

Application of data mining in multiobjective optimization problems

Amir Mosavi*

Faculty of Informatics, University of Debrecen, No. 1, Egyetem Str., Debrecen, Hungary

Received 17 September 2012 / Accepted 5 November 2013 / Published online 10 February 2014

Abstract – In the most engineering optimization design problems, the value of objective functions is not clearly defined in terms of design variables. Instead it is obtained by some numerical analysis such as FE structural analysis, fluid mechanic analysis, and thermodynamic analysis, etc. Usually, these analyses are considerably time consuming to obtain a value of objective functions. In order to make the number of analyses as few as possible a methodology is presented as a supporting tool for the meta-modeling techniques. Researches in meta-modeling for multiobjective optimization are relatively young and there is still much to do. It is shown that visualizing the problem on the basis of the randomly sampled geometrical data of CAD and CAE simulation results, in addition to utilizing classification tool of data mining could be effective as a supporting system to the available meta-modeling techniques. To evaluate the effectiveness of the proposed method a study case in 3D wing design is given. Along with this example, it is discussed how effective the proposed methodology could be in the practical engineering problems.

Key words: Multiobjective optimization, Meta-modeling, Data mining.

1. Introduction

The research field of considering decision problems with multiple conflicting objectives is known as multiple criteria decision making (MCDM) [1]. Solving a multiobjective optimization problem has been characterized as supporting the decision making (DM) in finding the best solution for the DM's problems. DM and optimization creates typically an interactive procedure for finding the most preferred solution. It has been tried to pay attention on improving all the defined objective functions instead of reduce or ignore some of them. For this reason the objective function are treated by trade-off analysis methods.

The complete process of multiobjective optimization has two parts (1) multiobjective optimization process which tries to find the Pareto frontier solutions (2) decision making process which tries to make the best decision out of the possible choices. This paper focuses on the first part which mostly deals with variables, constraints and objective functions.

1.1 Computational intelligence and multiobjective optimization

The methods for multiobjective optimization using computational intelligence along with real applications seem quite

new. However it has been observed that techniques of computational intelligence are effective in this regard [1]. Moreover, techniques of multiobjective optimization themselves can also be applied to develop effective methods in computational intelligence [2].

Currently there are many computational intelligence-based methods available to generate Pareto frontiers. However, it is still difficult to generate Pareto frontiers in the cases with more than three objectives. In this situation, methods of sequential approximate optimization of computational intelligence with meta-modeling are recognized to be very effective in many practical problems [1, 3].

1.2 Meta-modeling and multiobjective optimization; focusing on shape optimization

Meta-modeling is a method for building simple and computationally inexpensive models which replicate the complex relationships. However the research in meta-modeling for multiobjective optimization is relatively young and there is still much to do. So far there are few standards for comparisons of methods, and little is yet known about the relative performance and effectiveness of different approaches [3].

The most famous methods of Meta-modeling are known as response surface methods (RSM) and design of experiments (DOE). Although, as it is concluded in previous efforts

*e-mail: a.mosavi@math.klte.hu

[16, 18–20], in the future, scalability of methods in variable dimension and objective space dimension will become more important, as the methods need to be capable of dealing with higher computation cost, noise and uncertainties.

According to [1, 10], where the application of meta-modeling optimization methods in industrial optimization problems is discussed, some of the major difficulties in real-life engineering design problems counted as (1) there are too many objective functions involved and (2) the function form of criteria is a black box, in which cannot be explicitly given in terms of design variables (3) the huge number of unranked and non organized input variables.

Additionally in engineering design problems, the value of objective functions is not clearly defined in terms of design variables. Instead it is obtained by some numerical analysis such as FE structural analysis, fluid mechanic analysis, thermodynamic analysis, etc. These analyses to obtain a single value for the objective functions are often time consuming. Considering the high computation costs the number of CAE evaluations/calculations are subjected to minimization with the aid of meta-models [10].

In order to make the number of analyses as few as possible, sequential approximate optimization is one of the possible methods, utilizing machine learning techniques for identifying the form of objective functions and optimizing the predicted objective function [1]. Machine learning techniques have been applied for approximating the black-box of CAE function in many practical projects. Although the major problems in these realms would be (1) how to approach an ideal approximation of the objective function based on as few sample data as possible, (2) how to choose additional data effectively. The objective functions are modeled by fitting a function through the evaluated points. This model is then used to help the prediction the value of future search points. Therefore those high performance regions of design space can be identified more rapidly. Moreover the aspects of dimensionality, noise and expensive-ness of evaluations are related to method selection [20]. However according to Bruyneel et al. [10] for the multiobjective capable version of meta-modeling algorithms further aspects such as how to define the improvement in a Pareto approximation set and how to model each objective function must be considered.

Today, numerical methods make it possible to obtain models or simulations of quite complex and large scale systems. But there are difficulties when the system is being modeled numerically. In this situation modeling the simplified problems is an effective method, generating a simple mode that captures only the relevant input and output variables instead of modeling the whole design space [3].

The increasing desire to apply optimization method in expensive CAE domains is driving forward research in meta-modeling. The RSM is probably the most widely applied to meta-modeling. The process of a meta-model from data is related to classical regression methods and also to machine learning [3]. When the model is updated using new samples, classical DOE principles are not effective. In meta-modeling, the training sets will often highly correlated data, which can affect the estimation of goodness of fit and generalization performance. Meta-modeling brings together a number of different

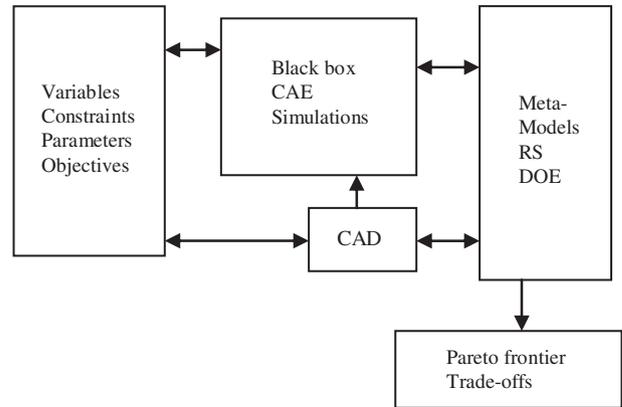


Figure 1. Meta-modeling tools in multiobjective optimization process.

fields to tackle the problem of how to optimize expensive functions. Classical DOE methods with employing evolutionary algorithms have delivered more advantage in this realm. Figure 1 describes the common arrangement of meta-modeling tools in multiobjective optimization processes. Worth mentioning that the other well-known CAD-Optimization integrations for shape optimization, e.g., [16, 18, 19] also follow the described arrangements.

2 Data mining classification in engineering applications

The particular advantage of Evolutionary Algorithms (EAs) in the multiobjective optimization applications (EMO) is that they work with a population of solutions. Therefore they can search for several Pareto optimal solutions providing the DM with a set of alternatives to choose from [9]. EMO-based techniques have application where mathematical-based methods have difficulties with. EMO are also helpful in knowledge discovery related tasks in particular for mining the data samples achieved from CAE and CAD systems [18, 19]. Useful mined information from the obtained EMO trade-off solutions have been discovered in many real-life engineering design problems.

2.1 Classifications

Finding useful information in large volumes of data drives the development of data mining procedure forward. Data mining classification process refers to the induction of rules that discriminate between data organized in several classes so as to gain predictive power [4].

There are some example applications of data mining classification in evolutionary multiobjective optimization available in the literature of [1, 5, 6, 11].

The goal of the classification algorithms is to discover rules by accessing the training sets. Then the discovered rules are evaluated using the test sets, which could not be seen during training [4].

In the classification procedures, the main goal is to use observed data to build a model, which is able predict the categorical or nominal the class of a dependent variable given the

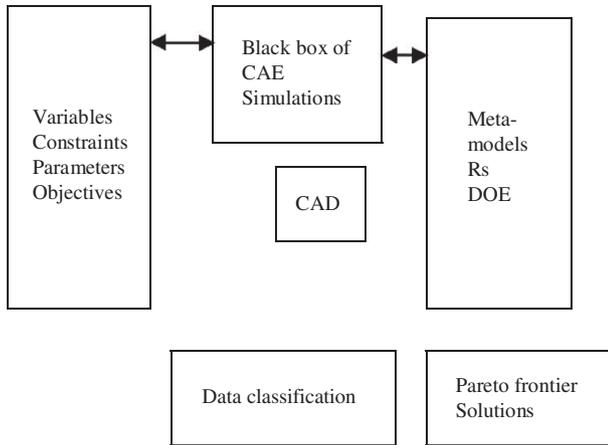


Figure 2. Supporting the meta-modeling process by mining the data.

value of the independent variables [4]. Obayashi [7] for the reason of data mining the engineering multiobjective optimization and visualization applied self-organizing maps (SOM) along with a data clustering method of data mining. Moreover Witkowski and Tushar [8] and Mosavi [12] used classification tools of data mining for decision making process of multiobjective optimization.

2.2 Modeling the problem

According to [1] before any optimization can be done, the problem must first be modeled. In this case identifying all the dimensions of the problem such as formulation of the optimization problem with specifying decision variables, objectives, constraints, and variable bounds is an important task. Mining the available sample data will help to better model the problem as it delivers more information about the importance of input variables and could rank the input variables. The proposed method of classification [12] which is presented in Figure 2 supposed to mine the input variables and resulted CAE data.

3. Three-objective and 42-variable optimization problem

The applications in engineering design have different disciplines to bring into the consideration. In mechanical engineering, the structural simulation is tightly integrated more than one discipline [10, 13, 14, 15, 16, 22]. Meanwhile, the trend nowadays is to utilize independent computational codes for each discipline [20]. In this situation, the aim of MCDM tools is to develop methods in order to guarantee that all physical variables be involved. Bo and Any [17] in aerodynamic optimization of a 3D wing has tried to utilize the multiobjective optimization techniques in a multidisciplinary environment.

In the similar cases [12, 16, 18, 20] in order to approach the optimal shape in an aerospace engineering optimization problem, the multiobjective optimization techniques are necessary to deal with all important objectives and variables efficiently. Here the optimization challenge is to identify as many optimal designs as possible to provide a choice of better decision. However the common task is very complicated with an increase

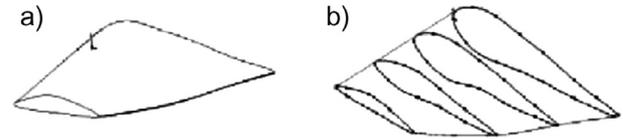


Figure 3. Airfoil geometry modeled by S-plines [12, 14]

in the number of design variables. Therefore the multiobjective optimization tasks become more difficult with the increasing number of variables [12, 21]. Although the recent advances in parametric CAD/CAE integrations [16, 18, 19] have reduced the complexity of the approach to some levels.

The airfoil of Figure 3a is subjected for shape improvement. The shape needs to be optimized in order to deliver minimum displacement distribution in terms of applied pressure on the surface. Figure 3b shows the basic curves of the surface modeled by S-plines. The utilized geometrical modeling methodology successfully implemented by Albers and Leon-Rovira [22]. Here for modeling the surface four profiles have been utilized with 42 points. The coordinates of all points are supplied by a digitizer in which each point includes three dimensions of X , Y and Z . Consequently there are 126 columns plus 3 objectives which are going to be more complicated by adding the variables constraints.

The objectives are listed as follow:

- Objective 1: Minimizing the displacements distribution in the airfoil for constant pressure value of α .
- Objective 2: Minimizing the displacements distribution in the airfoil for constant pressure value of 2α .
- Objective 3: Minimizing the displacements distribution in the airfoil in constant pressure value of 4α .

An optimal configuration of 42 variables supposed to satisfy the above three described objectives.

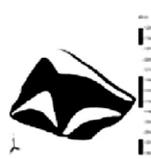
In the described multiobjective optimization problem the number of variables is subjected to minimization before the multiobjective optimization process took place in order to change the large scale design space to the smaller design space. Here the proposed and utilized model reduction methodology differs from the previous efforts of Filomeno et al. [21] in terms of applicability and ease of use in general multiobjective optimization design problems.

The datasets for data mining are supplied from Table 1. This table has gathered initial datasets including shapes' geometries and simulation results from five calculations, based on random initial values of variables which in the proposed method will be analyzed. In the next chapter it will be discussed how the data from five random calculations could be utilized for creating the smaller design space of multiobjective optimization.

4 Methodology and experimental results

The effectiveness of data mining tools in multiobjective optimization problems is presented by Coello et al. [2]. And earlier in [4] the classification rules for evolutionary multiobjective

Table 1. Training dataset including five calculations' results.

	Variables configuration: V_1-V_4	CAD model	Displacement distribution	Objective results
1	0,1,1.2,1,0.8,0.4,0.2,0,-0.4,-0.48,0.6,-0.8,-0.72,0,0.84,0.99,0.84,0.62,0.26,0,-0.20,-0.40,-0.36,-0.70,-0.58,0,0.59,0.78,0.56,0.30,0,-0.21,-0.24,-0.38,-0.38,0,0.26,0.50,0.39,-0.03,-0.10,-0.12,			$O_1 = c$ $O_2 = c$ $O_3 = c$
2	0,1,1,1,1.21,9,0.82,0.42,0.18,1,-0.41,-0.46,-0.62,-0.81,-0.70,0,0.86,0.1,0.82,0.60,0.2,5,0.01,-0.20,-0.39,-0.39,-0.70,-0.58,0,0.58,0.76,0.57,0.32,0,-0.21,-0.23,-0.37,-0.39,0.26,0.54,0.40,-0.03,-0.1,-0.1,0			$O_1 = b$ $O_2 = c$ $O_3 = b$
3	0,1,1.2,1,0.8,0.4,0.2,0,-0.4,-0.48,-0.6,-0.8,-0.72,0,88,0.99,0.84,0.62,0.2,6,0,-0.23,-0.35,-0.37,-0.70,-0.54,0,0.58,0.76,0.58,0.31,0,-0.23,-0.23,-0.37,-0.37,0,0.24,0.50,0.40,-0.03,-0.13,-0.10,			$O_1 = d$ $O_2 = c$ $O_3 = b$
4	0,1,3,1.23,1.06,0.83,0.4,1,0.28,0.07,-0.41,-0.48,-0.6,-0.8,-0.78,0,0.84,92,0.84,0.6,2,0.26,0,-0.23,-0.39,-0.37,-0.70,-0.54,0,0.58,0.76,0.58,0.31,0,-0.24,-0.22,-0.36,-0.38,0,0.24,0.52,0.38,-0.02,-0.12,-0.12,			$O_1 = d$ $O_2 = c$ $O_3 = b$
5	0,1,0.1,1.21,1,0.8,0.4,0.2,1,0,-0.41,-0.47,-0.59,-0.79,-0.69,0,0.80,1.01,0.86,0.64,0.26,-0.01,-0.20,-0.40,-0.40,-0.72,-0.56,0,0.58,0.76,0.58,0.31,0,-0.23,-0.23,-0.37,-0.37,0,0.24,0.52,0.38,-0.06,-0.10,-0.10,			$O_1 = c$ $O_2 = d$ $O_3 = e$

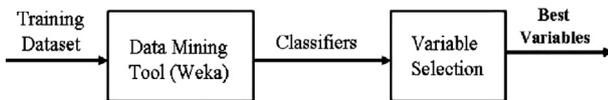


Figure 4. Proposed methodology workflow.

algorithms were well implemented in which along with the research work of Witkowski and Tushar [8] forms the proposed methodology via a novel workflow. The workflow of mining procedure methodology is described in Figure 4. In this method, the classification method is utilized to create several classifiers or decision trees. In the next steps the most important variables which have more effects on the objectives are selected.

Regressions and model trees are constructed by a decision tree to build an initial tree. However, most decision tree algorithms choose the splitting attribute to maximize the information gain. It is appropriate for numeric prediction to minimize the intra subset variation in the class values under each branch.

The splitting criterion is used to determine which variable is the better to split the portion T of the training data. Based on the treating the standard deviation of the objective values in T as a

measure of the error and calculation the expected reduction in error as a result of testing each variable is calculated. The variables which maximize the expected error reduction are chosen for splitting. The splitting process terminates when the objective values of the instances vary very slightly, that is, when their standard deviation has only a small fraction of the standard deviation of the original instance set. Splitting also terminates when just a few instances remain. Experiments show that the obtained results are not very sensitive to the exact choice of these thresholds. Data mining classifier package of Weka provides implementations of learning algorithms and dataset which could be preprocessed and feed into a learning scheme, and analyze the resulting classifier and its performance. The workbench includes methods for all the standard data mining problems such as regression, classification, clustering, association rule mining, and attribute selection. Weka also includes many data visualization facilities and data preprocessing tools. Three different data mining classification algorithms are applied (J48, BFTree, LADTree) and their performance are compared in order to choose attribute importance. The Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) of the class probability is estimated and assigned by the algorithm output. The RMSE is the square root of the average quadratic loss

Table 2. Variables importance ranking for three classification methods

Classification method	MAE	RMSE	Effective variables	Objectives
BFTree	0.370	0.517	15	O ₁
	0.412	0.519	23	O ₂
	0.418	0.555	41	O ₃
J48	0.309	0.514	15,24	O ₁
	0.482	0.642	13	O ₂
	0.378	0.590	35,41	O ₃
LADTree	0.277	0.500	15,24,2,32,41,39,3	O ₁
	0.604	0.769	23,22,18,15,42,2,17,	O ₂
	0.365	0.584	20 41,35,9,17,11,38,37, 16	O ₃

and the MAE is calculated in a similar way using the absolute instead of the squared difference.

The comparison between importance ranking results is obtained by our experiments listed in Table 2. It is concluded that in the worst case, more than 55% variable reduction is achieved. As one can see, BFTree and J48 algorithms have classified the datasets with less number of variables. While in LADTree algorithms, at least seven variables have utilized to classify dataset. Variables number 15 and 24 play much more important role in changing the first objective (O₁).

Variables number 41 and 35 also have effect on third objective (O₃) as well. According to the experimental results, it is possible to optimize the model by reducing the 45% number of variables. In Table 2, two types of classification error (MAE, RMSE) are shown for all algorithms corresponding to different class of objectives.

5 Conclusions

The modified methodology is demonstrated successfully in the framework. The author believes that the process is simple and fast. In order to deliver more information about the optimization variables in a reasonable way, data mining have been applied. Variables were ranked and organized utilizing three different classification algorithms. The presented results as reduced variables could speed up and scale up the process of optimization as a preprocessing step. Data mining tools has found to be effective in this regard. Additionally it is evidenced that the growing complexity can be easily handled by a preprocessing step utilizing data mining classification tools.

For future works studying the effectiveness of the introduced data reduction process is suggested. Also trying other tools of the data mining such as clustering, association rules, and comparison the results could be beneficial.

Acknowledgements. Author would like to thank Eesti Institute for funding this research with the generous grant of *Estophilus*.

References

1. Branke J, Deb K, Miettinen K, Słowiński R. 2008. Multi-objective optimization. Springer Berlin/Heidelberg, New York.

2. Coello C, Dehuri S, Ghosh S. 2009. Swarm intelligence for multi-objective problems in data mining. Springer, Berlin/Heidelberg, New York.
3. Knowles J, Nakayama H. 2008. Meta-modeling in multiobjective optimization. Springer, Berlin/Heidelberg, New York.
4. Kshetrapalapuram KK, Kirley M. 2005. Mining classification rules using evolutionary multi-objective algorithms. Knowledge-based intelligent, information and engineering systems, 3683, Springer, Berlin/Heidelberg, New York.
5. Freitas AA. 1998. On objective measures of rule surprisingness, in Principles of data mining and knowledge discovery, Springer Berlin, Heidelberg, pp. 1–9.
6. Rowland J. 2003. Generalisation and model selection in supervised learning with evolutionary computation, in Applications of Evolutionary Computing, Springer Berlin, Heidelberg, pp. 119–130.
7. Obayashi S. 2005. Evolutionary multiobjective optimization and visualization, in New Developments in Computational Fluid Dynamics, Springer Berlin, Heidelberg, pp. 175–185.
8. Witkowski K, Tushar M. 2009. Decision making in multiobjective optimization for industrial application-Data mining and visualization of Pareto, In Proceedings of 7th European LS-DYNA Conference, USA, 416–423.
9. Deb K. 2007. Current trends in evolutionary multi-objective optimization. Int. J. Simul. Multidisci. Des. Optim., 2, 1–8.
10. Bruyneel M, Colson B, Jetteur P, Raick C, Remouchamps A, Grihon S. 2008. Recent progress in the optimal design of composite structures: industrial solution procedures on case studies. Int. J. Simul. Multidisci. Des. Optim., 2, 283–288.
11. Bedingfield SE, Smith KA. 2003. Evolutionary Rule Generation classification and its Application to multi-class data. In Computational Science – ICCS 2003, Springer Berlin Heidelberg, 868–876.
12. Mosavi A. 2010. Multiple criteria decision-making pre-processing using data mining tools. IJCSI, International Journal of Computer Science Issues, 7, 26–34.
13. Arularasan V. 2008. Modeling and simulation of a parallel plate heat sink using computational fluid dynamics. Int. J. Adv. Manuf. Technol., 5, 172–183.
14. Esmaeili M, Mosavi A. 2010. Variable reduction for multi-objective optimization using data mining techniques; application to aerospace structures. Proceedings of the 2nd International IEEE Conference on Computer Engineering and Technology, 5, 333–337.
15. Olcer AI. 2007. A hybrid approach for multi-objective combinatorial optimization problems in ship design and shipping. Computers & Operations Research, 35, 2760–2775.
16. Toussaint L, Lebaal N, Schlegel D, Gomes S. 2010. Automatic optimization of air conduct design using experimental data and numerical results. Int. J. Simul. Multidisci. Des. Optim., 4, 77–83.
17. Bo Y, Any X. 2008. Aerodynamic optimization of 3D wing based on iSIGHT. Appl. Math. Mech. -Engl. Ed., 5, 603–610.
18. Bluntzer J-B, Gomes S, Bassir DH, Varret A, Sagot JC. 2008. Direct multi-objective optimization of parametric geometrical models stored in PLM systems to improve functional product design. Int. J. Simul. Multidisci. Des. Optim., 2, 83–90.
19. Vik P, Luís D, Guilherme P, Oliveira J. 2010. Automatic generation of computer models through the integration of production systems design software tools. Int. J. Simul. Multidisci. Des. Optim., 4, 141–148.
20. Mosavi A. 2009. Hydrodynamic design and optimization: application to design a general case for extra equipments on the

- submarine's hull. Proceedings of the International IEEE Conference on Computer Technology and Development, 2, 139–143.
21. Filomeno R, Coelho C, Breitkopf P, Knopf-Lenoir C. 2008. Model reduction for multidisciplinary optimization-application to a 2D wing. Struct. Multidisc. Optim., 7, 29–48.
22. Albers A, Leon-Rovira N. 2009. Development of an engine crankshaft in a framework of computer-aided innovation. Computers in Industry, 60, 604–612.

Cite this article as: Mosavi A: Application of data mining in multiobjective optimization problems. Int. J. Simul. Multisci. Des. Optim., 2014, 5, A15.