DATA SUMMARIZER FOR MOBILE DEVICES USING PSEUDO RELEVANCE FEEDBACK WITH SEMANTIC FEATURE

SUN PARK, BYUNGRAE CHA

1,2,3Department of Information and Communication Engineering, GIST, South Korea
E-mail: 1sunpark@gist.ac.kr, 2brcha@gist.ac.kr

Abstract - The proposed method provides an automatic relevance judgment to reformulate query for summarizing data. It also can minimize bias of query expansion by using semantic features of NMF and clustering. Our data summarizer consists of the data summarization module of server side and the mobile device module (i.e., android smartphone) of user side for reducing the overload of mobile device. The experimental results demonstrate that the proposed method achieves better performance than the other query based data summarization methods.

Keywords - Automatic summarizer, document summarization, NMF, PRF, clustering, query expansion, mobile device.

I. INTRODUCTION

As the World Wide Web (WWW) on the internet has been widely used, the amount of information with respect to accessible data has explosively increased and vast information has accumulated. In addition, as mobile devices are more and more popular, people are used to browsing and retrieving information through mobile devices. However, it is difficult to read and understand large numbers of individual information on the Internet by mobile devices which the mobile devices have limited in hardware resources, such as processor, memory, and display screen [1]. It also takes long time to retrieve the large amount of text from the internet by using the mobile devices. In such case, the process of automatic summarization aids the user to help quickly analyzing the key information in the internet contents [2].

Automatic data summarization is the essential technologies for information seeking and condensing goals. Besides, it becomes more and more important in many data applications of electronic and mobile devices [1, 2, 3, 4, 5, 6, 7].

In order to overcome the limitations of the previous data summarization for mobile devices, we propose a query-based data summarization method (i.e., automatic summarizer) using pseudo relevance feedback (PRF) based on semantic features of NMF and K-means clustering method [5, 6, 7].

II. DATA SUMMARIZER

We propose a new automatic data summarizer based on pseudo relevance feedback with semantic features by NMF and clustering method. The proposed summarizer consists of mobile device module of user side, and automatic summarization module of server side. The mobile device module is user interface for data summarization.

The mobile device module carries out the user interface of automatic summarizer for data summarization. User requests summarizing data by the initial query on mobile device. Automatic Summarization Module consists of the preprocessing phase, the PRF (pseudo relevance feedback) phase, and the data summarization phase. The preprocessing phase constructs term-frequency matrix corresponding sentences set of data for calculating the next phases. In PRF phase, subtopics of data are extracted by clustering method, and then query expansion is performed by using the extracted subtopics. The data summarization phase performs the summarizing data by using the expanded query and semantic features of documents by NMF.

CONCLUSIONS

The proposed method has the following advantages: First, it provides an automatic relevance judgment to expand query without the intervention of a user. Second, it can successful decompose semantic feature from extremely different mixing topics of data since the clustering based semantic features of NMF help us to minimize the biased semantic features by means of the clustered topics. Third, it can minimize bias of query expansion since the proposed method uses the clustered topics by K-means clustering. Therefore, it can select more meaningful sentences with respect to the closest sentences to the query. Forth, the method can improve the quality of data summarization since it can minimize a semantic gap between user’s concept for summarization and the vector representation of machine for the query by using the PRF based on NMF and clustering. Finally, in order to overcome the limited processing power capacity of mobile devices, our automatic summarizer is separated by the automatic summarization module of server side and the mobile device module of user side (i.e., android smartphone) regarding user interface.

ACKNOWLEDGMENTS

This work was supported by Basic Science Research Program through the National Research Foundation.
of Korea (NRF) funded by the Ministry of Education, Science and Technology (MEST) (2016R1D1A1B03934823). This research was one of KOREN projects supported by National Information Society Agency (18-951-00-001).

REFERENCES