World Competitive Contests (WCC) algorithm: A novel intelligent optimization algorithm for biological and non-biological problems

Yosef Masoudi-Sobhanzadeh [a,b,*], Habib Motieghader [b]

[a] Department of Computer Engineering, Payam Noor University, Tabriz, Iran
[b] Laboratory of Systems Biology and Bioinformatics (LBB), Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran

ARTICLE INFO

ABSTRACT

Since different sciences face lots of problems which cannot be solved in reasonable time order, we need new methods and algorithms for getting acceptable answers in proper time order. In the present study, a novel intelligent optimization algorithm, known as WCC (World Competitive Contests), has been proposed and applied to find the transcriptional factor binding sites (TFBS) and eight benchmark functions discovery processes. We recognize the need to introduce an intelligent optimization algorithm because the TFBS discovery is a biological and an NP-Hard problem. Although there are some intelligent algorithms for the purpose of solving the above-mentioned problems, an optimization algorithm with good and acceptable performance, which is based on the real parameters, is essential. Like the other optimization algorithms, the proposed algorithm starts with the first population of teams. After teams are put into different groups, they will begin competing against their rival teams. The highly qualified teams will ascend to the elimination stage and will play each other in the next rounds. The other teams will wait for a new season to start. In this paper, we're going to implement our proposed algorithm and compare it with five famous optimization algorithms from the perspective of the following: the obtained results, stability, convergence, standard deviation and elapsed time, which are applied to the real and randomly created datasets with different motif sizes. According to our obtained results, in many cases, the WCC's performance is better than the other algorithms'.

1. Introduction

Since there are a lot of problems with the theory of computational complexities, which need more and more time for being solved, we employ meta-heuristic optimization algorithms. Most of these algorithms can result in suitable solutions with less computational complexities. An evolutionary algorithm's good performance, its stability in different types of execution with minimum errors and its elapsed time to produce acceptable results, are the main criteria that must be considered.

Many different optimization algorithms have been offered and various methods and new algorithms have been proposed to search for the motif discovery sites. Related works are examinable from two aspects: 1- proposing a new optimization algorithm aspect, and 2- the motif discovery aspect. Some of the imminent optimization algorithms which were proposed during the last decades are listed as follows. The genetic algorithm was inspired by living organisms' evolutionary process [1]. A large number of modified genetic algorithms were proposed with a different crossover, mutation, and distinct selection operations [2]. A new optimization algorithm for a single hidden layer of feedback on a neural network, based on the genetic and convex combination algorithms, was introduced by Kwan et al. The most important characteristic of their algorithm was that it could get an acceptable answer in more reasonable time order compared with the others [3]. A new optimization algorithm, known as SOS, was introduced by Min-Yuan et al. to study living organisms. Obtained results proved the SOS's good performance [4]. A combination of imperialist competitive and bee colony algorithms which was recently introduced led to a new method for an optimization problem. Improving a global exploration of a combination of floods of bee colony algorithm is mentioned as method's goals. Moreover, it was applied to two real datasets [5]. The heat transfer optimization algorithm was inspired by both the law of heat transfer and thermodynamics [6]. Another optimization algorithm, inspired by the particle swarm optimization, was recently introduced by Gosciniak et al. Their algorithm was not influenced by the environment, but rather by the particle swarm optimization that was employed as a model of a predator and prey. Moreover, it

* Corresponding author at: Department of Computer Engineering, Payam Noor University, Tabriz, Iran.
E-mail address: yms3786@yahoo.com (Y. Masoudi-Sobhanzadeh).
URL: http://LBB.ut.ac.ir (Y. Masoudi-Sobhanzadeh).

proved that their approach can be practicable in computer games [7]. The particle swarm optimization is an algorithm in which the first populations were created randomly. It is based on the social interaction among flocks of birds or schools of fish. Zhang et al. enhanced the search ability using the Bayesian techniques through exploiting past particles' positions and using a new type of mutation in order to obtain better results. Their approach is known as the BPSO method [8]. Gomez et al. introduced new optimization algorithm for spatial variability analysis and showed high improvement in the NDT measurement position [9].

The other (The second) aspect of related works is the motif discovery. Binding sites, which indicate chemical specificity and affinity, are either part of a protein or part of DNA or RNA in ligands which form a chemical bond. One of the most important components of the biomolecules' general characteristics is binding sites that can identify the other proteins. Binding sites also exist in antibodies as coded regions. A non-covalent bond between the two polypeptide chains is formed when a binding site of one protein determines another protein's surface. Transcriptional factors (TF), are proteins which bind to specific DNA sequences and control the rate of the transcription of genetic information from DNA to RNA. TF are categorized as a sequence of specific DNA binding factors that are of great importance to the regulation of a gene expression. They contain one or more DNA-binding domains that bind to specific sequences of DNA that are contiguous to the genes which they regulate. DNA binding sites are distinguishable from the other algorithms because they include the targets of the other proteins and therefore are part of DNA- in other words, are bound by DNA-binding proteins. DNA binding sites divide into several categories that are as follows: transcriptional factor binding sites, recombination sites, and restriction sites. There is a need for the evolutionary algorithms because the motif discovery problem is an NP (nondeterministic polynomial) hard problem that cannot be solved in polynomial time order. Although these algorithms are not deterministic, they can get acceptable answers in reasonable time order.

We combine the protein motif search processes with the gene discovery methods in which the discovery of human new genes of cadherin superfamily proteins is the main purpose, to identify genes' encoding protein containing specific domains [10]. Proposing a genetic algorithm for the motif discovery based on the

statistical significance of the current research is another recent study carried out, along with the above-mentioned works. After their proposed algorithm was applied to several different datasets, the obtained motifs were compared with each other [11]. Research into the Bcl-2 family protein is among subjects that relate to the present discussion. The family protein comprises several homologous proteins as well as a collection of other proteins which consists of a sequence of similarities except for a Bcl-2 homology (BH)3 motif. One of the four BH motifs is required for membership in the Bcl-2 family. Abdel et al. evaluated Bcl-2 homology motifs [12]. Determining the cellular binding motif in the case of the acute respiratory syndrome coronavirus is essential because the respiratory syndrome coronavirus threatens human lives [13]. Parkinson, Alzheimer, and some other diseases are a result of aging and stress that are among the major causes of the protean damage. According to Bonham-Carter [14], oxidation of amino acids with a high rate of motifs is due to some protean damages. Most of the solutions for bioinformatics NP_hard problems are based on ACO and PSO [15,16]. Some of the other related works that we can refer to are as follows:

Specifications of the monocyte activating motif in the mycobacterium (mycobacterial) tuberculosis [17], motifs retrieval by the Secondary Structure terns occurrences [18], and the interaction of binding motif within the nucleocapsid protein of porcine reproductive and respiratory syndrome virus and the host cellular signaling proteins [19].

2. The proposed algorithm

Proposed optimization algorithm starts with the first population of teams [20]. Every team includes several players with random values. The mentioned teams are placed into different groups based on geographical distance and begin to compete with each other in it. This stage can be considered as a local optimization. Global optimization stage which is elimination stage will begin after grouping matches. Eliminated teams will not be removed from the contests, and they will wait for new season of matches. Here is a flowchart of WCC (Fig. 1).

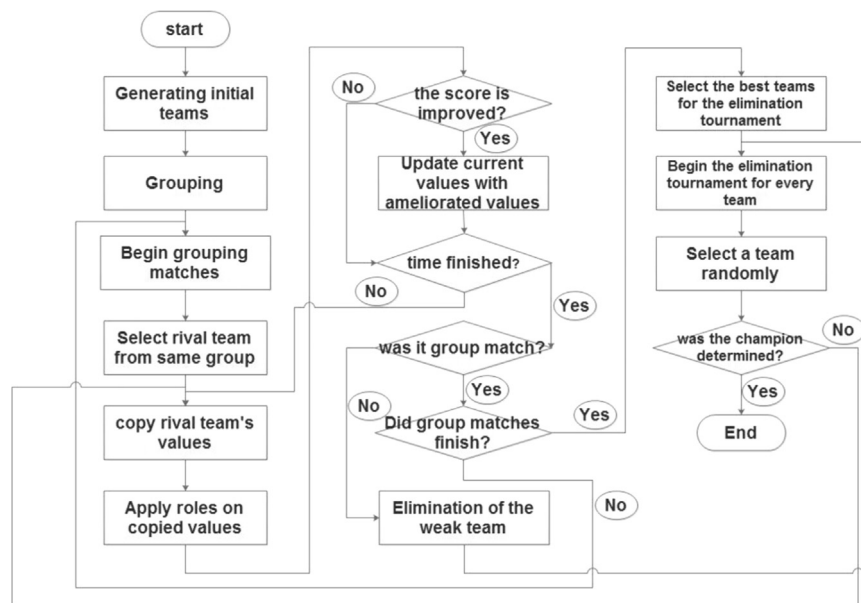WCC includes several main stages:

Fig. 1. The WCC flowchart.