

A Study of Optimized Summarization Models on Mobile Devices for Sight-Interpreters

Jianhua Wang^{1, a}, Ying Zhou^{2, b}

¹ Renmin University of China, Beijing 100872, China

² Beijing Open University, Beijing 100081, China

^awjhsfl@ruc.edu.cn, ^bjoyzhouying@163.com

Keywords: Optimized Summarization Model, Mobile Devices, Sight-Interpreters.

Abstract. To obtain key information quickly from long texts on small screens of mobile devices, sight-interpreters need to establish optimized summarization model for fast information retrieval. Four summarization models based on previous studies were studied including title+key words (TKW), title+topic sentences (TTS), key words+topic sentences (KWTS) and title+key words+topic sentences (TKWTS). Psychological experiments were conducted on the four models for three different genres of interpreting texts to establish the optimized summarization model for sight-interpreters. This empirical study shows that the optimized summarization model for sight-interpreters to quickly grasp the key information of the texts they interpret is title+key words (TKW) for cultural texts, title+key words+topic sentences (TKWTS) for economic texts and topic sentences+key words (TSKW) for political texts.

Introduction

Sight-interpreting is also called simultaneous-interpreting with text which can be in source-or-target language. Sight-interpreters used to take with them large piles of paper in source-or-target language to fulfill their duty as an interpreter. But, now sight-interpreters prefer to take with them a mobile device such as an Ipad or Mini-Ipad as an auxiliary device to help them finish their tasks of sight-interpreting. Sight-interpreters always interpret the materials from source language into target language or occasionally read the information in target language from mobile screens by keeping pace with the speaker with the proceeding the conference. They do have a strong desire to get the general information of the texts they sight-interpret for immediately after they receive the texts stored in the mobile device they hold with them. So, they need to have an efficient and effective strategy to get as much as possible information from the texts to facilitate their interpreting on work site. This study focuses on finding the optimized information extraction strategy from three different genres of texts sight-interpreters normally encounter.

Sight-interpreting as a type of special simul-interpreting demands high on interpreter's attention distribution and capacity balance studied by Sylvie Lambert^[1]. This study shows that sight-interpreting can be enhanced by interpreting techniques such as learning about the general idea of the interpreting material in advance as well as good efforts coordination. To quickly retrieve the general idea of a text by sight-interpreters, several ways of summarization on large scale of texts have been found and proved to be effective by researchers. The first way to extract English summary can be called sentence-based summarization which includes a multi-documental summarization system developed successfully by Rafael Ferreira et al.^[2] The general idea of a large text is automatically extracted through sentence clustering extraction technique by linguistic treatment on extracting generalization structure categorized by discourse relations studied by Wolf & Gibson^[3]. A centroid-based summarization of topic sentence clusterings was developed by Dragomir R. Radev et al.,^[4] through detecting and retrieving topic of large documents for summarization and archetypal sentences summarization proposed and studied by Ercan Canhasi^[5] to produce effective summaries. The second way for summarization focuses on analyzing English text structural hierarchical features to retrieve general topic sentences by making use of interrelationship between text units including the correlation between units calculated by hierarchical topic tree, the rhetorical relationship and temporal relationship

represented at different levels of structure units studied by Yong-Dong Xu.^[6] Summarization can be made with extracting idea clusters by taking advantage of the hypertext structure for summary proposed and proved by Davida Charney^[7]. The third way to obtain main idea through key word extraction from a single English document through frequency of the key words and the co-occurrence of the words designed and studied by Y. MATSUO and M. Ishizuka^[8]. While, for Chinese summary extraction, key word and sentence extraction for Chinese summary designed by Zheng Yi, Huang Xuanjing, Wu Lide^[9] relies on the words of the highest frequency in the Chinese text and paragraph topic sentences as well as the general topic sentence of the whole text in the technical texts developed by Wang Yongcheng^[10]. In many cases, the title of the text represents the general idea of the whole text. Automatic Chinese summarization was developed by Yang Yongtao^[11] based on feature words as key words with the highest frequency and co-occurrences in the text. General idea can also be automatically extracted on mutual information with topic sentences by Liu Xinghan and Huo Hua^[12]. From the summary extraction review, we can see that for English and Chinese text summarization, there are generally four models for sight-interpreters to retrieve the general idea of the interpreting materials including title+key words, key words+paragraph topic sentences, title+paragraph topic sentences and title+key words+paragraph topic sentences. But for different styles of texts, what is the optimized model for each interpreting material? This study will try to establish the optimized summarization model for different interpreting materials of different styles of texts through psychological experiments. Sight-interpreters can use the model established here to quickly retrieve the general idea of the interpreting materials.

Sight-interpreters are normally faced with interpreting economic texts, political ones as well as cultural ones. Therefore, in this study, we will design experiments to test and establish the optimized summarization model for the three different genres of interpreting materials.

Method

Experiment materials. Twenty-four texts are selected with each genre having eight ones as experiment materials from the BLCU (Beijing Language and Culture University) Chinese study corpus. Three genres for experiment design include economic texts, cultural ones and political ones. All the selected materials are evaluated by two field experts to ensure their similarity in difficulty. The 8 texts for each of the three genres are summarized with 4 different summaries for each text including title+key words (TKW), key words+topic sentences (KWTS), title+topic sentences (TTS) and title+key words+topic sentences (TKWTS).

Experiment Apparatus. All subjects conduct experiment on an Ipad screen with same window size of 190mmX250mm for each summary of the text experimented on. On the right upper corner of the small screen window, there is a button for subjects to press to start the experiment. At the beginning of the experiment, the subjects need to log in their personal information and read the instructions for consent confirmation. The Ipad with the small screen on for the experiment can store the reaction time automatically during the experiment which can be as precise as 1 msec.

Subjects. All the 48 subjects are college senior students from a certain university in Beijing at the spring semester. All the 48 subjects passed Chinese CET-6 and accomplished interpreting tasks at least 3 times. All the subjects are from 19 to 25 years old and have good eyesight strong enough to read the different summaries of the texts displayed on the Ipad small screen. All 48 subjects are divided into three groups for three genres of experiment materials with 16 for cultural texts, 16 for political ones and 16 for economic ones, respectively.

Experiment Design. To ensure the smooth and successful progress of the experiment, 4 students with the same qualification as the subjects selected are recruited to do the pretest to learn about whether the design of the experiment is scientific and needs to be improved or not. After the pretest, the experiment will start as designed. All subjects will sit tightly with an Ipad held on their left hand and press the start button to begin. Each subject can read the summary extracted on the screen only once with no back-page reading during which the reading time on each word as reaction time is stored

automatically by the Ipad. Each subject are asked to do right-or-wrong questionnaire on the summary comprehension they just read. After the right-wrong question, each subject are asked to read the full text of the summary they just read for information percentage evaluation. Subjects need to evaluate what percentage of the full text information was delivered by the summary. For the information percentage evaluation, the 7-degree measurement form needs to be filled in for evaluation.

Taken the whole experiment condition into account, this experiment has a Latin-design for being balanced and scientific. Every subject needs to read 1 of the 4 extracted summaries of the 4 selected texts for experiment and 4 other full texts so altogether each subject needs to read 8 experiment materials in the reaction time collection and 4 full texts for information percentage test. 16 subjects in one group can have 4 cycles of summary reading experiment to form a Latin-design experiment. After Latin-design experiment in each group, subject still needs to read the full text of the summaries extracted for information percentage evaluation. Accordingly, in one group of experiment, 16 subjects need to be recruited as participants for one of the three genre experiments. The detailed Latin-design can be seen in Table 1.

Table 1. Summary Text Presented in Latin Design

Items		Subject1	Subject2	Subject3	Subject4
Group 1 Team 1	One cycle of Latin design	Text1:TTS	Text2:TKW	Text3:TSKW	Text4:TKWTS
		Text5:fulltext	Text6:fulltext	Text7:fulltext	Text8:fulltext
		Text2:TKW	Text3:TSKW	Text4:TKWTS	Text1:TTS
		Text6: fulltext	Text7:fulltext	Text8:fulltext	Text5:fulltext
		Text3:TSKW	Text4:TKWTS	Text1:TTS	Text2:TKW
		Text7:fulltext	Text8:fulltext	Text5:fulltext	Text6: fulltext
		Text4:TKWTS	Text1:TTS	Text2:TKW	Text3:TSKW
Text8:fulltext	Text5:fulltext	Text6:fulltext	Text7:fulltext		
Items		Subject5	Subject6	Subject7	Subject8
Group 1 Team 2	One cycle of Latin design	Text5:TTS	Text6:TKW	Text7:TSKW	Text8:TKWTS
		Text1:fulltext	Text2:fulltext	Text3:fulltext	Text4:fulltext
		Text6:TKW	Text7:TSKW	Text8:TKWTS	Text5:TTS
		Text2:fulltext	Text3:fulltext	Text4:fulltext	Text1:fulltext
		Text7:TSKW	Text8:TKWTS	Text5:TTS	Text6:TKW
		Text3:fulltext	Text4:fulltext	Text1:fulltext	Text2: fulltext
		Text8:TKWTS	Text5:TTS	Text6:TKW	Text7:TSKW
Text4:fulltext	Text1:fulltext	Text2:fulltext	Text3:fulltext		

In order to counter-balance the influence of turning over the pages, the time of reading each page is recorded automatically on the screen. Besides, reaction time for each word is measured in milliseconds for being accurate. In each summary model, there include title, topic sentence or key words of the text under study. Key words are listed on the screen with a small blank space between every. Topic sentences and title of the text for study are listed completely on the small screen. Subjects are asked to read the given information on the screen to get the general information of the whole text through the help of these given informaion. There is no limit on the reading time of each subject. The 4 extracted summary models of 1 text displayed on each small screen with the size 190*250 can be illustrated in Fig. 1.



Figure 1. Demonstration of the Four Summarization Models of one Text

Questionnaire. Subjects will take right-or-wrong questionnaires which were designed for each text with 5 questions in which there are 4 for detailed information understanding tests and 1 for general information understanding tests. In total, each subject will do 20 right-or-wrong questionnaire judgments tests. If the subject does one judgment correct, he will receive 1 score otherwise 0. If all subjects can score 70% correct. The final data processing will only take reaction time into consideration.

Evaluation Measurement Form. For information percentage evaluation, we design a 7-degree evaluation measurement form in which 1 degree representing 10% of full text the information that the summary carries. The pretest conducted on 4 subjects has found that the percentage of information the summary carries displays the range we expected.

Results and Analysis

Reaction Time for Each Word for 4 different Summarization Models for 3 Different Genres.

Figure 2 shows that the reading reaction time for each word shows big differences for the 4 different extracted summaries of the 3 different genres of texts. TKWTS model takes the least time with 398 msec for each word for cultural texts and 300 msec for economic ones. Also, for political texts, TKW takes the least time of 456 msec for each word. While, the longest time for cultural texts is 598 msec with TTS model, 600 msec for economic ones with TSKW and 597 msec for political ones with TTS.

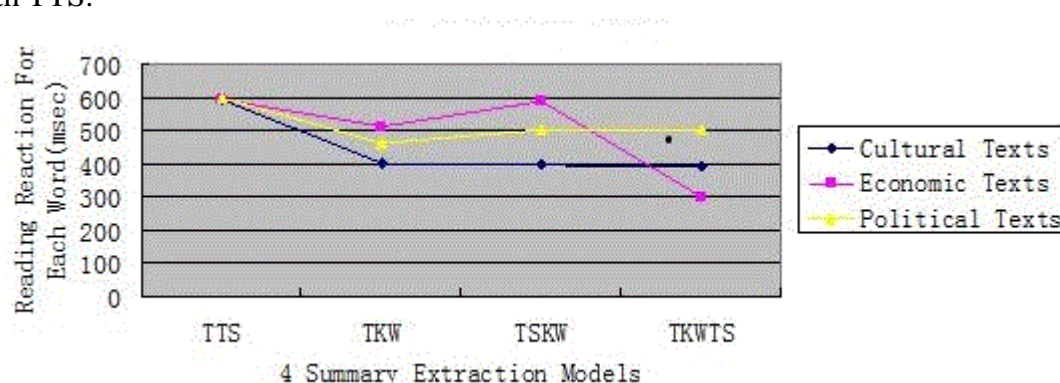


Figure 2. Reaction Time for Each Word in Four Different models of Three Different Genres

Information Percentage from the 7-degree Measurement.

Figure 3 shows that the biggest information percentage for cultural texts is 62% retrieved by subjects from the TKW summarization model, 68% for economic texts from TKWTS model and 67% for political ones from the TKWTS model. While, it is also shown clearly in the above diagram that the smallest information percentage for cultural texts is 52% from the TSKW summarization model, 48 % for political ones from TTS and 46% for economic ones from TTS.

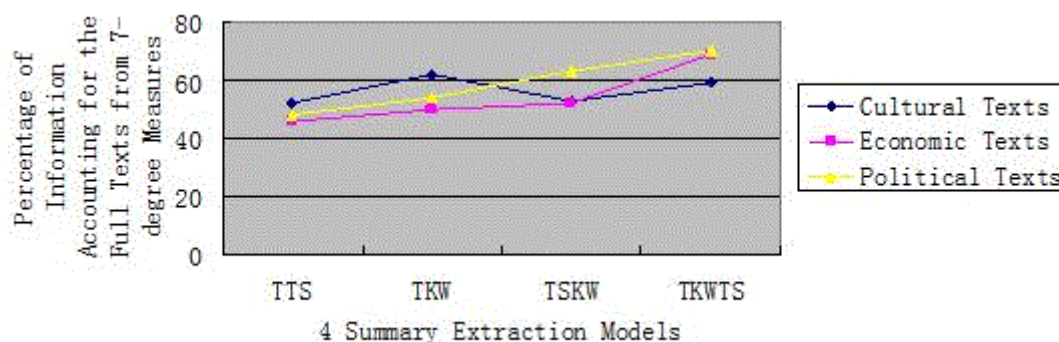


Figure 3. Information Percentage of the Four Different Summarization Models Accounting for the Full Corresponding Texts

Correction Percentage from Right-or-Wrong Questionnaire.

Figure 4 shows the highest correction percentage of right-or-wrong questionnaire for cultural texts is 60% with TKW summarization model, 60% for economic ones with TSKW and 60% for political ones with TTS model. While, the lowest correction percentage of right-or-wrong questionnaire is 40% for economic texts with TKW summarization model, 40% for political ones with TKW model and 50% for cultural ones with TSKW model.

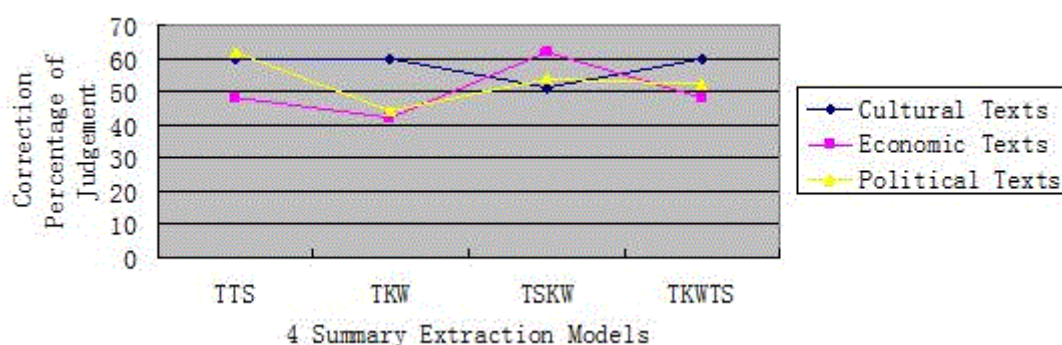


Figure 4. Correction Percentage for Questionnaire Judgement Based on Extracted Summary Readings

Discussions

Results illustrate that different genres of texts have different optimal summarization models.

For cultural texts, TKW is the optimized summary extraction model for sight-interpreters to learn about the key information of the interpreting materials quickly. TKW has the shortest reading reaction time among the four models with 400 msec for each word, the highest information percentage of the four models with 62% and the highest correction percentage of the right-or-wrong questionnaire judgment with 60%.

For economic texts, TKWTS is the optimized summary extraction model. TKWTS has the least reading reaction time among the 4 models being 300 msec for each word, the highest information percentage of the four models being 68% and the highest correction percentage of the right-or-wrong questionnaire judgment being 50%.

For the political texts, TSKW is the optimized summary extraction model. TSKW has the second least reading reaction time among the 4 models being 500 msec for each word, the second highest information percentage of the four models being 63% and the second highest correction percentage of the right-or-wrong questionnaire judgment being 50%.

According to textual structure and development, cultural texts are always written with topic sentences of different paragraphs just subcategorizing the title in a way of sense repetition. So, the title and key words of the texts can convey the key information of cultural texts. For economic texts, topic sentences of different paragraphs play supplementary roles to the title of these texts. Therefore, title,

topic sentences and key words together can convey the general idea and key information of these texts. For political texts, the title is very frequently repeated in the key words of the whole text. As a result, topic sentences and key words combined can transfer the general meaning and key information of the full text. Secondly, these summarization models can enable the sight-interpreters to judge the right-or-wrong questionnaire of the corresponding texts with the highest or second highest correction percentage with the least or next to least reaction time. Taking the three dimensions as reading reaction time, information percentage and correction percentage of the right-or-wrong questionnaire into consideration, we can find that the optimized summarization models for sight-interpreters to use are quite different for the three different genres of texts.

Conclusions

The study on the optimized summarization model is based on experimental studies. Based on previous studies, four different summarization models are tested for the establishment of the optimized model for each of three different genres of texts. Forty-eight subjects participated in the two rounds of Latin-design experiments for data collection. Data processing and analysis show that different genres of texts have different optimized summarization models for sight-interpreters to obtain the key information from long sight-interpreting texts effectively. The results show that title+key words (TKW) is the optimized summarization model for cultural texts, title+key words+topic sentences (TKWTS) for economic texts and topic sentences+key words (TSKW) for political texts. Besides, the establishment of summarization model for different genres of texts has hopeful value for machine translation in human intelligence.

Acknowledgements

This work was financially supported by the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China (15XNL026).

References

- [1] S. Lambert: Attention During Sight Translation, Sight Interpretation and Simultaneous Interpretation. *Meta: Translators' Journal*, 49. P. 294-306. (2004)
- [2] R. Ferreira et al.: A multi-document summarization system based on statistics and linguistic treatment. *Expert Systems with Applications*, 41. P. 5780–5787. (2014)
- [3] F. Wol, E. Gibson: Representing Discourse Coherence: A Corpus-Based Study. *Computational Linguistics*, 31(2). P. 249–288. (2005)
- [4] D. R. Radev, H. Jing, M. Budzikowska: Centroid-based Summarization of Multiple Documents. *Information Processing and Management*, 6. (2003)
- [5] E. Caihasi: Graph-based models for multi-document summarization. A Dissertation. In *Computer and Information Science*, Ljubljana. (2014)
- [6] Y. Xu, X. Zhang, G. Quan, Y. Wang: MRS for Multi-Documnt Summarization by Sentence Extraction. *Telecommun Syst*, 53. P. 91–98. (2013)
- [7] C. Davida: The Impact of Hypertext on Processes of Reading and Writing. New York: Modern Language Association. P. 238-263. (1994)
- [8] Y. MATSUO, M. Ishizuka: Keyword Extraction from a Single Document using Word Co-occurrence Statistical Information. *International Journal on Artificial Intelligence Tools*. (2004)

- [9] Y. Zheng, X. Huang, L. Wu: Research and Application of Automatic Summarization System. Computer Science Research and Development, 11. (2003). In Chinese.
- [10] Y. Wang, G. Chen: Automatic Summarization for Internet Information. Hi-tech Communication. (1999)
- [11] Y. Yang: Algorithm for Automatic Text Summarization. Journal of Chengdu University (Natural Science Edition), 6. P. 143-146. (2009). In Chinese.
- [12] X. Liu, H. Huo: Automatic Summarization for Text Based on Mutual Information. Journal of Hefei University of Technology, 10. P. 1198-1203. (2014). In Chinese.