

A Modified ACO-based Search Algorithm for Detecting Protein Functional Module From Protein Interaction Network

Jamaludin Sallim, Rozlina Mohamed, Roslina Abdul Hamid

Abstract— Recent high-throughput experiments have generated protein-protein interaction data on a genomic scale, yielding the complete protein-protein interaction network for several organisms. Various graph clustering algorithms have been applied to protein interaction networks for detecting protein functional modules. Although the previous algorithms are scalable and robust, their accuracy is still limited because of the complex connectivity found in protein interaction networks. The Ant Colony Optimization (ACO) Algorithm has been adapted for the protein functional module detection by modeling the problem as an optimization problem. The adapted ACO (ACO-PFMDA) has obtained feasible solution but not as magnificent as those reported in the literature. Some shortcomings were identified and addressed by proposing a Modified Ant Colony Optimization Algorithm (ACO-PFMDM), which introduces two new scheme for controlling the two main parameters of ACO to solve PFMDP. Experiments on one popular benchmark dataset namely "Saccharomyces cerevisiae" which taken from two popular databases DIP and MIPS has been performed. The experimental result have proved that ACO-PFMDM have improved the overall performance of protein functional module detection. The search process of ACO-PFMDM has converged effectively compared to some state-of-art algorithms. Moreover, the proposed dynamic update of the heuristic parameters based on entropy has generated high quality tours and it can guide ants toward the effective solutions space in the initial search stages.

Keywords- Ant Colony Optimization Algorithm; Searching Algorithm; Protein Functional Module; Protein Interaction Network.

I. INTRODUCTION

Major research efforts have been introduced to the field of bioinformatics in many disciplines such as: sequence alignment, gene finding, gene therapy, gene expression prediction, drug design and discovery, protein structure alignment, protein structure prediction, and protein-protein interactions. The scientific experiment has resulted the huge biological data which need fast data analysis and data management methods for processing them. One of the main approaches of analyzing the biological data is protein functional module detection based on protein-protein interaction data [2]. Protein-protein interaction data are considered as a vital resource in order to determine the organizations of molecular and biological processes. The evidence of interaction indicates both physical association and functional relatedness [3]. The protein-protein interaction data have been accumulated by high-throughput scientific

experimental methods such as mass spectrometry, affinity purification and yeast two-hybrid systems and has been referred collectively as the protein interaction network or *interactome* [4, 5]. Due to the rapid proceeding of the generation of model organism protein interaction network, the demand for analyzing the protein interaction network using advances computational techniques has been increasing and so far, clustering is the most popular approach [6,7]. Nowadays, this field has become a very active in research. There are many clustering algorithms [8] have been used to analyze the protein interaction networks and from this analysis, it can identify the protein complexes, cellular pathways, protein function prediction and functional modules. This thesis has mainly considered protein functional module detection (PFMD) problem. A protein functional module is defined as "a group of proteins that participate in the same biological process or perform the same molecular function while binding each other even at a different time and place" [9]. Though many algorithms are introduced to solve this problem these last few years [10], this research has focused on the optimization algorithms which deal with the protein functional module detection as an optimization problem such as such as Genetic Algorithm (GA), Honey Bees (HB) and Simulated Annealing (SA). Optimization can be defined as the process of finding the best value from many possible values under certain constraints (if any).

II. PROTEIN INTERACTION NETWORK & PROTEIN FUNCTIONAL MODULE DETECTION

A. Protein Interaction Network

Protein clustering (or classification) aims to discover a group of protein in a given data set (PIN) by identifying and quantifying the similarities between the proteins. In this work, we introduce a new approach that applies one of the combinatorial optimization algorithms, Ant Colony Optimization (ACO) together with the classic Traveling Salesman Problem (TSP) approach to cluster the proteins based on protein-protein interaction data which exists in PIN. Since there was no attempt to combine ACO algorithm with TSP approach, we limit this paper to the real issue of applying Ant Colony Optimization algorithm, on how to map the protein interaction network to a representation that can be used by artificial ants to build solutions. In other words, we try to