UNIVERSITY TEKNOLOGI MARA

ANT COLONY ALGORITHM FOR TEXT CLASSIFICATION IN MULTICORE-MULTITHREAD ENVIRONMENT

AHMAD NAZMI BIN FADZAL

Thesis submitted in fulfilment
of the requirements for the degree of
Master of Science

Faculty of Computer and Mathematical Sciences

January 2017

ABSTRACT

In the age of wide digital usage, text classification is one of the significant prominent attribute required in order to automatically arrange emails, articles, and other textual data in an organization. Unclassified data can lead to slower data retrieval thus a reliable method is required to effectively retrieve data efficiently and in systematic manner. Ant Colony Optimization (ACO) is a bio-inspired technique that was introduced to solve Non-Polynomial hard problem of high text data dimension that is similar to Traveling Salesman Problem (TSP) using probabilistic way. Pheromone concept is the main criterion that distinguish ACO to other algorithms. Based on the concept, pheromone saturation is used to combine stackable solution pattern that is discovered while straying to different term node to build a path. ACO classification accuracy is compared to Genetic Algorithm classifier which also a wrapper method. On integration of the technique, ACO is proposed to work in a multicore-multithread environment to gain additional execution time advantage. In multicore-multithread environment, the adjustment aims to make artificial ants communicate across the physical core of processor. As a trade to the investment for more computing power, the execution time reduction is expected to show an improvement without compromising the original classification accuracy. The unthreaded and multicore-multithreaded version of ACO was experimented and compared in term of accuracy and execution time. It was found that the result show a positive improvement.

ACKNOWLEDGEMENT

Firstly, I wish to thank God for giving me the opportunity to embark on my master programme and for completing this long and challenging journey successfully. My gratitude and thanks go to my supervisor Prof Madya Dr. Mazidah Puteh, with co-supervisor, Assoc. Prof. Dr. Adnan Ahmad and Dr. Norlela Samsudin. Thank you for the support, patience and ideas in assisting me with this project.

Finally, this thesis is dedicated to the loving memory of my very dear mother and late father for the vision and determination to educate me. This piece of victory is dedicated to both of you. Alhamdulillah.

TABLE OF CONTENTS