Comparison of Non-Deterministic Iterative Methods

*Éric D. Taillard**

* Computer Science Institute, EIVD, University of Applied Sciences of Western Switzerland
Route de Cheseaux 1, CH-1400 Yverdon-les-Bains, Switzerland
Email: Eric.Taillard@eivd.ch http://www.eivd.ch/ina/taillard

1 Introduction

Comparing two (or more) heuristic methods based on metaheuristic principles is a difficult task that is not solved yet in a satisfactory way. Indeed these methods are iterative, meaning that the longer they run, the better the solution they produce are. They are also almost all non-deterministic, since they use a pseudo-random number generator. This means that when running twice the same heuristic method, it is possible to obtain two different solutions. There are few deterministic taboo searches (all the other metaheuristic are based on probabilistic choices), but it would be easy to make them non deterministic since they made arbitrary choices, such as the order in which neighbour solutions are examined, the neighbour solution chosen in case of equivalent evaluation, etc.

It is clear that comparing iterative methods must be done considering both the quality of the solution produced and the computational effort. The last is traditionally the running time on a given machine. Unfortunately, this measure both not precise and volatile. Indeed, the running time strongly depends on the programming style, on the compiler, on the compiling options, etc. Moreover, the lifetime of computer equipment is very limited. Sometimes, it is already obsolete when the articles describing a new method are published!

If the combinatorial problem consists of optimizing a unique objective function, it is easy to compare the quality of two solutions. However, the comparison of non deterministic methods is not so easy since the solution quality not only depends on the computational effort but also on the pseudo-random number generator seed. The aim of this article is to present some tools for comparing non-deterministic iterative methods.

2 Numerical experiments

Let us suppose that heuristic method $\boldsymbol{A}$ (respectively : heuristic method $\boldsymbol{B}$ that has to be compared to $\boldsymbol{A}$) has been run n_A (respectively : n_B) times. During these runs, each improvement of solution quality is stored with the corresponding computational effort in such a way that, for a given time t, it is possible to know the quality of the $n_A + n_B$ solutions produced by both method. An intermediate question to be answered is to know whether $\boldsymbol{A}$ is better than $\boldsymbol{B}$ at time t.

The solution quality obtained by method $\boldsymbol{A}$ (respectively $\boldsymbol{B}$) at time t can be considered as a random variable $X_A(t)$ (respectively $X_B(t)$). The probability density of these random variables being $f_A(t)$ (respectively $f_B(t)$). Typically, one wants to compare the mathematical expectations $E(X_A(t))$ and $E(X_B(t))$ in order to know which is the lowest and therefore decide which method is best for time t. Unfortunately, $f_A(t)$ and $f_B(t)$ are unknown. Moreover, the number of runs n_A and n_B are typically limited to few dozens. So, it is not possible to try to determine $f_A(t)$ and $f_B(t)$ with the data available. Supposing (like often done in literature) that these curves are Gaussian is abusive (for instance, one knows that these functions are bounded by the value of the global optimum). Therefore, a comparison of both methods on the base of average and standard deviation of solution values observed is not valid in general.

Since there is no statistical test for comparing the average of two samples with arbitrary distributions, other hypothesis tests have to be used. The Mann-Whitney test allows an interesting comparison: Translated to the comparison of non deterministic heuristic method, the null hypothesis is $f_A = f_B$. If the probability that this hypothesis is true is lower than a given confidence level α, considering the n_A and n_B solutions obtained, then the null hypothesis is rejected and the alternate hypothesis is accepted, i.e. the probability that $\boldsymbol{A}$ produces a solution better than $\boldsymbol{B}$ is larger (or lower) than the probability that $\boldsymbol{B}$ produces a solution better than $\boldsymbol{A}$. If it is known that both distributions f_A and f_B are the same but translated, the test is also valid for the stronger hypothesis: $E(X_A) = E(X_B)$.

Very shortly, the Mann-Whitney works as follows: first, the $n_A + n_B$ solutions are ranked by decreasing quality and a rank between 1 and $n_A + n_B$ is attributed to each solution. If many solutions are the same, they all got the same rank, i.e. the average of the rank they would have obtained if there were no equality. Then, a statistic S_A is computed. This statistic corresponds to the sum of the ranks of solutions issued from method $\boldsymbol{A}$. If $S_A > T_\alpha(n_A, n_B)$, then the null hypothesis is rejected with a confidence level of α and it is accepted that $\boldsymbol{A}$ is worse than $\boldsymbol{B}$. The values of $T_\alpha(n_A, n_B)$ can be found in numeric tables; there are also books that provides tables that give α as a function of S_A and $n_A + n_B$. More details on the Mann-Whitney test can be found in [1] for instance.

Naturally, this test has to be repeated for different values of computational time t. A convenient way to compare methods $\boldsymbol{A}$ and $\boldsymbol{B}$ is to graphically draw the probability that $\boldsymbol{A}$ is better than $\boldsymbol{B}$ as a function of the computational effort. Naturally, such a comparison is possible only if reliable computational times can be obtained. However, it could happen that one method could be implemented in such a way that it runs faster. So, it is convenient to be particularly careful when comparing computing times, for instance by considering a multiplicative factor of security in the measure of computing times.

Finally, as mentioned above, the computing times is an essential criterion when comparing heuristic methods, but this criterion is volatile. If one wants to be independent from the computer used for presenting the comparisons, it is required to consider an absolute computational effort, such as the number of iterations. In case the mathematical complexity of one iteration can be derived, a much more reliable computational measure can be obtained.

Figure 1 illustrates these principles for comparing two iterative heuristic methods. The horizontal axis has a double gradation: computing time and iteration number. On the top diagram, the vertical axis indicates the average solution value that has to be minimized. On the bottom diagram, the vertical axis provides the probability that a method is better than the other on the base of a Mann-Whitney test. The confidence levels are indicated by two horizontal lines. This second diagram allows to immediately know whether one method (those in bold on the top diagram) is significantly better than the other. In such a case, the probability level must be higher than the higher confidence line. If the bold method is significantly worse than the other, the probability level must be lower than the lower confidence line. In case the probability is between the confidence lines, the comparison is not significant. In addition, a probability level in bold indicates that the test would be also significant even if one method could be implemented in such a way that it runs twice as fast.

With such a figure, it is possible to draw more reliable conclusions than those that can be drawn from a traditional table. For the example presented in Figure 1, when presenting the computational result with a table and by choosing an appropriate computational effort, it could be derived that FANT [3] is better than MMAS [2] — or vice versa! If the results are given under the form of the first diagram, a conclusion could be that FANT is better than MMAS for computational times higher than 1 second. However, this conclusion — which is not necessarily false — cannot be formally derived for the numerical experiments performed.

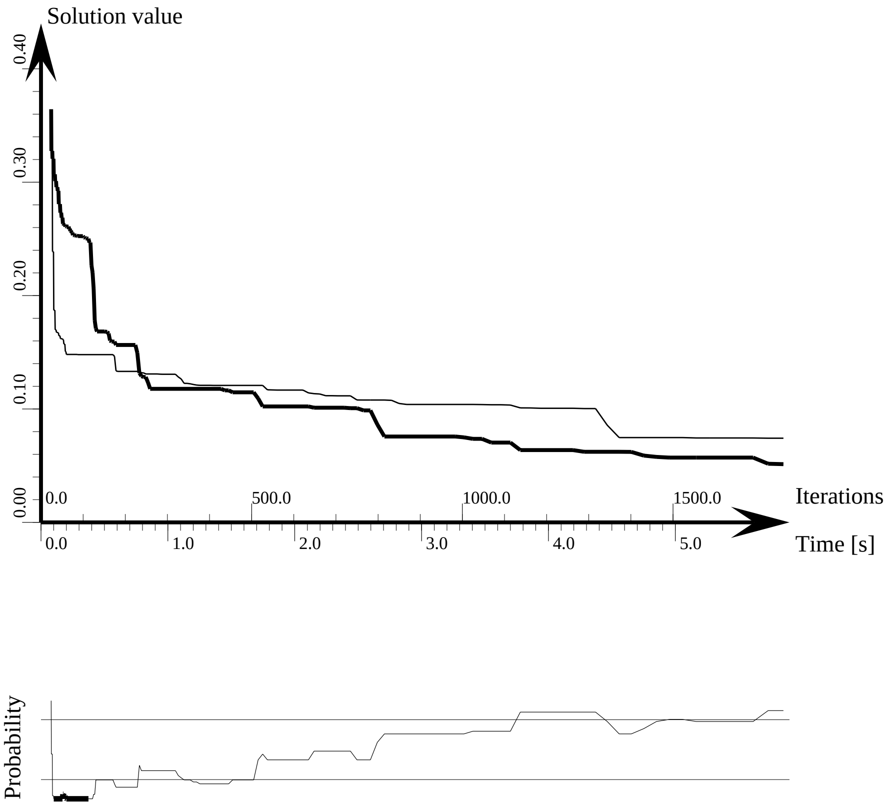

Figure 1: Comparison of heuristic methods FANT and MMAS for quadratic assignment problem instance bur26a. The higher diagram provides the average solution value produced by both method on 10 independent runs as a function of the computational time and number of iterations performed by a local search included in both methods. The lower diagram provides the probability that FANT is better than MMAS. significantly better than FANT. A probability in boldface indicates that the comparison would remain significant even if one method would run twice as fast.

Figure 1 has been automatically generated with a software programmed at the EIVD. A first version of this software is available on the web page: `http://www.eivd.ch/ina/taillard`.

3 Extensions

By drawing only probability diagrams, it is possible to compare many methods on the same figure. It is also possible to compare two heuristics executed on many problem instances. Such diagrams can also be very useful for parameter tuning.

References

[1] W. J. Conover. *Practical Nonparametric Statistics, 3. ed.* Wiley, 1999.

[2] Thomas Stützle. The Max-Min Ant System and Local Search for Combinatorial Optimization Problems. in S. Voss, S. Martello, I.H. Osman and C. Roucairol, *Meta-heuristics: Advances and Trends in Local Search paradigms for Optimization* Kluwer, 1999, 313–329.

[3] Éric D. Taillard. An Introduction to Ant Systems. in M. Lagura and J. L. Gonzalez-Velarde *Computing tools for modeling, optimization and simulation* Kluwer, 2000, 131–144.