Swarm and Evolutionary Computation 1 (2011) 3–18

Contents lists available at ScienceDirect

Swarm and Evolutionary Computation

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

Invited paper

A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms

Joaquín Derrac [a,*], Salvador García [b], Daniel Molina [c], Francisco Herrera [a]

[a] Department of Computer Science and Artificial Intelligence, CITIC-UGR (Research Center on Information and Communications Technology), University of Granada, 18071 Granada, Spain

[b] Department of Computer Science, University of Jaén, 23071 Jaén, Spain

[c] Department of Computer Engineering, University of Cadiz, 11003 Cadiz, Spain

ARTICLE INFO

Article history:
Received 18 October 2010
Received in revised form
22 December 2010
Accepted 8 February 2011
Available online 18 February 2011

Keywords:
Statistical analysis
Nonparametric statistics
Pairwise comparisons
Multiple comparisons
Evolutionary algorithms
Swarm intelligence algorithms

ABSTRACT

The interest in nonparametric statistical analysis has grown recently in the field of computational intelligence. In many experimental studies, the lack of the required properties for a proper application of parametric procedures – independence, normality, and homoscedasticity – yields to nonparametric ones the task of performing a rigorous comparison among algorithms.

In this paper, we will discuss the basics and give a survey of a complete set of nonparametric procedures developed to perform both pairwise and multiple comparisons, for multi-problem analysis. The test problems of the CEC'2005 special session on real parameter optimization will help to illustrate the use of the tests throughout this tutorial, analyzing the results of a set of well-known evolutionary and swarm intelligence algorithms. This tutorial is concluded with a compilation of considerations and recommendations, which will guide practitioners when using these tests to contrast their experimental results.

1. Introduction

In recent years, the use of statistical tests to improve the evaluation process of the performance of a new method has become a widespread technique in computational intelligence. Usually, they are employed inside the framework of any experimental analysis to decide when one algorithm is considered better than another. This task, which may not be trivial, has become necessary to confirm whether a new proposed method offers a significant improvement, or not, over the existing methods for a given problem.

Statistical procedures developed to perform statistical analyses can be categorized into two classes: parametric and nonparametric, depending on the concrete type of data employed [1]. Parametric tests have been commonly used in the analysis of experiments in computational intelligence. Unfortunately, they are based on assumptions which are most probably violated when analyzing the performance of stochastic algorithms based on computational intelligence [2,3]. These assumptions are known as independence, normality, and homoscedasticity. To overcome this problem, our interest is focused on nonparametric statistical procedures, which provide to the researcher a practical tool to use when the previous assumptions cannot be satisfied, especially in multi-problem analysis.

In this paper, the use of several nonparametric procedures for pairwise and multiple comparison procedures is illustrated. Our objectives are as follows.

- To give a comprehensive and useful tutorial about the use of nonparametric statistical tests in computational intelligence, using tests already proposed in several papers of the literature [2–5]. Through several examples of application, we will show their properties, and how the use of this complete framework can improve the way in which researchers and practitioners contrast the results achieved in their experimental studies.
- To analyze the lessons learned through their use, providing a wide list of guidelines which may guide users of these tests when selecting procedures for a given case of study.

For each kind of test, a complete case of application is shown. A contest held in the CEC'2005 special session on real parameter optimization defined a complete suite of benchmarking functions (publicly available; see [6]), considering several well-known domains for real parameter optimization. These benchmark functions will be used to compare several evolutionary and swarm intelligence continuous optimization techniques, whose differences will be contrasted through the use of nonparametric procedures.

* Corresponding author. Tel.: +34 958 240598; fax: +34 958 243317.
E-mail addresses: jderrac@decsai.ugr.es (J. Derrac), sglopez@ujaen.es (S. García), daniel.molina@uca.es (D. Molina), herrera@decsai.ugr.es (F. Herrera).

2210-6502/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.swevo.2011.02.002

To do so, this paper is organized as follows. Section 2 shows the experimental framework considered for the application of the statistical methods and gives some preliminary background. Section 3 describes the nonparametric tests for pairwise comparisons. Section 4 deals with multiple comparisons by designating a control method, whereas Section 5 deals with multiple comparisons among all methods. Section 6 surveys several recommendations and considerations on the use of nonparametric tests. Finally, Section 7 concludes this tutorial.

2. Preliminaries

In this section, the benchmark functions (Section 2.1) and the evolutionary and swarm intelligence algorithms considered for our case of study (Section 2.2) are presented. Furthermore, some basic concepts on inferential statistics are introduced (Section 2.3), providing the necessary background for properly presenting the statistical procedures included in this tutorial.

2.1. Benchmark functions: CEC'2005 special session on real parameter optimization

Thorough this paper, the results obtained in a experimental study regarding 9 well-known algorithms and 25 optimization functions will be used, illustrating the application of the different statistical methodologies considered. The nonparametric tests will be used to show significant statistical differences among the different algorithms of the study.

As benchmark suite, we have selected the 25 test problems of dimension 10 that appeared in the CEC'2005 special session on real parameter optimization [6]. This suite is composed of the following functions.

- 5 unimodal functions
  - F1: Shifted Sphere Function.
  - F2: Shifted Schwefel's Problem 1.2.
  - F3: Shifted Rotated High Conditioned Elliptic Function.
  - F4: Shifted Schwefel's Problem 1.2 with Noise in Fitness.
  - F5: Schwefel's Problem 2.6 with Global Optimum on Bounds.
- 20 multimodal functions
  - 7 basic functions.
    - F6: Shifted Rosenbrock's Function.
    - F7: Shifted Rotated Griewank Function without Bounds.
    - F8: Shifted Rotated Ackley's Function with Global Optimum on Bounds.
    - F9: Shifted Rastrigin's Function.
    - F10: Shifted Rotated Rastrigin's Function.
    - F11: Shifted Rotated Weierstrass Function.
    - F12: Schwefel's problem 2.13.
  - 2 expanded functions.
    - F13: Expanded Extended Griewank's plus Rosenbrock's Function (F8F2)
    - F14: Shifted Rotated Expanded Scaffers F6.
  - 11 hybrid functions. Each one (F15 to F25) has been defined through compositions of 10 out of the 14 previous functions (different in each case).

All functions are displaced in order to ensure that their optima can never be found in the center of the search space. In two functions, in addition, the optima cannot be found within the initialization range, and the domain of search is not limited (the optimum is out of the range of initialization).

2.2. Evolutionary and swarm intelligence algorithms

Our main case of study consist of the comparison of performance between 9 continuous optimization algorithms. Their main characteristics are described as follows.

- **PSO**: A classic Particle Swarm Optimization [7] model for numerical optimization has been considered. The parameters are $c1 = 2.8$, $c2 = 1.3$, and w from 0.9 to 0.4. Population is composed by 100 individuals.
- **IPOP-CMA-ES**: IPOP-CMA-ES is a restart Covariant Matrix Evolutionary Strategy (CMA-ES) with Increasing Population Size [8]. This CMA-ES variation detects premature convergence and launches a restart strategy that doubles the population size on each restart; by increasing the population size, the search characteristic becomes more global after each restart, which empowers the operation of the CMA-ES on multi-modal functions. For this algorithm, we have considered the default parameters. The initial solution is uniform randomly chosen from the domain, and the initial distribution size is a third of the domain size.
- **CHC**: The key idea of the CHC algorithm [9] concerns the combination of a selection strategy with a very high selective pressure and several components inducing a strong diversity. In [10], the original CHC model was extended to deal with real-coded chromosomes, maintaining its basis as much as possible. We have tested it using a real-parameter crossover operator, BLX-α (with $\alpha = 0.5$), and a population size of 50 chromosomes.
- **SSGA**: A real-coded Steady-State Genetic Algorithm specifically designed to promote high population diversity levels by means of the combination of the BLX-α crossover operator (with $\alpha = 0.5$) and the negative assortative mating strategy [11]. Diversity is favored as well by means of the BGA mutation operator [12].
- **SS-arit & SS-BLX**: Two instances of the classic Scatter Search model [13] have been included in the study: the original model with the arithmetical combination operator, and the same model using the BLX-α crossover operator (with $\alpha = 0.5$) [14].
- **DE-Exp & DE-Bin**: We have considered a classic Differential Evolution model [15], with no parameter adaptation. Two classic crossover operators proposed in the literature, *Rand/1/exp*, and *Rand/1/bin*, are applied. The *F* and *CR* parameters are fixed to 0.5 and 0.9, respectively, and the population size to 100 individuals.
- **SaDE**: Self-adaptive Differential Evolution [16] is a Differential Evolution model which can adapt its *CR* and *F* parameters for enhance its results. In this model, the population size has been fixed to 100 individuals.

All the algorithms have been run 50 times for each test function. Each run stops either when the error obtained is less than 10^{-8}, or when the maximal number of evaluations (100 000) is achieved. Table 1 shows the average error obtained for each one over the 25 benchmark functions considered.

2.3. Some basic concepts on inferential statistics

Single-problem and multi-problem analyses can usually be found contrasting the results of computational intelligence experiments, both in isolation [17] and simultaneously [18]. The first kind, single-problem analysis, deals with results obtained over several runs of the algorithms over a given problem, whereas multi-problem analysis considers a result per algorithm/problem pair.

Inside the field of inferential statistics, hypothesis testing [19] can be employed to draw inferences about one or more populations from given samples (results). In order to do that, two hypotheses, the null hypothesis H_0 and the alternative hypothesis H_1, are defined. The null hypothesis is a statement of no effect or no difference, whereas the alternative hypothesis represents the presence of an effect or a difference (in our case, significant differences between algorithms). When applying a statistical procedure to reject a hypothesis, a level of significance α is used to determine at which level the hypothesis may be rejected.