

Using Context Inference to Improve Sentence Ordering for Multi-document Summarization

Peifeng Li Guangxi Deng Qiaoming Zhu
School of Computer Science and Technology, Soochow University
1 Shizi St, Suzhou, China, 215006
{pfli, gxdeng, qmzhu}@suda.edu.cn

Abstract

In this paper, we propose a novel context inference-based approach for sentences ordering in multi-document summarization application. Our method first detects whether or not two summarization sentences should be adjacent according to the similarity between one summarization sentence and the context of the other one, and then it computes the reliability of the adjacent summarization sentences based on the similarity and their relative position. To be specific, the first sentence will be selected according to features of sentence, and the second sentence will be selected if and only if it has the maximum reliability with previous sentence. Evaluation result shows that our method outperforms the state-of-the-art ones on DUC2004 corpus.

1 Introduction

Multi-document summarization is an automatic procedure aimed at extraction of information from multiple texts written about the same topic. Whereas, sentence ordering, the last task of multi-document summarization, will finally affect the quality of summarization. Furthermore, the task of sentence ordering in multi-document summarization is harder than that in single document summarization because multiple documents are created by different authors who have different writing styles and backgrounds. Therefore, no natural order of texts can be extracted as the basis of sentence ordering judgment. How to conduct an efficient and effective method for sentences ordering is a difficult but important task for both multi-document summarization and other text processing job, e.g. Question Answering.

Currently, a variety of studies have been

reported on sentence ordering. Some methods adopted chronological information (McKeown et al., 1999; Lin et al., 2001; Barzilay et al., 2002; Okazaki et al., 2004) while others learned the natural order of sentences from source documents or large corpus (Lapata, 2003; Barzilay and Lee, 2004; Nie et al., 2006; Ji and Nie, 2008). However, chronological information cannot be easily extracted from those non-news documents and constructing a large corpus also is not so easy. Furthermore, those results of all above methods are far from satisfactory. Therefore, how to achieve coherent summarization still is an issue for us.

This paper proposes a novel method to infer the order of summarization sentences for multi-document summarization according to their context. We first judge whether or not two summarization sentences should be adjacent based on the similarity between one summarization sentence and the context of the other one, and then compute the reliability based on the similarity and relative position of the adjacent summarization. To be specific, the first sentence will be selected based on its features, and the second sentence will be selected if and only if it has the maximum adjacency reliability with previous sentence. The experimental results on DUC2004 corpus show that our method outperforms state-of-the-art ones.

The reminder of this paper is organized as follows. In Section 2, the related work is introduced. The motivation of our method is discussed in Section 3. In Section 4, the context inference-based sentence ordering algorithm is described. In Section 5, the experiments and evaluation are presented. Conclusions and future work are given in Section 6.

2 Related Work

So far many studies on sentence ordering have

been proposed. They can be divided into two categories: chronology-based method and natural order-based method. To be specific, chronology-based methods determine sentence ordering according to the chronological order of events while natural order-based methods infer sentence ordering according to some cues that are learned from source documents and corpus.

Barzilay et al. (2002) proposed a chronology-based method to sort sentence, they assumed the theme of sentences were the cues of sentence ordering. Based on it, they presented a strategy according to the first published date of the theme; then sorted sentences based on their order of presentation in the same article.

Okazaki et al. (2004) proposed an approach to improve the chronological sentence ordering method by using precedence relation technology. They assumed the presupposed information should be transferred to reader in advance before the sentence was interpreted. They first arranged sentences in a chronological order and then estimated the presupposed information for each sentence by using the content of the sentences located in before each sentence in its original article.

Generally speaking, the articles in news domain usually contain descriptions of date and events accompanying the publication sequences. Therefore, chronology-based method is practical for news domain. However, chronological information is not ubiquitous in a large number of multi-document summarization tasks. So, some studies began to focus on general domain, and they sorted sentences according to the source documents and corpus.

McKeown et al. (2001) and Barzilay et al. (2001) presented a majority ordering algorithm to sort sentences. They classified sentences of source documents into different themes or topics using summarization sentences. If sentences from theme 1 preceded sentences from theme 2 when they appeared in same text, then putted theme 1 was preceding theme 2. The order of summarization sentences was determined by the order of themes. However, there were some potential issues in this kind of method. Firstly, it was not easy to correctly cluster sentences into topics. Secondly, some summarization sentences may belong to a same topic. Finally, the relative order between two topics may be not steady.

Lapata (2003) provided an unsupervised probabilistic model for sentence ordering. The model assumed that sentences were represented as a set of features that could be automatically

extracted from the corpus without manual annotation. The conditional probability of sentence pairs can be learned from a training corpus. By computing conditional probability of each sentence pairs, the approximate optimal global sentence ordering can be achieved using a simple greedy algorithm.

Bollegala et al. (2005) combined three ordering methods together, said chronological ordering, probabilistic ordering and topic relatedness ordering, and adopted a machine learning approach for sentence ordering. The mixed system can achieve better performance than any of the three individual one. To be extended, they also proposed a bottom-up approach for sentence ordering (Bollegala et al., 2010). They defined four criteria (chronology, topical-closeness, precedence and succession) between two textual segments, and adopted SVM classifier to learn the relative order of them. They repeatedly concatenated two textual segments into one segment based on the relative order until all sentences were arranged.

Nie et al. (2006) adopted adjacency of sentence pairs to sort sentences. Sentence adjacency was calculated based on adjacency of features of each sentence pairs. Adjacency between two sentences represented how closely they should be putted together in a set of summarization sentences. Ji et al. (2008) extended the above adjacency model. They proposed a cluster-adjacency based method to map each sentence of source documents to a theme by using semi-supervised classification method. The adjacency of sentences pair was learned from source documents according to adjacency of clusters they belonged to.

Our method is different from above ones in two aspects. First, we try to find the indirect relations of summarization sentences according to source documents. Second, the source documents are topic-related and most summarization sentences or their similar ones appear in more than one source document. Meanwhile, there are more than two summarization sentences or their similar ones may co-occur in a source document. Therefore, we can infer the order of summarization sentences by using the indirect relation between summarization sentences.

3 Motivation

Different documents have different writing styles and backgrounds. Besides, sentences that used to

describe a same topic may have different forms. Therefore, source documents maybe cannot provide the direct information for sentence ordering. However, they still provide some indirect information which can be used to infer the order of summarization sentences.

Each source document for multi-document summarization is not an information island. They may have some overlapped information. For example, there are three documents to report a same news event. The first document describes its background and cause. The second one describes the origin, process and result. The third one describes the effect and comment. Each document has its own key point, but they have some overlapped information. The first and the second one report the cause in both while the second and the third one report the effect in both.

Each summarization sentence is extracted from a source document even though it is created or rewritten manually, it also can link to one or more sentences in source documents. Therefore, each summarization sentence is also not isolated and there are many sentences surrounding them in the source document. We call these sentences as the context of the summarization sentence. Furthermore, the source documents are in same topic and most of summarization sentences are presented in more than one document. In a word, a summarization sentence may have some similar sentences in source documents.

Accordingly, we provide a method to infer the order of summarization sentences by using their context. Let us consider an example: there are two documents, d_1 and d_2 , sketched in Figure 1, where ss_1 and ss_2 are two summarization sentences or their similar ones in source documents, s_{1i} and s_{2i} are the context sentence of ss_1 and ss_2 respectively.

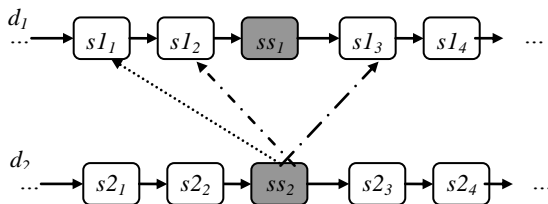


Figure 1. An example of summarization sentences and their context

If ss_2 is similar to s_{11} , we have reason to believe that ss_2 and ss_1 should be adjacent and ss_2 should be in front of ss_1 because s_{11} is front of ss_1 . The highly similarity between ss_2 and s_{11} , the highly probability that ss_2 is in front of ss_1 .

If ss_2 is not similar to s_{11} , but it is similar to s_{12} , we also consider ss_2 should be in front of ss_1 , but the probability will higher because s_{12} is closer to ss_1 than s_{11} . Similarly, if ss_2 is similar to s_{13} , ss_1 should be in front of ss_2 .

4 Context Inference-based Sentence Ordering Algorithm

From the analysis in section 3, we can infer the order of sentence by using the similarity between summarization sentences and their context. We compute the adjacent credibility of two sentences and then use a directed graph with weight to represent the order of summarization sentences. A vertex denotes a sentence. If two sentences are adjacent then an edge exists of them. The weight of edge is the adjacency credibility of two sentences. Figure 2 shows the order relation among summarization sentences, there is an edge from s_1 to s_3 and the weight is 0.5, which means s_1 and s_3 are adjacent, and the credibility of them is 0.5. From figure 2, we can find a path contained all vertexes, which has the maximum weight. The order of the path is considered as the best logical order of summarization sentences.

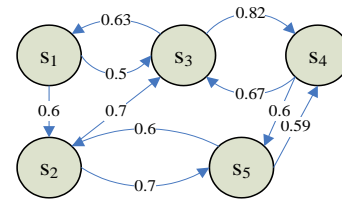


Figure 2. Graph of summarization sentences and their relation

Due to finding an optimal path in a graph is a typical NP problem; we will use an algorithm to find an approximate optimal path. We assume that the document set $D = \{d_1, d_2, \dots, d_n\}$, a document $d_j = \{s_{j,1}, s_{j,2}, \dots, s_{j,m}\}$ where $s_{j,i}$ is the i th sentence in d_j , the summarization sentences set $SS = \{ss_1, ss_2, \dots, ss_k\}$, the Graph $G = NULL$. For a sentence ss_i in d_j , its context sentence set is $\{d_j - ss_i\}$. The algorithm is described as follows.

Input: D , SS and G

Output: P , an ordered list that contains all summarization sentences.

BEGIN

Step 1: For each ss_i in SS , add a vertex to graph G ;

Step 2: For each ss_i in d_j , compute the similarity $sim(ss_k, s_{j,l})$ between each ss_k ($ss_k \in \{SS - ss_i\}$) and each $s_{j,l}$

($s_{j,l} \in \{d_j - ss_i\}$) respectively. ss_i and ss_k are directed adjacent if existing a $sim(ss_k, s_{j,l}) > \theta$ where $s_{j,l} \in \{d_j - ss_i\}$.

If $s_{j,l}$ is in front of ss_i , the credibility $weight(ss_i, ss_k)$ that ss_k is in front of ss_i is computed as follow:

$$weight(ss_i, ss_k) = w * sim(ss_k, s_{j,l}) + (1-w) * \frac{rank(s_{j,l})}{rank(ss_i)} \quad (1)$$

where $rank(ss_i)$ is the sequence of ss_i in the document it belongs to (e.g. the rank of second sentence of document is 2), w is the contribution of sentence similarity to the credibility, $sim(ss_k, s_{j,l})$ is the similarity between ss_k and $s_{j,l}$. If there is no edge from ss_k to ss_i in G , then add an edge to G ; else if $weight(ss_i, ss_k)$ is greater than the weight of existing edge, then update the edge.

If ss_i is in front of $s_{j,l}$, the credibility $weight(ss_i, ss_k)$ is computed as follows:

$$weight(ss_i, ss_k) = w * sim(ss_k, s_{j,l}) + (1-w) * \frac{Len(d_j) - rank(s_{j,l})}{Len(d_j) - rank(ss_i)} \quad (2)$$

where $len(d_j)$ is the total number of sentences contained in d_j . If there is no edge from ss_i to ss_k in G , then add an edge to G ; else if $weight(ss_i, ss_k)$ is greater than the weight of the existing edge, update the edge.

Sentence similarity between sentences s_1 and s_2 can be computed based on TF-IDF or WordNet (Achananuparp, 2008). Sentences similarity based TF-IDF can be calculated as follow:

$$sim(s_1, s_2) = \vec{V}_1 \cdot \vec{V}_2 = \frac{\sum_{i=1}^n x_i \cdot y_i}{\sqrt{\sum_{i=1}^n x_i^2} * \sqrt{\sum_{i=1}^n y_i^2}} \quad (3)$$

where $V_1 = \{x_1, x_2, \dots, x_n\}$, $V_2 = \{y_1, y_2, \dots, y_n\}$, V_1, V_2 are TF-IDF vectors of sentences s_1 and s_2 .

Sentence similarity $sim(s_1, s_2)$ based on WordNet is defined as

$$sim(s_1, s_2) = \frac{\sum_{w \in s_1} Sim(w, s_2) + \sum_{w \in s_2} Sim(w, s_1)}{length(s_1) + length(s_2)} \quad (4)$$

where $sim(w, s_i)$ is the maximum semantic similarity between word w and sentence s_i , $length(s_i)$ is the length of sentence s_i .

Step 3: We use a feature-based method to find the first sentence, assume that there is a null sentence at the beginning of each source document and it contains a null feature (Nie, 2006). The probability of a sentence ss_i is the first one is defined as follow:

$$Prob_i = \frac{1}{K} \sum_{i=1}^K \frac{freq(feature_i, nullfeature)}{freq(feature_i)} \quad (5)$$

where K is the number of features in sentence s_i . $freq()$ is the frequency function, $freq(feature_i)$ denotes the frequency of $feature_i$ in the source documents, $freq(feature_i, nullfeature)$ denotes the frequency of $feature_i$ and the *null feature* co-occurring in the source documents within a limited range (one or several sentences, we assign 3 to the range). Select the sentence with the maximum probability as the first sentence. Add the first sentence to P .

Step 4: Get the trail sentence of P , select the next sentence and add it to P . Given an already ordered sentence serial $P: ss_1, ss_2 \dots ss_i$ which is a subset of the summarization sentences set SS . The next sentence can be found by formula (6):

$$ss_j = \arg \max_{s_j \in SS-P} (weight(ss_i, ss_j)) \quad (6)$$

Step 5: If P contains all summarization sentences then exit; otherwise go to step 4.

END

5 Experiments and Evaluation

5.1 Test Set and Evaluation Metrics

Due to current methods provided their experiment results using DUC04 corpus, we also use it to conduct our experiment. DUC04 provide 50 source document sets and four manual summaries for each document set in its Task2. Each document set consists of 10 documents. In the DUC04, sentences in summaries are not directly come from the source documents and they are created by manually. Therefore, we need map them back to the sentences in source documents. For each manual summarization sentence, we find a sentence in source document sets that has the maximal similarity with it as its source sentence. These source sentences of each summarization are taken as inputs to ordering model, but sequential information is neglected. The output ordering of various models are compared with the specified ones in manual summaries. A number of metrics can be used to evaluate the difference between two orderings. In this paper, we adopt Kendall's τ (Lebanon, 2002), which was defined as:

$$\tau = 1 - \frac{2(number_of_inversions)}{N(N-1)/2} \quad (7)$$

where N is the number of objects to be ordered (i.e., sentences), *number_of_inversions* is the minimal number of interchanges of adjacent

objects to transfer an ordering into another. Intuitively, τ can be considered as how easily an ordering can be transferred to another. The value of τ varies from -1 to 1, where 1 denotes the best situation where two orderings are the same, and -1 denotes the worst situation where two orderings are completely reversed. Given a standard ordering, randomly produced orderings of the same objects would get an average τ of 0. For example, Table 1 lists three numbers of sequences, their natural sequences and the corresponding τ values.

Examples	Natural sequences	τ values
1 2 4 3	1 2 3 4	0.67
1 5 2 3 4	1 2 3 4 5	0.40
2 1 3	1 2 3	0.33

Table 1: Ordering Examples

5.2 Experimental Results

In order to get the best performance, we adjusted the parameters of our approach (θ and w). We adopted DUC02 as the development set to adjust parameters θ and w . In DUC02, it provide 60 document sets of approximately 10 documents each and one manual summarization of about 100 words for each document set in Task2. Figure 3 show the experiment results, from it, we can finally get the optimal value of θ and w (0.3 and 0.6 respectively).

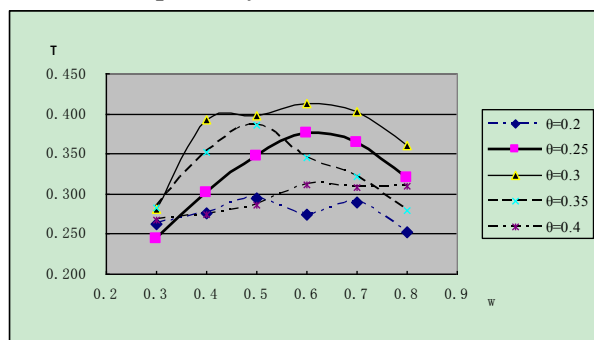


Figure 3. Graph of adjusting parameters

We assume Rd, Mo, Pr, Fa and Ca to denote random ordering, majority ordering, probabilistic model, feature-adjacency based model and cluster-adjacency based model which have been mentioned in section 2 respectively. Due to other methods didn't report their results on DUC04 corpus; we do not compare our method with them in this paper. We define our Context-based approach as Co. Table 2 shows the results of different methods. The τ value of our approach reaches 0.43, which outperforms other

approaches. Also, the performance of WordNet-based similarity measure is slightly better than the TF-IDF based one. The reason may be that some authors will express the same meaning by using some different but synonymous words.

Models	τ	Similarity Measure
Rd	-0.007	
Mo	0.143	
Pr	0.144	
Fa	0.316	
Ca	0.415	
Co	0.424	TF-IDF
Co	0.432	WordNet

Table 2: Experimental results ($\theta=0.3, w=0.6$ in Co)

We conduct another experiment, correction ratio of sentence inference from correctly ordered previous sentences, to see why our method gets better performance. The result is listed in table 3.

Models	1 st sentence \rightarrow 2 nd sentence	2 nd sentence \rightarrow 3 rd sentence
Mo	24.4%	10%
Pr	25.0%	25%
Fa	31.8%	50%
Ca	56.2%	61.6%
Co (TF-IDF)	62.5%	64.2%
Co (WordNet)	63.2%	65.4%

Table 3. Comparison of correct sentence inference

From table 3, our method has the highest correction ratio of sentence inference, which mean that our method's strategy to choose next sentence is reasonable than that of others. Probabilistic ordering, a statistical method, tries to find the ordering clue from the corpus, but they ignore the importance of source documents. Basically, we know that people can easily find the order of summarization sentences according to the source documents. But it's difficult to give the order for the sentence base on a corpus without the source documents even for a knowledgeable man. Therefore, we believe the source documents can give us some effective and efficient information. Fa and Ca models sort sentences by using sentence adjacency, which is calculated based on adjacency of feature pairs. However, how to choose effective features is a potential issue of this kind of method. Besides, it

doesn't always mean two sentences are likely to be adjacent when some words of them are adjacent. Our method is different from other ones in that we find the indirect relations between summarization sentences based on the context and make full use of the information which provided by source documents.

Table 4 shows a further comparison among all methods. "Positive ordering" means that the result of ordering is better than that of random ordering and "Negative ordering" denotes the output ordering which gets a negative τ . From it, we can review that our method generates the most positive orderings while with the least negative orderings. Precision of first sentence of our method is less than that of Ca. If we assume the first sentence of summarization is known in advance, experiments show that the average τ value of our method could reach 0.562.

Models	Precision of 1st sentence	Positive Orderings	Negative Orderings
Rd	14.0%	48.4%	44.6%
Mo	21.6%	61.8%	31.8%
Pr	40.8%	62.5%	29.5%
Fa	59.5%	71.5%	19.0%
Ca	65.5%	81.0%	15.5%
Co (TF-IDF)	59.5%	84.5%	10.5%
Co (WordNet)	59.5%	84.8%	10.1%

Table 4. Correction ratio of 1st sentence ranking

6 Conclusion and Future Work

This paper presents a novel method for sentence ordering in multi-document summarization application. The proposed method first computes the similarity between summarization sentences and context of other summarization sentences, judge whether or not two sentences should be adjacent, and then compute the reliability according to the similarity. The experimental results review that our method outperforms other sentence ordering methods on DUC04 corpora.

For further work, we will change the strategy to improve the accuracy of first summarization sentence; therefore, the τ value could have an obvious improvement. In addition, we will conduct more experiments to verify the effectiveness and efficient of our method using manual evaluation.

Acknowledgments The authors would like to thank the anonymous reviewers for their comments on this paper. This research was supported by the National Natural Science

Foundation of China under Grant No. 61070123 and No. 60970056, The Research Fund for the Doctoral Program of Higher Education of China under Grant No. 20093201110006, the Natural Science Major Fundamental Research Program of the Jiangsu Higher Education Institutions under Grant No. 08KJA520002.

References

- Danushka Bollegala, Naoaki Okazaki, Mitsuru Ishizuka. 2005. A Machine Learning Approach to Sentence Ordering for Multi-document Summarization and Its Evaluation. In *IJCNLP 2005, LNAI 3651*. pages 624-635.
- Danushka Bollegala, Naoaki Okazaki, Mitsuru Ishizuka. 2010. A bottom-up approach to sentence ordering for multi-document summarization, *Information Processing & Management*. 46:89-109.
- Ji Donghong and Nie Yu. 2008. Sentence Ordering based on Cluster Adjacency in Multi-Document Summarization. In *IJCNLP 2008*, pages 745-750.
- Guy Lebanon and John Lafferty. 2002. Cranking: Combining Rankings Using Conditional Probability Models on Permutations. In *ICML 2002*, Morgan Kaufmann Publishers, San Francisco, CA, pages 363-370.
- Kathleen R. McKeown, Regina Barzilay, David Evans, Vasileios Hatzivassiloglou, Simone Teufel. 2001. Columbia Multi-document Summarization: Approach and Evaluation. In *DUC01*.
- Mirella Lapata. 2002. Automatic Evaluation of Information Ordering: Kendall's Tau. *Computational Linguistics*, 32(4): 471-484.
- Mirella Lapata. 2003. Probabilistic Text Structuring: Experiments with Sentence Ordering. In *ACL 2003*. pages 545-552.
- Naoaki Okazaki, Yutaka Matsuo, Mitsuru Ishizuka. 2004. Improving chronological sentence ordering by precedence relation, In *COLING 2004*, pages 750-756.
- Nie Yu, Ji Donghong and Yang Lingpeng. 2006. An Adjacency Model for Sentence Ordering in Multi-document. In *AIRS 2006*.
- Palakom Achananuparp, Xiaohua Hu, Xiaojong Shen. 2008. The Evaluation of Sentence Similarity Measures. In *Proceedings of the 10th international conference on Data Warehousing and Knowledge Discovery*, Berlin, Heidelberg, Springer-Verlag, pages 305-316.
- Regina Barzilay, Noemie Elhadad, McKeown R. Kathleen. 2002. Inferring Strategies for Sentence Ordering in Multi-document News Summarization.

Journal of Artificial Intelligence Research, 17:(35–55).

Regina Barzilay, Noemie Elhadad and Kathleen R. McKeown. 2001. Sentence Ordering in Multi-document Summarization. In *HLT 2001*, San Diego, CA, pages 149-156.