On the Influence of Outliers on the Loss of Population Diversity of Multi–Objective Estimation of Distribution Algorithms

Luis Martí Jesús García Antonio Berlanga José M. Molina
Dept. of Informatics, Universidad Carlos III de Madrid.
Av. de la Universidad Carlos III, 22. Colmenarejo 28270 Madrid. Spain.
{lmarti,jgherrer}@inf.uc3m.es; {aberlan,molina}@ia.uc3m.es
http://www.giaa.inf.uc3m.es

Mesa: Ciencias de la Computación e Informática

1 Introduction

Multi–objective optimization problems (MOPs) have been successfully addressed using evolutionary approaches. The resulting multi–objective optimization evolutionary algorithms (MOEAs) [1] have attracted a great deal of attention because of their practical impact and their interesting theoretical aspects.

A MOP is an optimization problem where there are two or more functions (known as objective functions) that must be simultaneously optimized. Correspondingly, the solution is a set of trade–off points that represent equally good combinations of the values of the objective functions.

In spite of the favorable results obtained with MOEAs some issues remain to be properly handled. One of these issues is the scalability with respect to the number of objective functions. There are some experimental evidences that show that there is an exponential relationship between the number of objectives and the amount of resources required to correctly solve the problem (see [1] pp. 414–419).

A viable approach to solve this is to employ cutting–edge evolutionary algorithms that would deal more efficiently with the high–dimensional problems. Estimation of distribution algorithms (EDAs) [7] constitute good candidates for such task. EDAs have been claimed to be a paradigm shift in the evolutionary computation field. They replace the application of evolutionary operators with the creation of a statistical model of the fittest elements of the population. This model is then sampled to produce new elements.

The extension of EDAs to the multi–objective domain has lead to what has been called multi–objective EDAs (MOEDAs) [9]. However, although MOEDAs have yielded encouraging results, their introduction has not lived up to a priori expectations. This situation can be attributed in part to the off–the-shelf machine learning methods that are used for model–building.

The improvement of the model–building algorithm used by MOEDAs is, in our opinion, a fertile line of research. This can be attributed to the fact that current MOEDAs employ machine learning algorithms not specifically meant for the model–building task. Therefore, these algorithms do not fully address the requirements of this new task.

Most MOEDAs rely on well–known statistical representations, like graphical models, mixtures of distributions, etc., that are constructed using different methods that are not meant for this particular task.

In particular, those methods disregard data outliers as they assume that that data as noisy, irrelevant or invalid. Although this assumption is correct in the context of their original application scope, it is not valid for MOEDAs. In this last case, it is known beforehand that all elements in the dataset are important, as they are the best elements of the current population. Therefore, none of them should be disregarded. Furthermore, elements that are relatively isolated from the rest of the dataset deserve an special handling since they represent unexplored zones of the search space that are locally optimal solutions. As a consequence, in this case, outliers not only must not be disregarded but probably their role in the statistical model should be reinforced.

In this work we assess some of the main machine learning algorithms used for or suitable for model–building in a controlled environment and under equal conditions. We analyze the possible causes of this underachievement and propose a set of measures that should be taken in order to overcome the current situation.

The rest of this contribution proceeds as we first briefly introduce the theoretical aspects that support our paper. After that, we describe the model–building algorithms under analysis and the MOEDA framework we propose for the tests. Subsequently, a set of experiments, using community–accepted DTLZ3, DTLZ6 and DTLZ7 scalable test problems [2] are presented. Every problem is addressed with a scaling degree of complexity, using 2, 3 and 4 objective functions. Finally some concluding remarks and lines for future work are put forward.

2 Theoretical Aspects

The multi–objective optimization problem (MOP) can be expressed as the problem in which a set of objective functions $f_1(\boldsymbol{x}),\ldots,f_M(\boldsymbol{x})$ should be jointly optimized;

$$\min\ \boldsymbol{F}(\boldsymbol{x}) = \langle f_1(\boldsymbol{x}),\ldots,f_M(\boldsymbol{x})\rangle\,,\ \boldsymbol{x}\in\mathcal{D}\,, \tag{1}$$

where $\mathcal{D}$ is known as the decision space. The image set, $\mathcal{O}$, resulting of the projection $\boldsymbol{F}:\mathcal{D}\to\mathcal{O}$ is called objective space.

In this class of problems the optimizer must find one or more feasible solutions that jointly minimizes (or maximizes) the objective functions. Therefore, the solution to this type of problem is a set of trade–off points. The adequacy of a solution can be expressed in terms of the Pareto dominance relation:

Definition 1 *For the optimization problem specified in (1) and having $\boldsymbol{x}_1,\boldsymbol{x}_2\in\mathcal{D}$. $\boldsymbol{x}_1$ is said to dominate $\boldsymbol{x}_2$ (expressed as $\boldsymbol{x}_1\prec\boldsymbol{x}_2$) iff $\forall f_j$, $f_j(\boldsymbol{x}_1)\leq f_j(\boldsymbol{x}_2)$ and $\exists f_i$ such that $f_i(\boldsymbol{x}_1)<f_i(\boldsymbol{x}_2)$.*

The solution of (1) is the Pareto–optimal set, $\mathcal{D}^*$; which is the subset of $\mathcal{D}$ that contains elements that are not dominated by other elements of $\mathcal{D}$. Its image in objective space is called Pareto–optimal front, $\mathcal{O}^*$.

Finding the explicit formulation of $\mathcal{D}^*$ is often impossible. Generally, an algorithm solving (1) yields a discrete local Pareto–optimal set, $\mathcal{P}^*$, that approximates $\mathcal{D}^*$. The image of $\mathcal{P}^*$ in objective space, $\mathcal{PF}^*$, is known as the local Pareto–optimal front.

2.1 Multi–Objective Estimation of Distribution Algorithms

MOPs have been addressed with a variety of methods. Among them, evolutionary algorithms (EAs) have proven themselves as a valid and competent approach from theoretical and practical points of view. This has led to what has been called multi–objective optimization evolutionary algorithms (MOEAs). Their success is due to the fact that EAs do not make any assumptions about the underlying fitness landscape. Therefore, it is believed they perform consistently well across a wide range of problems.

Estimation of distribution algorithms (EDAs) have been claimed as a paradigm shift in the field of evolutionary computation. Like EAs, EDAs are population based optimization algorithms. However in EDAs the step where the evolutionary operators are applied to the population is substituted by construction of a statistical model of the most promising subset of the population. This model is then sampled to produce new individuals that are merged with the original population following a given substitution policy.

The introduction of machine learning techniques implies that these new algorithms lose the biological plausibility of its predecessors. In return, they gain the capacity of scalably solving many challenging problems, often significantly outperforming standard EAs and other optimization techniques.

Multi–objective optimization EDAs (MOEDAs) [9] are the extensions of EDAs to the multi–objective domain. Most of MOEDAs consist of a modification of existing EDAs whose fitness assignment strategy is substituted by a previously existing one used by MOEAs.

2.2 Outliers, Diversity Loss and Model–Building

The cause of MOEDAs underachievement in many–objective problems can be traced back to their model–building algorithms. Not until recently, EDAs practitioners have failed to recognize that machine learning approaches can't be extrapolated as–is to the model–building task. In particular, there are properties shared by most machine–learning approaches that could be preventing MOEDAs from yielding a substantial improvement over MOEAs. They are:

- the incorrect treatment of data outliers, and;
- tendency towards the loss of population diversity.

These issues can be traced back to the single–objective predecessor of most MOEDAs and their corresponding model–building algorithms. Most model–building schemes used so far by EDAs use off–the–shelf machine learning methods.

In the statistical and machine learning areas the data instances that are relatively isolated or diverse from the greater masses of data are known as outliers. Historically, these outliers are handled as not representative, noisy or bogus data. On the other hand, in model–building it is known beforehand that all the available data is valid as it represents the best part of the current population. Therefore, no points must be disregarded. Instead, these outliers are essential, as they represent unexplored or recently discovered areas of the current Pareto–optimal front.

Another drawback of most MOEDAs, and most EDAs, for that matter, is the lost of diversity lost of population diversity. This fact has been already pointed out and some proposals have been laid out for addressing it [11]. This loss of diversity could also be traced back to the nature of the model–building algorithm.

The root cause that makes current methods disregard outliers and lose population diversity can be traced to the global error–based learning that take place in those methods. In that type of learning a dataset–wise error is minimized. Because of that, infrequent or poorly represented elements are sacrificed in order to achieve a better overall error.

3 Experiments

Having properly dealt with the theoretical constituents of our study we will proceed with the study of the performance of each model builder under different circumstances. In particular, we will gauge the randomized leader algorithm [4], the k–means algorithm [8], the expectation maximization algorithm [3], Bayesian networks [5] and the kernel k–means algorithm [10]. The first four algorithms were previously used by MOEDAs and the fifth is an advanced clustering algorithm that, in our opinion, could yield interesting results.

The problems to be addressed are the DTLZ3, DTLZ6 and DTLZ7 scalable continuous test problems [2]. The accuracy and adequacy of the solutions was determined with a commonly accepted indicators of multi–objective optimization performance: the unary additive ϵ–indicator [6].

The outcome of the experiments measured with the ϵ–indicator is summarized on Figs. 1. These figures display the results in as box–plots. The results for the lowest dimension are relatively similar across the different algorithms. This can be explained by the fact that in relatively low dimensions these problems are relatively tractable and it is not required that model builder play a fundamental role promoting search.

However, when we analyze the outcome for the $M = 4$ problems some interesting conclusions emerge. First, it is noticeable that statistically sound approaches like Bayesian networks yield poorer results when compared to others. It is most interesting how an statistically unsound algorithm like the leader algorithm produces better results than classical ones like Bayesian networks. The methods based on k–means also performed adequately with a slight advantage in some cases for the kernel version.

This leads to the conclusion that the model–building problem has its own particularities that do not conform to the typical statistical or machine learning scheme. This might open a line of research for creating new model builders that deviate themselves from the current approaches.

Although more studies are required to achieve a comprehensive understanding of the nature of the model–building problem these results cast some light on the matter.

4 Concluding Remarks

In this work we have taken the first steps towards the understanding of the nature of the model–building problem of MOEDAs and their diversity loss. We have found that non–rigorous or inexact approaches performed better than more robust methods. This leads to the conclusion that the model–building issue have its own set of requirements that hinders the application of “classical” methods. Reflecting on this we can see that, for example, outliers must be treated differently as they represent newly discovered local optima that should be explored –not left aside.

However, in order to gain a better comprehension more experiments are necessary. On one hand, different test problems must be addressed to realize if the results obtained here can be generalized. On the other, it is also of interest to further scale the problems to more objective functions. The analysis of the behavior of the algorithms in those situations might lead to their adaptation to the problem.

Acknowledgements

This work was supported by projects CICYT TIN2008–06742–C02–02/TSI, CICYT TEC2008–06732–C02–02/TEC, SINPROB, CAM MADRINET S–0505/ TIC/0255 and DPS2008–07029–C02–02.

References

[1] Kalyanmoy Deb. *Multi-Objective Optimization using Evolutionary Algorithms*. John Wiley & Sons, Chichester, UK, 2001. ISBN 0-471-87339-X.

[2] Kalyanmoy Deb, Lothar Thiele, Marco Laumanns, and Eckart Zitzler. Scalable test problems for evolutionary multiobjective optimization. In Ajith Abraham, Lakhmi Jain, and Robert Goldberg, editors, *Evolutionary Multiobjective Optimization: Theoretical Advances and Applications*, Advanced Information and Knowledge Processing, pages 105–145. Springer Verlag, 2004.

[3] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, B(39):1–38, 1977.

[4] J. A. Hartigan. *Clustering Algorithms*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York, 1975.

[5] David Heckerman and Michael P. Wellman. Bayesian networks. *Communications of the ACM*, 38(3):27–30, 1995.

[6] Joshua Knowles, Lothar Thiele, and Eckart Zitzler. A Tutorial on the Performance Assessment of Stochastic Multiobjective Optimizers. 214, Computer Engineering and Networks Laboratory (TIK), ETH Zurich, Switzerland, February 2006. revised version.

[7] P. Larrañaga and J. A. Lozano, editors. *Estimation of Distribution Algorithms. A new tool for Evolutionary Computation*. Genetic Algorithms and Evolutionary Computation. Kluwer Academic Publishers, Boston/Dordrecht/London, 2002.

[8] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical*, volume 1, pages 281–297, 1967.

[9] Martin Pelikan, Kumara Sastry, and David E. Goldberg. Multiobjective estimation of distribution algorithms. In Martin Pelikan, Kumara Sastry, and Erick Cantú-Paz, editors, *Scalable Optimization via Probabilistic Modeling: From Algorithms to Applications*, Studies in Computational Intelligence, pages 223–248. Springer–Verlag, 2006.

[10] Bernhard Scholköpf, Alexander Smola, and Klaus-Robert Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5):1299–1319, 1998.

[11] Jonathan Shapiro. Diversity loss in general estimation of distribution algorithms. In *Parallel Problem Solving from Nature - PPSN IX*, pages 92–101, 2006.

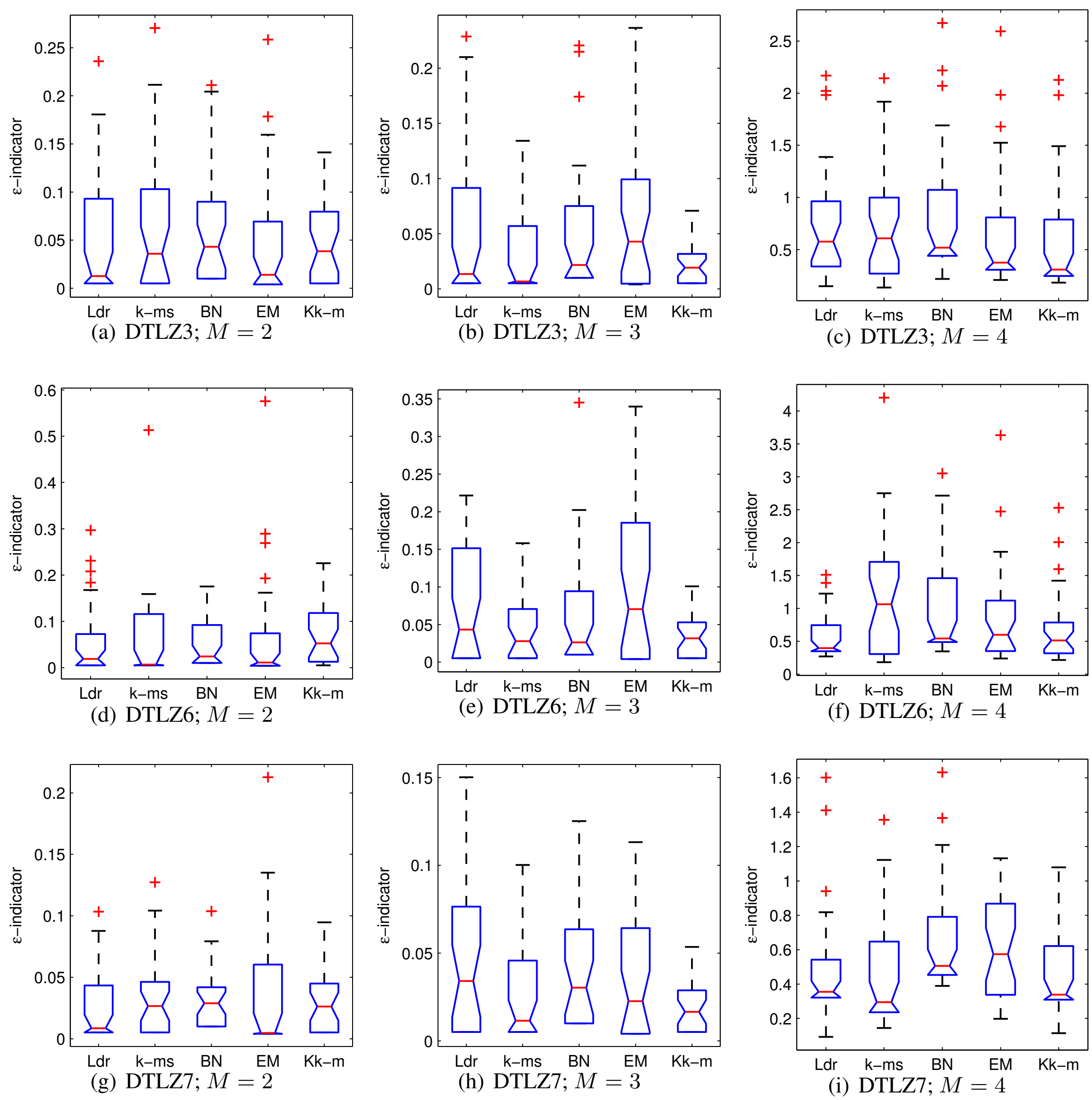

Figure 1: Unary Additive ϵ–indicator values for DTLZ3, DTLZ6 and DTLZ7 problems. Each column show the results obtained with different amount of objectives ($M = 2, 3, 4$) using the Leader algorithm (Ldr), the k–means algorithm (k-ms), Bayesian networks (BN), expectation maximization (EM), and the kernel k–means algorithm (Kk-m).