Chapter 15
Comparative Study Between Two Swarm Intelligence Automatic Text Summaries:
Social Spiders vs. Social Bees

Mohamed Amine Boudia
Dr. Tahar Moulay University of Saida, Algeria

ABSTRACT

This chapter is a comparative study between two bio-inspired approach based on the swarm intelligence for automatic text summaries: Social Spiders and Social Bees. The authors use two techniques of extraction, one after the other: scoring of phrases and similarity that aims to eliminate redundant phrases without losing the theme of the text. While the optimization uses the bio-inspired approach to perform the results of the previous step, the objective function of the optimization is to maximize the sum of similarity between phrases of the candidate summary in order to keep the theme of the text and minimize the sum of scores in order to increase the summarization rate. This optimization will also give a candidate's summary where the order of the phrases changes compared to the original text. For the third and final step concerning choosing a best summary from all candidate summaries generated by optimization layer, the authors opted for the technique of voting with a simple majority.

INTRODUCTION AND PROBLEM

The quantity of electronic textual information increase day after day making more difficult the access to relevant information without the use of special tools. The software and hardware infrastructure to access to the content of information, are an obstacle, and the major problem is the exponential quantity of textual information electronically. This requires the use of more specific tools i.e. access to the content of texts by rapid and effective means has become a necessary task. Current requirements of users continue to increase in addition to the quality of the results, he wants to get these results very quickly.

A summary of a text is an effective way to represent the contents of the texts and allow quick access to their semantic content (Boudia et al, 2016). The purpose of a summarization is to produce an abridged text covering most of the content from the source text. As a matter of fact, a summary of the text has rewritten the text in smaller way under constraint kept the semantics of a document that is minimized entropy semantics to help the reader to identify interesting information for him/her without reading the entire document (Boudia et al, 2016).

Automatic summaries can be used to reduce the search time to find the relevant documents or to reduce the treatment of long texts by identifying the key information e(Boudia et al, 2016). Automatic summaries can be used to reduce the search time to find the relevant documents or to reduce the treatment of long texts by identifying the key information.

To make an automatic summary, the current literature presents three approaches: by extraction, by understanding and by classification. Our current work uses automatic summarization by extraction as it is a simple method to implement that gives good results; Noting that they are three mains techniques in this approach: Scoring, similarity and prototype.

Our study focuses on comparative study between two approaches based on bio-inspired and more precisely a swarm intelligence that we have already proposed: Social spiders for automatic summarization by extraction vs Social bees for automatic summarization by extraction

Before stating in a comparative study, we will give a brief explanation of each method, and then we will identify the experiment environment and present the results of each method. It concludes with a comparison of four methods' using the same corpus, this comparison is based on execution time, confusion matrix for each method, that will be used to construct a podium according to Recall, Precision, F-measure and ROUGE.

The first part contribution of our work is to use two methods of summarization at the same time on the quality of summary. The second part and the most important of this work is the proposition to use a bio-inspired based on the swarm intelligence whether the social spiders or social bees to the automatic summary. We aim to evaluate the impact of them on the quality of summary.

STATE OF THE ART

Automatic summarization appeared earlier as a field of research in computer science from the axis of NLP (automatic language processing), HP Luhn (1958) proposed in 1958 a first approach to the development of automatic abstracts from extracting phrases.

In the early 1960s, HP Edmundson and other participants in the project TRW (Thompson Ramo Wooldridge Inc.) (Edmundson et al, 1960) Proposed a new system of automatic summarization where it combined several criteria to assess the relevance of phrases to extract.

These works were made to identify the fundamental ideas around the automatic summarization, such as problems caused by extraction to build summaries (problems of redundancy, incompleteness, break, etc..), the theoretical inadequacy of the use of statistics, or the difficulties to understand a text (from semantic analysis) to summarize.

Other projects were based on the same technique to build the abstract; otherwise, they were based on the extraction of phrases from the evaluation of the relevance of each of these phrases. Consideration of the evaluating of a phrase was a plus: its relationship with the rest of the text has grown steadily, and

25 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/comparative-study-between-two-swarm-intelligence-automatic-text-summaries/197706

Related Content

Networking E-Learning Hosts Using Mobile Agents
Jon T.S. Quah, Y. M. Chenand Winnie C.H. Leow (2004). *Intelligent Agents for Data Mining and Information Retrieval (pp. 263-294).*
www.irma-international.org/chapter/networking-learning-hosts-using-mobile/24169

Digital Pedagogies of Academic Librarians in the Fourth Industrial Revolution
Mousin Omarsaib, Mogiveny Rajkoomarand Nalindren Naicker (2022). *Innovative Technologies for Enhancing Knowledge Access in Academic Libraries (pp. 247-270).*
www.irma-international.org/chapter/digital-pedagogies-of-academic-librarians-in-the-fourth-industrial-revolution/306441

Interactive IR Framework
Iris Xie (2008). *Interactive Information Retrieval in Digital Environments (pp. 215-262).*
www.irma-international.org/chapter/interactive-framework/24529

Proximity-Based Good Turing Discounting and Kernel Functions for Pseudo-Relevance Feedback
Ilyes Khennakand Habiba Drias (2017). *International Journal of Information Retrieval Research (pp. 1-21).*
www.irma-international.org/article/proximity-based-good-turing-discounting/181723

Hamming Distance based Clustering Algorithm
Ritu Vijay, Prerna Mahajanand Rekha Kandwal (2012). *International Journal of Information Retrieval Research (pp. 11-20).*
www.irma-international.org/article/hamming-distance-based-clustering-algorithm/72704