GuideRank: A Guided Ranking Graph Model for Multilingual Multi-document Summarization

Haoran Li, Jiajun Zhang, Yu Zhou and Chengqing Zong
National Laboratory of Pattern Recognition, Institute of Automation Chinese Academy of Sciences, Beijing, China
University of Chinese Academy of Sciences, Beijing, China
{haoran.li, jjzhang, yzhou, cqzong}@nlpr.ia.ac.cn

Abstract. Multilingual multi-document summarization is a task to generate the summary in target language from a collection of documents in multiple source languages. A straightforward approach to this task is automatically translating the non-target language documents into target language and then applying monolingual summarization methods, but the summaries generated by this method is often poorly readable due to the low quality of machine translation. To solve this problem, we propose a novel graph model based on guided edge weighting method in which both informativeness and readability of summaries are taken into consideration fully. In methodology, our model attempts to choose from the target language documents the sentences which contain important shared information across languages, and also retains the salient sentences which cannot be covered by documents in other language. The experimental results on our manually labeled dataset [1] show that our method significantly outperforms other baseline methods.

1 Introduction

The explosion of multilingual news in the Internet provides users with the opportunity to capture richer information about a specific topic but also increases the difficulty to focus on the important information. Multilingual multi-document summarization aims to provide users with the summary in their own language from multilingual documents of the same topic, which will help users to obtain clear and brief information in a short time.

In this work, English and Chinese documents are considered as the input, and we perform extractive summarization experiments on two tasks: one is generating English summaries and the other is generating Chinese summaries. As the models for these two tasks are the same, we just introduce the model producing English summaries from English and Chinese documents.

A simple approach to this task is first translating the Chinese documents into English by machine translation (MT) and then regarding it as a general monolingual multi-document summarization task. However, as MT is still far from being perfect, translation errors are propagated to the summarization task and

[1] It will be released to the public.

can lead to less readable summaries. While in fact, the information of translated documents is also necessary.

This paper proposes a guided edge weighting graph model for multilingual multi-document summarization (GuideRank). An important component of our method is edge weights which can be learned. Through controlling the weights flow, we can guide the system to choose from target language documents more sentences which contain important shared information of documents in both languages, without ignoring the translated sentences which cannot be covered by target language documents. Our model is mostly inspired by CoRank model [16] which was proposed for cross-lingual summarization and the edges in CoRank model are equal in both the directions. Different from CoRank model, in our model, the cross-lingual edges connecting the related sentences in different languages are unidirection which invalidate the direction from original English sentences to translated sentences. In this way, the translated English sentences will contribute the weights to their related original English sentences but the opposite is not the case. This transformation brings two advantages: one is that the sentences in original English documents which contain the shared information expressed in both languages will tend to be chosen as the summary, the other is that the important translated sentences which cannot be covered by the original English documents also have the opportunity to appear in the summary. Note that the original English sentences sharing little information with Chinese sentences are not affected in our model. We also employ different measures to re-weight these cross-lingual edges in different languages.

We use Figure 1 to illuminate our GuideRank model furtherly. Sentence S_1 is extracted from English documents and the sentence T_1 from Chinese documents. S_1 and T_1 express the information about the plane crash site, but the quality of T_1_mt, the machine translation version of T_1, is far from satisfactory. We tend to extract original English sentence S_1 rather than translated sentence T_1_mt considering the readability of the summary. Our GuideRank model attempts to achieve this goal through modifying the direction of the weight propagation in the process of random walk of graph model.

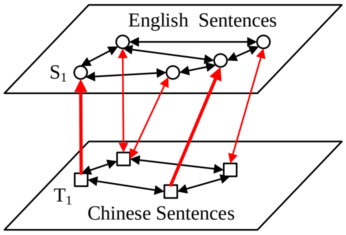
We make the following contributions:

- We propose GuideRank model which can generate the target language summaries with more target language sentences which contain shared information across language to enhance the readability.
- We employ several approaches to measure relevance between sentences across language and to re-weight the edges connecting cross-lingual sentences.
- The experiments show that we can outperform baselines on our dataset.

2 Related Work

2.1 Multilingual multi-document summarization

The Multilingual Summarization Evaluation (MSE) 2005 and 2006 aimed to create a 100-word English summary on documents consisting of English and Arabic



S_1: The plane crashed on to the Syria side of the Turkish-Syrian border.

T_1: 俄罗斯一架苏-24战机24日在土耳其和叙利亚边境叙利亚一侧坠毁。

T_1_mt: A Russian Su-24 fighter crashed 24 side of the Syrian border in Turkey and Syria.

Fig. 1. The simplified illustration of our GuideRank model. The vertices denote sentences and the edges reflect the relationships between sentences. The thickness of the edges connecting two parts indicates the strength of the relationships in which the strong connections are converted to unidirection.

news. Many researchers [3, 13, 22, 1, 20] participated the evaluation and they regarded the task as a general summarization from original English documents along with English documents translated from Arabic. Daumé III and Marcu [3] achieved the first place in MSE 2005 and they got the better performance when never extracting sentences from the Arabic MT documents. Although MT sentences are often largely incomprehensible, they failed to access effective means to take advantage of the Arabic documents which can provide useful information beyond English documents. Our GuideRank model can make full use of the information in both language documents.

The Text Analysis Conference (TAC) 2011 MultiLing [6] posed a multilingual summarization task which aimed to generate a summary from a set of documents in seven languages. The MultiLing task required language-independent summarization methods that the language of output summary is the same as input documents, which is different from our task. Cross-lingual document summarization [17, 16, 21] aims to produce a summary in a different target language for a set of documents in a source language, which is also different from our task.

2.2 Graph-based Extractive Summarization Models

Graph-based methods [12, 11, 18, 4, 16, 2] have been widely used to rank sentences for general document summarization. Documents are represented as a graph and sentences are represented as nodes. The edges reflect relations between nodes. The importance of the sentences are decided through random walk. Graph-based methods have the advantage in that they do not require training data and can be easily adapted to any languages, which is suitable for our task.

CoRank [16] is a graph model which is proposed to address cross-language summarization, in which the different language sentences are ranked simultaneously using a unified graph-based algorithm.

3 Methods

3.1 CoRank Model

The CoRank algorithm first needs to calculate three similarity matrices: M^{en} which denotes affinity matrix between the original English sentences, M^{c2e} which denotes affinity matrix between the translated English sentences from Chinese and $M^{en\text{-}c2e}$ which denotes affinity matrix between the original and the translated English sentences. The similarity matrices are computed as follows:

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

where s_i^{en} denotes English sentence which can be represented by TF-IDF vectors or averaging the embeddings of words (except stop-words) contained in the sentence. $sim(\cdot)$ denotes the similarity between two sentences, which is calculated with cosine measure. M^{c2e} and $M^{en\text{-}c2e}$ are computed in the same way. Note that these matrices are normalized to make the sum of each row equal to 1.

Next, the salience scores for the original and the translated English sentences, which are denoted by $u(s_i^{en})$ and $v(s_j^{c2e})$, are calculated iteratively until convergence using the following equations:

$$u(s_j^{en}) = \alpha \sum_i M_{ij}^{en} u(s_i^{en}) + (1-\alpha) \sum_i M_{ij}^{en\text{-}c2e} v(s_i^{c2e})$$
$$v(s_i^{c2e}) = \alpha \sum_j M_{ji}^{c2e} v(s_j^{c2e}) + (1-\alpha) \sum_j M_{ji}^{en\text{-}c2e} u(s_j^{en})$$

Finally, re-ranking is employed to remove the redundant information in the summary as Wan et al. [19] did and summary is generated by the sentences with highest scores.

3.2 GuideRank Model

The affinity matrices in CoRank model are symmetric, while for the task of multilingual document summarization, in consideration of the unsatisfactory quality of the translated English sentences, the symmetric affinity matrices is inappropriate. Specifically, for a translated English sentence, if there are some original English sentences which are related to it, we would prefer to choosing these original English sentences instead of the translated English sentence. In other words, the summarization system should be guided to control the direction of sentence salience score updating: when an original sentence is related to a translated sentence, the symmetric weighted edge between them should be transformed into unidirection in which we invalidate the direction from original sentences to translated sentences.

We use $M_{ij}^{en\text{-}c2e}$ to represent the weight pointing from the original English sentences to the translated English sentences, and use $M_{ij}^{c2e\text{-}en}$ to represent the weight pointing from the translated English sentences to the original English sentences. The similarity matrices representing relations between sentences across languages are changed in our GuideRank model as follows:

$$M_{ij}^{en\text{-}c2e} = \begin{cases} 0, & \text{if } \mathrm{s_i^{en}} \text{ is related to } \mathrm{s_j^{c2e}} \\ sim(s_i^{en}, s_j^{c2e}), & \text{otherwise} \end{cases}$$

$$M_{ij}^{c2e\text{-}en} = \begin{cases} relevance(s_i^{c2e}, s_j^{en}), & \text{if } \mathrm{s_i^{c2e}} \text{ is related to } \mathrm{s_j^{en}} \\ sim(s_i^{c2e}, s_j^{en}), & \text{otherwise} \end{cases}$$

where $relevance(\cdot)$ denotes the semantic relevance between two sentences from different languages. The motivation of our GuideRank model is that if there are some English sentences which are related to a Chinese sentence, we should guide the weight of this Chinese sentence to be transformed to its corresponding English sentences. Towards this objective, a requirement is to identify whether s_i^{en} is related to s_j^{c2e} and how to measure the semantic relevance between sentences across the languages. We propose the following three methods to achieve this goal.

Similarity (Sim) Evaluation This method is a simple decision mechanism where cosine similarity of two sentence is leveraged. We propose three heuristic approaches to search for the related s_i^{en} for s_j^{c2e} using similarity evaluation.

The Maximum Similarity (SimMax).

$$s_i^{en} = \operatorname*{argmax}_{s_i^{en^*}} sim(s_i^{en^*}, s_j^{c2e})$$

The Top-five Similarity (SimTop5).

$$s_i^{en} \in \left\{ s_i^{en^*} \,|\, \operatorname*{arg\,top5}_{s_i^{en^*}} sim(s_i^{en^*}, s_j^{c2e}) \right\}$$

where top5 denotes five highest values.

Higher than the Average Similarity (SimAve).

$$s_i^{en} \in \left\{ s_i^{en^*} | sim(s_i^{en^*}, s_j^{c2e}) > \frac{\sum_k sim(s_k^{en}, s_i^{c2e})}{N} \right\}$$

where s_k^{en} denotes the original English sentences and N is the total number of them.

$relevance(\cdot)$ in this method is equal to $sim(\cdot)$ which is introduced in Section 3.1.

Textual Entailment (TE) Evaluation This method regards identification of semantic relevance as recognizing textual entailment (RTE) task where entailment and non-entailment relations are seen as judgments about semantic relevance.

RTE is a task to recognize, given two text fragments, whether one can be inferred by the other. For the following text-hypothesis pair:

Text: ... Obasanjo invited him to step down as president ... and accept political asylum in Nigeria.

Hypothesis: Charles G. Taylor was offered asylum in Nigeria.

After reading **T** we can infer that **H** is true, which means **T** entails **H**.

We use BiuTee [14], a transformation-based TE system using various types of knowledge resources, to determine textual entailment. We train BiuTee with 800 entailment or non-entailment text-hypothesis pairs of the RTE-3 [5] dataset for our task, and the possible inputs to BiuTee are pairs of sentences consisting of any two sentences in which one is extracted from original English documents and the other from translated English documents. Since the size of the inputs is very large, in order to keep the whole summarization system efficient, we eliminate the sentence pairs which have no token (except stop-words) overlap. For the remaining pairs, the longer sentence is regarded as **T** and the other as **H**.

relevance$(\cdot)$ in this method is represented by textual entailment score, obtained by BiuTee, between pairs of sentences determined as TE relation.

Translation (Trans) Evaluation This method regards identification of semantic relevance as a translation evaluation where the probability of fully or partially translating a sentence into the other is seen as judgments about semantic relevance.

The translation probability is obtained based on the word alignment model. We use TsinghuaAligner [2] to perform word alignment. First, we train word alignment model on a large English-Chinese parallel corpus A which consists of two million sentence pairs. Then, another corpus B consisting of fully or partially translation English-Chinese pair run the word alignment model. Next, several features based on the word alignment are extracted for sentence pairs in B. Taking the features as input, an SVM classifier for determining translation relations is trained to fit the data B. The last step is to use the classifier to predict the translation probability for the candidate sentence pairs in summarization dataset in which the English sentence is from original English documents and the Chinese sentence is from original Chinese documents. The candidate sentence pairs are also obtained by the approach introduce in TE evaluation.

A preprocessing step to this method is to build the dataset to train the model detecting fully or partially translation English-Chinese pair. We construct the dataset in a straightforward way as follows:

We use an English-Chinese parallel corpus in FBIS corpus, which contains around 236 thousand English and Chinese sentence pairs as primary data. They come from the domain of news which is same as our summarization task. Then, we parse all the English and Chinese sentences using Stanford parser [9, 7], and last, we randomly remove one of the verb phrases in the sentences except the following conditions:

[2] http://nlp.csai.tsinghua.edu.cn/~ly/systems/TsinghuaAligner/TsinghuaAligner.html

(1) There will be no verb phrases in the sentence after removing the selected verb phrase;

(2) The length of the verb phrase is 1;

(3) The length of the verb phrase is longer than the half of sentence.

The constraints described above are expected to guarantee the removing operations effective and keep the generated sub-sentences meaningful. Note that both of the sentences in a English-Chinese parallel sentence pair have the chances to randomly remove a verb phrase or keep unchanged. We generate corrupted sentence pairs by random sampling.

To train the model for detecting fully or partially translation relation between the sentence pair (S_i, S_j), we extract 6 features based on proportions of aligned unigram and bigram as follows:

$$maxang = \max\left\{\frac{ang(S_i)}{L(S_i)}, \frac{ang(S_j)}{L(S_j)}\right\}$$

$$minang = \min\left\{\frac{ang(S_i)}{L(S_i)}, \frac{ang(S_j)}{L(S_j)}\right\}$$

$$aveang = \frac{ang(S_i) + ang(S_j)}{L(S_i) + L(S_j)}$$

where $ang(S_i)$ and $L(S_i)$ denotes the number of the aligned n-gram (n=1, 2) and the number of the words (except stop-words) in the sentence S_i, respectively.

We evaluate the effectiveness of these features on our fully or partially translation English-Chinese pair dataset through the 10-fold cross-validation, and find that the F-scores using unigram and bigram features are 69% and 78%, respectively. When we combine unigram and bigram features, the F-score reaches up to 96%, which proves the robustness of these simple features to detect whether one text segment can be fully or partially translated by the other.

$relevance(\cdot)$ in this method is represented by the estimated probability of mutual translation, obtained by the SVM classifier, between pair of sentences determined as translation relation. Note that the *relevance* scores are normalized to make the sum for each sentence equal to 1.

4 Experiment

4.1 Dataset

There is no benchmark dataset for multilingual multi-document summarization (the datasets in MSE 2005 and 2006 were not released by the organizers), and we construct a dataset as follows.

We first select 15 news topics in 2015, and collect 20 articles in Chinese and 20 in English about each topic within the same period using Google News [3]. The statistics of the corpus is shown in Table 1.

We employ 9 graduate students to write the English and Chinese reference summaries for the 15 topics after reading both English and Chinese documents

[3] http://news.google.com/

Table 1. Corpus Statistics.

	topic number	article number	Average sentence number per topic	Average word number per article
English	15	300	513.3	590.5
Chinese	15	300	447.7	556.6

for each topic. There are 3 reference summaries for each topic. For English reference summaries, we set the length limit to 250 words, and for Chinese the limit to 400 characters. The different length limits are set for considering the ratio of the lengths of translation English and Chinese text. We perform sentences and words tokenization and all the Chinese sentences are segmented by Stanford Chinese Word Segmenter [15].

4.2 Baseline Models

We compare our GuideRank model with the following baseline CoRank models without guidance.

Baseline-EN. This model generates summaries only using the original English documents.

Baseline-CN. This model generates summaries only using the translated English documents.

Baseline-ENCN. This models generate summaries using all the multilingual documents.

Replacement Models. Replacement strategy is adopted in the process of re-ranking. If a translated English sentence is chosen as summary, we will replace it with original English sentence with highest *relevance* score which is determined by Sim, TE and Trans evaluations introduced in Section 3.2.

4.3 Experimental Results

We use the ROUGE-1.5.5 [10] toolkit to evaluate the output summaries. Table 2 and 3 show the averaged ROUGE-2 and ROUGE-SU4 scores regarding to the three reference summaries for each topic. The value of α is set to 0.5.

To evaluate the effectiveness of proposed GuideRank model, we conduct experiments using different sentence representations, i.e., TF-IDF vectors and averaging word embeddings.

The results for English summaries. For the first three lines in Table 2, *Baseline-EN* outperforms *Baseline-CN* and even *Baseline-ENCN*, which may due to the translation errors. This phenomenon has been also verified by Daumé III and Marcu [3].

For *Replacement* models, when we replace the translated English sentences in summaries with original sentences using *Sim* evaluation, the system does not achieve the desirable results. The reason is that this simple strategy cannot accurately capture the sentences which are semantically related to the translated sentences. The performances of the *Replacement TE* and *Trans* models are much better which means better related English sentences to Chinese sentences are obtained.

GuideRank Trans model achieves the highest Rouge score, and then GuideRank TE model. *GuideRank SimMax* and *GuideRank SimTop*5 model do not perform well. The reason is that the strengths of the guidance for these two models are weak: for a certain topic document set which contains thousands of sentences, only changing one or five edges for every translated sentence seems to be negligible. By contrast, *GuideRank SimAve* model performs much better than other *GuideRank Sim* models. We also conduct experiments regarding different proportion of sentences with highest similarity score as related, and we get the similar results when the proportion ranging from 10%-50%.

The advantage of *GuideRank* models over *Replacement* models is that the algorithms optimize the problem globally, which take the interactions between sentences across languages into account during the process of calculating the sentence weights. While *Replacement* models are post-processing methods which will prevent some important translated sentences which cannot be covered by English sentences.

Table 2. Experimental results (F-score) for English summaries. * denotes statistically significant better than the baselines, p <0.01, t-test.

		TF-IDF		Embeddings	
		Rouge-2	Rouge-SU4	Rouge-2	Rouge-SU4
Baselines	EN	0.13477	0.18904	0.13326	0.18071
	CN	0.10272	0.15812	0.09067	0.14504
	ENCN	0.11325	0.16671	0.10709	0.16156
Replacement Models	Sim	0.12780	0.17977	0.14052*	0.18963*
	TE	0.13883	0.18832	0.14457*	0.19411*
	Trans	0.16471*	0.20798*	0.16805*	0.20813*
GuideRank Models	SimMax	0.11279	0.16688	0.10749	0.16105
	SimTop5	0.11315	0.16699	0.11035	0.16304
	SimAve	0.12757	0.18104	0.13814	0.18384
	TE	0.13261	0.18357	0.14447	0.19477
	Trans	0.16863*	0.21122*	**0.18360***	**0.22215***

Table 3. Experimental results (F-score) for Chinese summaries.

Models		Word Level Evaluation				Character Level Evaluation			
		TF-IDF		Embeddings		TF-IDF		Embeddings	
		Rouge -2	Rouge -SU4	Rouge -2	Rouge -SU4	Rouge -2	Rouge -SU4	Rouge -2	Rouge -SU4
Baselines	CN	0.12045	0.17581	0.11802	0.16145	0.23513	0.23314	0.22054	0.21797
	EN	0.07542	0.13666	0.05800	0.11497	0.17205	0.17946	0.14885	0.15827
	ENCN	0.09256	0.15273	0.08041	0.13528	0.19856	0.20189	0.18258	0.18388
Replacement Models	Sim	0.12015	0.17094	0.10453	0.15212	0.22473	0.22312	0.20700	0.20792
	TE	0.11653	0.16634	0.09129	0.14234	0.22423	0.22369	0.19329	0.19694
	Trans	0.11253	0.16435	0.10186	0.15023	0.21950	0.21906	0.20402	0.20596
GuideRank Models	SimMax	0.09386	0.15427	0.08094	0.13592	0.20110	0.20451	0.18256	0.18839
	SimTop5	0.09665	0.15698	0.08229	0.13640	0.20491	0.20816	0.18356	0.18883
	SimAve	0.12238	0.17680	0.10812	0.15664	0.22886	0.23175	0.21984	0.20182
	TE	0.11180	0.15908	0.11778	0.16365	0.22542	0.22250	0.23189	0.22666
	Trans	**0.13958***	**0.18999***	0.12323*	0.17349*	**0.25332***	**0.25148***	0.23507*	0.23343*

The results for Chinese summaries. We evaluate the Chinese summaries on word and character level, and the results are similar to English summaries that

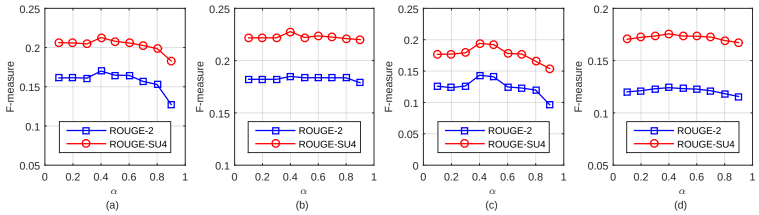


Fig. 2. Experimental results of *GuideRank Trans* models with different values of α. (a) English summaries taking TF-IDF as sentence feature. (b) English summaries taking embedding as sentence feature. (c) Word level evaluation for Chinese summaries taking TF-IDF as sentence feature. (d) Word level evaluation for Chinese summaries taking embedding as sentence feature.

GuideRank Trans model achieves the best performance. The performance using averaging word embeddings as sentence representation is worse than TF-IDF partially because we train the Chinese word embeddings with a relative small size corpus compared to English. When we use TE evaluation as the sentence relevance detection approach, the ROUGE scores are lower than *GuideRank SimAve* model for the reason that the input to BIUTEE toolkit must be two English sentences (Chinese sentences are translated into English), which will influence the TE recognition for Chinese sentences.

The influence of the parameter α. We evaluate the influence of the parameter α for *GuideRank Trans* model. The results are shown in Figure 2. Note that larger α means the model relies more on the information of the same side of language. We can conclude that our model benefits from both sides of language and relies more on cross-language information from the observation that ROUGE scores first increase with α, and after reach the peak value ROUGE scores where α is 0.4, decrease with α. This conclusion accords with the motivation of our GuideRank model that we take advantage of the interaction between different languages to guide the system to generate better summaries.

5 Analysis

To explore the differences of the three methods in evaluating the relevance between sentences across languages, we show two examples about the three methods to search for the most related sentence in English documents for a given Chinese sentence in Table 4. We can come into the following conclusions:

The results for similarity evaluation are the sentences with some words overlap with the translated English sentences, which suggests similarity is not sufficient enough for evaluating the semantic relevance no matter what kind of sentence representations.

TE evaluation obtains real related sentence, but the rigidity of TE may restrain the further improvement upon the Sim evaluation. The partial TE [8] may remedy this problem.

Table 4. Examples for searching for the most related English sentence to the Chinese sentences.

	Example 1	Example 2
Original	目前尚不清楚到周六早上是否还有其他袭击者逍遥法外。	1961年1月5日，美国宣布与古巴断绝外交关系。
Translation	It is unclear whether there are other Saturday morning to the attackers go unpunished.	January 5, 1961, the United States announced the severance of diplomatic relations with Cuba.
Sim Evaluation (embeddings)	It was unclear whether that term meant the terrorists were dead .	July 20 is the date when the United States and Cuba officially restore diplomatic ties
Sim Evaluation (tf-idf)	Paris terror attack: Everything we know on Saturday afternoon.	Obama announces re-establishment of U.S. Cuba diplomatic ties.
TE Evaluation	Many questions remain unanswered, including whether any accomplices are at large, who coordinated the attacks, and whether counterterrorism efforts could have foiled the plot.	In January of 1961 , the year I was born , when President Eisenhower announced the termination of our relations with Cuba , he said.
Trans Evaluation	It was not clear if all the attackers were accounted for.	The U.S. and Cuba broke ties in 1961.

Translation evaluation performs much better than other methods. To some degree, for Sim and TE evaluation, we need to translate the Chinese sentences into English, which will influence the downstream relevance detection. While there is no influence on this aspect for Translation evaluation.

For Trans, TE and SimAve evaluation, the proportion of target language sentences in the summaries is around 68%, 73% and 91%, which suggests that the more original target language sentences doesn't stand for higher performance. The best performance of Guide Trans model illuminates its ability to balance the informativeness and readability of summaries.

6 Conclusion

In this paper, we constructed a multilingual summarization dataset and propose GuideRank model by considering the interaction between sentences in different languages. Our model is designed to generate the target language summaries by selecting sentences from the target language documents which contain shared information across languages, and also remaining the salient translated sentences beyond the content of target language documents. The experimental results show the effectiveness of our method.

Acknowledgments

The research work has been funded by the Natural Science Foundation of China under Grant No. 61333018 and supported by the Open Project Program of the State Key Laboratory of Mathematical Engineering and Advanced Computing.

References

1. Dalli, A., Catizone, R., Wilks, Y.: Clustering-based language independent multiple-document summarizer at mse 2006. Proceedings of MSE (2006)
2. Daraksha Parveen, H.M.R., Strube, M.: Topical coherence for graph-based extractive summarization. In: EMNLP15 (2015)

3. Daumé III, H., Marcu, D.: Bayesian multidocument summarization at mse. In: Proceedings of MSE (2005)
4. Erkan, G., Radev, D.R.: Lexrank: Graph-based lexical centrality as salience in text summarization. Journal of Qiqihar Junior Teachers College 22, 2004 (2011)
5. Giampiccolo, D., Magnini, B., Dagan, I., Dolan, B.: The third pascal recognizing textual entailment challenge. In: Acl-Pascal Workshop on Textual Entailment and Paraphrasing. pp. 1–9 (2007)
6. Giannakopoulos, G., El-Haj, M., Favre, B., Litvak, M., Steinberger, J., Varma, V.: Tac2011 multiling pilot overview. Contribution in Book/report/proceedings (2011)
7. Klein, D., Manning, C.D.: Accurate unlexicalized parsing. In: Meeting on Association for Computational Linguistics. pp. 423–430 (2003)
8. Levy, O., Zesch, T., Dagan, I., Gurevych, I.: Recognizing partial textual entailment. In: Meeting of the Association for Computational Linguistics. pp. 451–455 (2013)
9. Levy, R., Manning, C.: Is it harder to parse chinese, or the chinese treebank? In: IN PROCEEDINGS OF THE 41ST ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. pp. 439–446 (2003)
10. Lin, C.Y., Hovy, E.: Automatic evaluation of summaries using n-gram co-occurrence statistics. In: Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology (2003)
11. Mihalcea, R.: Graph-based ranking algorithms for sentence extraction, applied to text summarization. In: Proceedings of the ACL 2004 on Interactive Poster and Demonstration Sessions. ACLdemo '04 (2004)
12. Mihalcea, R., Tarau, P.: Textrank: Bringing order into texts. Unt Scholarly Works pp. 404–411 (2004)
13. Siddharthan, A., Evans, D.: Columbia university at mse 2005
14. Stern, A., Dagan, I.: Biutee: a modular open-source system for recognizing textual entailment. In: ACL 2012 System Demonstrations. pp. 73–78 (2012)
15. Tseng, H., Chang, P., Andrew, G., Jurafsky, D., Manning, C.: A conditional random field word segmenter (2005)
16. Wan, X.: Using bilingual information for cross-language document summarization. In: ACL2011. pp. 1546–1555 (2011a)
17. Wan, X., Li, H., Xiao, J.: Cross-language document summarization based on machine translation quality prediction. In: ACL 2010, Proceedings of the Meeting of the Association for Computational Linguistics, July 11-16, 2010, Uppsala, Sweden. pp. 917–926 (2010)
18. Wan, X., Yang, J.: Improved affinity graph based multi-document summarization. In: Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 4-9, 2006, New York, New York, USA. pp. 181–184 (2006a)
19. Wan, X., Yang, J., Xiao, J.: Using cross-document random walks for topic-focused multi-document. In: 2006 IEEE / WIC / ACM International Conference on Web Intelligence (WI 2006), 18-22 December 2006, Hong Kong, China. pp. 1012–1018 (2006b)
20. Wei Xu, C.Y.: The thu/polyu system at mse 2006: An event-relevance based approach. Proceedings of MSE 2006 (2006)
21. Yao, J.G., Wan, X., Xiao, J.: Phrase-based compressive cross-language summarization. In: Conference on Empirical Methods in Natural Language Processing. pp. 1546–1555 (2015)
22. Zajic, D., Dorr, B., Lin, J., Schwartz, R., Zajic, D., Dorr, B., Lin, J.: Umd/bbn at mse2005. Proceedings of MSE (2005)