Chapter 2

The Development of Single-Document Abstractive Text Summarizer During the Last Decade

Amal M. Al-Numai
King Saud University, Saudi Arabia

Aqil M. Azmi
King Saud University, Saudi Arabia

ABSTRACT

As the number of electronic text documents is increasing so is the need for an automatic text summarizer. The summary can be extractive, compression, or abstractive. In the former, the more important sentences are retained, more or less in their original structure, while the second one involves reducing the length of each sentence. For the latter, it requires a fusion of multiple sentences and/or paraphrasing. This chapter focuses on the abstractive text summarization (ATS) of a single text document. The study explores what ATS is. Additionally, the literature of the field of ATS is investigated. Different datasets and evaluation techniques used in assessing the summarizers are discussed. The fact is that ATS is much more challenging than its extractive counterpart, and as such, there are a few works in this area for all the languages.

INTRODUCTION

The digital world has become more complex and crowded with massive volumes of digital data. In 2018, the size of the indexed World Wide Web is over 5.22 billion pages (Kunder, 2018, Dec 15), spread over 1.8 billion websites (Fowler, 2018, Feb 20). As the number of electronic text documents is growing so is the need for an automatic text summarizer. It is difficult to read and gain knowledge from vast number of texts. In some fields, reading and understanding long texts consume time and effort. Consequently, automatic text summarization can be seen as a viable solution which is used in different domains and applications. It can decrease the time taken to summarize huge texts in many areas and media. It extracts and identifies the important information from a text; which can provide concise information with less effort and time. In addition, it can solve the information storage problem by reducing the document's size. Text summarization can support different applications and usage such as news feed, reports abstract, meeting, email and email threads, digest web pages and blogs, recap large amount of web opinions, helping doctors to get an overview about their patients' medical history. Also, students can use the summarization as a helping tool for quick overviewing their studying materials. Web crawler bots can be used to browse the web systematically according to a specific field; news for example, and summarize their contents in a meaningful way. Text summarization can be used in various stand-alone applications or combined with other systems, such as information retrieval, text clustering, data mining applications, web documents and pages, tweet, and opinion summarization.

Automatic text summarization is not a new idea, but there is a huge room for improvement. Simulating how human summaries texts lead to a major innovation in the field of artificial intelligence, abstractive text summarization becomes a necessity in the field of Natural Language Processing. It needs multiple tools to run together in order to extract knowledge and generate a new text. Many researchers have focused on extractive method due to its simplicity. Even though the extractive summarization is quite advanced, there are still researchers working on single document summarization, and those working on multi-document summarization. Now, the abstractive summarization itself is a challenging area. There are few research studies about abstractive summarization in different languages which are still immature due to the difficulties and the challenges of the natural language generation. More effort is necessary to advance this important research field.

In the literature, abstractive text summarization has been applied on several languages; such as English, Arabic, Hindi, Kannada, Malayalam, Telugu, and Vietnamese. Different methods have been conducted to achieve abstractive summary; such as discourse structure, graph-base, semantic-base, linguistic-based, information extraction rules, statistical model, machine learning techniques which include deep

27 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/the-development-of-single-document-abstractive-text-summarizer-during-the-last-decade/235740

Related Content

Restful Web Service and Web-Based Data Visualization for Environmental Monitoring
Sungchul Lee, Ju-Yeon Joand Yoohwan Kim (2016). *Big Data: Concepts, Methodologies, Tools, and Applications (pp. 690-710).*
www.irma-international.org/chapter/restful-web-service-and-web-based-data-visualization-for-environmental-monitoring/150189

Robust Clustering with Distance and Density
Hanning Yuan, Shuliang Wang, Jing Geng, Yang Yuand Ming Zhong (2017). *International Journal of Data Warehousing and Mining (pp. 63-74).*
www.irma-international.org/article/robust-clustering-with-distance-and-density/181884

Fusion Cubes: Towards Self-Service Business Intelligence
Alberto Abelló, Jérôme Darmont, Lorena Etcheverry, Matteo Golfarelli, Jose-Norberto Mazón, Felix Naumann, Torben Pedersen, Stefano Bach Rizzi, Juan Trujillo, Panos Vassiliadisand Gottfried Vossen (2013). *International Journal of Data Warehousing and Mining (pp. 66-88).*
www.irma-international.org/article/fusion-cubes-towards-self-service/78287

A Telecommunications Model for Managing Complexity of Voice and Data Networks and Services
Bahador Ghahramani (2003). *Managing Data Mining Technologies in Organizations: Techniques and Applications (pp. 202-218).*
www.irma-international.org/chapter/telecommunications-model-managing-complexity-voice/25767

Boat Detection in Marina Using Time-Delay Analysis and Deep Learning
Romane Scherrer, Erwan Aulnette, Thomas Quiniou, Joël Kasarherou, Pierre Kolband Nazha Selmaoui-Folcher (2022). *International Journal of Data Warehousing and Mining (pp. 1-16).*
www.irma-international.org/article/boat-detection-in-marina-using-time-delay-analysis-and-deep-learning/298006