A Comparative Study of Sentiment-Based Graphs of Text Summaries

Stephanie Chua and Narayanan Kulathuramaiyer
Faculty of Computer Science and Information Technology
Universiti Malaysia Sarawak
Kota Samarahan, Sarawak 94300, Malaysia
{chlstephanie & nara}@unimas.my

Bali Ranaivo-Malançon and Hazimah Iboi
Faculty of Computer Science and Information Technology
Universiti Malaysia Sarawak
Kota Samarahan, Sarawak 94300, Malaysia
{malanconbali & ihazimah11}@gmail.com

Abstract— Sentiment included in a sentence can indicate whether a sentence may have positive, negative or neutral polarity. Polarity of the sentences is deemed important in text summarization, especially when summarizing narrative texts. This paper proposes to discover the patterns and sentiment scores of the summaries generated by established summarization methods: Luhn, Latent Semantic Analysis (LSA) and LexRank. This is done by conducting a study and comparison on the generated sentiment-based graphs of the summaries. A comparative study is conducted on the sentiment-based graph of the generated summaries with two different sentiment lexicons, namely SentiWordNet and VADER. The analysis is conducted by comparing the patterns of the sentiment-based graph and their sentiment scores as well. In the experiments conducted, there is an obvious pattern for the two sentiment lexicons. This implies that sentiment-based graph's pattern and score are helpful in generating a compact summary. The analysis will alleviate future research on sentiment-based summarization and motivates a new method which can be considered as a graph-based summarization to extract a summary based on its sentiment score.

Keywords - extractive summarization; sentiment analysis; Luhn method; LexRank method; LSA method

I. INTRODUCTION

With the growth in the number of digital documents, there is an important need for text summarization. Text summarization is able to reduce human effort in order to provide the reader with compact and meaningful information in just a few lines of texts. When reading a text, the reader usually tends to skim through the text for the first time to grab the general idea of the text. Hence, a summary can provide just that as needed.

Text summarization can generally be described as the process of generating a summary. A summary is defined as a text that is created from one or more texts, that delivers important information in the original text, and that is no longer than half of the original text [1]. The original text can be very lengthy and can consume more of reader's time to read. Thus, automatic text summarization (ATS) can assist the reader to understand the gist of the text in just a short time by providing a compact summary. ATS is helpful when a meaningful summary is needed from a very lengthy text. The sentiment portrayed in the text is useful as an indicator of the polarity of the sentences.

Sentiment analysis is used to identify the polarity of the text. It has been a popular method in gauging sentiments on the Web and social media. Sentiment analysis distinguishes and extracts subjective or emotion information in texts by using NLP, text analysis and computational linguistics [1]. It focuses on the expressed opinion of a text, disregarding the topic of the text itself.

The question that remains to be answered is whether the pattern of the summary can be extracted using sentiment analysis? This study was undertaken to study and compare the sentiment patterns and scores of the summaries generated by the established summarization methods: Luhn, LSA and LexRank.

II. RELATED WORKS

Text summarization can generally be described as the process of forming a summary out of the textual elements of a text narrative. [2] defined a summary as "a text that is produced from one or more texts, that conveys important information in the original text(s), and that is no longer than half of the original text(s) and usually significantly less than that". The original text can be very long and this may put the casual reader off. Having a summary of an original text usually helps the reader to understand the gist of the text in just a fraction of the time.

The scene of text summarization research has evolved over the years. The earliest works on summarization largely made use of statistical-based techniques based on word frequency [3], [4] and sentence position [4], [5]. These techniques form the foundation of feature extraction in text summarization and are still largely adopted in most text summarization approaches. Subsequently, machine learning and natural language processing techniques for text summarization followed. Machine learning techniques are used for selecting the best feature to extract in text summarization [6], [7], [8], [9] while natural language processing techniques allows elements of the natural language such as text structure [10], concepts in documents [11] and lexical chains [12], [13] to be exploited for text summarization. The major approaches to text summarization are also summarized in [2], highlighting the literature for summarization through extraction and abstraction. More recent approaches to text summarization looks at sentence ordering [14], [15], extracting salient sentences in given document(s) by modelling text