan, J. Yang and J. Xiao. 2007a. Towards an Iterative Reinforcement Approach for Simultaneous Document Summarization and Keyword Extraction. In *Proceedings of ACL2007*.

X. Wan, J. Yang and J. Xiao. 2007b. CollabSum: Exploiting Multiple Document Clustering for Collaborative Single Document Summarizations. In *Proceedings of SIGIR2007*.

X. Wan, J. Yang and J. Xiao. 2007c. Manifold-ranking based topic-focused multi-document summarization. In *Proceedings of IJCAI-07*.

F. Wei, W. Li, Q. Lu and Y. He. 2008. Query-sensitive mutual reinforcement chain and its application in query-oriented multi-document summarization. In *Proceedings of SIGIR-08*.

K.-F. Wong, M. Wu and W. Li. 2008. Extractive summarization using supervised and semi-supervised learning. In *Proceedings of COLING-08*.

H. Y. Zha. 2002. Generic Summarization and Keyphrase Extraction Using Mutual Reinforcement Principle and Sentence Clustering. In *Proceedings of SIGIR2002*, 113-120.