

This document is published in:

Coello Coello, Carlos A. (ed.), 2011. *Learning and Intelligent Optimization: 5th International Conference, LION 5, Rome, Italy, January 17-21, 2011. Selected Papers* (Lecture Notes in Computer Science, 6683), Springer, pp. 458-472.

DOI: 10.1007/978-3-642-25566-3_36

© 2011 Springer-Verlag Berlin Heidelberg

Multi-Objective Optimization with an Adaptive Resonance Theory-Based Estimation of Distribution Algorithm: A Comparative Study

Luis Martí, Jesús García, Antonio Berlanga, and José M. Molina

Universidad Carlos III de Madrid, Group of Applied Artificial Intelligence
Av. de la Universidad Carlos III, 22. Colmenarejo, Madrid 28270, Spain
{lmarti,jgherrer}@inf.uc3m.es, {aberlan,molina}@ia.uc3m.es
<http://www.giaa.inf.uc3m.es/>

Abstract. The introduction of learning to the search mechanisms of optimization algorithms has been nominated as one of the viable approaches when dealing with complex optimization problems, in particular with multi-objective ones. One of the forms of carrying out this hybridization process is by using multi-objective optimization estimation of distribution algorithms (MOEDAs). However, it has been pointed out that current MOEDAs have an intrinsic shortcoming in their model-building algorithms that hamper their performance.

In this work we argue that error-based learning, the class of learning most commonly used in MOEDAs is responsible for current MOEDA underachievement. We present adaptive resonance theory (ART) as a suitable learning paradigm alternative and present a novel algorithm called multi-objective ART-based EDA (MARTEDA) that uses a Gaussian ART neural network for model-building and an hypervolume-based selector as described for the HypE algorithm. In order to assert the improvement obtained by combining two cutting-edge approaches to optimization an extensive set of experiments are carried out. These experiments also test the scalability of MARTEDA as the number of objective functions increases.

1 Introduction

Multi-objective optimization has received a lot of attention by the evolutionary computation community leading to *multi-objective evolutionary algorithms* (MOEAs) (cf. [1]). A *multi-objective optimization problem* (MOP) can be expressed as the problem in which a set of M *objective functions* $f_1(\mathbf{x}), \dots, f_M(\mathbf{x})$ with should be jointly optimized;

$$\min \mathbf{F}(\mathbf{x}) = \langle f_1(\mathbf{x}), \dots, f_M(\mathbf{x}) \rangle; \mathbf{x} \in \mathcal{D}; \quad (1)$$

where $\mathcal{D} \subseteq \mathbb{R}^n$ is known as the *feasible set* and could be expressed as a set of restrictions over the decision set, \mathbb{R}^n . The image set of \mathcal{D} produced by function vector $\mathbf{F}(\cdot)$, $\mathcal{O} \subseteq \mathbb{R}^M$, is called *feasible objective set* or criterion set (see [2] for details on notation).

The solution to this problem is a set of trade-off points. The adequacy of a solution can be expressed in terms of the *Pareto dominance relation* [3]. The solution of (1) is the Pareto-optimal set, \mathcal{D}^* ; which is the subset of \mathcal{D} that contains all elements of \mathcal{D} that are not dominated by other elements of \mathcal{D} . Its image in objective space is called Pareto-optimal front, \mathcal{O}^* .

There is a class of MOPs that are particularly appealing because of their inherent complexity: the so-called *many-objective problems* [4]. These are problems with a relatively large number of objectives. It has been shown that “established” approaches fail to yield adequate solutions because of the exponential relation between the dimension of the objective space and the amount of resources, in particular population size, required to solve the problem correctly. Although somewhat counterintuitive and hard to visualize for a human decision maker, these problems are not uncommon in real-life engineering practice. For example, [5] details some relevant real problems of this type.

Many-objective problems have been addressed from three main fronts:

1. the design of better fitness assignment (selection) functions;
2. the use of objective reduction strategies, and;
3. application of better search (variation) methods

There has been a relatively large body of work on the first two issues. For example, it has been shown that the use performance indicators and some forms of relaxed Pareto dominance for the fitness assignment task allows the resulting algorithm to cope with higher dimension problems (cf. [6, 7, 8]). Similarly, some works have focused on the reduction of the amount of objectives to a minimum by eliminating redundant or irrelevant objectives (cf. [9, 10, 11]).

The third direction remains to be properly explored. Here, a viable approach is to employ cutting-edge evolutionary algorithms that could effectively deal with high-dimensional problems more efficiently.

The incorporation of learning as part of the search processes has been nominated as a viable way of dealing with that third issue [12]. There are some approaches that perform this task by providing hybrid evolutionary/machine learning method, like, for example, the *learnable evolution model* (LEM) [13]. However, these efforts seem to have been concentrated on single-objective optimization (c. f. [14, 15]).

Another form of carrying out this task is to resort to *estimation of distribution algorithms* (EDAs) [16]. This is because of EDAs capacity of learning the problem structure. EDAs replace the application of evolutionary operators with the creation of a statistical model of the fittest elements of the population in a process known as model-building. This model is then sampled to produce new elements. Nevertheless, the so-called *multi-objective EDAs* (MOEDAs) [17] have not live up to their a priori expectations. This is can be attributed to the fact that most MOEDAs have limited themselves to transforming single-objective EDAs into a multi-objective formulation by including an existing multi-objective fitness assignment function.

This straightforward extrapolation has prompted the existence of a number of shortcomings en current MOEDAs. We have recognized three of them, in

particular, those derived from the incorrect treatment of outstanding but isolated elements of the population (outliers); the loss of population diversity, and that too much computational effort is being spent on finding an optimal population model.

The performance issue of current MOEDAs has been traced back to the their underlying learning paradigm: the dataset-wise error minimization learning, or error-based learning, for short [18]. This class of learning, in different forms, is shared by most machine learning algorithms. It implies that model is tuned in order to minimize a global error measured across the dataset. In this type of learning isolated data is not taken into account because of their little contribution to the overall error and therefore they do not take an active part of learning process. This assertion is in part supported by the fact that most the approaches that had a better performance in comparative experiments like [18] do not exactly conform to the error-based scheme. That is why, other learning paradigms should be assessed.

Adaptive resonance theory (ART) [19] is a theory of human cognition that has seen a realization as a family of neural networks. It relies on a learning scheme denominated match-based learning and on intrinsic topology self-organization. These features make it interesting as a case study as model-building approach. Match-based learning equally weights isolated and clustered data [20], and, therefore, the algorithm does not disregard outliers. Similarly, self-organization makes possible the on-the-fly determination the model complexity required to correctly represent the data set, thus eliminating the need of an external algorithm for that task.

In this work we argue that error-based learning, the class of learning most commonly used in MOEDAs is responsible for current MOEDA underachievement. We discuss in detail ART-based learning as a viable alternative and present a novel algorithm called *multi-objective ART-based EDA* (MARTEDA) that uses a Gaussian ART neural network [21] for model-building and an hypervolume-based selection as described for the hypervolume estimation algorithm for multiobjective optimization (HypE) [8]. We experimentally show that thanks to MARTEDA's novel model-building approach and an indicator-based population ranking the algorithm it is able to outperform similar MOEDAs and MOEAs. Elements of MARTEDA have been discussed in some preliminary works [22], but it has not yet been presented in detail.

The remaining part of the work proceeds as we discuss the model-building issue. Following that we describe the Gaussian ART network that is used as start point for our model-building algorithm. Subsequently, MARTEDA is introduced, describing how the HypE selection and Gaussian ART are blended together in a MOEDA framework. Section 5 presents and discusses the results of the comparative experiments involving MARTEDA and a selection of other current state-of-the-art algorithms dealing with a set of community accepted problems. These problems are configured with an progressive number of objectives (3, 6, 9 and 12) in order to assess the performance of our proposal in the context of many-objective optimization. Finally, some conclusive remarks and future lines of research are outlined.

2 The Model-Building Issue

Notwithstanding the diverse efforts dedicated to providing usable model-building methods for EDAs the nature of the problem itself has received relatively low attention. An analysis of the results yielded by current multi-objective EDAs and their scalability with regard to the number of objective leads the identification of certain issues that might be hampering the obtention of substantially better results with regard to other evolutionary approaches.

Data outliers issue is a good example of insufficient comprehension of the nature of the model-building problem. In machine-learning practice, outliers are handled as noisy, inconsistent or irrelevant data. Therefore, outlying data is expected to have little influence on the model or just to be disregarded.

However, that behavior is not adequate for model-building. In this case, is known beforehand that all elements in the data set should be take into account as they represent newly discovered or candidate regions of the search space and therefore must be explored. Therefore, these instances should be at least equally represented by the model and perhaps even reinforced.

Another weakness of most MOEDAs (and most EDAs, for that matter) is the loss of population diversity. This is a point that has already been made, and some proposals for addressing the issue have been laid out [23,24,25,26,27]. This loss of diversity can be traced back to the above outliers issue of model-building algorithms.

The incorrect treatment of outliers and the loss of population diversity can be attributed the error-based learning approaches that take place in the most MOEDAs. Error-based learning is rather common in machine learning algorithms. It implies that model topology and parameters are tuned in order to minimize a global error measured across the learning data set. This type of learning isolated data is not taken into account because of their little contribution to the overall error and therefore they do not take an active part of learning process. In the context of many problems this behavior makes sense, as isolated data can be interpreted as spurious, noisy or invalid data.

That is not the case of model-building, as we have already argued. In model-building all data is equally important and, furthermore, isolated data might have a bigger significance as they represent unexplored zones of the current optimal search space. This assessment is supported by the fact that most the approaches that had a better performance do not follow the error-based scheme, like the k -means algorithm, randomized leader algorithm and the growing neural gas network [18]. That is why, perhaps another classes of learning, like instance-based learning or match-based learning would yield a sizable advantage.

3 Model Building with Adaptive Resonance Theory

Adaptive Resonance Theory (ART) neural networks are capable of fast, stable, on-line, unsupervised or supervised, incremental learning, classification, and prediction following a match-based learning scheme [19]. Match-based learning is complementary to error-based learning. During training, ART networks adjust previously-

learned categories in response to familiar inputs, and create new categories dynamically in response to inputs different enough from those previously seen. A vigilance test allows to regulate the maximum tolerable difference between any two input patterns in a same category. It has been pointed out that ART networks are not suitable for some classes of classical machine-learning applications [20], however, what is an inconvenience in that area is a feature in our case.

3.1 Gaussian ART for Model-Building

There are many variations of ART networks. Among them, the Gaussian ART [21] is most suitable for model-building since it capable of handling continuous data. The result of applying Gaussian ART is a set of nodes each representing a local Gaussian density. These nodes can be combined as a Gaussian mixture that can be used to synthesize new individuals.

Gaussian ART creates classes of similar inputs. A *match tracking mechanism* induces the creation of more specific classes when the prediction of the network differs from the expected output at some degree.

Gaussian ART has a layer of afferent or input nodes, F1, and a classification layer, F2. The F2 layer stores classes of inputs. Its activation is a combined measure of the similarity of the input and the prototype of each class, and the size of the given class.

For the model-building task we have modified the original formulation of the network to make it more suited for the task. When an input $\mathbf{x} \in \mathbb{R}^n$ is presented to the input layer it is propagated to the F2 layer. F2 has N^* units, with N of them committed. Each committed unit models a local density of the input space using Gaussian receptive fields with mean $\boldsymbol{\mu}_j$ and standard deviation $\boldsymbol{\sigma}_j$. A unit is activated if it satisfies the match criterion. That is, the match function,

$$G_j = \exp\left(-\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu_{ji}}{\sigma_{ji}}\right)^2\right), \quad j = 1, \dots, N, \quad (2)$$

must be greater than the F2 vigilance parameter, ρ ; according to this, the input strength of a unit is computed as

$$g_j = \begin{cases} \frac{\eta_j}{\prod_{i=1}^n \sigma_{ji}} G_j, & \text{if } G_j > \rho \\ 0 & \text{otherwise} \end{cases}, \quad \rho > 0, \quad (3)$$

where η_j is a measure of the unit a priori activation probability. This is different from the original Gaussian ART network where only one unit was allowed be active after an input presentation.

After the presentation of an input, if no F2 unit is active, then an uncommitted unit must be committed. The task of detecting when an input is not sufficiently coded in F2 is accomplished by the F2 gain control, G, that fires if no committed units are active. The signal

$$\Gamma = \begin{cases} 1 & \text{if } \max_{j=1, \dots, N} g_j = 0 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

is used to commit an uncommitted unit.

The activation of each unit is then calculated normalizing the unit’s input strength,

$$v_j = \frac{g_j}{\sum_{l=1}^N g_l}. \quad (5)$$

As other ART networks, this model is an on-line learning neural network. Therefore, all adaptation processes have local rules. In F2, μ_j and σ_j are updated using a learning rule based on the gated steepest descent learning rule. η_j is updated to represent the cumulative category activation,

$$\eta_j(t+1) = \eta_j(t) + v_j, \quad (6)$$

and, therefore, the amount of training that has taken place in the j th unit. The use of η_j equally weights inputs over time with the intention to measure their sample statistics.

Learning the first and second moments of the input is

$$\mu_{ji}(t+1) = (1 - \eta_j^{-1}v_j) \mu_{ji}(t) + \eta_j^{-1}v_j x_i, \quad (7)$$

$$\lambda_{ji}(t+1) = (1 - \eta_j^{-1}v_j) \lambda_{ji}(t) + \eta_j^{-1}v_j x_i^2. \quad (8)$$

The standard deviation,

$$\sigma_{ji}(t+1) = \sqrt{\lambda_{ji}(t+1) - \mu_{ji}(t+1)^2}, \quad (9)$$

is calculated using (7) and (8).

Gaussian ART is initialized with all units uncommitted ($N = 0$). Learning takes place in active ($v_j > 0$) F2 units following (7)–(9). However if no F2 units becomes active an uncommitted unit is committed and therefore N is incremented. The new unit is indexed by N and initialized with $v_N = 1$, $\eta_N = 0$. Learning will proceed as usual but a constant γ_i^2 will be added to each λ_{Ni} to set $\sigma_{Ni} = \gamma_i$. The value of γ_i has a direct impact on the quality of learning. A larger γ_i slows down learning in its corresponding input feature but warrants a more robust convergence.

The local Gaussian densities resulting from the described algorithm can be combined to synthesize a Gaussian mixture. This Gaussian mixture is then used can be used by the EDA to generate new individuals.

4 Multi-Objective ART-Based EDA

The multi-objective ART-based EDA (MARTEDA) is a MOEDA that uses the previously described Gaussian ART network as its model-building algorithm. Although it intends to deal with the issues raised by the previous discussion, it was also designed with scalability in mind, since it is expected to cope with many-objective problems. It also exhibits an elitist behavior, as it has proved itself a very advantageous property. Finally, thanks to the combination of fitness assignment and model-building it promotes diversity preservation.

MARTEDA maintains a population, \mathcal{P}_t , of n_{pop} individuals; where t is a given iteration. The algorithm’s workflow is similar to other EDAs. It starts with a random initial population \mathcal{P}_0 of individuals.

At a given iteration t the algorithm determines the set $\hat{\mathcal{P}}_t$ containing the best $\lfloor \alpha |\mathcal{P}_t| \rfloor$ elements.

$$\left| \hat{\mathcal{P}}_t \right| = \lfloor \alpha |\mathcal{P}_t| \rfloor = \lfloor \alpha n_{\text{pop}} \rfloor . \quad (10)$$

Different selection strategies can be applied. However, indicator-based selection seems to have a superior performance in complex and many-objective problems. The hypervolume-based selection have many theoretical features, like being the only indicator that have the properties of a metric and the only to be strictly Pareto monotonic [28] but has the drawback of being computationally intensive to compute.

A lot of research has focused on improving the computational complexity of this indicator [29, 30, 31, 32]. The exact computation of the algorithm has been shown to be $\#\mathbf{P}$ -hard [33] in the number of objectives. $\#\mathbf{P}$ problems are the analogous of \mathbf{NP} for counting problems [34]. Therefore, all algorithms calculating a hypervolume must have an exponential runtime with regard to the number of objectives if $\mathbf{P} \neq \mathbf{NP}$, something that seems to be true [35].

The HypE algorithms attempt to circumvent this problem by estimating the value of the hypervolume by means of a Monte Carlo simulation. The detailed description of this procedure is out of the scope of this paper, and, therefore we invite the interested reader to consult the corresponding paper.

A Gaussian ART network is then trained using $\hat{\mathcal{P}}_t$ as its training data set. In order to have a controlled relation between size of $\hat{\mathcal{P}}_t$ and the maximum size of the network, N_{max} , these two sizes are bound by the rate $\gamma \in (0, 1]$,

$$N_{\text{max}} = \left\lceil \gamma \left| \hat{\mathcal{P}}_t \right| \right\rceil = \lceil \gamma \lfloor \alpha n_{\text{pop}} \rfloor \rceil . \quad (11)$$

The trained GNG network is a model of $\hat{\mathcal{P}}_t$. The network can be interpreted as a Gaussian mixture, as explained in the previous section. Therefore it can be used to sample new individuals. In particular, $\lfloor \omega |\mathcal{P}_t| \rfloor$ new individuals are synthesized.

The local Gaussian densities resulting from the described algorithm can be combined to synthesize the Gaussian mixture with parameters Θ ,

$$P(\mathbf{x}|\Theta) = \frac{1}{N} \sum_{i=1}^N P(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\sigma}_i) . \quad (12)$$

Each Gaussian density is formulated as

$$P(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\sigma}_i) = \frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma}_i|^{1/2}} \exp \left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) \right) , \quad (13)$$

with the covariance matrices Σ_i defined as a diagonal matrix with its non-zero elements set to the values of the deviations σ_i . The Gaussian mixture can be used by the EDA to generate new individuals. These new individuals are created by sampling the $P(\mathbf{x}|\Theta)$. The generation of randomly distributed numbers that follow a given distribution has been dealt in depth by many authors. In our case, we applied the Box-Muller transformation [36].

Each one of these individuals substitute a randomly selected ones from the section of the population not used for model-building $\mathcal{P}_t \setminus \hat{\mathcal{P}}_t$. The set obtained is then united with the best elements, $\hat{\mathcal{P}}_t$, to form the population of the next iteration \mathcal{P}_{t+1} . Some other substitution strategies could be used in this step. For example, the new individuals could substitute the worst individuals of $\mathcal{P}_t \setminus \hat{\mathcal{P}}_t$. We have chosen the previously described approach because it promotes diversity and avoids stagnation.

Iterations are repeated until a given stopping criterion is met. The output of the algorithm is a subset of \mathcal{P}_t that contains the non-dominated solutions, \mathcal{P}_t^* .

5 Experimental Study

The results of the experiments involving MARTEDA, some current state-of-the-art MOEDAs and MOEAs in a selection of current community-accepted problems are reported in this section.

The Walking Fish Group (WFG) problem toolkit [37] is a toolkit for creating complex synthetic multi-objective test problems. The WFG test suite exceeds the functionality of previous existing test suites. These include: non-separable problems, deceptive problems, a truly degenerate problem, a mixed shape Pareto front problem, problems scalable in the number of position related parameters, and problems with dependencies between position- and distance-related parameters. The WFG test suite provides a better form of assessing the performance of optimization algorithms on a wide range of different problems.

From the set of nine problems WFG4 to WFG9 were selected because of the simplicity of their Pareto-optimal front that lies on the first orthant of a unit hypersphere. This decision was also caused by the high computational cost of the experiments being carried out and by the length restriction imposed upon this contribution. Each problem was configured with 3, 6, 9 and 12 objective functions. For all cases the decision space dimension was fixed to 30.

Besides applying MARTEDA to the aforementioned problems some other MOEDAs and MOEAs are also assessed in order to provide a comparative ground for the tests. One algorithm is of particular interest, the MONEDA [38] algorithm. This approach was previously proposed by the authors to deal with the model-building issue of MOEDAs and MARTEDA is supposed to be an improvement over it. However, as MONEDA used the less-performing NSGA-II selection, we have also tested MONEDA with the HypE selection, in order to have some basis for comparison.

Besides MONEDA, we also tested the naïve MIDEA [39], and MrBOA [40] MOEDAs and the SMS-EMOA [41], HypE [8] and NSGA-II [1] MOEAs. One

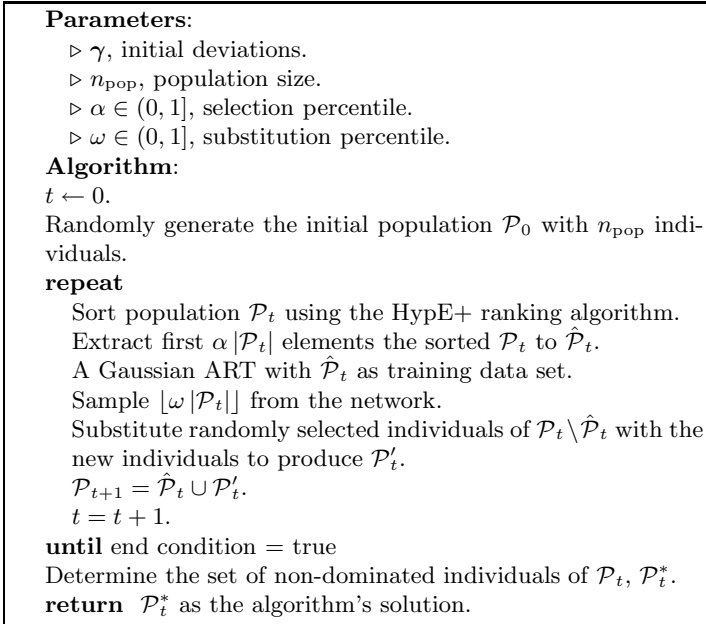


Fig. 1. Algorithmic representation of MARTEDA

of the purposes of this study is to assess the parameter robustness of the algorithms. That is why the same parameter values have been for all problems, only increasing the population size as the number of objectives grows. For each problem/dimension pair each algorithm was executed 30 times.

The quality of the solutions is determined by the use of the hypervolume indicator [42].

The stochastic nature of evolutionary algorithms prompts the use of statistical tools in order to reach a valid judgement of the quality of the solutions and how different algorithms compare with each other. Box plots [43] are one of such representations and have been repeatedly applied in our context. Although box plots allows a visual comparison of the results and, in principle, some conclusions could be deduced out of them. Nevertheless, in order to reach a substantiated judgement it is necessary go beyond reporting the descriptive statistics of the performance indicators. For this task is required to carry out a set of statistical inferences that would support any judgements made from the data.

The statistical validity of the judgment of the results calls for the application of statistical hypothesis tests. It has been previously remarked by different authors that the Mann-Whitney-Wilcoxon U test [44] is particularly suited for experiments in the context of multi-objective evolutionary optimization [42]. This test is commonly used as a non-parametric method for testing equality of population medians. In our case we performed pair-wise tests on the significance of the difference of the indicator values yielded by the executions of the algorithms. A significance level, α , of 0.05 was used for all tests.

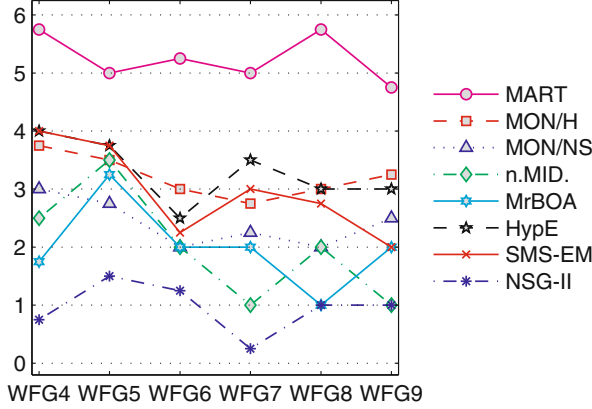


Fig. 2. Mean values of the performance index of MARTEDA (MART), MONEDA with HypE (MON/H) or NSGA-II selection (MON/NS), naïve MIDEA (n.MID), MrBOA, HypE, SMS-EMOA (SMS-EM) and NSGA-II (NSG-II) across the different problems, $\bar{P}_p(\cdot)$.

The visual analysis of the results is rather difficult as it implies cross-examining and comparing the results presented separately. That is why we decided to adopt a more integrative representation such as the one proposed in [45]. That is, for a given set of algorithms A_1, \dots, A_K , a set of P test problem instances $\Phi_{1,m}, \dots, \Phi_{P,m}$, configured with m objectives, the function $\delta(\cdot)$ is defined as

$$\delta(A_i, A_j, \Phi_{p,m}) = \begin{cases} 1 & \text{if } A_i \gg A_j \text{ solving } \Phi_{p,m} \\ 0 & \text{in other case} \end{cases}, \quad (14)$$

where the relation $A_i \gg A_j$ defines if A_i is significantly better than A_j when solving the problem instance $\Phi_{p,m}$, as computed by the statistical tests previously described.

Relying on $\delta(\cdot)$, the performance index $P_{p,m}(A_i)$ of a given algorithm A_i when solving $\Phi_{p,m}$ is then computed as

$$P_{p,m}(A_i) = \sum_{j=1; j \neq i}^K \delta(A_i, A_j, \Phi_{p,m}). \quad (15)$$

This index intends to summarize the performance of each algorithm with regard to its peers.

Figs. 2 and 3 exhibit the results computing the performance indexes grouped by problems and dimensions.

Fig. 2 represents the mean performance indexes yielded by each algorithm when solving each problem in all of its configured objective dimensions,

$$\bar{P}_p(A_i) = \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} P_{p,m}(A_i). \quad (16)$$

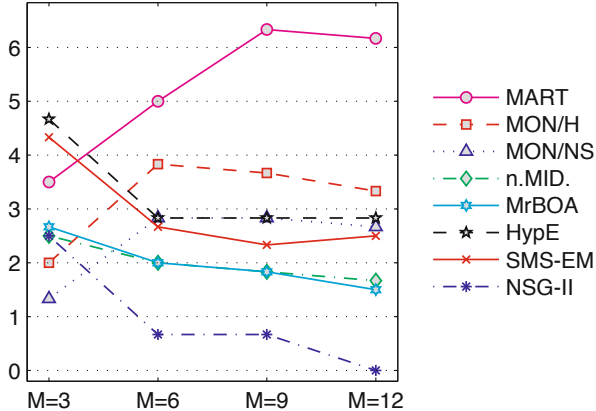


Fig. 3. Mean values of the performance index across the different space dimensions, \bar{P}_m . See Fig. 2 for a description of the acronyms.

It is worth noticing that MARTEDA has better overall results with respect to the other algorithms in all problems. As it could be expected, the use of indicator-based selection in MONEDA has yielded better results than the original MONEDA. Indicator-based MONEDA and the indicator-based MOEAs have a similar performance. It can be hypothesized that these results can be biased by the three objective problems, having dramatic differences in their results with respect to the rest of the dimensions considered.

This situation is clarified in Fig. 3, which presents the mean values of the index computed for each dimension

$$\bar{P}_m(A_i) = \frac{1}{P} \sum_{p=1}^P P_{p,m}(A_i). \quad (17)$$

In this case MARTEDA is shown to clearly outperform the rest in more than three dimensions. Still, another important conclusion can be extracted. For more than three objectives, the MOEDAs that attempt to tackle the model-building issue (MONEDA and MARTEDA) and that also exploit indicator-based selection have outperformed the rest of the methods. This is very important, as it transcends the particular results of a given algorithm but instead casts some light on what should be the proper trend of development in this field.

Finally, the above experiments lead us to hypothesize that thanks to the treatment of the outliers in the model-building data-set, the MARTEDA approach manages to overcome the difficulties that hampers the rest of the methods. Another important result is that MARTEDA was able to yield good results across a varied set of problems without tuning its parameters in every case. This implies that MARTEDA has a certain degree of robustness regarding its parameters.

6 Final Remarks

In this paper we have explored the model-building issue of MOEDAs and the requirements it imposes on its supporting learning paradigm. We put forward adaptive resonance theory as a alternative learning paradigm. Based on it, we introduced a novel algorithm called multi-objective ART-based EDA (MARTEDA) that uses a Gaussian ART neural network for model-building and the hyper volume-based selection described for the HypE algorithm. We showed that by using this novel model-building approach and an indicator-based population ranking the algorithm is able to outperform similar MOEDAs and MOEAs.

Still, the main conclusion of this work is that we provide strong evidences that further research must be dedicated to the model-building issue in order to make current MOEDAs capable of dealing with complex multi-objective problems with many objectives. In spite of the fact that obviously further studies are necessary, these extensive experiments have provided solid ground for the use of MARTEDA in a real-world application context.

Acknowledgements

This work was supported by projects CICYT TIN2008-06742-C02-02/TSI, CICYT TEC2008-06732-C02-02/TEC, CAM CONTEXTS (S2009/TIC-1485) and DPS2008-07029-C02-02.

References

1. Coello Coello, C.A., Lamont, G.B., Van Veldhuizen, D.A.: Evolutionary Algorithms for Solving Multi-Objective Problems. In: Genetic and Evolutionary Computation, 2nd edn. Springer, New York (2007)
2. Miettinen, K.: Nonlinear Multiobjective Optimization. International Series in Operations Research & Management Science, vol. 12. Kluwer, Norwell (1999)
3. Pareto, V.: Cours D'Économie Politique. F. Rouge, Lausanne (1896)
4. Purshouse, R.C., Fleming, P.J.: On the evolutionary optimization of many conflicting objectives. *IEEE Transactions on Evolutionary Computation* 11(6), 770–784 (2007)
5. Stewart, T., Bandte, O., Braun, H., Chakraborti, N., Ehrgott, M., Göbelt, M., Jin, Y., Nakayama, H., Poles, S., Di Stefano, D.: Real-world applications of multiobjective optimization. In: Branke, J., Deb, K., Miettinen, K., Słowiński, R. (eds.) *Multiobjective Optimization*. LNCS, vol. 5252, pp. 285–327. Springer, Heidelberg (2008)
6. Wagner, T., Beume, N., Naujoks, B.: Pareto-, aggregation-, and indicator-based methods in many-objective optimization. In: Obayashi, S., Deb, K., Poloni, C., Hiroyasu, T., Murata, T. (eds.) *EMO 2007*. LNCS, vol. 4403, pp. 742–756. Springer, Heidelberg (2007)
7. Bader, J., Deb, K., Zitzler, E.: Faster hypervolume-based search using Monte Carlo sampling. In: Beckmann, M., Küenzi, H.P., Fandel, G., Trockel, W., Basile, A., Drexler, A., Dawid, H., Inderfurth, K., Kürsten, W., Schittko, U., Ehrgott, M., Naujoks, B., Stewart, T.J., Wallenius, J. (eds.) *Multiple Criteria Decision Making for Sustainable Energy and Transportation Systems*. LNEMS, vol. 634, pp. 313–326. Springer, Berlin (2010)

8. Bader, J., Zitzler, E.: HypE: An Algorithm for Fast Hypervolume-Based Many-Objective Optimization. TIK Report 286, Computer Engineering and Networks Laboratory (TIK), ETH Zurich (2008)
9. Deb, K., Saxena, D.K.: Searching for Pareto-optimal solutions through dimensionality reduction for certain large-dimensional multi-objective optimization problems. In: 2006 IEEE Conference on Evolutionary Computation (CEC 2006), pp. 3352–3360. IEEE Press, Piscataway (2006)
10. Brockhoff, D., Zitzler, E.: Dimensionality reduction in multiobjective optimization: The minimum objective subset problem. In: Waldmann, K.H., Stocker, U.M. (eds.) Operations Research Proceedings 2006, pp. 423–429. Springer, Heidelberg (2007)
11. Brockhoff, D., Saxena, D.K., Deb, K., Zitzler, E.: On handling a large number of objectives a posteriori and during optimization. In: Knowles, J., Corne, D., Deb, K. (eds.) Multi-Objective Problem Solving from Nature: From Concepts to Applications. Natural Computing Series, pp. 377–403. Springer, Heidelberg (2008)
12. Corne, D.W.: Single objective = past, multiobjective = present, ??? = future. In: Michalewicz, Z. (ed.) 2008 IEEE Conference on Evolutionary Computation (CEC), Part of 2008 IEEE World Congress on Computational Intelligence (WCCI 2008). IEEE Press, Piscataway (2008)
13. Michalski, R.S.: Learnable evolution model: Evolutionary processes guided by machine learning. *Machine Learning* 38, 9–40 (2000)
14. Sheri, G., Corne, D.W.: The simplest evolution/learning hybrid: LEM with KNN. In: IEEE World Congress on Computational Intelligence, pp. 3244–3251. IEEE Press, Hong Kong (2008)
15. Sheri, G., Corne, D.W.: Learning-assisted evolutionary search for scalable function optimization: LEM(ID3). In: IEEE World Congress on Computational Intelligence. IEEE Press, Barcelona (2010)
16. Lozano, J.A., Larrañaga, P., Inza, I., Bengoetxea, E. (eds.): Towards a New Evolutionary Computation: Advances on Estimation of Distribution Algorithms. Springer, Heidelberg (2006)
17. Pelikan, M., Sastry, K., Goldberg, D.E.: Multiobjective estimation of distribution algorithms. In: Pelikan, M., Sastry, K., Cantú-Paz, E. (eds.) Scalable Optimization via Probabilistic Modeling: From Algorithms to Applications. SCI, pp. 223–248. Springer, Heidelberg (2006)
18. Martí, L., García, J., Berlanga, A., Coello Coello, C.A., Molina, J.M.: On current model-building methods for multi-objective estimation of distribution algorithms: Shortcomings and directions for improvement. Technical Report GIAA2010E001, Grupo de Inteligencia Artificial Aplicada, Universidad Carlos III de Madrid, Colmenarejo, Spain (2010)
19. Grossberg, S.: Studies of Mind and Brain: Neural Principles of Learning, Perception, Development, Cognition, and Motor Control. Reidel, Boston (1982)
20. Sarle, W.S.: Why statisticians should not FART. Technical report, SAS Institute, Cary, NC (1995)
21. Williamson, J.R.: Gaussian ARTMAP: A neural network for fast incremental learning of noisy multidimensional maps. *Neural Networks* 9, 881–897 (1996)
22. Martí, L., García, J., Berlanga, A., Molina, J.M.: Moving away from error-based learning in multi-objective estimation of distribution algorithms. In: Branke, J., Alba, E., Arnold, D., Bongard, J., Brabazon, A., Butz, M.V., Clune, J., Cohen, M., Deb, K., Engelbrecht, A., Krasnogor, N., Miller, J., O’Neill, M., Sastry, K., Thierens, D., Vanneschi, L., van Hemert, J., Witt, C. (eds.) GECCO 2010: Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation, pp. 545–546. ACM Press, New York (2010)

23. Ahn, C.W., Ramakrishna, R.S.: Multiobjective real-coded Bayesian optimization algorithm revisited: Diversity preservation. In: GECCO 2007: Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation, pp. 593–600. ACM Press, New York (2007)
24. Shapiro, J.: Diversity loss in general estimation of distribution algorithms. In: Runarsson, T.P., Beyer, H.-G., Burke, E.K., Merelo-Guervós, J.J., Whitley, L.D., Yao, X. (eds.) PPSN 2006. LNCS, vol. 4193, pp. 92–101. Springer, Heidelberg (2006)
25. Yuan, B., Gallagher, M.: On the importance of diversity maintenance in estimation of distribution algorithms. In: GECCO 2005: Proceedings of the 2005 Conference on Genetic and Evolutionary Computation, pp. 719–726. ACM Press, New York (2005)
26. Peña, J.M., Robles, V., Larrañaga, P., Herves, V., Rosales, F., Pérez, M.S.: GA-EDA: Hybrid evolutionary algorithm using genetic and estimation of distribution algorithms. In: Orchard, B., Yang, C., Ali, M. (eds.) IEA/AIE 2004. LNCS (LNAI), vol. 3029, pp. 361–371. Springer, Heidelberg (2004)
27. Zhang, Q., Sun, J., Tsang, E.: An evolutionary algorithm with guided mutation for the maximum clique problem. *IEEE Transactions on Evolutionary Computation* 9(2), 192–200 (2005)
28. Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C.M., Grunert da Fonseca, V.: Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Transactions on Evolutionary Computation* 7(2), 117–132 (2003)
29. While, L., Hingston, P., Barone, L., Huband, S.: A faster algorithm for calculating hypervolume. *IEEE Transactions on Evolutionary Computation* 10(1), 29–38 (2006)
30. Fonseca, C.M., Paquete, L., López-Ibáñez, M.: An improved dimension–sweep algorithm for the hypervolume indicator. In: 2006 IEEE Congress on Evolutionary Computation (CEC 2006), pp. 1157–1163 (2006)
31. Beume, N., Rudolph, G.: Faster S–metric calculation by considering dominated hypervolume as Klee’s measure problem. In: Kovalechuk, B. (ed.) Proceedings of the Second IASTED International Conference on Computational Intelligence, pp. 233–238. IASTED/ACTA Press (2006)
32. Beume, N.: S–metric calculation by considering dominated hypervolume as Klee’s measure problem. *Evolutionary Computation* 17(4), 477–492 (2009); PMID: 19916778
33. Bringmann, K., Friedrich, T.: Approximating the volume of unions and intersections of high–dimensional geometric objects. *Computational Geometry* 43(6–7), 601–610 (2010)
34. Papadimitriou, C.M.: Computational Complexity. Addison-Wesley, Reading (1994)
35. Deolalikar, V.: $P \neq NP$. Technical report, Hewlett Packard Research Labs, Palo Alto, CA, USA (2010)
36. Box, G.E.P., Muller, M.E.: A note on the generation of random normal deviates. *Annals of Mathematical Statistics* 29, 610–611 (1958)
37. Huband, S., Hingston, P., Barone, L., While, L.: A review of multiobjective test problems and a scalable test problem toolkit. *IEEE Transactions on Evolutionary Computation* 10(5), 477–506 (2006)

38. Martí, L., García, J., Berlanga, A., Molina, J.M.: Introducing MONEDA: Scalable multiobjective optimization with a neural estimation of distribution algorithm. In: Keizer, M., Antoniol, G., Congdon, C., Deb, K., Doerr, B., Hansen, N., Holmes, J., Hornby, G., Howard, D., Kennedy, J., Kumar, S., Lobo, F., Miller, J., Moore, J., Neumann, F., Pelikan, M., Pollack, J., Sastry, K., Stanley, K., Stoica, A., Talbi, E.G., Wegener, I. (eds.) *GECCO 2008: 10th Annual Conference on Genetic and Evolutionary Computation*, pp. 689–696. ACM Press, New York (2008); EMO Track “Best Paper” Nominee
39. Bosman, P.A.N., Thierens, D.: The naive MIDEA: A baseline multi-objective EA. In: Coello Coello, C.A., Hernández Aguirre, A., Zitzler, E. (eds.) *EMO 2005*. LNCS, vol. 3410, pp. 428–442. Springer, Heidelberg (2005)
40. Ahn, C.W.: *Advances in Evolutionary Algorithms*. In: *Theory, Design and Practice*. Springer, Heidelberg (2006) ISBN 3-540-31758-9
41. Beume, N., Naujoks, B., Emmerich, M.: SMS-EMOA: Multiobjective selection based on dominated hypervolume. *European Journal of Operational Research* 181(3), 1653–1669 (2007)
42. Knowles, J., Thiele, L., Zitzler, E.: A tutorial on the performance assessment of stochastic multiobjective optimizers. TIK Report 214, Computer Engineering and Networks Laboratory (TIK), ETH Zurich (2006)
43. Chambers, J., Cleveland, W., Kleiner, B., Tukey, P.: *Graphical Methods for Data Analysis*. Wadsworth, Belmont (1983)
44. Mann, H.B., Whitney, D.R.: On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics* 18, 50–60 (1947)
45. Bader, J.: *Hypervolume-Based Search for Multiobjective Optimization: Theory and Methods*. PhD thesis, ETH Zurich, Switzerland (2010)