A Survey of Performance Assessment for Multiobjective Optimizers

Peng Cheng[1,2], *Jeng-Shyang Pan*[2], *Li Li*[1], *Yan Tang*[1], *Chunlun Huang*[1]

1) College of Computer and Information Science, Southwest University, Chongqing, 400715, China
2) Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, 518055, China

e-mail: chengp@swu.edu.cn

Abstract—Various randomized search heuristics have been proposed for multiobjective optimization problems. We need evaluate and compare the performance of these optimizers in order to make good use of them. This paper reviews the theory and methods proposed in the past decade and summarize their characteristics based on relevant literature. However, we didn't list and analyze many methods proposed before because the relevant literature has done it. We look at them from the classification perspective. These assessment methods are classified as quality indicator based approaches and statistic test approach here. For quality indicators, we further classify them Pareto dominance compliant and non-compliant, unary and binary parameters. This enables us to choose suitable assessment measures in practice.

Keywords—performance assessment; quality indicator; attainment function; evolutionary algorithm; multiobjective optimization

1. Introduction

Various multi-objective optimization heuristics are available to provide Pareto-optimal approximation. However, in order to choose the suitable one for a given problem, we need compare these algorithms empirically or theoretically. It means that we need assess the performance of multi-objective algorithms.

The notion of performance includes both the quality of Pareto-optimal solutions as well as the computational resources needed to generate this outcome. In this paper, we mainly focus on the quality aspect. The measurement of outcomes of various methods is simple if a single objective problem is considered. We can utilize the value of objective function to compare approximate solutions. But there exist no such easy ways for multiple objective cases.

Zitzler et al [1] suggested three criteria to measure the quality of non-dominated solutions: 1) the distance of resulting non-dominated set the Pareto-optimal front should be minimized; 2) a good distribution of the solutions; 3) a wide range of value should be present. These criteria are suitable to general problems. For some optimization problems with particular Pareto-optimal fronts, we need alter some criteria to fit them.

It has been common in literature for performance to be indicated simply by way of a graphic plot. However, we need quantitative measures to evaluate approximations more precisely. Many quality measures have been proposed to evaluate obtained approximations. However, they are not consistent with each other. There is no common agreement on which measure should be used.

What are quality measures? As Zitzler et al defined [3], quality measures are used to compare the outcomes of multiobjective optimizers in a quantitative manner. The underlying idea is to quantify quality differences between approximation sets by applying common metrics. Some measures map approximation sets to the set of real numbers. We call this kind of quality measures as quality indicators and discuss them in section 2. Other measures are statistical test procedures based on multiple runs. We discuss them in section 3, and especially for attainment function.

Table 1: The relation between Pareto front approximations

relation		**interpretation in objective space**
Strictly dominates	$A \prec\prec B$	every $z^2 \in B$ is strictly dominated by at least one $z^1 \in A$
dominates	$A \prec B$	every $z^2 \in B$ is dominated by at least one $z^1 \in A$

978-0-7695-4281-2/10 $26.00 © 2010 IEEE
DOI 10.1109/ICGEC.2010.91

better	$A \lhd B$	every $z^2 \in B$ is weakly dominated by at least one $z^1 \in A$ and $A \prec \neq$
Weakly dominates	$A \preccurlyeq B$	every $z^2 \in B$ is weakly dominated by at least one $z^1 \in A$
incomparable	$A \mid\mid B$	neither $A \preccurlyeq B$ nor $B \preccurlyeq A$
indifferent	$A \sim B$	$A \preccurlyeq B$ and $B \preccurlyeq A$

2 APPROACHES BASED ON QUALITY INDICATORS

Many quality indicators have been proposed and used in the literature. Here we only recommend several of them. The characteristics and comparisons of more indicators can be found in the literatures [2][3][5][7]. First we introduce some basis to classify these indicators.

A. Pareto compliant indicator versus Pareto non-compliant indicator

In light of the consistency of quality indicators with the Pareto domination relation (refer to Table 1 [3]), indicators can be classified as Pareto compliant and Pareto non-compliant. As Knowels et al defined [7], Whenever an approximation set A is preferable to B with respect to Pareto dominance, the indicator value for A should be at least as good as that for B. We call such indicators Pareto compliant. If not, they are Pareto non-compliant. Many indicators are Pareto non-compliant. It does not mean that Pareto non-compliant indicators are useless. For instance, they may be used to refine the preference structure of approximation sets which have the same Pareto compliant indicator values.

Zitzler et al [3] analyzed assessment indicators more rigorously. They used the terms of compatibility and completeness to characterize a comparison method. From this perspective, they examined the connection of many indicators with each of dominance relations listed in Table 1. For example, a quality indicator is $\rhd -compatible$ if the indicator value of approximation set A is preferable than B infers $A \lhd B$; and a quality indicator is $\rhd -complete$ if $A \lhd B$ infers that the indicator value of approximation set A is preferable than B.

B. Unary quality indicator versus binary quality indicator

Unary indicators map an approximation set Ω to a real number R, $I : \Omega \to R$. The difference of indicator values between I(A) and I(B) represents the difference of the quality of approximation set A and B. In practice several unary indicators are often combined to indicate the quality. Binary indicators map an ordered pair of approximation sets to a real number, $I : \Omega \times \Omega \to R$.

Zitzler et al [3] analyzed and contrasted unary and binary indicators theoretically in detail. They proved the theoretical limitation of unary indicators and pointed out existing unary indicators at best to infer that an approximation set is not worse than another; at the same time, they indicated that binary indicators do not possess the limitations of unary indicators, but binary indicators also have a drawback. For instance, when we compare t optimizers using a unary indicator, we obtain t values in contrast to t(t-1) values in the case of a binary indicator.

C. Some quality indicators

About the choice of indicators, if the preference is not considered, the good choice may be the indicator which is compatible and complete with respect to as many of dominance relation as possible. According to such criteria and popularity, we recommend the following indicators.

The hypervolume indicator

It is also called as the S metric. The definition of it is given in [4]. It calculates the hypervolume of the multi-dimensional region enclosed by non-dominated set and a reference point, hence calculates the size of the region A dominates. It is widely studied and used. It is complete regarding $\rhd$ and is able to detect that A is not worse than B for all pairs of $A \rhd B$.

The binary ε-indicator

It was proposed by Zitzler et al [3]. We say a vector z^1 ε-dominate another vector z^2, if we can multiply each objective dimension value in z^2 by a factor of ε and the resulting objective vector is still weakly dominated by z^1.

The coverage indicator

This indicator gives the fraction of solutions in one approximation set that are weakly dominated by at least one solution in another [6]. If the indicator value is 1, it is equivalent to weakly domination.

The above three recommended indicators are Pareto compliant and the latter two are both $\triangleright$ *−compatible* and $\triangleright$ *−complete*. In recent years some new quality indicators [8][9][10][11] have been proposed. Most of them are about the measure of diversity and spread. It needs more work to compare these diversity indicators.

No matter approximation set A and B are comparable (with respect to Pareto dominance relation) or incomparable, indicator values for them always can be obtained. Thus indicator-based approach is able to compare any pair of approximation sets on the same problem. This characteristic goes beyond the Pareto dominance relation. Further, a quality measure is not only used to evaluate whether one approximation set is better than the other, but also it is utilized to show how much better it is. However, most indicators contain preference information. When we compare the same optimizers using different indicators, different outcomes will be returned, and sometimes these outcomes may disagree.

3 APPROACHES BASED ON PROBABILITY

Most multiobjective optimizers are variants of randomized search algorithms and they are stochastic in nature. The outcomes are different even if the same algorithm is applied several times on the same problem. So maybe it is more scientific to assess or compare different optimizers from a statistic perspective.

However, it is difficult to know exactly the underlying probability density of approximation sets by means of theoretical analysis. Therefore, most statistical assessment approaches are based on empirical studies. The following are three basic approaches in the literature.

A. Static methods on quality indicators

By running a specific algorithm several times on the same optimization problem, one obtains a sample of approximation sets and corresponding quality indicator values. And then these indicator values are analyzed and compared based on statistical test procedures.

B. Attainment surface and attainment function

C. M. Fonseca and P. J. Fleming [12] proposed a quantitative, non-parametric approach for performance assessment of stochastic multiobjective optimizers. This approach evaluates an optimizer based on multiple runs. In each run, a corresponding attainment surface is obtained. The attainment surface reflects two types of information: the approximation to Pareto front and the distribution of solutions. Thus, after a large number of runs, the superposition of multiple attainment surfaces provide a basis on which we can identify the family of objective vectors which are obtained by the optimizer at a certain percent. For instance, a 50%-attainment surface means that the surface is obtained in 50% of the runs, but is not obtained in the other 50% of runs.

Viviane Grunert da Fonseca and et al [13] define and analyze the attainment function in the light of classical statistics. The attainment function of a point in the objective space describes the probability of the approximation set attaining (dominating or being equal to) that particular point. They demonstrate us the attainment function is a generalization of the multivariate cumulative distribution function. So many classical concepts and methods applied to cumulative distribution function can be generalized here for the attainment function. The true attainment function is often unknown, and in practice the empirical attainment function is used. It's estimation from the outputs of several independent runs of the same algorithm.

In order to address the dependence structure of individual solutions within the approximation set, C. M. Fonseca and et al [14] use more informative second-order moment measures for evaluation and comparison of multi-objective optimizers. They introduced the concept of second-order attainment function and covariance function and apply them on two example problems. The experiments results show that this method can evaluate optimizers more effectively compared with using the first-order attainment function alone.

Visualization of the attainment surfaces is important. It eases the readers to assess or compare algorithms. J. D. Knowles [15] presented an algorithm to plot summary attainment surfaces for any objective dimension space (especially for three objective spaces). The summary attainment surface is defined as the union of points that have been attained in precisely s of n runs, for any $s \in 1..n$.

M. López-Ibáñez and et al [16] proposed a convenient approach to compare two different algorithms. It is difficult for readers to compare outputs of the algorithms even if their attainment surfaces have been plotted in the figures side by side. This method emphasizes to plot the difference of empirical attainment functions of two algorithms at each point in objective space. If a large difference happens at a certain point, it indicates the probability of attaining this point with one algorithm is greater than the other. As they point out, this method not only reveals the magnitude of differences, but also where the differences locate in the objective space. It is impossible to most of scalar quality indicator approaches. They also introduce graphical tools to visualize the attainment surface [17].

Compared to only consider a single run, the approach of attainment function may be more reliable to reflect the performance of an optimizer. However, the attainment surfaces of different optimizers on the same problem can't be compared likely. For example, some sections of attainment surface of one optimizer are better than the other optimizer, but some sections are not. Approaches based on quality indicators are less troubled by this way. Visualization of differences of attainment functions of two algorithms may be an effective and fine-grained method to solve this problem.

C. *Statistic test based on dominance- ranking distribution of approximation sets*

Another approach [7] is that all approximation sets generated by different optimizers are pooled, and then each approximation set is assigned a rank reflecting the number of approximation sets in the pool that are better (cf. Table 1, A $\lhd$ B) than it. Thereby, a set of ranks for each algorithm is obtained and can be statistically verified. A good idea is to use quality indicators or attainment function to characterize further the differences, if there is not a significant difference when we first utilize the ranking of approximation sets alone.

4 Other related topics

The above discussed approaches are empirical, and the issue of their adequacy to sustain performance statements needs more attention. S. C. Chiam [18] presented a conceptual framework to address it. In addition, to solve multi-objective problems, it often comprise two stages: 1) to find the approximation set to the Pareto front; 2) to select a suitable solution that best reflects the decision maker's preference form the approximation set. These two stages are not necessarily sequential. The past sections we discussed focus on the performance of the former, but the performance of latter is also important. In other words, there exist some methods which map the decision's preference or the parameter values of an optimizer to the corresponding solutions of interest in the approximation set, and the comparison of the performance of these methods need further research. Ferreira and et al. [16] proposed a theoretical measure of the quality of such a mapping.

5 Summary

If the decision maker's preference is not considered, we recommend Pareto dominance compatible and complete indicators, such as ε-indicators or binary hypervolume indicator. However, if the preference information is known, we need consider a specific scenario to choose or devise suitable indicators. Despite some indicators proposed before, which reflect certain preferences, such as convergence, uniformity and spread, scalability and so on, it still needs further research on how to choose or devise suitable indicators to reflect the preference. As for the approaches based on statistics, it well deserves attention in this direction. The method of attainment function and the method of ranking distribution introduced above are helpful attempt. How to combine quality indicators with statistic test procedure effectively is also a valuable research topic in the future. Zitzler and Knowles [20] proposed to use dominance ranking approaches firstly, and then apply the quality indicators for preferences and empirical attainment function for visualization to further the differences. In addition, we need take typical example experiments to compare these methods in the near future.

Acknowledgment

This work has been supported by the Fundamental Research Funds for the Central Universities of China (No: XDJK2009C030).

REFERENCES

[1] E. Zitzler, K. Deb, and L. Thiele, “Comparison of multiobjective evolutionary algorithms: Empirical results,” Evolutionary Computation, Vol. 8, No. 2, pp. 173-195, 2000.

[2] J. Knowles and D. Corne, “On metrics for comparing nondominated sets,” IEEE Congress on Evolutionary Computation (CEC 2002). Piscataway, NJ: IEEE Press, 2002, pp. 711-716.

[3] E. Zitzler, L. Thiele and et al., “Performance assessment of multiobjective optimizers: an analysis and review,” IEEE Transaction on Evolutionary Computation, Vol. 7, No. 2, pp. 117-132, 2003.

[4] E. Zitzler, “Evolutionary algorithms for multiobjective optimization: methods and application,” PhD thesis, Swiss federal institute of technology (ETH), Zurich, Switzerland, 1999.

[5] Hansen, P. and A. Jaszkiewiecz, “Evaluating the quality of approximations to the non-dominated set,” Institute of mathematical modeling, Technical University of Denmark, Lyngby. Technical Report 1998, IMM-REP-1998-7.

[6] E. Zitzler and L. Thiele, “Multiobjective optimization using evolutionary algorithms - A comparative case study,” Parallel Problem Solving from Nature - PPSN-V, pp. 292-301, Springer, Berlin, September 1998.

[7] J. D. Knowels, L. Thiele, and E. Zitzler, “A tutorial on the performance assessment of stochastic multiobjective optimizaers,” TIK-Report No. 214, Computer engineering and networks laboratory, ETH Zurich, 2006.

[8] V. L. Silva, E. F. Wanner, S. A. Cerqueira, and R. H. Takahashi, "A new performance metric for multiobjective optimization: the integrated sphere counting," Proceedings of the 2007 IEEE Congress on Evolutionary Computation, pp. 3625-3630.

[9] F. G. Guimaraes, E. F. Wanner, R. H. C. Takahashi, “A quality metric for multi-objective optimization based on Hierarchical Clustering Techniques,” Proceedings of the 2009 IEEE Congress on Evolutionary Computation(CEC 2009), pp. 3292-3299.

[10] M. Q. Li and J. H. Zheng, “Spread assessment for evolutionary multi-objective optimization,” Proceedings of the 5th International Conference on Evolutionary Multi-Criterion Optimization (EMO 2009), LNCS 5467, Berlin: Springer, pp. 216-230, 2009.

[11] S. Y. Zeng, G. Chen, R. Wang, and et al., "A new technique for assessing the diversity of close-Pareto-optimal front," In IEEE Congress on Evolutionary Computation (CEC 2008), Hongkong, 2008, pp. 344-349.

[12] C. M. Fonseca and P. J. Fleming, “On the performance assessment and comparison of stochastic multiobjective optimizers,” in Parallel Problem Solving from Nature - PPSN IV, vol. 1141 of Lecture Notes in Computer Science, Berlin: Springer, 1996, pp. 584-593.

[13] V. Grunert da Fonseca, C. M. Fonseca, and A. O. Hall, “Inferential performance assessment of stochastic optimizer and the attainment function,” in Evolutionary Multi-Criterion Optimization (EMO 2001), Berlin, Germany: Springer, 2001, pp. 213-225.

[14] C. M. Fonseca, V. Grunert da Fonseca, and L. Paquete, “ Exploring the performance of stochastic multiobjective optimisers with the second-order attainment function,” Proceedings of the Third Evolutionary Multi-Criterion Optimization(EMO 2005), pp. 250--264.

[15] J. D. Knowles. A summary-attainment-surface plotting method for visualizing the performance of stochastic multiobjective optimizers. In Proceedings of the 5th International Conference on Intelligent Systems Design and Applications, pages 552--557, 2005.

[16] M. López-Ibáñez, T. Stützle, L. Paquete, "Graphic tools for the analysis of bi-objective optimization algorithms," Proc. of the Genetic and Evolutionary Computation Conference (GECCO 2010), ACM, pp. 1959-1962, 2010.

[17] M. López-Ibáñez, L. Paquete and T. Stützle, “Exploratory analysis of stochastic local search algorithms in biobjective optimization,” In T. Bartz-Beielstein, et al., editors, Experimental Methods for the Analysis of Optimization Algorithms, pp. 209–233, 2010.

[18] S. C. Chiam, C. K. Goh, and K. C. Tan, “Adequacy of empirical performance assessment for multiobjective evolutionary optimizer,” Proceedings of the 4th international conference on Evolutionary multi-criterion optimization(EMO 2007), LNCS 4403, Berlin: Springer, pp. 893-907, 2007.

[19] J. C. Ferreira, C. M. Fonseca, and A. Gaspar-Cunha, “Assessing the quality of the relation between scalarizing function parameters and solutions in multiobjective optimization,” in IEEE Congress on Evolutionary Computation (CEC 2009), Trondheim, Norway, pp. 1131-1136, July 2009.

[20] E. Zitzler, J. D. Knowles, L. Thiele, “Quality assessment of Pareto set approximations,” Multiobjective Optimization: Interactive and Evolutionary Approaches, vol. 5252 of Lecture Notes in Computer Science, Berlin: Springer, 2008, pp. 373–404.