# Using Bilingual Information for Cross-Language Document Summarization

**Xiaojun Wan**

Institute of Compute Science and Technology, Peking University, Beijing 100871, China
Key Laboratory of Computational Linguistics (Peking University), MOE, China
wanxiaojun@icst.pku.edu.cn

## Abstract

Cross-language document summarization is defined as the task of producing a summary in a target language (e.g. Chinese) for a set of documents in a source language (e.g. English). Existing methods for addressing this task make use of either the information from the original documents in the source language or the information from the translated documents in the target language. In this study, we propose to use the bilingual information from both the source and translated documents for this task. Two summarization methods (SimFusion and CoRank) are proposed to leverage the bilingual information in the graph-based ranking framework for cross-language summary extraction. Experimental results on the DUC2001 dataset with manually translated reference Chinese summaries show the effectiveness of the proposed methods.

## 1 Introduction

Cross-language document summarization is defined as the task of producing a summary in a different target language for a set of documents in a source language (Wan et al., 2010). In this study, we focus on English-to-Chinese cross-language summarization, which aims to produce Chinese summaries for English document sets. The task is very useful in the field of multilingual information access. For example, it is beneficial for most Chinese readers to quickly browse and understand English news documents or document sets by reading the corresponding Chinese summaries.

A few pilot studies have investigated the task in recent years and exiting methods make use of either the information in the source language or the information in the target language after using machine translation. In particular, for the task of English-to-Chinese cross-language summarization, one method is to directly extract English summary sentences based on English features extracted from the English documents, and then automatically translate the English summary sentences into Chinese summary sentences. The other method is to automatically translate the English sentences into Chinese sentences, and then directly extract Chinese summary sentences based on Chinese features. The two methods make use of the information from only one language side.

However, it is not very reliable to use only the information in one language, because the machine translation quality is far from satisfactory, and thus the translated Chinese sentences usually contain some errors and noises. For example, the English sentence "*Many destroyed power lines are thought to be uninsured, as are trees and shrubs uprooted across a wide area.*" is automatically translated into the Chinese sentence "许多破坏电源线被认为是保险的，因为是连根拔起的树木和灌木，在广泛的领域。" by using Google Translate[1], but the Chinese sentence contains a few translation errors. Therefore, on the one side, if we rely only on the English-side information to extract Chinese

---

[1] http://translate.google.com/. Note that the translation service is updated frequently and the current translation results may be different from that presented in this paper.

summary sentences, we cannot guarantee that the automatically translated Chinese sentences for salient English sentences are really salient when these sentences may contain many translation errors and other noises. On the other side, if we rely only on the Chinese-side information to extract Chinese summary sentences, we cannot guarantee that the selected sentences are really salient because the features for sentence ranking based on the incorrectly translated sentences are not very reliable, either.

In this study, we propose to leverage both the information in the source language and the information in the target language for cross-language document summarization. In particular, we propose two graph-based summarization methods (SimFusion and CoRank) for using both English-side and Chinese-side information in the task of English-to-Chinese cross-document summarization. The SimFusion method linearly fuses the English-side similarity and the Chinese-side similarity for measuring Chinese sentence similarity. The CoRank method adopts a co-ranking algorithm to simultaneously rank both English sentences and Chinese sentences by incorporating mutual influences between them.

We use the DUC2001 dataset with manually translated reference Chinese summaries for evaluation. Experimental results based on the ROUGE metrics show the effectiveness of the proposed methods. Three important conclusions for this task are summarized below:

1) The Chinese-side information is more beneficial than the English-side information.
2) The Chinese-side information and the English-side information can complement each other.
3) The proposed CoRank method is more reliable and robust than the proposed SimFusion method.

The rest of this paper is organized as follows: Section 2 introduces related work. In Section 3, we present our proposed methods. Evaluation results are shown in Section 4. Lastly, we conclude this paper in Section 5.

## 2    Related Work

### 2.1    General Document Summarization

Document summarization methods can be extraction-based, abstraction-based or hybrid methods. We focus on extraction-based methods in this study, and the methods directly extract summary sentences from a document or document set by ranking the sentences in the document or document set.

In the task of single document summarization, various features have been investigated for ranking sentences in a document, including term frequency, sentence position, cue words, stigma words, and topic signature (Luhn 1969; Lin and Hovy, 2000). Machine learning techniques have been used for sentence ranking (Kupiec et al., 1995; Amini and Gallinari, 2002). Litvak et al. (2010) present a language-independent approach for extractive summarization based on the linear optimization of several sentence ranking measures using a genetic algorithm. In recent years, graph-based methods have been proposed for sentence ranking (Erkan and Radev, 2004; Mihalcea and Tarau, 2004). Other methods include mutual reinforcement principle (Zha 2002; Wan et al., 2007).

In the task of multi-document summarization, the centroid-based method (Radev et al., 2004) ranks the sentences in a document set based on such features as cluster centroids, position and TFIDF. Machine Learning techniques have also been used for feature combining (Wong et al., 2008). Nenkova and Louis (2008) investigate the influences of input difficulty on summarization performance. Pitler et al. (2010) present a systematic assessment of several diverse classes of metrics designed for automatic evaluation of linguistic quality of multi-document summaries. Celikyilmaz and Hakkani-Tur (2010) formulate extractive summarization as a two-step learning problem by building a generative model for pattern discovery and a regression model for inference. Aker et al. (2010) propose an A* search algorithm to find the best extractive summary up to a given length, and they propose a discriminative training algorithm for directly maximizing the quality of the best summary. Graph-based methods have also been used to rank sentences for multi-document summarization (Mihalcea and Tarau, 2005; Wan and Yang, 2008).

## 2.2 Cross-Lingual Document Summarization

Several pilot studies have investigated the task of cross-language document summarization. The existing methods use only the information in either language side. Two typical translation schemes are document translation or summary translation. The document translation scheme first translates the source documents into the corresponding documents in the target language, and then extracts summary sentences based only on the information on the target side. The summary translation scheme first extracts summary sentences from the source documents based only on the information on the source side, and then translates the summary sentences into the corresponding summary sentences in the target language.

For example Leuski et al. (2003) use machine translation for English headline generation for Hindi documents. Lim et al. (2004) propose to generate a Japanese summary by using Korean summarizer. Chalendar et al. (2005) focus on semantic analysis and sentence generation techniques for cross-language summarization. Orasan and Chiorean (2008) propose to produce summaries with the MMR method from Romanian news articles and then automatically translate the summaries into English. Cross language query based summarization has been investigated in (Pingali et al., 2007), where the query and the documents are in different languages. Wan et al. (2010) adopt the summary translation scheme for the task of English-to-Chinese cross-language summarization. They first extract English summary sentences by using English-side features and the machine translation quality factor, and then automatically translate the English summary into Chinese summary. Other related work includes multilingual summarization (Lin et al., 2005; Siddharthan and McKeown, 2005), which aims to create summaries from multiple sources in multiple languages.

## 3 Our Proposed Methods

As mentioned in Section 1, existing methods rely only on one-side information for sentence ranking, which is not very reliable. In order to leveraging both-side information for sentence ranking, we propose the following two methods to incorporate the bilingual information in different ways.

## 3.1 SimFusion

This method uses the English-side information for Chinese sentence ranking in the graph-based framework. The sentence similarities in the two languages are fused in the method. In other words, when we compute the similarity value between two Chinese sentences, the similarity value between the corresponding two English sentences is used by linear fusion. Since sentence similarity evaluation plays a very important role in the graph-based ranking algorithm, this method can leverage both-side information through similarity fusion.

Formally, given the Chinese document set $D^{cn}$ translated from an English document set, let $G^{cn}=(V^{cn}, E^{cn})$ be an undirected graph to reflect the relationships between the sentences in the Chinese document set. $V^{cn}$ is the set of vertices and each vertex $s^{cn}_i$ in $V^{cn}$ represents a Chinese sentence. $E^{cn}$ is the set of edges. Each edge $e^{cn}_{ij}$ in $E^{cn}$ is associated with an affinity weight $f(s^{cn}_i, s^{cn}_j)$ between sentences $s^{cn}_i$ and $s^{cn}_j$ ($i{\neq}j$). The weight is computed by linearly combining the similarity value $sim_{cosine}(s^{cn}_i, s^{cn}_j)$ between the Chinese sentences and the similarity value $sim_{cosine}(s^{en}_i, s^{en}_j)$ between the corresponding English sentences.

$$f(s^{cn}_i, s^{cn}_j) = \lambda \cdot sim_{cosine}(s^{cn}_i, s^{cn}_j) + (1-\lambda) \cdot sim_{cosine}(s^{en}_i, s^{en}_j)$$

where $s^{en}_j$ and $s^{en}_i$ are the source English sentences for $s^{cn}_j$ and $s^{cn}_i$. $\lambda \in [0, 1]$ is a parameter to control the relative contributions of the two similarity values. The similarity values $sim_{cosine}(s^{cn}_i, s^{cn}_j)$ and $sim_{cosine}(s^{en}_i, s^{en}_j)$ are computed by using the standard cosine measure. The weight for each term is computed based on the TFIDF formula. For Chinese similarity computation, Chinese word segmentation is performed. Here, we have $f(s^{cn}_i, s^{cn}_j)=f(s^{cn}_j, s^{cn}_i)$ and let $f(s^{cn}_i, s^{cn}_i)=0$ to avoid self transition. We use an affinity matrix $M^{cn}$ to describe $G^{cn}$ with each entry corresponding to the weight of an edge in the graph. $M^{cn}=(M^{cn}_{ij})_{|V^{cn}|\times|V^{cn}|}$ is defined as $M^{cn}_{ij}=f(s^{cn}_i,s^{cn}_j)$. Then $M^{cn}$ is normalized to $\tilde{M}^{cn}$ to make the sum of each row equal to 1.

Based on matrix $\tilde{M}^{cn}$, the saliency score *InfoScore($s^{cn}_i$)* for sentence $s^{cn}_i$ can be deduced from those of all other sentences linked with it and it can be formulated in a recursive form as in the PageRank algorithm:

$$InfoScore(s_i^{cn}) = \mu \cdot \sum_{all\ j \neq i} InfoScore(s_j^{cn}) \cdot \widetilde{M}_{ji}^{cn} + \frac{(1-\mu)}{n}$$

where $n$ is the sentence number, i.e. $n = |V^{cn}|$. $\mu$ is the damping factor usually set to 0.85, as in the PageRank algorithm.

For numerical computation of the saliency scores, we can iteratively run the above equation until convergence.

For multi-document summarization, some sentences are highly overlapping with each other, and thus we apply the same greedy algorithm in Wan et al. (2006) to penalize the sentences highly overlapping with other highly scored sentences, and finally the salient and novel Chinese sentences are directly selected as summary sentences.

## 3.2 CoRank

This method leverages both the English-side information and the Chinese-side information in a co-ranking way. The source English sentences and the translated Chinese sentences are simultaneously ranked in a unified graph-based algorithm. The saliency of each English sentence relies not only on the English sentences linked with it, but also on the Chinese sentences linked with it. Similarly, the saliency of each Chinese sentence relies not only on the Chinese sentences linked with it, but also on the English sentences linked with it. More specifically, the proposed method is based on the following assumptions:

**Assumption 1**: A Chinese sentence would be salient if it is heavily linked with other salient Chinese sentences; and an English sentence would be salient if it is heavily linked with other salient English sentences.

**Assumption 2**: A Chinese sentence would be salient if it is heavily linked with salient English sentences; and an English sentence would be salient if it is heavily linked with salient Chinese sentences.

The first assumption is similar to PageRank which makes use of mutual "recommendations" between the sentences in the same language to rank sentences. The second assumption is similar to HITS if the English sentences and the Chinese sentences are considered as authorities and hubs, respectively. In other words, the proposed method aims to fuse the ideas of PageRank and HITS in a unified framework. The mutual influences between the Chinese sentences and the English sentences are incorporated in the method.

Figure 1 gives the graph representation for the method. Three kinds of relationships are exploited: the CN-CN relationships between Chinese sentences, the EN-EN relationships between English sentences, and the EN-CN relationships between English sentences and Chinese sentences.

Formally, given an English document set $D^{en}$ and the translated Chinese document set $D^{cn}$, let $G=(V^{en}, V^{cn}, E^{en}, E^{cn}, E^{encn})$ be an undirected graph to reflect all the three kinds of relationships between the sentences in the two document sets. $V^{en} = \{s_i^{en} \mid 1 \leq i \leq n\}$ is the set of English sentences. $V^{cn} = \{s_i^{cn} \mid 1 \leq i \leq n\}$ is the set of Chinese sentences. $s_i^{cn}$ is the corresponding Chinese sentence translated from $s_i^{en}$. $n$ is the number of the sentences. $E^{en}$ is the edge set to reflect the relationships between the English sentences. $E^{cn}$ is the edge set to reflect the relationships between the Chinese sentences. $E^{encn}$ is the edge set to reflect the relationships between the English sentences and the Chinese sentences. Based on the graph representation, we compute the following three affinity matrices to reflect the three kinds of sentence relationships:
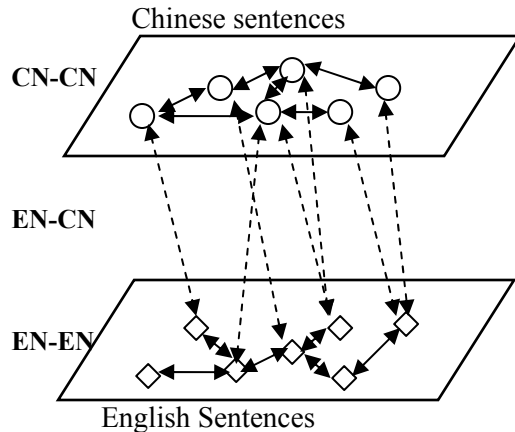


Figure 1. The three kinds of sentence relationships

1) $M^{cn} = (M_{ij}^{cn})_{n \times n}$: This affinity matrix aims to reflect the relationships between the Chinese sentences. Each entry in the matrix corresponds to the cosine similarity between the two Chinese sentences.

$$M_{ij}^{cn} = \begin{cases} sim_{cosine}(s_i^{cn}, s_j^{cn}), & if\ i \neq j \\ 0, & otherwise \end{cases}$$

Then $M^{cn}$ is normalized to $\widetilde{M}^{cn}$ to make the sum of each row equal to 1.

2) $M^{en}=(M^{en}{}_{i,j})_{n \times n}$: This affinity matrix aims to reflect the relationships between the English sentences. Each entry in the matrix corresponds to the cosine similarity between the two English sentences.

$$M_{ij}^{en} = \begin{cases} sim_{cosine}(s_i^{en}, s_j^{en}), & if\ i \neq j \\ 0, & otherwise \end{cases}$$

Then $M^{en}$ is normalized to $\widetilde{M}^{en}$ to make the sum of each row equal to 1.

3) $M^{encn}=(M^{encn}{}_{ij})_{n \times n}$: This affinity matrix aims to reflect the relationships between the English sentences and the Chinese sentences. Each entry $M^{encn}{}_{ij}$ in the matrix corresponds to the similarity between the English sentence $s^{en}{}_i$ and the Chinese sentence $s^{cn}{}_j$. It is hard to directly compute the similarity between the sentences in different languages. In this study, the similarity value is computed by fusing the following two similarity values: the cosine similarity between the sentence $s^{en}{}_i$ and the corresponding source English sentence $s^{en}{}_j$ for $s^{cn}{}_j$, and the cosine similarity between the corresponding translated Chinese sentence $s^{cn}{}_i$ for $s^{en}{}_i$ and the sentence $s^{cn}{}_j$. We use the geometric mean of the two values as the affinity weight.

$$M_{ij}^{encn} = \sqrt{sim_{cosine}(s_i^{en}, s_j^{en}) \times sim_{cosine}(s_i^{cn}, s_j^{cn})}$$

Note that we have $M^{encn}{}_{ij}=M^{encn}{}_{ji}$ and $M^{encn}=(M^{encn})^{\mathrm{T}}$. Then $M^{encn}$ is normalized to $\widetilde{M}^{encn}$ to make the sum of each row equal to 1.

We use two column vectors $u=[u(s^{cn}{}_i)]_{n \times 1}$ and $v=[v(s^{en}{}_j)]_{n \times 1}$ to denote the saliency scores of the Chinese sentences and the English sentences, respectively. Based on the three kinds of relationships, we can get the following four assumptions:

$$u(s_i^{cn}) \propto \sum_j \widetilde{M}_{ji}^{cn} u(s_j^{cn})$$

$$v(s_j^{en}) \propto \sum_i \widetilde{M}_{ij}^{en} v(s_i^{en})$$

$$u(s_i^{cn}) \propto \sum_j \widetilde{M}_{ji}^{encn} v(s_j^{en})$$

$$v(s_j^{en}) \propto \sum_i \widetilde{M}_{ij}^{encn} u(s_i^{cn})$$

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

$$u(s_i^{cn}) = \alpha \sum_j \widetilde{M}_{ji}^{cn} u(s_j^{cn}) + \beta \sum_j \widetilde{M}_{ji}^{encn} v(s_j^{en})$$

$$v(s_j^{en}) = \alpha \sum_i \widetilde{M}_{ij}^{en} v(s_i^{en}) + \beta \sum_i \widetilde{M}_{ij}^{encn} u(s_i^{cn})$$

And the matrix form is:

$$u = \alpha(\widetilde{M}^{cn})^T u + \beta(\widetilde{M}^{encn})^T v$$

$$v = \alpha(\widetilde{M}^{en})^T v + \beta(\widetilde{M}^{encn})^T u$$

where $\alpha$ and $\beta$ specify the relative contributions to the final saliency scores from the information in the same language and the information in the other language and we have $\alpha+\beta=1$.

For numerical computation of the saliency scores, we can iteratively run the two equations until convergence. Usually the convergence of the iteration algorithm is achieved when the difference between the scores computed at two successive iterations for any sentences and words falls below a given threshold. In order to guarantee the convergence of the iterative form, $u$ and $v$ are normalized after each iteration.

After we get the saliency scores $u$ for the Chinese sentences, we apply the same greedy algorithm for redundancy removing. Finally, a few highly ranked sentences are selected as summary sentences.

## 4 Experimental Evaluation

### 4.1 Evaluation Setup

There is no benchmark dataset for English-to-Chinese cross-language document summarization, so we built our evaluation dataset based on the DUC2001 dataset by manually translating the reference summaries.

DUC2001 provided 30 English document sets for generic multi-document summarization. The average document number per document set was 10. The sentences in each article have been separated and the sentence information has been stored into files. Three or two generic reference English summaries were provided by NIST annotators for each document set. Three graduate students were employed to manually translate the reference English summaries into reference Chinese summaries. Each student manually translated one third of the reference summaries. It was much easier and more reliable to provide the reference Chinese summaries by manual translation than by manual summarization.

|  | ROUGE-2 Average_F | ROUGE-W Average_F | ROUGE-L Average_F | ROUGE-SU4 Average_F |
|---|---|---|---|---|
| **Baseline(EN)** | 0.03723 | 0.05566 | 0.13259 | 0.07177 |
| **Baseline(CN)** | 0.03805 | 0.05886 | 0.13871 | 0.07474 |
| **SimFusion** | 0.04017 | 0.06117 | 0.14362 | 0.07645 |
| **CoRank** | 0.04282 | 0.06158 | 0.14521 | 0.07805 |

Table 1: Comparison Results

All the English sentences in the document set were automatically translated into Chinese sentences by using *Google Translate*, and the Stanford Chinese Word Segmenter[2] was used for segmenting the Chinese documents and summaries into words. For comparative study, the summary length was limited to five sentences, i.e. each Chinese summary consisted of five sentences.

We used the ROUGE-1.5.5 (Lin and Hovy, 2003) toolkit for evaluation, which has been widely adopted by DUC and TAC 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. We showed three of the ROUGE F-measure scores in the experimental results: ROUGE-2 (bigram-based), ROUGE-W (based on weighted longest common subsequence, weight=1.2), ROUGE-L (based on longest common subsequences), and ROUGE-SU4 (based on skip bigram with a maximum skip distance of 4). Note that the ROUGE toolkit was performed for Chinese summaries after using word segmentation.

Two graph-based baselines were used for comparison.

**Baseline(EN)**: This baseline adopts the summary translation scheme, and it relies on the English-side information for English sentence ranking. The extracted English summary is finally automatically translated into the corresponding Chinese summary. The same sentence ranking algorithm with the SimFusion method is adopted, and the affinity weight is computed based only on the cosine similarity between English sentences.

**Baseline(CN)**: This baseline adopts the document translation scheme, and it relies on the Chinese-side information for Chinese sentence ranking. The Chinese summary sentences are directly extracted from the translated Chinese documents. The same sentence ranking algorithm with the SimFusion method is adopted, and the affinity

weight is computed based only on the cosine similarity between Chinese sentences.

For our proposed methods, the parameter values are empirically set as $\lambda=0.8$ and $\alpha=0.5$.

## 4.2 Results and Discussion

Table 1 shows the comparison results for our proposed methods and the baseline methods. Seen from the tables, Baseline(CN) performs better than Baseline(EN) over all the metrics. The results demonstrate that the Chinese-side information is more beneficial than the English-side information for cross-document summarization, because the summary sentences are finally selected from the Chinese side. Moreover, our proposed two methods can outperform the two baselines over all the metrics. The results demonstrate the effectiveness of using bilingual information for cross-language document summarization. It is noteworthy that the ROUGE scores in the table are not high due to the following two reasons: 1) The use of machine translation may introduce many errors and noises in the peer Chinese summaries; 2) The use of Chinese word segmentation may introduce more noises and mismatches in the ROUGE evaluation based on Chinese words.

We can also see that the CoRank method can outperform the SimFusion method over all metrics. The results show that the CoRank method is more suitable for the task by incorporating the bilingual information into a unified ranking framework.

In order to show the influence of the value of the combination parameter $\lambda$ on the performance of the SimFusion method, we present the performance curves over the four metrics in Figures 2 through 5, respectively. In the figures, $\lambda$ ranges from 0 to 1, and $\lambda=1$ means that SimFusion is the same with Baseline(CN), and $\lambda=0$ means that only English-side information is used for Chinese sentence ranking. We can see that when $\lambda$ is set to a value larger than 0.5, SimFusion can outperform the two baselines over most metrics. The results show that $\lambda$ can be set in a relatively wide range. Note that

λ>0.5 means that SimFusion relies more on the Chinese-side information than on the English-side information. Therefore, the Chinese-side information is more beneficial than the English-side information.

In order to show the influence of the value of the combination parameter α on the performance of the CoRank method, we present the performance curves over the four metrics in Figures 6 through 9, respectively. In the figures, α ranges from 0.1 to 0.9, and a larger value means that the information from the same language side is more relied on, and a smaller value means that the information from the other language side is more relied on. We can see that CoRank can always outperform the two baselines over all metrics with different value of α. The results show that α can be set in a very wide range. We also note that a very large value or a very small value of α can lower the performance values. The results demonstrate that CoRank relies on both the information from the same language side and the information from the other language side for sentence ranking. Therefore, both the Chinese-side information and the English-side information can complement each other, and they are beneficial to the final summarization performance.

Comparing Figures 2 through 5 with Figures 6 through 9, we can further see that the CoRank method is more stable and robust than the SimFusion method. The CoRank method can outperform the SimFusion method with most parameter settings. The bilingual information can be better incorporated in the unified ranking framework of the CoRank method.

Finally, we show one running example for the document set D59 in the DUC2001 dataset. The four summaries produced by the four methods are listed below:

Baseline(EN): *周日的崩溃是 24 年来第一次乘客在涉及西北飞机事故中丧生。有乘客和观察员的报告，这架飞机的右翼引擎也坠毁前失败。在坠机现场联邦航空局官员表示不会揣测关于崩溃或在飞机上的发动机评论的原因。美国联邦航空局的记录显示，除了那些涉及的飞机坠毁，与 JT8D 涡轮路段-200 系列发动机问题的三个共和国在过去四年的航班发生的事件。1988 年 7 月，一个联合国的 DC-10 坠毁的苏城，艾奥瓦州后，发动机在飞行中发生外，造成 112 人。*

Baseline(CN): *第二，在美国历史上最严重的事故是 1987 年 8 月 16 日，坠毁，造成 156 人时，美国西北航空公司飞机上的底特律都市机场起飞时坠毁。据美国联邦航空管理局的纪录，麦道公司的 MD-82 飞机在 1985 年和 1986 年紧急降落后，在其两个引擎之一是失去权力。周日的崩溃是 24 年来第一次乘客在涉*

*及西北飞机事故中丧生。今年 4 月，国家运输安全委员会敦促美国联邦航空局后进行一些危险，发动机故障，飞机的一个发动机的 200 系列 JT8D 安全调查。目前，机组人员发现了一个黑人师谁说，他可以引导飞机在附近的人们听到了他们的区域。*

SimFusion: *第二，在美国历史上最严重的事故是 1987 年 8 月 16 日，坠毁，造成 156 人时，美国西北航空公司飞机上的底特律都市机场起飞时坠毁。周日的崩溃是 24 年来第一次乘客在涉及西北飞机事故中丧生。在坠机现场联邦航空局官员表示不会揣测关于崩溃或在飞机上的发动机评论的原因。有乘客和观察员的报告，这架飞机的右翼引擎也坠毁前失败。据美国联邦航空管理局的纪录，麦道公司的 MD-82 飞机在 1985 年和 1986 年紧急降落后，在其两个引擎之一是失去权力。*

CoRank：*周日的崩溃是 24 年来第一次乘客在涉及西北飞机事故中丧生。第二，在美国历史上最严重的事故是 1987 年 8 月 16 日，坠毁，造成 156 人时，美国西北航空公司飞机上的底特律都市机场起飞时坠毁。在坠机现场联邦航空局官员表示不会揣测关于崩溃或在飞机上的发动机评论的原因。最严重的航空事故不断，在美国是一个在芝加哥的美国航空公司客机 1979 年崩溃。有乘客和观察员的报告，这架飞机的右翼引擎也坠毁前失败。*

## 5 Conclusion and Future Work

In this paper, we propose two methods (SimFusion and CoRank) to address the cross-language document summarization task by leveraging the bilingual information in both the source and target language sides. Evaluation results demonstrate the effectiveness of the proposed methods. The Chinese-side information is validated to be more beneficial than the English-side information, and the CoRank method is more robust than the SimFusion method.

In future work, we will investigate to use the machine translation quality factor to further improve the fluency of the Chinese summary, as in Wan et al. (2010). Though our attempt to use GIZA++ for evaluating the similarity between Chinese sentences and English sentences failed, we will exploit more advanced measures based on statistical alignment model for cross-language similarity computation.
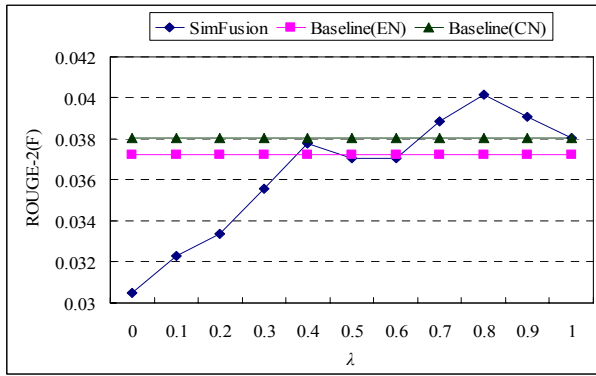
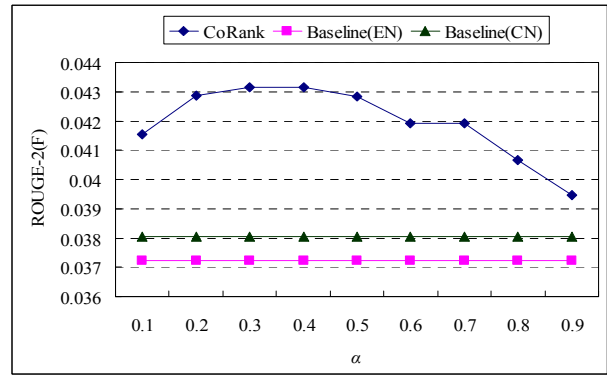Figure 2. ROUGE-2(F) vs. $\lambda$ for SimFusion



Figure 6. ROUGE-2(F) vs. $\alpha$ for CoRank
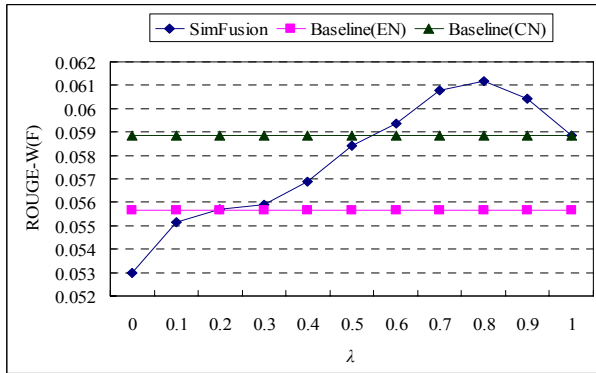


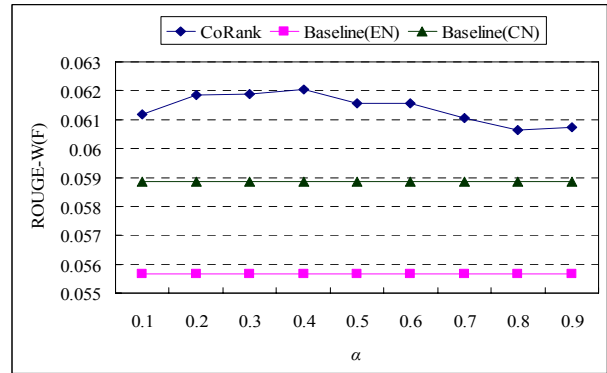Figure 3. ROUGE-W(F) vs. $\lambda$ for SimFusion



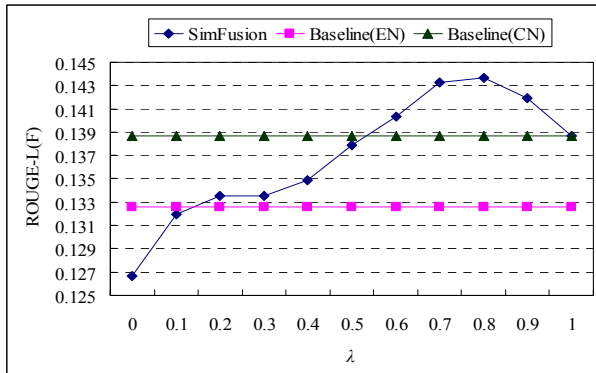Figure 7. ROUGE-W(F) vs. $\alpha$ for CoRank


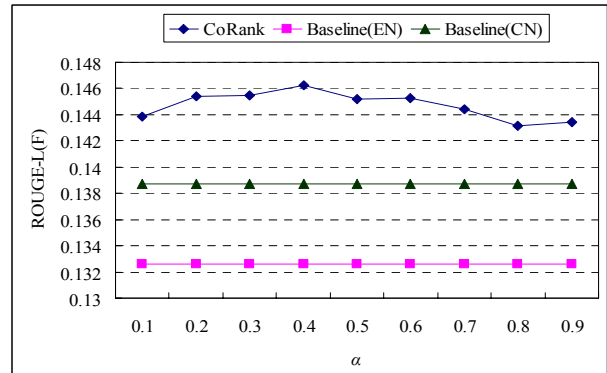
Figure 4. ROUGE-L(F) vs. $\lambda$ for SimFusion
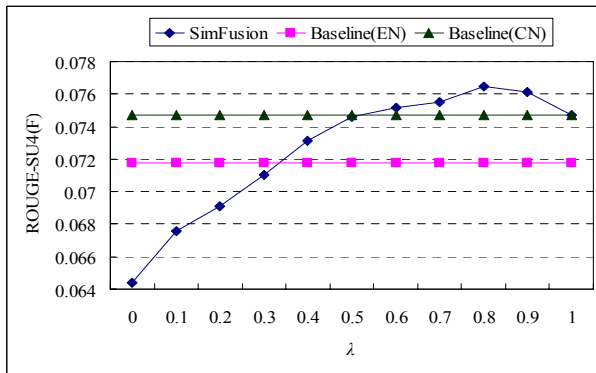


Figure 8. ROUGE-L(F) vs. $\alpha$ for CoRank
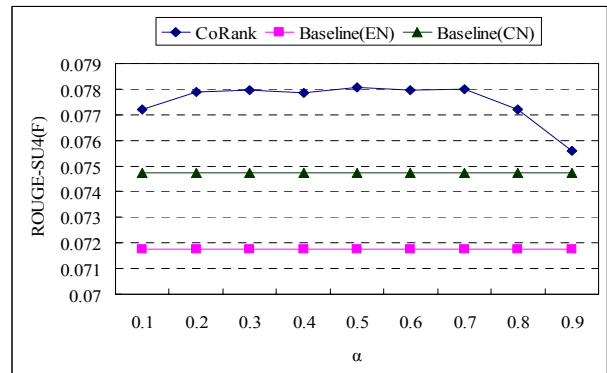


Figure 5. ROUGE-SU4(F) vs. $\lambda$ for SimFusion



Figure 9. ROUGE-SU4(F) vs. $\alpha$ for CoRank

# References

A. Aker, T. Cohn, and R. Gaizauskas. 2010. Multi-document summarization using A* search and discriminative training. In *Proceedings of EMNLP2010*.

M. R. Amini, P. Gallinari. 2002. The Use of Unlabeled Data to Improve Supervised Learning for Text Summarization. In *Proceedings of SIGIR2002*.

G. de Chalendar, R. Besançon, O. Ferret, G. Grefenstette, and O. Mesnard. 2005. Crosslingual summarization with thematic extraction, syntactic sentence simplification, and bilingual generation. In *Workshop on Crossing Barriers in Text Summarization Research, 5th International Conference on Recent Advances in Natural Language Processing (RANLP2005)*.

A. Celikyilmaz and D. Hakkani-Tur. 2010. A hybrid hierarchical model for multi-document summarization. In *Proceedings of ACL2010*.

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

D. Klein and C. D. Manning. 2002. Fast Exact Inference with a Factored Model for Natural Language Parsing. In *Proceedings of NIPS2002*.

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

A. Leuski, C.-Y. Lin, L. Zhou, U. Germann, F. J. Och, E. Hovy. 2003. Cross-lingual C*ST*RD: English access to Hindi information. *ACM Transactions on Asian Language Information Processing*, 2(3): 245-269.

J.-M. Lim, I.-S. Kang, J.-H. Lee. 2004. Multi-document summarization using cross-language texts. In *Proceedings of NTCIR-4*.

C. Y. Lin, E. Hovy. 2000. The Automated Acquisition of Topic Signatures for Text Summarization. In *Proceedings of the 17th Conference on Computational Linguistics*.

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

C.-Y. Lin, L. Zhou, and E. Hovy. 2005. Multilingual summarization evaluation 2005: automatic evaluation report. In *Proceedings of MSE (ACL-2005 Workshop)*.

M. Litvak, M. Last, and M. Friedman. 2010. A new approach to improving multilingual summarization using a genetic algorithm. In *Proceedings of ACL2010*.

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 document summarization. In *Proceedings of IJCNLP-05*.

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

A. Nenkova, R. Passonneau, and K. McKeown. 2007. The Pyramid method: incorporating human content selection variation in summarization evaluation. *ACM Transactions on Speech and Language Processing (TSLP)*, 4(2).

C. Orasan, and O. A. Chiorean. 2008. Evaluation of a Crosslingual Romanian-English Multi-document Summariser. In *Proceedings of 6th Language Resources and Evaluation Conference (LREC2008)*.

P. Pingali, J. Jagarlamudi and V. Varma. 2007. Experiments in cross language query focused multi-document summarization. In *Workshop on Cross Lingual Information Access Addressing the Information Need of Multilingual Societies in IJCAI2007*.

E. Pitler, A. Louis, and A. Nenkova. 2010. Automatic evaluation of linguistic quality in multi-document summarization. In *Proceedings of ACL2010*.

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.

A. Siddharthan and K. McKeown. 2005. Improving multilingual summarization: using redundancy in the input to correct MT errors. In *Proceedings of HLT/EMNLP-2005*.

X. Wan, H. Li and J. Xiao. 2010. Cross-language document summarization based on machine translation quality prediction. In *Proceedings of ACL2010*.

X. Wan, J. Yang and J. Xiao. 2006. Using cross-document random walks for topic-focused multi-documetn summarization. In *Proceedings of WI2006*.

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. 2007. Towards an Iterative Reinforcement Approach for Simultaneous Document Summarization and Keyword Extraction. In *Proceedings of ACL2007*.

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*.