Information Sciences 372 (2016) 446–469

Contents lists available at ScienceDirect

Information Sciences

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

Parameter tuning with Chess Rating System (CRS-Tuning) for meta-heuristic algorithms

Niki Veček[a,*], Marjan Mernik[a], Bogdan Filipič[b], Matej Črepinšek[a]

[a] University of Maribor, Faculty of Electrical Engineering and Computer Science, Smetanova ulica 17, 2000 Maribor, Slovenia
[b] Jožef Stefan Institute, Department of Intelligent Systems, Jamova cesta 39, 1000 Ljubljana, Slovenia

ARTICLE INFO

Article history:
Received 2 April 2015
Revised 16 August 2016
Accepted 19 August 2016
Available online 20 August 2016

Keywords:
Parameter settings
Tuning methods
F-Race
Revac
Chess rating system for evolutionary algorithms

ABSTRACT

Meta-heuristic algorithms should be compared using the best parameter values for all the involved algorithms. However, this is often unrealised despite the existence of several parameter tuning approaches. In order to further popularise tuning, this paper introduces a new tuning method CRS-Tuning that is based on meta-evolution and our novel method for comparing and ranking evolutionary algorithms Chess Rating System for Evolutionary Algorithms (CRS4EAs). The utility or performance a parameter configuration achieves in comparison with other configurations is based on its rating, rating deviation, and rating interval. During each iteration significantly worse configurations are removed and new configurations are formed through crossover and mutation. The proposed tuning method was empirically compared to two well-known tuning methods F-Race and Revac through extensive experimentation where the parameters of Artifical Bee Colony, Differential Evolution, and Gravitational Search Algorithm were tuned. Each of the presented methods has its own features as well as advantages and disadvantages. The configurations found by CRS-Tuning were comparable to those found by F-Race and Revac, and although they were not always significantly different regarding the null-hypothesis statistical testing, CRS-Tuning displayed many useful advantages. When configurations are similar in performance, it tunes parameters faster than F-Race and there are no limitations in tuning categorical parameters.

1. Introduction

Current practice of comparing meta-heuristic algorithms is seriously flawed [32] as very often a users preferred algorithm is compared to other, not necessarily state-of-the-art, meta-heuristic algorithms. Furthermore, different parameter settings are experimented to a certain extent for a user's preferred algorithm, whilst for other algorithms included within a comparison parameter settings are more or less arbitrary and often taken from previous papers (despite these settings have not been tuned previously or were used for different problems). This situation was well paraphrased by [15]: "*An EA with good parameter values can be orders of magnitude better than one with poorly chosen parameter values.*" and also discussed in recently published guidelines [7] for replicating and comparing experiments in Evolutionary Computation (EC) where fair comparisons amongst algorithms are promoted. All the algorithms used in the comparisons, not only the users' preferred,

* Corresponding author.
E-mail addresses: niki.vecek@student.um.si, niki.vecek@gmail.com (N. Veček), marjan.mernik@um.si (M. Mernik), bogdan.filipic@ijs.si (B. Filipič), matej.crepinsek@um.si (M. Črepinšek).

http://dx.doi.org/10.1016/j.ins.2016.08.066

should be using the best parameter settings [29]. Hence, performing extensive parameter tuning or control [13,26] for all algorithms involved in an experiment is a prerequisite for a fairer comparison. However, extensive parameter tuning using full factorial design [27] is often too expensive. Therefore, users should apply already-available tuning approaches (e.g., F-Race [2], Revac [35,36], SPO [1], ParamILS [24]) for setting parameters or investigate some parameter control approaches (e.g., driven by diversity [46], entropy [30], exploration and exploitation measures [31], self-adaptation [4,5]). However, current practices often diverge from these guidelines. In this paper we concentrate solely on parameter tuning. It is our speculation that the aforementioned tuning approaches are not yet well accepted amongst practitioners and more experience is needed on how to use them in practice. One attempt in this direction is the recent work [34].

The objectives of this paper are twofold. Firstly, to further report on experiences in using F-Race [2] and Revac [35,36] tuning approaches and, secondly, to propose a new tuning approach called Tuning with Chess Rating System (CRS-Tuning), which is based on the Chess Rating System for Evolutionary Algorithms (CRS4EAs). This is a new method for comparing and ranking evolutionary algorithms [47], which is used within an open-source framework, the Evolutionary Algorithms Rating System (EARS) [12]. Note that comparisons between different meta-heuristic algorithms are inevitably necessary regardless of whether dealing with the scientific testing approach [23] (the aim of which is to learn on which kinds of problems and why one algorithm performs better) or with the horse racing approach [14,40] (the aim of which is to outperform other algorithms). Even during the scientific testing approach, simply understanding parameter interactions and placing emphasis on the analysis of robustness may be insufficient if an algorithm under investigation performs poorly. The obvious question is whether a new tuning approach is needed, in particular, because due to the No Free Lunch theorem [48] no method exists that would solve all possible optimisation problems better than random search. Hence, there is no guarantee that one tuning method will outperform others in all cases. There is, however, a need amongst EC researchers to compare proposed methods for certain types of optimisational problems and appropriate parameter values are always welcomed. Nevertheless, it is one of our desires that the EARS framework, which is a consistent methodology based on chess rating (CRS4EAs), becomes widely used amongst EC researchers. As tuning is an important aspect that should be supported within EARS, in this paper we propose a new tuning approach, which is consistent with CRS4EAs. It can be used also when a new meta-heuristic algorithm is developed and appropriate parameter setting is required.

The paper is organised as follows. Section 2 overviews two common tuning methods: F-Race and Revac. In Section 3 our novel tuning method CRS-Tuning is presented and explained in detail. Section 4 is divided into three parts. The first two describe an extensive experiment the goal of which was to compare CRS-Tuning with F-Race and Revac. The first part shows the experimental results of tuning two parameters of an Artifical Bee Colony, and the second part the results of tuning four parameters of Differential Evolution. The last part of this section briefly presents two additional scenarios that were studied using the CRS-Tuning method, i.e., tuning of the Gravitational Search Algorithm parameters and tuning for practical industrial problems. Section 5 discusses the main findings, and the advantages and disadvantages of all three methods. The paper concludes with Section 6.

2. Related work

Given the algorithm A with k parameters, a tuning method is a procedure that searches for (and finds) the configuration $c = \{p_1, p_2, \ldots, p_k\}$ for which A performs best. The performance is usually referred to as utility or cost and can be measured in different ways. The usual measurement is the mean value, however, median, mode, or different types of rankings can also be used. The numbers and types of parameters depend on the algorithm. The larger the number of parameters to be tuned, the harder the tuning process. The parameter values in EC are distinguishable as either categorical or numerical. The former are usually a function or a procedure implemented within an algorithm (e.g., type of crossover operator), and the latter are usually integer or real values (e.g., population size or crossover rate). Often, these two types of parameters are interrelated (type of crossover operator and value of crossover rate) and the selection of a categorical parameter affects the appropriate value of a numerical parameter. The more interesting and harder to tune are numerical parameters, mainly because the values have to be chosen from infinitely many possibilities (the sizes of $\mathbb{N}$ and $\mathbb{R}$). However, even though there are infinitely many possible values, the numerical parameters are usually limited regarding the lower and the upper bounds, forming an interval from which the best parameter value can be chosen (e.g., crossover rate $\in [0, 1]$). Parameter tuning, which is a complex optimisation problem itself, faces several difficulties: (a) tuning is a time-consuming task, (b) the best parameter values depend on the characteristics of the optimisation problem, (c) parameters are interrelated, and (d) randomness in meta-heuristics usually leads to different outputs even if the initial parameter values are the same. Thus, several different approaches to parameter tuning have been developed. Eiben et al. overviewed the field [15], compared different tuning approaches [41], discussed the costs and benefits of tuning parameters [37], and showed that with parameter tuning the performance of an algorithm can be improved enormously [42,43]. In the meantime, Smit et al. applied a multi-objective approach [44] to parameter tuning. One of the latest overviews was conducted by Montero et al. [34], who compared some of the well-known approaches to parameter tuning for meta-heuristics. They presented five popular tuning methods and applied them to a simple genetic algorithm on a benchmark that consisted of eight minimisation problems. Their results showed that tuning methods are more preferable than using the recommended parameter values. For instance, they showed that even the configuration found by a simple blind search is, on average, significantly better than a default or random configuration. It is explained in [34] that not all tuning methods are appropriate for all types of parameters and that the tuning process itself usually has some parameters that need to be set before the tuning process starts. Our research cannot