Cohesion-based Sentence Ordering for Multi-document Summarization

Xiaoyu Jiang

School of Business, Beijing Institute of Fashion Technology, Beijing, 100029, China
sxyjiangxiaoyu@bift.edu.cn

Keywords: Multi-document summarization; sentence ordering; cohesion; local topic; Majority Ordering

Abstract. Sentence ordering is a difficult but an important task for multi-document summarization. A wrong order of sentences not only makes it difficult to be understood but also convey entirely different ideas to the reader. In this paper, we describe two naive ordering algorithms and give the reasons why they do not perform well. We propose a new sentence ordering algorithm, which combines constraints from majority order of themes and cohesion. This algorithm ensures: if the cohesion between two themes is bigger than the threshold value, then the sentences of the two themes would be placed to the adjacent location in the final summary. Experimental results show that summary generated by our improved sentence ordering algorithm is more coherent and more readable than the two naive algorithms.

1 Introduction

Multi-document summarization (MDS) is the task of generating a human readable summary from a given set of documents (Ani Nenkova et al.). It can be considered as a two-stage process. On the first stage we must extract a set of sentences from the given document set. The second stage of MDS is creating a comprehensible summary from this extract. A good ordering of sentences improves coherence of a summary, so the summarizer must decide in which order to present them so that the summary make sense.

Two naive ordering algorithms are introduced and the reasons why they do not yield satisfactory results are analyzed. Majority Ordering algorithm is critically linked to the level of similarity of the information ordering across the input texts. But many times input texts have different structure, and therefore, this algorithm does not always perform well. Chronological Ordering algorithm can produce good results when the information is event-based and can be ordered based on temporal occurrence. However, texts do not always refer to events. R. Barzilay observed that cohesion as an important constraint is a necessary feature for a generated summary. In this paper, we give an operational way to ensure cohesion for ordering sentences in an output summary. We augment the MO algorithm with a cohesion constraint, and compare it to the two naive algorithms.

2 Naive Ordering Algorithm

Multi-document summarization is the technology of natural languages processing, which extracts important sentences from input texts which are composed by some themes with similar information (Bollegala et al.). Themes are sets of sentences from different documents that contain repeated information. So we first identify similar information across the input document set using statistical techniques and shallow text analysis, and cluster them into themes. For each theme, we select the intersection sentences of the theme as part of the summary. So how these sentences are ordered to produce a coherent text can be attributed to how to order these themes.

2.1 Chronological Ordering

2.1.1 Algorithm

Chronological Ordering (CO) uses time related features to order sentences. Since the sentence ordering can be attributed to how to order all themes, we need to mark time for each theme. D. Bollegala et al. assumes that each document of the entire news sets marked by date, hours and
minutes of publication time, and there are no any two documents that are marked with the same time. Based on the above assumptions, the sequencing steps of CO algorithm are as follows:

1. Each sentence of the theme will be marked with the time, and the time is the release time of document in which the sentence appears;
2. The earliest marking time of all sentences of the theme will be the marking time of this theme;
3. If the marking time of two themes is the same, then the two themes first appeared in the same document, and their ranking is consistent with the orders of appearance in the document;

2.1.2 Algorithm Deficiency

From the application of CO algorithm in the Chinese multi-document summarization, we have found four deficiencies, such as:

1. The marking time of theme is too dependent on the time of the following sentences that requires particularly high accuracy of division of themes;
2. Most of the time of Chinese-language news is not accurate to hours and minutes, more simply refers to the date of issue, which causes problems for the comparison of marking time of the theme;
3. Though the marking time of two themes is the same, their first appearance is not necessarily in the same documents, the assumption of steps (3) is not established;
4. CO algorithm is applicable only for multi-document summarization system with time information in given document set.

2.2 Majority Ordering

2.2.1 Algorithm

We define notation A ≻ B to represent that theme A always precedes theme B and sentence A_i of A and sentence B_j of B appear in a same document, but when A ≻ B is not established, we define notation A ∼ B to represent the number of texts in input documents which meets the demand of A ≻ B.

For each pair of themes, A and B, we keep two counts: C_{AB} and C_{BA}, in which C_{AB}=A ∼ B, C_{BA}=B ∼ A. Taking theme as the vertex, C_{AB} and C_{BA} as the weights of the sides connecting with the vertex, these vertices and sides can compose a directed graph, as shown in Figure 1. The location of each theme in the summary can be decided by the optimal path in the directed graph. And the path must meet: (1) iterate through each vertex once and only once; (2) with the largest weight.

Unfortunately, the search for the optimal path is a typical NP problem, R. Barzilay et al. give a proximate algorithm to get the optimal path P, as following the steps:

1. Set the initial path P as null;
2. Calculate the weight for each vertex;
3. Choose the vertex with the greatest weight and insert it into the path P, and delete the vertex and all the sides connecting with it in the directed graph, and recalculate the weights of the rest vertexes;
4. If the directed graph is not empty, turn to step 2 and go on running, otherwise, with the end, path P is the location of each theme in the final summary.

Figure 1. Directed Graph of theme relation.

2.2.2 Algorithm Deficiency

In the experimental process, we found that the MO sentence ordering algorithm performed not very well from an overall perspective. There are still some problems that cannot be ignored; the questions are as follows:

1. In the process of searching path, when the weight of the two vertexes are the same, MO algorithm fails to provide sufficient constraint to decide on which vertex will be better, but only selects one of them randomly, this is quite blind;
2) If there are any two themes that do not appear at the same time in any document, MO algorithm cannot solve their ordering problem well, such as the orders of \( B \) and \( D \) in Figure 1;
3) When the position of each theme is relatively fixed, the summary has a better readability, but when the relative positions of the themes frequently change, the coherence and readability of the summary will be poor.

### 3 The Improved Algorithm

If the relative position of the same theme changes frequently in given document sets, the summary ordered by MO algorithm is always not coherent and not easy to be understood. As shown in Figure 2: \( T_1, T_2, T_3 \) and \( T_4 \) are the document set of input, \( A_1, A_2, A_3 \) and \( A_4 \) are sentences which belong to the same theme \( A \), similarly, \( B, C \) and \( D \) are several other themes. \( S_1 \) is the summary generated basically by the sequence means of MO. \( S_2 \) is the one generated by human. We can find that: although only the sequence of \( B \) and \( C \) appears upside down, but on the whole summary, the effect of \( S_2 \) is obviously superior to that of \( S_1 \).

![Figure 2. Input texts \( T_i \) are summarized by the Majority Ordering (\( S_1 \)) or by Human (\( S_2 \)).](image)

As shown in Figure 2, based on the decision of themes and extraction intersection sentences for each theme, the sequencing process of the summary by using MO is shown in Figure 3. By analyzing the figure, MO algorithm can be found to choose one of the biggest weights vertex in the directed graph then delete the vertex and all the sides connected with it from the directed graph, and recalculate the weights of the rest vertexes, and it results in the choice of next vertex only related to the rest vertexes when one vertex is chosen and unrelated to the vertexes having been chosen, and it’s easy for those themes which is with low cohesion, or even documents simply do not exist in the order of those themes, are located in proximity in the summary, thus it leads to the summary is not coherent an hard to be understood. For example, there is a contiguous sequence \( A \) and \( B \) in the three documents of the four of the document set, to some extent, it can illuminate that the two themes are closely relevant. So it’s a more ideal result of order to put \( A \) and \( B \) in proximity in the summary, and it’s the same result as the manual sequence, but in the summary \( S_1 \) generated by MO, \( A \) isn’t connected with \( B \).

![Figure 3. Sequencing Process of Majority Ordering.](image)

From the analysis, we can know that the shortage of MO algorithm lies in only considering the prior relations between the themes in the sequencing process without taking account of the degree of cohesion between them, and it makes some sentences of themes with high cohesion be in the discrete state and leads to the result that the summary is incoherent and hard to understand. So, on the basis of MO algorithm, we propose the method of approaching the cohesion between the themes into the sequence process of MO. As following the sequencing steps:
1) Use the MO algorithm to compose a theme relation directed graph:(more details in 2.2)
2) Calculate the cohesion between every two local themes, that is, the degree of close connection;
3) Ordering for these themes by combining MO algorithm and the cohesion.
3.1 Cohesion Calculation

Assume the themes $A$ and $B$, where $A$ contains sentences $(A_1, A_2, \ldots, A_n)$, and $B$ contains sentences $(B_1, B_2, \ldots, B_m)$. Recall that a theme is a set of sentences conveying similar information drawn from different input text. We denote $A_iB_j=1$ to represent the pair of sentences $(A_i, B_j)$ which appear in the same text, otherwise, $A_iB_j=0$, and $A_iB_j^+=1$ to represent pair of sentences $(A_i, B_j)$ which appear in the same or adjacent segment, otherwise, $A_iB_j^+=0$. We denote $\#_{AB} = \sum_{i=1}^{n} \sum_{j=1}^{m} A_i B_j$ to be the number of pairs of sentences $(A_i, B_j)$ which appear in the same text, and $\#_{AB}^+ = \sum_{i=1}^{n} \sum_{j=1}^{m} A_i B_j^+$ to be the number of sentence pairs which appear in the same or adjacent segment (R. Barzilay et al.).

For each pair of themes $(A, B)$, $R_{AB}=\#_{AB}/\#_{AB}^+$ is computed to measure the cohesion of two themes. For the opposite order, the cohesion of $(B, A)$ is $R_{BA}=\#_{BA}/\#_{BA}^+$. This measure takes into account both positive and negative evidence. When $R_{AB}$ is bigger than the threshold value, it means that $A$ and $B$ are highly topically related, otherwise, we regard that $A$ and $B$ are not topically related. Based on the calculation and analysis of cohesion on Chinese-language document sets which contain 540 different topics, we adopt the experience value 0.7 as the threshold value. By this method, we can calculate the cohesion between each pair of themes from the input texts, to make an $N \times N$ asymmetric cohesion array; $N$ is the number of themes. As the input texts shown in Figure 2, the cohesion array is shown in Figure 4.

![Fig. 4. Cohesion array between themes](image)

3.2 Sentence Ordering

After the directed graph is built and the cohesion array between themes is calculated, the path $P$ can be attained by the following steps:

1) Set the initial path $P$ as null and calculate the weight of each vertex;
2) Choose the vertex with the greatest weight and insert it into the path $P$, and delete the vertex and all the sides connecting with it in the directed graph, and recalculate the weights of rest vertexes;
3) Search the right vertex which has the largest cohesion with vertex $i$ from all remaining vertices, Assume vertex $j$ is the right vertex which has the largest cohesion $R_{ij}$, if $R_{ij}$ is less than the threshold value, then turn to step 3;
4) Insert vertex $j$ into the after of vertex $i$ in the path $P$ (if the largest cohesion is $R_{ij}$, then insert it into the before of vertex $i$), and delete the vertex and all the sides connecting with it in the directed graph, and recalculate the weights of the rest vertexes;
5) If the graph is empty, then output path $P$; otherwise, $j$ is deemed to current vertex, turn to step 4;
6) Sentence Ordering according to the path $P$.

As the input texts shown in Figure 2, after the decision of themes and extraction intersection sentences for each theme is completed, the sequencing process of the summary generated from the Improved Majority Ordering Algorithm is shown in Figure 5.

![Fig. 5. Sequencing Process of the Improved Algorithm](image)
4 Experiments and Discussion

In order to study which algorithm makes an order cohesive, we collect a corpus including 20 different topics sets by search engine, each topic set has 5-8 documents about the same event. For every set, when the themes are clustered and the intersection sentences are extracted from these themes, we use MO, CO and the improved algorithm to order these sentences respectively. Each algorithm generates 20 summaries.

4.1 Manual Evaluation

To allow every summary can be objectively evaluated, we adopt the method proposed by Barzilay which evaluates the order of summary into three categories: Poor, Fair and Good. A Poor summary is poor readability, but its readability would be significantly improved by reordering its sentences. A Fair summary is the text which makes sense but reordering of some sentences can yield a better readability. Finally, the summary which cannot be further improved by any sentence reordering is considered a Good summary.

On the basis of not reading original text, five experts were asked to give these summaries a score only considering the order in which the information is presented. Through analyzing the experts’ grades of each summary, we find that experts had strong agreement to every summary, Table 1 shows the grades assigned to the summaries using above three algorithms.

<table>
<thead>
<tr>
<th>Ordering Algorithm</th>
<th>Poor</th>
<th>Fair</th>
<th>Good</th>
</tr>
</thead>
<tbody>
<tr>
<td>Majority Ordering</td>
<td>4</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>Chronological Ordering</td>
<td>5</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>Improved Ordering</td>
<td>2</td>
<td>8</td>
<td>10</td>
</tr>
</tbody>
</table>

4.2 ROUGE Evaluation

ROUGE (Recall-Oriented Understudy of Gisting Evaluation) is a collection of measures to automatically evaluate the summaries by comparing them with ideal summaries, without much of human intervention (Chin-Yew Lin, 2004). Firstly, we choose three themes from test corpus randomly; secondly, we invite five experts to sort the extracted sentences for each themes, and the five sort result are regarded as ideal ordering, and five summaries generated using these ideal ordering are ideal summaries; finally, summaries are evaluated automatically from six perspective (ROUGE-1, ROUGE-2, ROUGE-3, ROUGE-4, ROUGE-L and ROUGE-W) which are generated using three ordering algorithms(Majority Ordering, Chronological Ordering and Improved Ordering), the results as shown in Table 2.

4.3 Experimental Analysis

From the experimental data in Table 1 and Table 2, we find that manual evaluation has the same results with ROUGE evaluation, summaries generated through improved sentence ordering algorithm are significantly better than the summary from the methods of MO and CO. From ROUGE-L and ROUGE-W, we find the summary generated using Improved Ordering is more coherent and readable. According to the degree of cohesion between themes, those with high cohesion are placed in an adjacent location in the summary, which could effectively avoid the possibility of decentralization of some themes to the summary during using MO algorithm and greatly enhance the readability of the summary. However, through the analysis of two summaries with poor grade and its related original text, we found that when the theme in the position performs cohesion closely, but not very coherent in the logical sense, the summary is still poor readable, so the method has also some dependence on the structure of each document.
Table 2. Result of Rouge Evaluation of Three Ordering Algorithms.

<table>
<thead>
<tr>
<th>Theme</th>
<th>Ordering Algorithm</th>
<th>ROUGE-1</th>
<th>ROUGE-2</th>
<th>ROUGE-3</th>
<th>ROUGE-4</th>
<th>ROUGE-L</th>
<th>ROUGE-W</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MO</td>
<td>0.37320</td>
<td>0.04646</td>
<td>0.03752</td>
<td>0.00230</td>
<td>0.18822</td>
<td>0.12236</td>
</tr>
<tr>
<td></td>
<td>CO</td>
<td>0.37320</td>
<td>0.04646</td>
<td>0.03755</td>
<td>0.00270</td>
<td>0.10647</td>
<td>0.12348</td>
</tr>
<tr>
<td></td>
<td>AO</td>
<td>0.37320</td>
<td>0.04673</td>
<td>0.03963</td>
<td>0.00766</td>
<td>0.20288</td>
<td>0.13563</td>
</tr>
<tr>
<td>2</td>
<td>MO</td>
<td>0.35317</td>
<td>0.06479</td>
<td>0.01527</td>
<td>0.01029</td>
<td>0.31179</td>
<td>0.10189</td>
</tr>
<tr>
<td></td>
<td>CO</td>
<td>0.35317</td>
<td>0.06479</td>
<td>0.02022</td>
<td>0.01054</td>
<td>0.31326</td>
<td>0.06537</td>
</tr>
<tr>
<td></td>
<td>AO</td>
<td>0.35317</td>
<td>0.06863</td>
<td>0.02510</td>
<td>0.01365</td>
<td>0.41630</td>
<td>0.12065</td>
</tr>
<tr>
<td>3</td>
<td>MO</td>
<td>0.26238</td>
<td>0.03859</td>
<td>0.00993</td>
<td>0.00321</td>
<td>0.22335</td>
<td>0.09325</td>
</tr>
<tr>
<td></td>
<td>CO</td>
<td>0.26238</td>
<td>0.03861</td>
<td>0.00799</td>
<td>0.00489</td>
<td>0.23748</td>
<td>0.09499</td>
</tr>
<tr>
<td></td>
<td>AO</td>
<td>0.26238</td>
<td>0.03936</td>
<td>0.01204</td>
<td>0.01017</td>
<td>0.26303</td>
<td>0.12256</td>
</tr>
</tbody>
</table>

5 Conclusion

In this paper we describe limitations of Chronological Ordering and Majority Ordering, and give a new sentence ordering algorithm in which cohesion between themes and the major order of themes are associated into the sentence ordering. It put these sentences with high cohesion to the adjacent location in the final summary. Experimental results show that summary generated by our improved sentence ordering algorithm is more coherent and more readable.

References


