

Summarizing the Differences in Multilingual News

Xiaojun Wan, Houping Jia, Shanshan Huang and Jianguo Xiao
Institute of Computer Science and Technology, Peking University, Beijing 100871, China
Key Laboratory of Computational Linguistics (Peking University), MOE, China
{wanxiaojun, jiahouping, huangshanshan, xiaojianguo}@icst.pku.edu.cn

ABSTRACT

There usually exist many news articles written in different languages about a hot news event. The news articles in different languages are written in different ways to reflect different standpoints. For example, the Chinese news agencies and the Western news agencies have published many articles to report the same news of “Liu Xiaobo’s Nobel Prize” in Chinese and English languages, respectively. The Chinese news articles and the English news articles share something about the news fact in common, but they focus on different aspects in order to reflect different standpoints about the event. In this paper, we investigate the task of multilingual news summarization for the purpose of finding and summarizing the major differences between the news articles about the same event in the Chinese and English languages. We propose a novel constrained co-ranking (C-CoRank) method for addressing this special task. The C-CoRank method adds the constraints between the difference score and the common score of each sentence to the co-ranking process. Evaluation results on the manually labeled test set with 15 news topics show the effectiveness of our proposed method, and the constrained co-ranking method can outperform a few baselines and the typical co-ranking method.

Categories and Subject Descriptors

H.3.1 [Information Storage and Retrieval]: Content Analysis and Indexing – *abstracting methods*; I.2.7 [Artificial Intelligence]: Natural Language Processing – *text analysis*

General Terms

Algorithms, Experimentation, Design.

Keywords

Multi-document summarization, multilingual summarization, constrained co-ranking.

1. INTRODUCTION

With the advent of a hot news event, many news articles are written and published by different news agencies in different languages. The news articles in different languages are usually written in different ways. The contents in these news articles are overlapping with each other, but they usually cover a few different aspects about the event (or they have different focuses), because

the agencies write the articles with different backgrounds, standpoints and purposes. For example, given the news of “Liu Xiaobo’s Nobel Prize”, in addition to the common content of event introduction, most Chinese news articles aim to reflect Chinese policies and standpoints about the event, but most English news articles aim to reflect the western countries’ understanding about the event. The reading of only the news articles in a single language will be biased. The summary of the differences between the two sets of news articles in the two languages can be used to help people understand the event in a more comprehensive and unbiased way.

We focus on summarizing the major differences in multilingual news (Chinese vs. English) in this study. Given a Chinese document set and an English document set about the same event, a Chinese summary is required to be created to reflect the major differential aspects in the Chinese news articles, and an English summary is required to be created to reflect the major differential aspects in the English news articles. In other words, the content in the Chinese summary is not the focus of the English documents, and the content in the English summary is not the focus of the Chinese documents. Our task can be considered as an extension of the update summarization task [7] in the following two ways: 1) our task aims to simultaneously create two summaries for two document sets to reflect the major differences in the two sets, while the update summarization task aims to create a summary for one document set after reading the other document set; 2) our task is defined in a multilingual setting, while the update summarization task is defined in a monolingual setting.

In this study, we present two graph-based ranking methods for addressing this special summarization task. The first method is named CoRank, which can simultaneously extract a Chinese summary and an English summary from the two document sets in a unified graph-based ranking process. The second method is named Constrained CoRank (C-CoRank), which improves the CoRank method by introducing a new factor (common score) for each sentence and adding important constraints into the graph-based ranking process. We manually labeled the Chinese and English difference summaries for 15 news topics, and the experimental results on the data set show the effectiveness of our proposed method. The C-CoRank method can significantly outperform the baselines and the CoRank method.

The rest of this paper is organized as follows: Section 2 reviews related work. Our proposed method is presented in Section 3. Sections 4 and 5 show the experiments and results. Lastly, we conclude this paper in Section 6.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SIGIR ’11, July 24–28, 2011, Beijing, China.

Copyright 2011 ACM 978-1-4503-0757-4/11/07...\$10.00.

2. RELATED WORK

2.1 Generic 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 [16, 21]. Machine learning techniques have been used for sentence ranking [2, 13]. Litvak et al. [20] 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 [9, 23]. Other methods include mutual reinforcement principle [33, 38].

In the task of multi-document summarization, the centroid-based method [28] ranks the sentences in a document set based on such features as cluster centroids, position and TFIDF. NeATS [17] makes use of new features such as topic signature to select important sentences. Machine Learning techniques have also been used for feature combining [37]. Themes (or topics, clusters) discovery in documents has been used for sentence selection [11]. Nenkova and Louis [26] investigate the influences of input difficulty on summarization performance. Celikyilmaz and Hakkani-Tur [6] 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. [1] 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 [24, 30, 32, 34].

2.2 Update Summarization

Update summarization is an emerging new summarization task in the very recent years. Piloted in DUC2007, the task aims to create a short summary of a set of news articles, under the assumption that the user has already read a given set of earlier articles [7]. The update summary is expected to be relevant to the given query or topic, and contain salient and new content in the current articles. Many existing multi-document summarization methods can be adapted for this task by considering the information in the earlier documents. For example, Fisher and Roark [10] present a supervised sentence ranking approach for extractive update summarization. Nastase et al. [25] present an update summarization technique based on query expansion with encyclopedic knowledge and activation spreading in a large document graph. Boudin et al. [3] revise the MMR criteria for update summarization. Li et al. [15] propose a novel graph-based sentence ranking algorithm named PNR² for update summarization, and it models both the positive and negative mutual reinforcement between sentences in the ranking process. Du et al. [8] propose a novel approach based on manifold ranking with sink points for update summarization. Wang and Li [2010] propose a new summarization method based on an incremental hierarchical clustering framework to update summaries when a new document arrives. In [36], the task is a little bit different from the above update summarization task, and it aims to address the

problem of incremental summarization of a periodically updated document set in a timely way.

As mentioned earlier, the update summarization task is closely related to our task. Our task can be considered as an extension of the update summarization task by requiring creating difference summaries for two document sets in a multilingual setting.

2.3 Contrastive Summarization

Though there is no common definition for contrastive summarization, the general aim of the task is to create a comparative summary for two comparable document sets. Most previous work focuses on generating comparative summaries for opinionated texts, because the opinionated texts usually contain a few explicit comparable aspects. For example, Kim and Zhai [12] formulate the problem of contrastive opinion summarization as an optimization problem and present two general methods for generating a comparative summary using the framework. Paul et al. [27] present a two-stage approach with an unsupervised probabilistic model and Comparative LexRank to summarizing multiple contrastive viewpoints in opinionated text. The most closely related work is [22, 35]. Mani and Bloedorn [22] propose an approach based on term graph to highlight the similarities and differences in related documents. They first find the common and distinct terms and then pinpoint the similarities and differences based on the terms and align text extracts across documents. Wang et al. [35] study the problem of summarizing the differences between document groups in a monolingual setting. They propose a discriminative sentence selection method to extract the most discriminative sentences from each document group. Other related work includes correlated summarization [39], which aims to summarize the correlation or similarity for a pair of news articles.

2.4 Multilingual and Cross-Lingual Summarization

Multilingual summarization [19, 29] aims to create a summary in a specific language from multiple documents in multiple languages, which is very different from our task of summarizing the differences in multilingual documents.

Cross-lingual document summarization [5, 14, 31] aims to produce a summary in a different target language for a set of documents in a source language, which is also very different from our task. One popular method first translates the source documents into the corresponding documents in the target language, and then extracts summary sentences based on the information on the target side. The other popular method first extracts summary sentences from the source documents based on the information on the source side, and then translates the summary sentences into the corresponding summary sentences in the target language.

3. OUR PROPOSED METHOD

The task of summarizing the differences in multilingual news is defined as follows: Given a set of Chinese news articles and a set of English news articles about the same news topic, the task aims to extract a few sentences from the Chinese news articles to reflect the important but differential aspects about the topic, and also extract a few sentences from the English news articles to reflect the important but differential aspects about the topic. The content of the Chinese summary is not the focus of the English news articles, and the content of the English summary is not the focus of the Chinese news articles.

In order to address the above summarization task, we present a co-ranking method (CoRank) and a constrained co-ranking method (C-CoRank) within the graph-based ranking framework. The CoRank method is inspired by the recent graph-based ranking methods for document summarization [15, 30, 33], and it can extract the difference summaries from the two document sets in a unified ranking process. The C-CoRank method is an improvement of the CoRank method by incorporating a new factor into the graph model, and adding important constraints in the ranking process.

3.1 CoRank

In this method, the Chinese sentences and the English sentences are simultaneously ranked in a unified graph-based algorithm. We assign each sentence a difference score to indicate how much the sentence contains important but differential information. The difference score of each Chinese sentence relies not only on the Chinese sentences in the Chinese document set, but also on the English sentences in the English document set. Similarly, the difference score of each English sentence relies not only on the English sentences in the English document set, but also on the Chinese sentences in the Chinese document set. In particular, the CoRank method is based on the following four assumptions:

Assumption 1: The difference score of a Chinese sentence would be high if the sentence is heavily correlated with other Chinese sentences with high difference scores in the Chinese document set.

Assumption 2: The difference score of a Chinese sentence would be high if the sentence is very unrelated to the English sentences with high difference scores in the English document set.

Assumption 3: The difference score of an English sentence would be high if the sentence is heavily correlated with other English sentences with high difference scores in the English document set.

Assumption 4: The difference score of an English sentence would be high if the sentence is very unrelated to the Chinese sentences with high difference scores in the Chinese document set.

Formally, given an English document set D^en and a Chinese document set D^{cn} , let $G=(V^{en}, V^{cn}, E^{en}, E^{cn}, E^{encn})$ be an undirected graph for the sentences in the two document sets. $V^{en}=\{s_i^{en} \mid 1 \leq i \leq m\}$ is the set of English sentences. $V^{cn}=\{s_i^{cn} \mid 1 \leq i \leq n\}$ is the set of Chinese sentences. m, n are the sentence numbers in the two document sets, respectively. Each sentence is represented by a term vector in the VSM model¹. The term weight is computed by TFIDF. E^{en} is the edge set to reflect the relationships between the English sentences in the English document set. E^{cn} is the edge set to reflect the relationships between the Chinese sentences in the Chinese document set. 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 matrices to reflect the three kinds of sentence relationships:

$M^{en}=(M_{ij}^{en})_{m \times m}$: This matrix aims to reflect the similarity relationships between the English sentences. Each entry in the matrix corresponds to the cosine similarity between the two English sentences.

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

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

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

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

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

$M^{encn}=(M_{ij}^{encn})_{m \times n}$: This matrix aims to reflect the difference relationships between the English sentences and the Chinese sentences. Each entry M_{ij}^{encn} in the matrix corresponds to the distance between the English sentence s_i^{en} and the Chinese sentence s_j^{cn} . It is hard to directly compute the distance between the sentences in different languages. In this study, the distance value is computed by fusing the following two distance values: the normalized Euclidean distance between the English sentence s_i^{en} and the translated English sentence $s_j^{transen}$ for s_j^{cn} , and the normalized Euclidean distance between the translated Chinese sentence $s_i^{transcn}$ for s_i^{en} and the Chinese sentence s_j^{cn} ². We use the geometric mean of the two values as the difference weight.

$$M_{ij}^{encn} = \sqrt{\text{Dis}_{NEuclidean}(s_i^{en}, s_j^{transen}) \times \text{Dis}_{NEuclidean}(s_i^{transcn}, s_j^{cn})}$$

The normalized Euclidean distance between two sentences (s_i and s_j) is computed by normalizing the Euclidean distance between the two term vectors (\vec{s}_i and \vec{s}_j) by dividing the sum of the norms of the vectors.

$$\text{Dis}_{NEuclidean}(s_i, s_j) = \frac{\|\vec{s}_i - \vec{s}_j\|}{\|\vec{s}_i\| + \|\vec{s}_j\|}$$

Then M^{encn} is normalized to \tilde{M}^{encn} to make the sum of each row equal to 1.

In addition, we use $M^{cnen}=(M_{ij}^{cnen})_{n \times m}$ to denote the transpose of M^{encn} , i.e., $M^{cnen}=(M^{encn})^T$. Then M^{cnen} is normalized to \tilde{M}^{cnen} to make the sum of each row equal to 1.

We use two column vectors $u=[u(s_i^{cn})]_{n \times 1}$ and $v=[v(s_i^{en})]_{m \times 1}$ to denote the difference scores of the Chinese sentences and the English sentences, respectively. The above four assumptions can be rendered as follows:

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

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

¹ For Chinese sentences, a term means a Chinese word after applying Chinese word segmentation.

² We use the Google Translate service for automatic Chinese-to-English and English-to-Chinese sentence translation. The normalized Euclidean performs better than the cosine-based distance in our study.

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

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

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

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

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

And the matrix form is:

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

$$\mathbf{v} = \alpha(\tilde{\mathbf{M}}^{en})^T \mathbf{v} + \beta(\tilde{\mathbf{M}}^{cnen})^T \mathbf{u}$$

where α and β specify the relative contributions to the final difference 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 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 falls below a given threshold. In order to guarantee the convergence of the iterative form, \mathbf{u} and \mathbf{v} are normalized after each iteration.

After we obtain the difference scores \mathbf{u} and \mathbf{v} for the Chinese and English sentences, we apply the same greedy algorithm [34] for redundancy removing in each language. Finally, a few highly ranked Chinese sentences are selected as the difference summary for the Chinese document set, and a few highly ranked English sentences are selected as the difference summary for the English document set.

3.2 C-CoRank

This method improves the CoRank method by introducing a new factor for each sentence and adding strict constraints in the ranking process. In addition to the difference score, we introduce a common score for each sentence to indicate how much a sentence contains important and common information in the two document sets. Similarly, the common scores of the sentences can be computed based on the following four assumptions:

Assumption 5: The common score of a Chinese sentence would be high if the sentence is heavily correlated with other Chinese sentences with high common scores in the Chinese document set.

Assumption 6: The common score of an English sentence would be high if the sentence is heavily correlated with other English sentences with high common scores in the English document set.

Assumption 7: The common score of a Chinese sentence would be high if the sentence is heavily correlated with the English sentences with high common scores in the English document set.

Assumption 8: The common score of an English sentence would be high if the sentence is heavily correlated with the Chinese sentences with high common scores in the Chinese document set.

In addition to the above assumptions, we assume that the difference score and the common score of each sentence is mutually exclusive.

Assumption 9: The sum of the difference score and the common score of each sentence is fixed to a particular value.

The above assumption is reasonable because if a sentence is a good sentence to be selected into the difference summary, it is not suitable to be selected into the common summary, and vice versa. This assumption can be used to adjust the inappropriately assigned difference scores for some sentences.

In order to compute the common scores for the sentences, we need to compute the following new matrices:

$\mathbf{W}^{encn} = (\mathbf{W}^{encn}_{ij})_{m \times n}$: This matrix aims to reflect the similarity relationships between the English sentences and the Chinese sentences. Each entry \mathbf{W}^{encn}_{ij} in the matrix corresponds to the similarity value between the English sentence s_i^{en} and the Chinese sentence s_j^{cn} . Similarly, the similarity value is computed by fusing the following two values: the cosine similarity value between the English sentence s_i^{en} and the translated English sentence $s_j^{transcn}$ for s_j^{cn} , and the cosine similarity value between the translated Chinese sentence $s_i^{transcn}$ for s_i^{en} and the Chinese sentence s_j^{cn} . We use the geometric mean of the two values as the affinity weight.

$$\mathbf{W}^{encn}_{ij} = \sqrt{\text{Sim}_{\cosine}(s_i^{en}, s_j^{transcn}) \times \text{Sim}_{\cosine}(s_i^{transcn}, s_j^{cn})}$$

Then \mathbf{W}^{encn} is normalized to $\tilde{\mathbf{W}}^{encn}$ to make the sum of each row equal to 1.

In addition, we use $\mathbf{W}^{cnen} = (\mathbf{W}^{cnen}_{ij})_{n \times m}$ to denote the transpose of \mathbf{W}^{encn} , i.e., $\mathbf{W}^{cnen} = (\mathbf{W}^{encn})^T$. Then \mathbf{W}^{cnen} is normalized to $\tilde{\mathbf{W}}^{cnen}$ to make the sum of each row equal to 1.

We use two column vectors $\mathbf{y} = [y(s_i^{cn})]_{n \times 1}$ and $\mathbf{z} = [z(s_j^{en})]_{m \times 1}$ to denote the common scores of the Chinese sentences and the English sentences, respectively. Based on the assumptions 5-8, we can similarly obtain the following iterative forms:

$$y(s_i^{cn}) = \alpha \sum_j \tilde{M}_{ji}^{cn} y(s_j^{cn}) + \beta \sum_j \tilde{W}_{ji}^{encn} z(s_j^{en})$$

$$z(s_j^{en}) = \alpha \sum_i \tilde{M}_{ij}^{en} z(s_i^{en}) + \beta \sum_i \tilde{W}_{ij}^{cnen} y(s_i^{cn})$$

And the matrix form is:

$$\mathbf{y} = \alpha(\tilde{\mathbf{M}}^{cn})^T \mathbf{y} + \beta(\tilde{\mathbf{W}}^{encn})^T \mathbf{z}$$

$$\mathbf{z} = \alpha(\tilde{\mathbf{M}}^{en})^T \mathbf{z} + \beta(\tilde{\mathbf{W}}^{cnen})^T \mathbf{y}$$

In order to simplify the parameter tuning, we use the same parameters of α and β as in the CoRank algorithm, and we have $\alpha + \beta = 1$.

Till now the common scores and the difference scores are computed by using the CoRank algorithm separately. Based on assumption 9, we can add the following constraints at each iteration:

$$u(s_i^{cn}) + y(s_i^{cn}) = \delta(s_i^{cn})$$

$$v(s_j^{en}) + z(s_j^{en}) = \delta(s_j^{en})$$

where $\delta(s_i^{cn})$ and $\delta(s_j^{en})$ are the specified constraint value for the sentences s_i^{cn} and s_j^{en} . It is obvious that the values for different sentences should be different, because not all sentences are equally important in the document sets. For example, the constraint value for a trivial sentence should be smaller than that

for a salient sentence. In this study, we use the generic saliency score of each sentence as the constraint value for the sentence. The generic saliency score of a Chinese sentence is computed by using the basic graph-based ranking algorithm based on \tilde{M}^{cn} in the Chinese document set. It can be formulated in a recursive form as in the PageRank algorithm:

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

where μ is the damping factor usually set to 0.85, as in the PageRank algorithm.

The generic saliency score of an English sentence is computed based on \tilde{M}^{en} in a similar way.

In order to add the constraints to the ranking process, all the difference and common scores are initialized to 1, and the following steps are iteratively performed:

1) Perform the following two equations to compute the difference scores of the sentences:

$$\begin{aligned} u^{(t+1)}(s_i^{cn}) &= \alpha \sum_j \tilde{M}_{ji}^{cn} u^{(t)}(s_j^{cn}) + \beta \sum_j \tilde{M}_{ji}^{encn} v^{(t)}(s_j^{en}) \\ v^{(t+1)}(s_j^{en}) &= \alpha \sum_i \tilde{M}_{ij}^{en} v^{(t)}(s_i^{en}) + \beta \sum_i \tilde{M}_{ij}^{cnen} u^{(t)}(s_i^{cn}) \\ \mathbf{u}^{(t+1)} &= \mathbf{u}^{(t+1)} / \|\mathbf{u}^{(t+1)}\| \quad \mathbf{v}^{(t+1)} = \mathbf{v}^{(t+1)} / \|\mathbf{v}^{(t+1)}\| \end{aligned}$$

2) Perform the following two equations to compute the common scores of the sentences:

$$\begin{aligned} y^{(t+1)}(s_i^{cn}) &= \alpha \sum_j \tilde{M}_{ji}^{cn} y^{(t)}(s_j^{cn}) + \beta \sum_j \tilde{W}_{ji}^{encn} z^{(t)}(s_j^{en}) \\ z^{(t+1)}(s_j^{en}) &= \alpha \sum_i \tilde{M}_{ij}^{en} z^{(t)}(s_i^{en}) + \beta \sum_i \tilde{W}_{ij}^{cnen} y^{(t)}(s_i^{cn}) \\ \mathbf{y}^{(t+1)} &= \mathbf{y}^{(t+1)} / \|\mathbf{y}^{(t+1)}\| \quad \mathbf{z}^{(t+1)} = \mathbf{z}^{(t+1)} / \|\mathbf{z}^{(t+1)}\| \end{aligned}$$

3) Normalize the difference score and the common score of each sentence to meet the constraints (η, ξ are temporary vectors):

$$\begin{aligned} \eta^{(t+1)}(s_i^{cn}) &= u^{(t+1)}(s_i^{cn}) + y^{(t+1)}(s_i^{cn}) \\ \xi^{(t+1)}(s_j^{en}) &= v^{(t+1)}(s_j^{en}) + z^{(t+1)}(s_j^{en}) \\ u^{(t+1)}(s_i^{cn}) &= \delta(s_i^{cn}) \cdot \frac{u^{(t+1)}(s_i^{cn})}{\eta^{(t+1)}(s_i^{cn})} \\ y^{(t+1)}(s_i^{cn}) &= \delta(s_i^{cn}) \cdot \frac{y^{(t+1)}(s_i^{cn})}{\eta^{(t+1)}(s_i^{cn})} \\ v^{(t+1)}(s_j^{en}) &= \delta(s_j^{en}) \cdot \frac{v^{(t+1)}(s_j^{en})}{\xi^{(t+1)}(s_j^{en})} \\ z^{(t+1)}(s_j^{en}) &= \delta(s_j^{en}) \cdot \frac{z^{(t+1)}(s_j^{en})}{\xi^{(t+1)}(s_j^{en})} \end{aligned}$$

Note that the difference scores and the common scores are linearly normalized in the third step. For numerical computation of the scores, we can iteratively run the above steps until convergence or the iteration count reaches a preset number.

Finally, we obtain the difference scores \mathbf{u} and \mathbf{v} for the Chinese and English sentences, and we apply the same greedy algorithm [34] for redundancy removing in each language. The difference summaries in the two languages are produced by selecting the highly ranked sentences, respectively.

4. EVALUATION SETUP

4.1 Data Set

There is no benchmark dataset for difference summarization of multilingual news, so we built the evaluation dataset as follows:

We first selected 15 hot news topics (e.g. ‘‘G20 summit in 2010’’, ‘‘North Korea’s attack on South Korea’’, ‘‘Liu Xiaobo’s Nobel Prize’’, etc.), and then collected a few news articles about each topic in the Chinese and English languages by using both the Chinese and English versions of *Google News Search*³. For each topic, we selected the articles in the two languages within the same period. We collected the news articles published by major Chinese and Western news agencies. The statistics of the corpus are summarized in Table 1.

Table 1: Corpus statistics

	Average article number per topic	Average length per article
Chinese	36.3	498 words 885 characters
English	28.3	606 words

The articles were first split into sentences. The Chinese sentences were automatically translated into English sentences by using *Google Translate*⁴, and the English sentences were automatically translated into Chinese sentences by using *Google Translate*. All the Chinese sentences (including the translated Chinese sentences) were segmented into words by using the Stanford Chinese Word Segmenter⁵.

Two graduate students were employed to manually label the reference summaries for the 15 topics. Each student was asked to read both the Chinese and English documents for each topic, and then selected five Chinese sentences and five English sentences as reference difference summaries for the two document sets. The labeling process lasted two weeks. Finally, there were totally 30 Chinese reference summaries and 30 English reference summaries. The Chinese reference summaries were segmented into words by using the same word segmentation tool.

³ The English Google News is <http://news.google.com/>

The Chinese Google News is <http://news.google.com.hk/>

⁴ <http://translate.google.com/>

⁵ <http://nlp.stanford.edu/software/segmenter.shtml>

4.2 Evaluation Metric

In the experiments, the summary length was fixed to five sentences in both languages. We used the ROUGE-1.5.5 [18] toolkit for comparing the system summaries with the reference summaries, 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 the most popular two ROUGE F-measure scores in the experimental results: ROUGE-2 (bigram-based) and ROUGE-SU4 (based on skip bigram with a maximum skip distance of 4). For Chinese summaries, we reported two types of ROUGE scores: character-based and word-based. Character-based ROUGE scores were computed based on Chinese characters, and word-based ROUGE scores were computed based on Chinese words after using word segmentation. Character-based evaluation will not be influenced by the word segmentation performance, but Chinese character is less meaningful than Chinese word.

5. EVALUATION RESULTS

The CoRank and C-CoRank methods are compared with the following three baselines:

Centroid: This baseline uses the centroid-based method [28] to compute the saliency scores of the sentences in each language separately. The saliency score of each sentence is computed by linearly combining the following three features: the cosine similarity between the sentence and the whole document set, the cosine similarity between the sentence and the first sentence within the same document, and the position-based weight. The highly ranked sentences are selected into the difference summary after removing redundancy. Note that this baseline does not make any use of the cross-lingual information between the two document sets.

Centroid++: This baseline is an improvement of the centroid-based method by integrating the feature of the similarity between each sentence and the sentences in the other language. The difference score of each sentence is the subtraction of the centroid-based weight and the cross-language similarity value.

MMR++: This baseline uses the update summarization method in [3]. It revises the MMR criteria [4] by incorporating the cross-language sentence similarity.

Table 2 shows the comparison results for Chinese difference summaries, and Table 3 shows the comparison results for English difference summaries. The parameter value of α for the CoRank and C-CoRank methods is simply set to 0.5 without tuning. We can see that both the CoRank and C-CoRank methods can significantly outperform the baselines methods over all metrics in both languages. The C-CoRank method can significantly outperform the CoRank method over most metrics. The results demonstrate the good effectiveness of the proposed C-CoRank method. The reason is that the introduction of the new factor of common score and the constraints can make the computation of the difference scores more reliable.

Table 2. Comparison results (F-measure) for Chinese difference summaries

	Word-Based		Character-Based	
	ROUGE-2	ROUGE-SU4	ROUGE-2	ROUGE-SU4
Centroid	0.01589	0.02676	0.08808	0.08372
Centroid++	0.01393	0.02943	0.07990	0.07823
MMR++	0.00959	0.02215	0.06364	0.07604
CoRank	0.03433*	0.04564*	0.17894*	0.17954*
C-CoRank	0.04484*#	0.05314*#	0.18794*#	0.18578*

Table 3. Comparison results (F-measure) for English difference summaries

	ROUGE-2	ROUGE-SU4
Centroid	0.02568	0.04368
Centroid++	0.03079	0.04828
MMR++	0.01759	0.05342
CoRank	0.11505*	0.15421*
C-CoRank	0.13193*#	0.16712*#

(* indicates that the improvement over the three baselines are all statistically significant by using t-test. # indicates that the improvement over the CoRank method is statistically significant by using t-test.)

In order to show the influence of the parameter α on the performance of our proposed method, we present the performance curves of the CoRank and C-CoRank method over different metrics when α ranges from 0.1 to 0.9. Figures 1-4 show the performance curves for Chinese difference summaries. Figures 5-6 show the performance curves for English difference summaries.

Seen from the figures, the performance values of the CoRank method first increase with α , and then reach the peak values when α is set to a large value, and after that, the performance values almost do not change any more. The results demonstrate that the CoRank method relies more on the monolingual sentence links and the cross-language sentence links do not contribute much as expected. By analysis of the results, the CoRank method has the potential to select some sentences containing much common information into the summaries, which is not very appropriate for difference summarization.

However, the performance values of the C-CoRank method first increase with α , and then decrease when α is set to a large value. The C-CoRank method can well incorporate both the monolingual information and cross-language information for the summarization task. Overall speaking, the C-CoRank method can outperform the CoRank method over almost all the metrics when α is set within a wide range. The results further demonstrate the robustness of the C-CoRank method. The use of the constraints in the C-CoRank method makes the difference score and the common score for each sentence mutually exclusive, and thus the difference score for a sentence containing much common information will not be high, when the common score for the sentence is high.

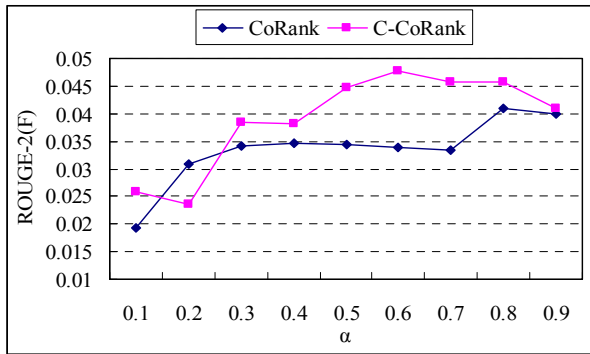


Fig.1. Word-based ROUGE-2(F-measure) vs. α for Chinese difference summaries

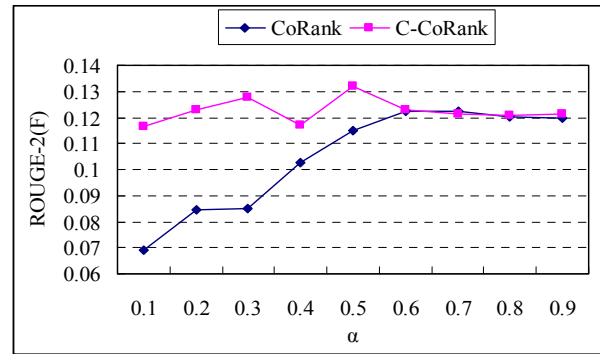


Fig.5. ROUGE-2(F-measure) vs. α for English difference summaries

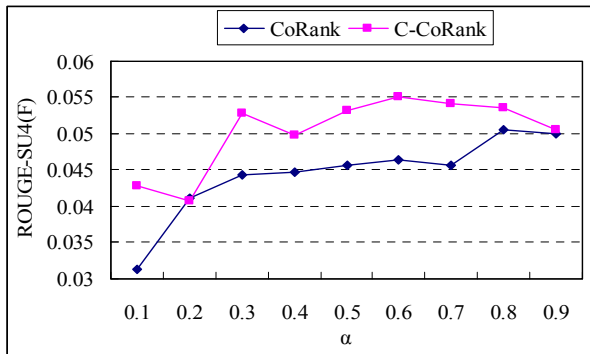


Fig.2. Word-based ROUGE-SU4(F-measure) vs. α for Chinese difference summaries

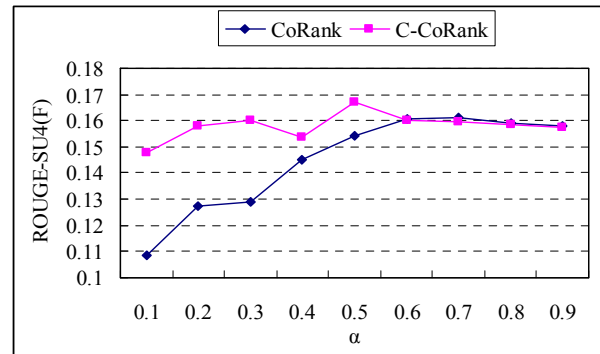


Fig.6. ROUGE-SU4(F-measure) vs. α for English difference summaries

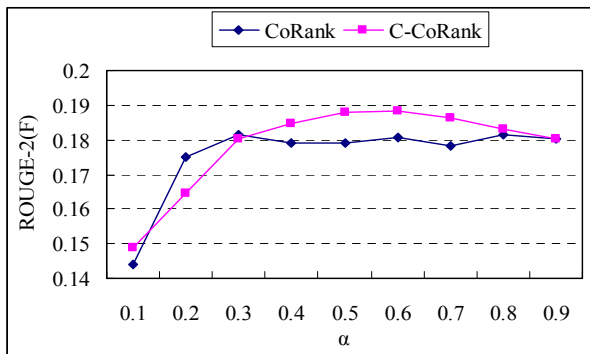


Fig.3. Character-based ROUGE-2(F-measure) vs. α for Chinese difference summaries

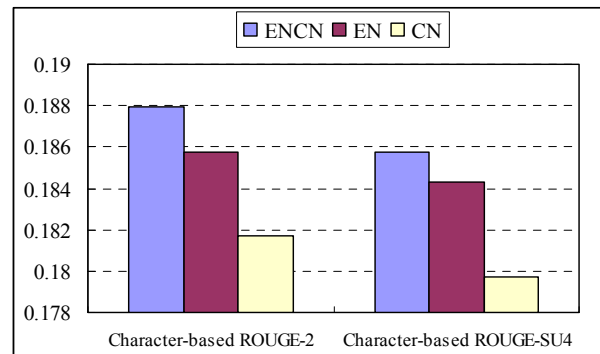


Fig.7. C-CoRank vs. Different cross-language sentence distance/similarity weights for Chinese difference summaries

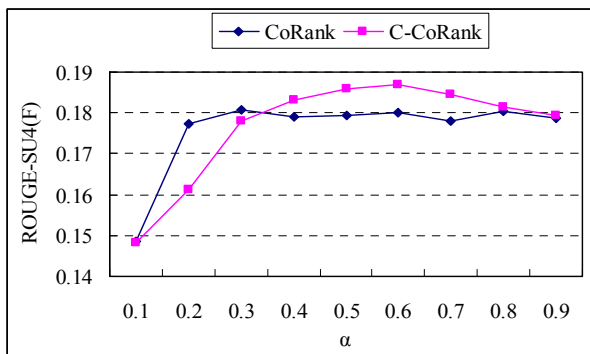


Fig.4. Character-based ROUGE-SU4(F-measure) vs. α for Chinese difference summaries

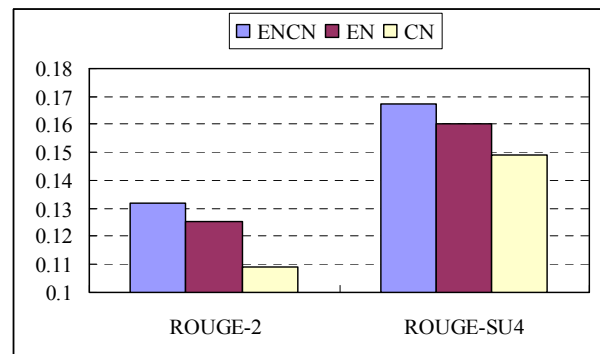


Fig.8. C-CoRank vs. Different cross-language sentence distance/similarity weights for English difference summaries

In the above experiments, the cross-language sentence distance weight in matrix M^{encn} and the cross-language sentence similarity weight in matrix W^{encn} in the C-CoRank method are computed based on the geometric mean of the Chinese-side value and the English-side value. In order to investigate whether the two values really contribute to the final performance, we compare the original weight computation method (ENCN) with the following two methods using only one-side value:

The first method (EN) uses only the English-side value, and the weights in the two matrices are computed as follows:

$$M_{ij}^{encn} = Dis_{NEuclidean}(s_i^{en}, s_j^{transcn}) \quad W_{ij}^{encn} = Sim_{cosine}(s_i^{en}, s_j^{transcn})$$

The second method (CN) uses only the Chinese-side value, and the weights in the two matrices are computed as follows:

$$M_{ij}^{encn} = Dis_{NEuclidean}(s_i^{transcn}, s_j^{cn}) \quad W_{ij}^{encn} = Sim_{cosine}(s_i^{transcn}, s_j^{cn})$$

The comparison results for the Chinese summaries and the English summaries are shown in Figures 7 and 8, respectively. We can see that the original weight computation method (ENCN) can outperform the two methods (EN and CN) using only one-side value, and both the Chinese-side information and the English-side information contribute to the final performance. We also observe that the English-side value is more reliable than the Chinese-side value for computing the cross-language sentence similarity or distance, and the reason lies in that the step of Chinese word segmentation may introduce some noise for the distance/similarity computation.

Finally, we show one running example for the topic of “Liu Xiaobo's Nobel Prize”. The difference summaries produced by the C-CoRank method are given below. The manual English translation of the Chinese summary is also given. We can see that the Chinese summary focuses on the opposition and accuse of Liu Xiaobo's Nobel Prize, while the English summary focuses on the support of Liu Xiaobo's Nobel Prize and China's arrest of Liu Xiaobo. The difference summaries reflect the different standpoints of the Chinese news agencies and the Western news agencies.

The Chinese difference summary is as follows:

诺贝尔和平奖是西方给刘晓波的政治“犒赏”。挪威诺贝尔委员会将今年诺贝尔和平奖授予刘晓波，再次暴露了西方国家的双重标准。诺贝尔委员会 8 日把今年的诺贝尔和平奖授予中国“异见人士”刘晓波。迄今诺贝尔和平奖授予过两个中国人，一个是达赖，另一个是刘晓波。挪威诺委会这个出于意识形态偏见和政治需要的决定，将诺贝尔和平奖沦为西方一些势力的政治工具，严重损害了诺贝尔和平奖的公信力，也玷污了诺贝尔先生的荣誉。

(Nobel Peace Prize is a political “reward” to Liu Xiaobo by the West. Norwegian Nobel Committee awarded the Nobel Peace Prize to Liu Xiaobo in this year, which once again exposed the double standards of Western countries. Nobel Committee awarded this year's Nobel Peace Prize to the Chinese “dissident” - Liu Xiaobo on day 8. To date, the Nobel Peace Prize has been awarded to two Chinese persons: one is the Dalai Lama and the other is Liu Xiaobo. Norwegian Nobel Committee made the decision for ideological bias and political needs, and the Nobel Peace Prize will become a political tool of some Western forces, and it seriously undermined the credibility of the Nobel Peace Prize, but also tarnished the honor of Mr. Nobel.)

The English difference summary is as follows:

As the West applauds Liu Xiaobo's Nobel, China sees another attempt to impose Western values on it. Norway's Nobel Peace Prize committee has done the right thing in awarding this year's prize to Chinese dissident Liu Xiaobo. Liu was represented at Friday's Nobel ceremony by an empty chair because China would not release him from prison - only the fifth time in the 109-year history of the prize that the winner was not in attendance. It tried to bully the Nobel committee into not awarding Mr. Liu this year's Nobel Peace Prize. China has prohibited Liu and his family members from leaving China to attend Friday's ceremony in Oslo. Two days after the prize was announced, Mr Liu's wife, Liu Xia, met with him at the prison in northeastern China where he is serving his sentence, but she was escorted back to Beijing and placed under house arrest, a human rights group said.

6. CONCLUSION AND FUTURE WORK

In this paper we address a novel summarization problem of summarizing the differences in multilingual news. We present the CoRank and the C-CoRank methods. In addition to the difference score of each sentence, the C-CoRank method incorporates the common score of each sentence into the ranking process by adding strict constraints. Evaluation results on the manually labeled test set demonstrate the effectiveness and robustness of the C-CoRank method.

In future work, we will adapt the proposed C-CoRank method for the update summarization task to further investigate its robustness. We will also summarize the differences between the news articles written by different news agencies in a monolingual setting. In addition to two languages, we will extend the summarization framework to summarize the differences in three or more languages.

7. ACKNOWLEDGMENTS

This work was supported by NSFC (60873155), Beijing Nova Program (2008B03) and NCET (NCET-08-0006). We thank the anonymous reviewers for their helpful comments.

8. REFERENCES

- [1] A. Aker, T. Cohn, and R. Gaizauskas. Multi-document summarization using A* search and discriminative training. In Proceedings of EMNLP2010.
- [2] M. R. Amini, P. Gallinari. The Use of Unlabeled Data to Improve Supervised Learning for Text Summarization. In Proceedings of SIGIR2002.
- [3] F. Boudin, M. El-Bèze, J.-M. Torres-Moreno. The LIA update summarization systems at TAC-2008. In Proceedings of TAC2008.
- [4] J. Carbonell and J. Goldstein. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In Proceedings of SIGIR1998.
- [5] G. de Chalendar, R. Besançon, O. Ferret, G. Grefenstette, and O. Mesnard. 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).

- [6] A. Celikyilmaz and D. Hakkani-Tur. A hybrid hierarchical model for multi-document summarization. In Proceedings of ACL2010.
- [7] H. T. Dang and K. Owczarzak. Overview of the TAC 2008 update summarization task. In Proceedings of TAC2008.
- [8] P. Du, J. Guo, J. Zhang, X. Cheng. Manifold ranking with sink points for update summarization. In Proceedings of CIKM2010.
- [9] G. ErKan, D. R. Radev. LexPageRank. Prestige in Multi-Document Text Summarization. In Proceedings of EMNLP2004.
- [10] S. Fisher and B. Roark. Query-focused supervised sentence ranking for update summaries. In Proceeding of TAC2008.
- [11] S. Harabagiu and F. Lacatusu. Topic themes for multi-document summarization. In Proceedings of SIGIR2005.
- [12] H. D. Kim and C. Zhai. Generating comparative summaries of contradictory opinions in text. In Proceedings of CIKM2009.
- [13] J. Kupiec, J. Pedersen, F. Chen. A Trainable Document Summarizer. In Proceedings of SIGIR1995.
- [14] A. Leuski, C.-Y. Lin, L. Zhou, U. Germann, F. J. Och, E. Hovy. Cross-lingual C*ST*RD: English access to Hindi information. *ACM Transactions on Asian Language Information Processing*, 2(3): 245-269, 2003.
- [15] W. Li, F. Wei Q. Lu and Y. He. PNR²: Ranking sentences with positive and negative reinforcement for query-oriented update summarization. In Proceedings of COLING2008.
- [16] C. Y. Lin, E. Hovy. The Automated Acquisition of Topic Signatures for Text Summarization. In Proceedings of COLING2000.
- [17] C.-Y. Lin and E. Hovy. From Single to Multi-document Summarization: A Prototype System and its Evaluation. In Proceedings of ACL2002.
- [18] C.-Y. Lin and E.H. Hovy. Automatic Evaluation of Summaries Using N-gram Co-occurrence Statistics. In Proceedings of HLT-NAACL -2003.
- [19] C.-Y. Lin, L. Zhou, and E. Hovy. Multilingual summarization evaluation 2005: automatic evaluation report. In Proceedings of MSE (ACL2005 Workshop).
- [20] M. Litvak, M. Last, and M. Friedman. A new approach to improving multilingual summarization using a genetic algorithm. In Proceedings of ACL2010.
- [21] H. P. Luhn. The Automatic Creation of literature Abstracts. *IBM Journal of Research and Development*, 2(2), 1969.
- [22] I. Mani and E. Bloedorn. Summarizing similarities and differences among related documents. *Information Retrieval*, 1: 35-67, 1999.
- [23] R. Mihalcea, P. Tarau. TextRank: Bringing Order into Texts. In Proceedings of EMNLP2004.
- [24] R. Mihalcea and P. Tarau. A language independent algorithm for single and multiple document summarization. In Proceedings of IJCNLP-2005.
- [25] V. Nastase, K. Filippova, S. P. Ponzetto. Generating update summaries with spreading activation. In Proceedings of TAC2008.
- [26] A. Nenkova and A. Louis. Can you summarize this? Identifying correlates of input difficulty for generic multi-document summarization. In Proceedings of ACL-2008:HLT.
- [27] M. J. Paul, C. Zhai, and R. Girju. Summarizing contrastive viewpoints in opinionated text. In Proceedings of EMNLP2010.
- [28] D. R. Radev, H. Y. Jing, M. Stys and D. Tam. Centroid-based summarization of multiple documents. *Information Processing and Management*, 40: 919-938, 2004.
- [29] A. Siddharthan and K. McKeown. Improving multilingual summarization: using redundancy in the input to correct MT errors. In Proceedings of HLT/EMNLP-2005.
- [30] X. Wan. Towards a unified approach to simultaneous single-document and multi-document summarizations. In Proceedings of COLING2010.
- [31] X. Wan, H. Li and J. Xiao. Cross-language document summarization based on machine translation quality prediction. In Proceedings of ACL2010.
- [32] X. Wan and J. Yang. Multi-document summarization using cluster-based link analysis. In Proceedings of SIGIR-2008.
- [33] X. Wan, J. Yang and J. Xiao. Towards an Iterative Reinforcement Approach for Simultaneous Document Summarization and Keyword Extraction. In Proceedings of ACL2007.
- [34] X. Wan, J. Yang and J. Xiao. Manifold-ranking based topic-focused multi-document summarization. In Proceedings of IJCAI-2007.
- [35] D. Wang, S. Zhu, T. Li, and Y. Gong. Comparative document summarization via discriminative sentence selection. In Proceedings of CIKM2009.
- [36] D. Wang, T. Li. Document update summarization using incremental hierarchical clustering. In Proceedings of CIKM2010.
- [37] K.-F. Wong, M. Wu and W. Li. Extractive summarization using supervised and semi-supervised learning. In Proceedings of COLING-2008.
- [38] H. Y. Zha. Generic Summarization and Keyphrase Extraction Using Mutual Reinforcement Principle and Sentence Clustering. In Proceedings of SIGIR2002.
- [39] Y. Zhang, X. Ji, C.-H. Chu, and H. Zha. Correlating summarization of multi-source news with K-way graph bi-clustering. *SIGKDD Explorations*, 6(2), 2004.