Chemometrics and Intelligent Laboratory Systems 159 (2016) 97–107

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

Hierarchy cuckoo search algorithm for parameter estimation in biological systems

Hojjat Rakhshani[a], Effat Dehghanian[b,*], Amin Rahati[a]

[a] Department of Computer Science, Faculty of Mathematics, University of Sistan and Baluchestan, Zahedan 98135-674, Iran
[b] Department of Chemistry, Faculty of Sciences, University of Sistan and Baluchestan, Zahedan 98135-674, Iran

ARTICLE INFO

Keywords:
Cuckoo search
Hybrid algorithm
Optimization
Biological systems
Parameter estimation

ABSTRACT

Mathematical models play a critical role in describing behavior of biomedical and biological processes. These models have unknown parameters that cannot be measured directly and researchers need reliable tools to estimate them from available experimental data. We present a new modification of cuckoo search (CS) algorithm, named hybrid hierarchical cuckoo search (HHCS), as a powerful computational tool to help with this task. The proposed algorithm uses an information sharing mechanism which allows cuckoo eggs to share their best search information through a heterogeneous hierarchical structure. The HHCS takes advantage of the cooperation between cuckoo eggs to maintain the diversity of solutions and to avoid premature convergence. This approach is also hybridized with a local search method in order to reduce the computational time required for determining the optimal solution. First, the generalization properties of HHCS is evaluated against a set of benchmark functions. Thereafter, we investigate the application of HHCS on two highly non-linear parameter estimation problems of biological systems, namely isomerization of α-Pinene and inhibition of HIV proteinase. Statistical analysis of experimental results shows that the proposed algorithm provides very competitive results compared to CS and several state-of-the-art algorithms in terms of solution accuracy, convergence speed, and robustness. Notably, the HHCS improves the best known solution in the literature for the inhibition of HIV proteinase.

1. Introduction

Modeling of chemical and biological processes has gained wider attention in recent years and various modeling approaches have been employed so far by researchers for this aim, such as mathematical modeling [1], response surface methodology [2], and artificial neural network [3]. Mathematical modeling has shown to be a very reliable method since it develops models based on the correct understanding of reactions that occur in these processes. More precisely, in this modeling approach, the dynamical behavior of chemical or biological processes is predicted by some nonlinear ordinary differential equations (ODEs) with several adjustable parameters. This approach plays an important role in effective implementation of some famous industrial processes such as Fischer-Tropsch synthesis via modeling the chemical kinetic of their chemical reactions [1]. In addition, mathematical models provide a new way towards the analysis of experimental data in medicine and bioinformatics industries. For instance, they can be used in drug studies to investigate the effects of possible new drugs on biochemical pathways and physiology [4,5]. Parameter estimation of these models is of great importance in computational biology.

Generally speaking, parameter estimation can be stated as a numerical optimization problem which aims to find the parameters settings of mathematical models by minimizing residuals between their measured and simulated data. Deterministic approaches like Newton and Levenberq–Maquardt are able to yield accurate solutions within a reasonable computation time [1]. However, these methods have their own limitations. They often rely heavily on the initial guess of parameters settings and topology of the fitness landscape. Also, due to the inherent complexity of chemical and biological processes which is originated from their multi-dimensional and nonlinear characteristics, they are easily trapped in local optimum [6]. In the case of biological systems, such inaccurate methods may lead to the models that are unable to predict precisely the performance of simulated systems[7].

Bio-inspired algorithms, on the other hand, have shown to be very promising and successful approaches for tackling nonlinear optimization problems in natural sciences. For example, Stivala et al. adopted simulated annealing (SA) for accurate protein substructure searching [8]. Interestingly, Krause et al. employed particle swarm optimization (PSO) for online monitoring of biotechnological processes to insure

* Corresponding author.

http://dx.doi.org/10.1016/j.chemolab.2016.10.011
Received 21 June 2016; Received in revised form 20 October 2016; Accepted 20 October 2016
Available online 21 October 2016
0169-7439/

optimal processing as well as sensor failure detection [9]. In another study, PSO has been coupled with a prediction model to identify protein arginine methylation sites [10]. Moreover, Sa'idi et al. applied artificial bee colony algorithm (ABC) for continuous catalytic regeneration reforming process optimization [11]. In addition, genetic algorithm (GA) has been considered for estimating the kinetic parameters of hydrogenation reactions [12], finding kinetic and energetic parameters of chemical models [13], and determining the parameters of Peleg model for water absorption kinetics [6]. Furthermore, differential evolution (DE) [14] has been adopted to estimate kinetic parameters of oxidation in supercritical water [15] and mercury oxidation [16].

Despite the efficiency and wide use of the bio-inspired algorithms, they are very time consuming. Therefore, some researchers believe that proper combination of a deterministic exploiter and a stochastic explorer can increase both the convergence rate and the chance of finding the global optimum, respectively. Thus, they have attempted to solve the parameter estimation problem for dynamic systems using hybrid methods [17,18]. This motivated us to investigate a novel hybrid algorithm which uses the exploration capability of a bio-inspired algorithm and the exploitation ability of a local search method.

As a stochastic explorer, cuckoo search (CS) algorithm [19] can be a good choice due to its simplicity and efficiency. It performs well in searching for global optimum on various types of optimization problems [20,21]. A previous study by Civicioglu and Besdok [22] showed that CS provides more robust results than PSO and ABC. The standard CS algorithm combines the obligate brood parasitism of some cuckoo species with the advantages of Lévy flights and Markov chain random walk. The long jump length distributions provided by Lévy flights is effective for optimization problems where the search space is highly complicated [19]. The common conclusion from all these studies is that CS can be a potential tool for finding initial estimates in large parameter spaces. However, despite all these advantages, CS is also vulnerable to both slow and premature convergence problems. To address these problems, we extend CS by using a heterogeneous hierarchical structure that helps to increase the diversity of solutions through transferring knowledge from the best solutions to other solutions.

As a deterministic exploiter, we incorporate the N2FB local search method via a self-adaptive procedure into the extended CS algorithm. The introduced mechanism provides a complementary searching technique that can reduce the computational time of finding the optimal solution.

The rest of this paper is organized as follows. Section 2 explains the parameter estimation problem and Section 3 reviews the standard CS algorithm. Section 4 presents our proposed hybrid algorithm and its technical details. Thereafter, performance of this algorithm is analyzed in Section 5 by carrying out some numerical experiments on some well-known benchmark functions and comparing the results with several state-of-the-art algorithms. In Section 6, we apply the proposed algorithm to estimate parameters of the isomerization of α-Pinene and the inhibition of HIV proteinase problems such that the simulated responses are as close as possible to the experimental data. Finally, conclusion is presented in the last section.

2. Parameter estimation problem statement

Parameter estimation problem of nonlinear dynamical systems can be stated as minimizing a cost function with respect to some experimental data and differential equality constraints. Given this theoretical definition, the problem can be mathematically formulated as below [4]:

Find $\mathbf{p}$ to minimize:

$$J=\int_0^{t_f}(y_{msd}(t)-y(\mathbf{p},\ t))^T w(t)(y_{msd}(t)-y(\mathbf{p},\ t))dt \tag{1}$$

Subject to:

$$f\left(\frac{dx}{dt},\ \mathbf{x},\ \mathbf{y},\ \mathbf{p},\ \mathbf{v},\ t\right)=0 \tag{2}$$

$$x(t_0)=\mathbf{x}_0 \tag{3}$$

$$h(\mathbf{x},\ \mathbf{y},\ \mathbf{p},\ \mathbf{v})=0 \tag{4}$$

$$g(\mathbf{x},\ \mathbf{y},\ \mathbf{p},\ \mathbf{v})\leq 0 \tag{5}$$

$$\mathbf{p}^{\mathbf{l}}\leq\mathbf{p}\leq\mathbf{p}^{\mathbf{U}} \tag{6}$$

Where J is the cost function to be optimized, $t\in$ R denotes time, $\mathbf{p}$ is the vector of model parameters to be estimated, $\mathbf{y_{msd}}$ is the experimental measure of state variables, $y(\mathbf{p},t)$ is the model prediction for the state variables, $w(t)$ is the weighting matrix, $\mathbf{x}$ is the differential state variables, $\mathbf{v}$ is the vector of time independent parameters, f determines the nonlinear process model, $\boldsymbol{x}_0$ is the initial state. h and g are equality and inequality algebraic constraints. Finally, $\mathbf{p}^{\mathbf{l}}$ and $\mathbf{p}^{\mathbf{U}}$ are upper and lower inequality constraints bounds for $\mathbf{p}$. Standard procedure to solve the aforementioned problem is based on the minimum mean square error criterion. This procedure may be very sensitive to variations in the data [23]. In addition, experimental data inevitably contains some kind of noise. Furthermore, biological models themselves are often highly nonlinear and multimodal. Therefore, there is a great need for further research in this area to achieve more efficient and robust approaches.

3. Cuckoo search

Some cuckoo species like *ani* and *guira* are brood parasites. In other words, they lay their eggs in the nests of other host birds. If a host bird finds such eggs, it will either throw alien eggs away or abandon its nest and build a new one [19]. They reproduce in this fashion to maximize survival rate of their eggs. CS is a stochastic optimization algorithm that mimics this obligate brood parasitism of cuckoo birds. In this model, each egg in a nest shows a solution, each cuckoo egg represents a new solution, and the aim is to replace some of solutions in the host nests with the new and potentially better solutions. The standard CS steps can be expressed as the following three idealized rules [20]:

- Each cuckoo puts one egg (solution) at a time and dumps it in a randomly chosen nest (crossover operator),
- Number of nests that contain eggs with high quality will be transferred to the next generation (elitism),
- The number of available host nests is fixed, and a host can discover alien eggs with probability $p_a\in[0,\ 1]$ (mutation operator).

To generate new solutions in the first rule, CS models a random walk. A random walk is a Markov chain in which the next move only depends on the current location and the transition probability to the next location. Among other potentially interesting approaches, Lévy flights random walk is more efficient for searching than regular random walk or Brownian motions. CS uses Mantegna's algorithm to produce a symmetric Lévy stable distribution. Outline of generating new solutions $\mathbf{x}_\mathbf{i}^{\mathbf{t}+1}$ via Lévy flights can be summarized as below [20]:

$$\mathbf{x}_\mathbf{i}^{\mathbf{t}+1}=\mathbf{x}_\mathbf{i}^{\mathbf{t}}+\mathbf{s}\otimes\mathbf{L\acute{e}vy}(\partial) \tag{7}$$

and

$$\mathbf{s}=s_0\otimes(\mathbf{x}_\mathbf{j}^{\mathbf{t}}-\mathbf{x}_\mathbf{i}^{\mathbf{t}}) \tag{8}$$

$$\mathbf{L\acute{e}vy}(\partial)=\frac{u}{|v|^{\frac{1}{\partial}}} \tag{9}$$

here, $\mathbf{s}$ is the step size parameter, s_0 is the step size scaling factor, $\mathbf{x}_\mathbf{i}^{\mathbf{t}}$ and $\mathbf{x}_\mathbf{j}^{\mathbf{t}}$ are two different solutions, $\otimes$ shows entry-wise multiplications, $\mathbf{L\acute{e}vy}(\partial)$ is the step length generated according to the Mantegna's algorithm, and ∂ is Lévy flights exponent. Also, u and v are two random