Chapter 4
Data Text Mining Based on Swarm Intelligence Techniques: Review of Text Summarization Systems

Mohamed Atef Mosa
Institute of Public Administration, Department of Information Technology, Riyadh, Saudi Arabia
National Authority for Remote Sensing and Space Sciences, Cairo, Egypt

ABSTRACT

Due to the great growth of data on the web, mining to extract the most informative data as a conceptual brief would be beneficial for certain users. Therefore, there is great enthusiasm concerning the developing automatic text summary approaches. In this chapter, the authors highlight using the swarm intelligence (SI) optimization techniques for the first time in solving the problem of text summary. In addition, a convincing justification of why nature-heuristic algorithms, especially ant colony optimization (ACO), are the best algorithms for solving complicated optimization tasks is introduced. Moreover, it has been perceived that the problem of text summary had not been formalized as a multi-objective optimization (MOO) task before, despite there are many contradictory objectives in needing to be achieved. The SI has not been employed before to support the real-time tasks. Therefore, a novel framework of short text summary has been proposed to fulfill this issue. Ultimately, this chapter will enthuse researchers for further consideration for SI algorithms in solving summary tasks.

DOI: 10.4018/978-1-5225-9373-7.ch004

INTRODUCTION

The need for text summarization has often evolved with the growth of information and publishing on the web. An immense amount of data is being produced day by day as a result of the interactions and interchange of knowledge among users on internet platforms. Which makes it an extremely difficult process to nominate the most relevant, important information. To outdo the problems of information explosion, automatic text summary has become a necessity. The task of summarization minimizes the exertion time required to highlight the most informative sentences. Generally, a summary can be characterized as a text that is generated from one or more texts that reveals the important information in the original text while being short. The field of automatic summarization is over fifty years of age[1].

Since the summarization issue had been introduced a long time ago, it was solved by several algorithms. Nonetheless, few researchers have been persisted for solving text summarization using swarm intelligence (SI) algorithms. This chapter aims to motivate swarm intelligence techniques to be employed in the future for solving summarization tasks which have proven their effectiveness in other several areas. In addition, the authors have highlighted the state-of-the-art papers that have been used in summarizing the content of social media effectively (Mosa, Hamouda, &Marei, 2017a; Gambhir & Gupta, 2017).

On the other hand, as far as we know, there is not any reviews or surveys about automatic summarization using SI techniques presented antecedently. Only some surveys have reviewed some of the approaches that have presented summarization tasks based on other conventional techniques except SI4, 5. It intends to show a general figure of an automatic summary of text by investigating many existing studies that have been developed based on SI techniques. Additionally, the authors have been addressed the utilized evaluation strategies, the outcomes, and the differently related corpora. Besides, they suggested a generic framework for future work.

TYPES OF TEXT SUMMARIZATION

Single/multi-documents and short text are the three important categories of the summary. The task of generating a brief from many documents is more complicated than the extraction the information from the single document. The main problem appears in summarizing several documents together, particularly in a huge amount of short text. Some researchers suggested with regard to the redundancy to pick the sentences that are at the beginning of the paragraph and then measuring the similarity with the later sentences to select the best one (Sarkar, 2019). Therefore, the Maximal Marginal Relevance approach (MMR) is proposed by (Mosa, Hamouda, & Marei,