Information Sciences 289 (2014) 41–58

Contents lists available at ScienceDirect

Information Sciences

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

Analyzing convergence performance of evolutionary algorithms: A statistical approach

Joaquín Derrac [a], Salvador García [b], Sheldon Hui [c], Ponnuthurai Nagaratnam Suganthan [c,*], Francisco Herrera [d,e]

[a] School of Computer Science & Informatics, Cardiff University, Cardiff CF24 3AA, United Kingdom
[b] Department of Computer Science, University of Jaén, 23071 Jaén, Spain
[c] School of Electrical and Electronic Engineering, Nanyang Technological University, 50 Nanyang Avenue, 639798 Nanyang, Singapore
[d] Department of Computer Science and Artificial Intelligence, University of Granada, 18071 Granada, Spain
[e] Faculty of Computing and Information Technology – North Jeddah, King Abdulaziz University, 21589 Jeddah, Saudi Arabia

ARTICLE INFO

Article history:
Received 21 November 2013
Received in revised form 21 April 2014
Accepted 1 June 2014
Available online 8 August 2014

Keywords:
Page's trend test
Nonparametric tests
Convergence-based algorithmic comparison
Evolutionary algorithms

ABSTRACT

The analysis of the performance of different approaches is a staple concern in the design of Computational Intelligence experiments. Any proper analysis of evolutionary optimization algorithms should incorporate a full set of benchmark problems and state-of-the-art comparison algorithms. For the sake of rigor, such an analysis may be completed with the use of statistical procedures, supporting the conclusions drawn.

In this paper, we point out that these conclusions are usually limited to the final results, whereas intermediate results are seldom considered. We propose a new methodology for comparing evolutionary algorithms' convergence capabilities, based on the use of Page's trend test. The methodology is presented with a case of use, incorporating real results from selected techniques of a recent special issue. The possible applications of the method are highlighted, particularly in those cases in which the final results do not enable a clear evaluation of the differences among several evolutionary techniques.

1. Introduction

An analysis based on final results is the most popular way in which the performance of Computational Intelligence search methods is assessed. For example, in the field of evolutionary optimization, algorithms are usually evaluated with respect to the quality of the best result obtained, over a predefined set of benchmark functions. However, there are other traits of evolutionary algorithms that are worthy of analysis, beyond the quality of the final solution reached: Efficiency, applicability to different domains, diversity management and convergence [2].

Convergence is usually acknowledged to be a desirable capability for every new search algorithm designed today. In the case of Evolutionary Algorithms (EAs), this is a staple concern in the sense that good convergence is a must-have for any new technique to be accepted by the research community [4,7,30]. However, it is common to see convergence analyzed only as the *capability of the technique to reach the final*, regardless of how quickly such a result is reached.

* Corresponding author. Tel.: +65 6790 5404; fax: +65 6793 3318.
E-mail addresses: jderrac@decsai.ugr.es (J. Derrac), sglopez@ujaen.es (S. García), SHUI1@e.ntu.edu.sg (S. Hui), epnsugan@ntu.edu.sg (P.N. Suganthan), herrera@decsai.ugr.es (F. Herrera).

http://dx.doi.org/10.1016/j.ins.2014.06.009
0020-0255/© 2014 Published by Elsevier Inc.

In this sense, the development of a methodology to assess the convergence performance of several algorithms – that is, which algorithm converges faster – is important, particularly in cases in which a benchmark problem is unable to differentiate algorithms using the final results achieved.

The conclusions obtained after analyzing the final results of the algorithms are often backed up by using statistical techniques. Nonparametric tests [8,22] are preferred for this task due to the absence of strong limitations regarding the kind of data to analyze (in contrast with parametric tests, for which the assumptions of normality, independence and homoscedasticity of the data are necessary for the sake of reliability) [18,15,31,41].

Throughout this paper, we show how Page's trend statistical test [27] can be applied to the analysis of pairwise convergence. It is a nonparametric test for multiple classification, which allows trends to be detected among the results of the treatments if the null hypothesis of equality is rejected. In our case, if the treatments are chosen as the differences between the fitness values of two algorithms, computed at several points of the run (*cut-points*), the test can be used to detect increasing and decreasing trends in the differences as the search goes on. The study of these trends, representing the evolution of the algorithms during the search, enables us to develop a new methodology for comparing algorithms' convergence performance.

The description of our approach is completed with the inclusion of an alternative version for computing the ranks of the test. This second version allows the test to be applied safely should one of the algorithms reach the optimum of some of the benchmark functions before the end of the run (which would prevent it from progressing further, thereby preventing the proper evaluation of its convergence in the last stages of the search).

To demonstrate the usefulness of both the basic and the alternative versions of the test, a full case study is presented. The study compares the performance of several EAs for continuous optimization, namely advanced versions of the Differential Evolution evolutionary technique [32,11]. It is based on the submissions accepted for the Special Issue on Scalability of Evolutionary Algorithms and other Metaheuristics for Large Scale Continuous Optimization Problems [21] in the Soft Computing journal.

As will be shown in the study, the use of Page's trend test can be very useful when analyzing the performance of the algorithms throughout the search. Its use provides the researchers with a new perspective for assessing how the algorithms behave, considering intermediate results instead of just the final results in each function. This can reveal very illustrative information when comparing the methods, particularly in cases where the final results are statistically similar.

A further contribution presented in this work is the development of a Java program to implement our approach. The program processes the intermediate results of two or more algorithms. After that, Page's trend test is carried out for every pair of algorithms, and the results are output in TeX format. It can be downloaded at the following URL: http://sci2s.ugr.es/sicidm/pageTest.zip.

The rest of this paper is organized as follows: Section 2 provides some background regarding the use of nonparametric tests to contrast the results of evolutionary optimization experiments. Section 3 presents our approach, detailing how Page's trend test can be applied to compare the convergence performance of two algorithms. Section 4 describes the case study chosen to illustrate the application of the test. Section 5 presents the results obtained and the related discussions. Section 6 concludes the paper. Three appendices are also included, respectively providing a guide to obtaining and using the software used to run the test (A), detailed final results of the case of study (B) and the full results of the application of Page's trend test (C).

2. Background

The assessment of the performance of algorithms is an important task when performing experiments in Computational Intelligence. When comparing EAs, it is necessary to consider the extent to which the *No Free Lunch* theorem [39] limits the conclusions: Under no specific knowledge, any two algorithms are equivalent when their performance is averaged across all possible problems.

Therefore, assuming that EAs take advantage of the available knowledge in one way or another, it is advisable to focus interest on efficiency and/or effectiveness criteria. When theoretical developments are not available to check such criteria, the analysis of empirical results can help to discern which techniques perform more favorably for a given set of problems.

In the literature, it is possible to find different viewpoints on how to improve the analysis of experiments [23]: The design of test problems [13] (for example, the design of complex test functions for continuous optimization [14,38]), the use of advanced experimental design methodologies (for example, methodologies for adjusting the parameters of the algorithms depending on the settings used and results obtained [1,2] or for performing Exploratory Landscape Analysis [3,26]) or the analysis of the results [9] (to determine whether the differences between algorithms' performances are significant or not). Another example is [35], where a method inspired in chess rating systems is adapted to rank the performance of evolutionary algorithms.

From the statistical analysis perspective, the use of statistical tests enhances the conclusions drawn, by determining whether there is enough *evidence* to reject null hypotheses based on the results of the experiments. For this task, it is possible to find applications of both parametric [29,10] and, more recently, nonparametric [18,24,12] statistical procedures.

Nonparametric tests are used to compare algorithms' final results, represented as average values for each problem (using the same criterion: average, median, etc. over the same number of runs for each algorithm and problem). This usually enables practitioners to rank differences among algorithms and determine which ones are significant, thus leading to a characterization of which algorithms behave better than the rest.

However, a drawback of this methodology is that it only takes into consideration the final results obtained at the end. When analyzing EAs, this often overshadows interesting conclusions which could be drawn by analyzing the performance of the algorithms during the whole run.