

Efficient Global Optimization (EGO) for Multi-Objective Problem and Data Mining

Shinkyu Jeong

Institute of Fluid Science, Tohoku University
Katahira2-1-1, Aoba-ku
Sendai 980-8577, Japan
jeong@edge.ifs.tohoku.ac.jp

Shigeru Obayashi

Institute of Fluid Science, Tohoku University
Katahira2-1-1, Aoba-ku
Sendai 980-8577, Japan
obayashi@ieee.org

Abstract- In this study, a surrogate model is applied to multi-objective aerodynamic optimization design. For the balanced exploration and exploitation with the surrogate model, objective functions are converted to the Expected Improvements (EI) and these values are directly used as fitness values in the multi-objective optimization. Among the non-dominated solutions about EIs, additional sample points for the update of the Kriging model are selected. The present method is applied to a transonic airfoil design. In order to obtain the information about design space, two data mining techniques are applied to design results. One is Analysis of Variance (ANOVA) and the other is Self-Organizing Map (SOM).

1 Introduction

Recently, in order to ease computational burden of optimization, surrogate models [1] are widely used in the field of engineering design. However, most of surrogate models have a problem with fidelity because the function value predicted by surrogate models contains an uncertainty in it. Especially, in the aerodynamic design field [2, 3] where the response functions are often nonlinear and multimodal, the surrogate models cannot provide the sufficient fidelity for the optimization design.

In order to make up for the low fidelity of surrogate model, Jones et al. suggested the efficient global optimization (EGO) algorithm [4]. It makes use of the Kriging model, which is developed in the field of spatial statistic and geostatic, as a surrogate model. The Kriging model predicts not the function value itself but the distribution of the function value. From the distribution of the function value, one can predict not only the function value but also its uncertainty. The uncertainty information plays a key role in EGO. In EGO, the exploration is based on the potential of being the optimum instead of the objective function value itself. According to the concept of EGO, the solution that has a high predicted function value with a large error may be a more promising solution than the solution that has a low predicted function value with a small error in the minimization problem. By using EGO, the balanced exploitation and exploration is possible. EGO is successfully applied to the single objective aerodynamic optimization design [5]. In this

study, the EGO is applied to multi-objective aerodynamic design problem. For the multi-objective problem, Knowles et al. suggested ParEGO (Pareto EGO) which converts the objective functions into a single objective function by using a parameterized weighting vector [6]. Although ParEGO showed a good performance on several tested functions, its ability to find the Pareto front largely depends on the selection of the weighting vector. Thus, in this study, all objective functions are converted into the EIs of objective functions and these values are directly used as the fitness values in the multi-objective optimization. This makes it possible to obtain the correct non-dominated solutions about EIs. From these non-dominated solutions, designer can select the additional sample points for the update of the Kriging model. This method prevents an incorrect Pareto front exploration caused by the weighting vector.

In the multi-objective optimization, finding correct non-dominated solutions is very important. However, in the field of engineering design, the most important thing is the determination of the final design from the non-dominated solutions. Thus, it is preferable to supply the non-dominated solutions with the helpful information for the selection of the final design. The information about the design space, such as trade-offs between objective functions and relations between design variables and objective functions, is one of the useful information for the determination of the final design. Furthermore, the information of the relations between design variables and objective functions makes it possible to simplify the design space by eliminating the design variables which do not have a significant influence on the objective functions.

The process to find the information from the design results may be called 'data mining'. Recently, Parmee et al. applied the projection onto design variable hyperplane and the parallel co-ordinate box plot to the data obtained by the Cluster Oriented Genetic Algorithms (COGAs) [7] for the data mining. It makes possible to identify the trade-off relation between objective functions and the effect of design variable on the objective function. In this study, more advanced data mining techniques, Analysis of Variance (ANOVA) and Self-Organizing Map (SOM), are introduced. The former uses a variance of the objective functions due to design variables on surrogate models to obtain the information about design space. This method

can identify not only the effect of each design variable but also the effect of interaction between design variables on objective function. ANOVA expresses the information quantitatively. On the other hand, the latter employs an advanced variant of unsupervised neural networks and clustering technique. This method can not only visualize the trade-off between objective functions but also the relation between the design variables and the objective functions. The method expressed the information in the qualitative way.

2 EGO for Multi-Objective Problem: Multi-EGO

2.1 Kriging Model

The present Kriging model is composed of a constant global model μ and the Gaussian stochastic process $Z(\mathbf{x})$ representing a deviation from the global model:

$$y(\mathbf{x}) = \mu + Z(\mathbf{x}) \quad (1)$$

where \mathbf{x} is an m -dimensional vector (m design variables). The correlation between $Z(\mathbf{x}^i)$ and $Z(\mathbf{x}^j)$ is strongly related to the distance between the two corresponding points, \mathbf{x}^i and \mathbf{x}^j . In the Kriging model, a weighted distance is used instead of the Euclidean distance because the Euclidean distance weighs all design variables equally. The distance function between the point at \mathbf{x}^i and \mathbf{x}^j is expressed as

$$d(\mathbf{x}^i, \mathbf{x}^j) = \sum_{k=1}^m \theta_k |x_k^i - x_k^j|^2 \quad (2)$$

where θ_k ($0 \leq \theta_k \leq \infty$) is the k_{th} element of the correlation vector parameter $\boldsymbol{\theta}$. The correlation between the point \mathbf{x}^i and \mathbf{x}^j is defined as

$$Corr[Z(\mathbf{x}^i), Z(\mathbf{x}^j)] = \exp[-d(\mathbf{x}^i, \mathbf{x}^j)] \quad (3)$$

in the present model.

The Kriging predictor [8] is

$$\hat{y}(\mathbf{x}) = \hat{\mu} + \mathbf{r}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\hat{\mu}) \quad (4)$$

where $\hat{\mu}$ is the estimated value of μ , \mathbf{R} denotes the $n \times n$ matrix whose (i, j) entry is $Corr[Z(\mathbf{x}^i), Z(\mathbf{x}^j)]$ and $\mathbf{1}$ denotes an n -dimensional unit vector. \mathbf{r} is vector whose i_{th} element is

$$r_i(\mathbf{x}) \equiv Corr[Z(\mathbf{x}), Z(\mathbf{x}^i)] \quad (5)$$

and $\mathbf{y} = [y(\mathbf{x}^1), \dots, y(\mathbf{x}^n)]$.

The unknown parameter, $\boldsymbol{\theta}$, for the Kriging model can be estimated by maximizing the following likelihood function

$$Ln(\hat{\mu}, \hat{\sigma}^2, \boldsymbol{\theta}) = -\frac{n}{2} \ln(\hat{\sigma}^2) - \frac{1}{2} \ln(|\mathbf{R}|) \quad (6)$$

Maximizing the likelihood function is an m -dimensional unconstrained non-linear optimization problem. In this paper, GA is adopted to solve this problem. Solving this problem requires several thousand of matrix inversion which consumes a lot of computational time. However, compared with computational time of high-fidelity evaluation of the objective function, its computational time is negligible.

The uncertainty of the predicted value can be estimated by using the mean squared error of predicted value as follow:

$$s^2(\mathbf{x}) = \hat{\sigma}^2 \left[1 - \mathbf{r}'\mathbf{R}^{-1}\mathbf{r} + \frac{(1 - \mathbf{1}'\mathbf{R}^{-1}\mathbf{r})^2}{\mathbf{1}'\mathbf{R}^{-1}\mathbf{1}} \right] \quad (7)$$

2.2 Expected Improvement

EI means the potential of being smaller than current minimum (maximum) in minimization (maximization) problem. EI considers the predicted function value and its uncertainty at the same time. Thus, the solution that has a large objective function value and a large uncertainty may be a more promising solution than the solution that has a small objective function value and a small uncertainty in the minimization problem. This feature makes it possible to explore the design space globally. EI is calculated by using following equation [9]:

$$E[I(\mathbf{x})] = E[\max(f_{\min} - \hat{y}, 0)] \\ = (f_{\min} - \hat{y})\Phi\left(\frac{f_{\min} - \hat{y}}{s}\right) + s\phi\left(\frac{f_{\min} - \hat{y}}{s}\right) \quad (8)$$

Φ and ϕ are the standard distribution and normal density, respectively.

In this study, EIs of objective functions are directly used as fitness values in the multi-objective optimization. GA maximizes EIs of each objective function to find the non-dominated solutions about EIs and several points are selected from the non-dominated solutions for the update of the Kriging model. Overall procedure of EGO used in the study is shown in Fig. 1.

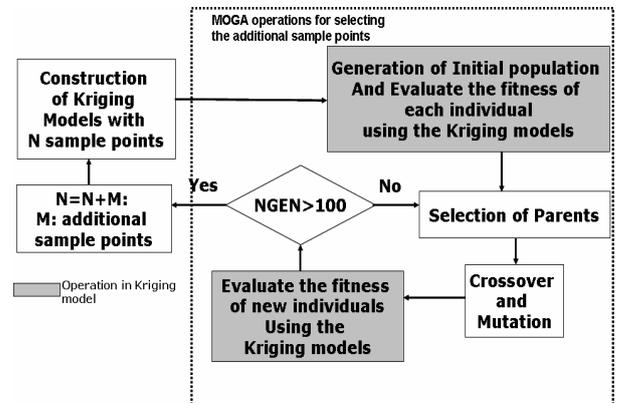


Figure 1. Overall procedure of EGO used in this study

3 Data Mining

3.1 Analysis of Variance (ANOVA)

ANOVA is one of the data mining techniques showing the effect of each design variable on the objective functions in a quantitative way. ANOVA uses the variance of the function due to the design variables on surrogate models. By decomposing the total variance of function into the variance due to each design variable, the influence of each design variable on the objective

function can be calculated. The decomposition is accomplished by integrating out the variables of function \hat{y} .

The total mean ($\hat{\mu}_{total}$) and variance ($\hat{\sigma}_{total}^2$) of model \hat{y} are as follows:

$$\hat{\mu}_{total} \equiv \int \cdots \int \hat{y}(x_1, \dots, x_n) dx_1 \cdots dx_n \quad (9)$$

$$\hat{\sigma}_{total}^2 = \int \cdots \int [\hat{y}(x_1, \dots, x_n) - \hat{\mu}_{total}]^2 dx_1 \cdots dx_n \quad (10)$$

The main effect of variable x_i is

$$\hat{\mu}_i(x_i) \equiv \int \cdots \int \hat{y}(x_1, \dots, x_n) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n - \hat{\mu}_{total} \quad (11)$$

The variance due to the design variable x_i is

$$\int [\hat{\mu}_i(x_i)]^2 dx_i \quad (12)$$

The proportion of the variance due to design variable x_i to the total variance of the function can be expressed by dividing Eq. (10) by Eq. (12).

$$\frac{\int [\hat{\mu}_i(x_i)]^2 dx_i}{\int \cdots \int [\hat{y}(x_1, \dots, x_n) - \hat{\mu}_{total}]^2 dx_1 \cdots dx_n} \quad (13)$$

This value indicates how much effect design variable x_i gives on the objective function \hat{y} .

3.2 Self-Organizing Map (SOM)

SOM is an unsupervised learning, nonlinear projection algorithm [10] from high to low-dimensional space. This projection is based on self-organization of a low-dimensional array of neurons. In the projection algorithm, the weights between the input vector and the array of neurons are adjusted to represent features of the high dimensional data on the low-dimensional map. The closer two patterns are in the original space, the closer is the response of two neighboring neurons in the low-dimensional space. Thus, SOM reduces the dimension of input data while preserving their features. While ANOVA shows the relation between the design variables and the objective function quantitatively, SOM shows it qualitatively.

A neuron used in SOM is associated with weight vector $\mathbf{m}_i = [m_{i1}, m_{i2}, \dots, m_{in}]$ ($i=1, \dots, M$) where n is equal to the dimension of input vector \mathbf{x} and M is number of neuron. Each neuron is connected to adjacent neurons by a neighborhood relation. The Batch-SOM algorithm is used for SOM training in this study. The algorithm consists of two steps that are iterated until no more significant changes occur. First, the best-matching unit (BMU) c_i whose weight vector is closest to the input vector \mathbf{x}_i is determined as follow:

$$c_i = \arg \min_j \|\mathbf{x}_i - \mathbf{m}_j\| \quad (14)$$

And the weight vectors of c_i and it neighbor are updated by using following equation:

$$\mathbf{m}_j^* = \frac{\sum_i h_{jc_i} \mathbf{x}_i}{\sum_{i'} h_{jc_{i'}}} \quad (15)$$

where \mathbf{m}_j^* is the updated weight vector.

The neighborhood relationship between two neurons j and k is defined by the following Gaussian-like function:

$$h_{jk} = \exp\left(-\frac{d_{jk}^2}{r_t^2}\right) \quad (16)$$

where d_{jk} denotes the Euclidean distance between the neuron k and the neuron j on the map, and r_t denotes the neighborhood radius which is decreased with the iteration steps t . Compared with the standard SOM, which updates the weight vector when each record is read and matched, the Batch-SOM takes a ‘batch’ of data (typically all records), and performs a ‘collected’ update of the weight vectors after all records have been matched. This is much like ‘epoch’ learning in supervised neural networks. The Batch-SOM is a more robust approach, since it mediated over a large number of learning steps. By adopting the Batch-SOM and linear initialization for input data, the uniqueness of the map can be ensured.

In this study, commercial software Viscovary® SOMine [11] produced by Eudaptics GmbH in Austria is used. Much like other SOM, SOMine creates a map in a two-dimensional hexagonal grid. Starting from numerical, multivariate data, the neurons on the grid gradually adapt to the intrinsic shape of the data distribution. Since the order on the grid reflects the neighborhood within the data, features of the data distribution can be read off from the emerging map on the grid. The trained SOM is systematically converted into visual information.

It is efficient to group all neurons by the similarity to facilitate SOM for the qualitative analysis, because number of neurons on the SOM is large as a whole. This process of grouping is called ‘clustering’. Hierarchical agglomerative algorithm is used for the clustering here. First, each node itself forms a single cluster, and two clusters, which are adjacent in the map, are merged in each step. The distance between two clusters is calculated by using the SOM-ward distance. The number of clusters is determined by the hierarchical sequence of clustering. A relatively small number of clusters are used for visualization, while a large number of clusters are used for the generation of weight vectors for respective design variables.

4 Application to Aerodynamic Design

4.1 Definition of Optimization Problem

The present method is applied to a transonic airfoil design. In the transonic region, the flowfield is drastically changed even with a small fluctuation. For the robust transonic airfoil design, the objective functions are defined as follows:

$$\begin{aligned} \text{Minimize} \quad & \text{obj1: Drag at fixed lift of 0.75 (Mach=0.70)} \\ & \text{obj2: Drag at fixed lift of 0.67 (Mach=0.74)} \end{aligned}$$

subject to

$$t/c > 11\%$$

First objective function is for the design condition and the other is for the off-design condition. The design constraint is the maximum thickness of the airfoil.

Geometry of airfoil is defined by using Non-uniform rational B-spline (NURBS) [12] to represent free-form shape. Total 26 design variables are used to represent the exact shape of airfoil as shown in Fig. 2. The search region of each design variable is determined to avoid unrealistic airfoil geometry such as a fish-tailed airfoil.

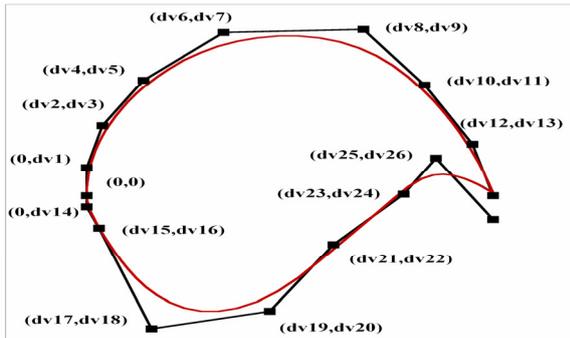


Figure 2. Airfoil geometry definition using the NURBS

4.2 Construction of Kriging Model

Sample points to fit the Kriging model are selected by using the Latin Hypercube Sampling (LHS) [13] with the constraint evaluation. Once LHS selects a point (airfoil), the point is checked whether it satisfies the design constraint or not. If the point satisfies the constraint, the point is selected as sample point, if not, the point is rejected. Total 26 sample points are selected with 50 divisions in LHS. Design variable distribution of 26 sample points are shown in Fig. 3. Sample points are uniformly spread in the search region of all design variables except dv7. This means that airfoil with small value of dv7 doesn't satisfy design constraint.

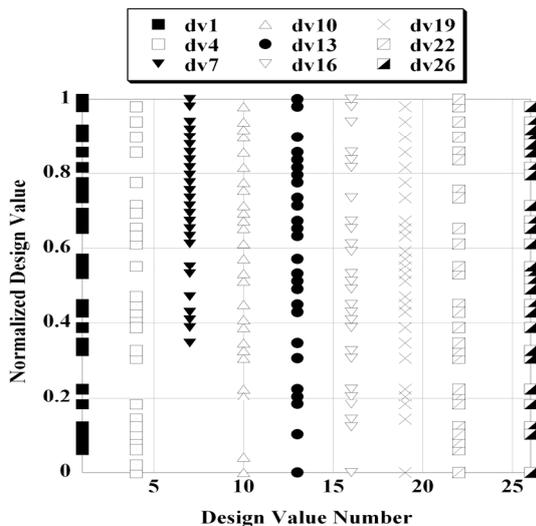


Figure 3. Distribution of initial sample points in the search region

In case of LHS without the constraint evaluation, only 13 points satisfy the design constraint with same number of division. It means that 37 sample points are selected from the infeasible design space. By using LHS coupled with the constraint evaluation, it is possible to select the sample points only from feasible design space.

Drag performances of 26 sample airfoils are evaluated by using a Navier-Stokes calculation. With the data obtained from the Navier-Stokes analysis, the Kriging parameter (θ) is determined by solving maximization problem of Eq. (6).

4.3 Design Results

On the Kriging models, EIs of two objective functions are maximized by using MOGA to find the non-dominated solutions about EIs. In the present MOGA, number of populations and generations are 512 and 100, respectively. Figure 4 shows the non-dominated solutions about EIs after the first EGO loop. EIs of objective functions are in the trade-off relation.

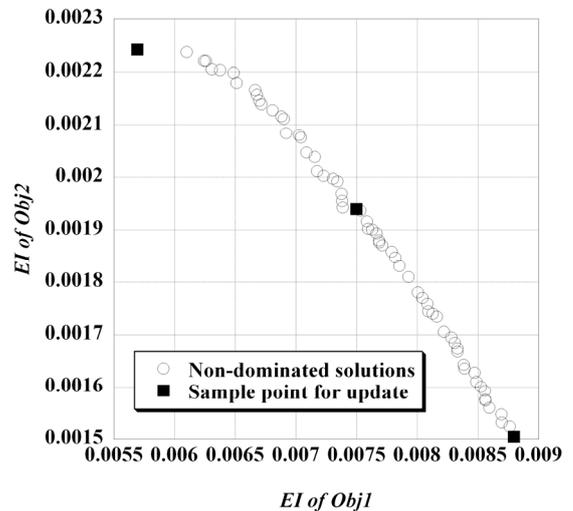


Figure 4. Non-dominated solutions about EI

Among these non-dominated solutions, three points are selected for the update of the Kriging models: i) the point whose EI of *obj1* is the largest, ii) the point whose EI of *obj2* is the largest, iii) mid point in the non-dominated solutions.

Figure 5 shows the non-dominated solutions about EIs obtained after the several EGO loops. The values of EI are gradually decreased as EGO loop iterated. It means that the accuracy near the Pareto front is improved by adding the additional sample points.

Figure 6 shows the initial sample points and the additional points selected by using EGO. Total 89 points are selected in this design. The additional points selected by EGO show better performances than the initial sample points. It means that EGO algorithm selected the additional points for update of the Kriging models correctly.

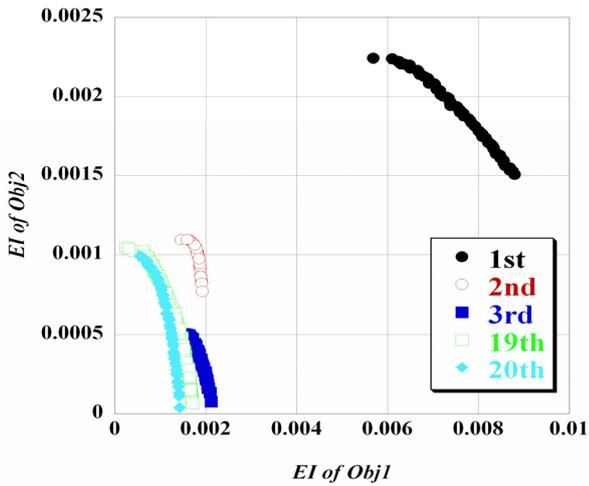


Figure 5. Non-dominated solutions of EIs

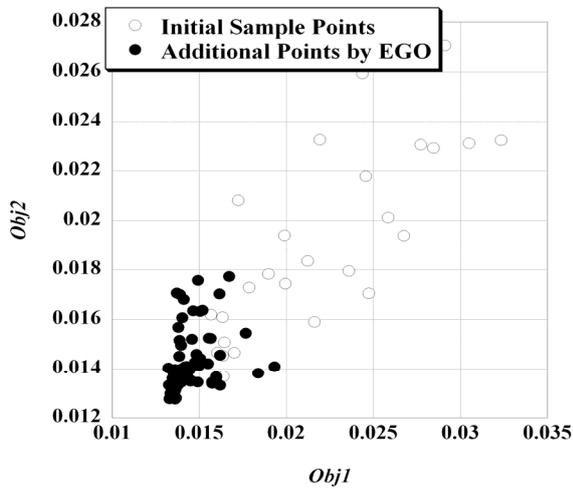


Figure 6 Initial sample points and additional points selected by EGO

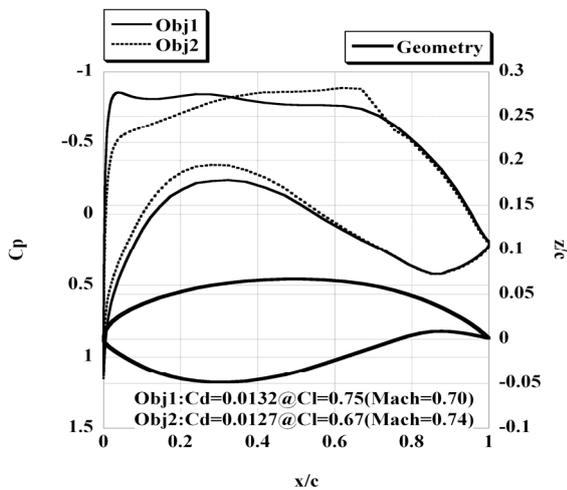


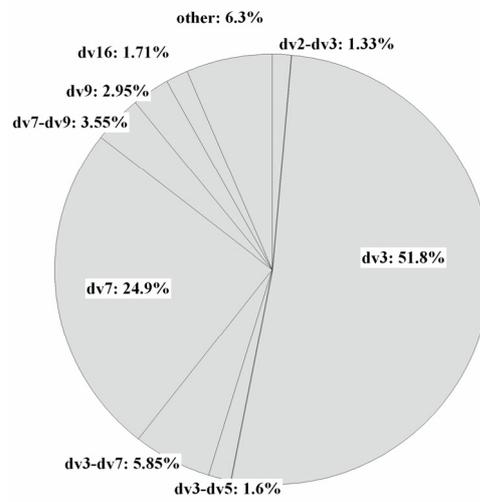
Figure 7. Geometry and pressure distributions of designed airfoil

Figure 7 shows the geometry and the pressure distributions of airfoil that shows the best performances about both objective functions among 89 points. At both conditions, the designed airfoil shows the good drag performances.

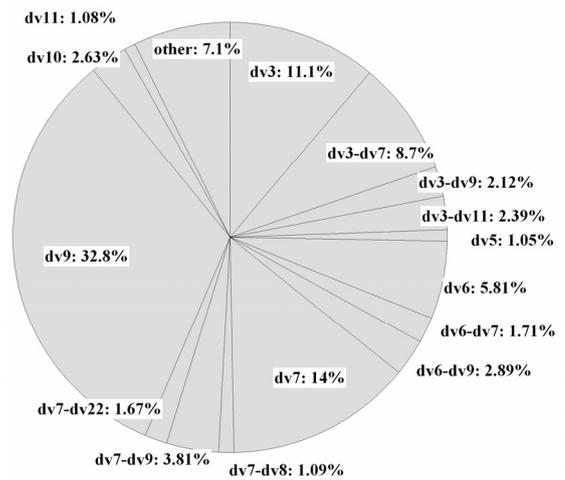
4.4 Result of Data Mining

4.4.1 ANOVA

Total variances of functions were decomposed into the variance due to each design variable. Design variables and their interactions whose proportion to total variable is over 1.0% are shown in Fig. 8. According to the results, dv3 gives the largest effect on the *obj1* and dv9 gives the largest effect on the *obj2*. dv7 and the interaction of dv3-dv7, dv3-dv9 and dv7-dv9 give a large effect on objective functions.



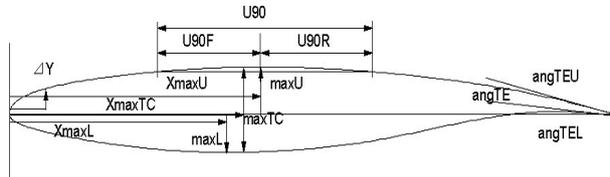
(a) *obj1*



(b) *obj2*

Figure 8. The proportion to the total variance of model

However, it is difficult to understand the relation between aerodynamic performance and airfoil geometry intuitively, because design variables defined by NURBS are not for the data mining but for the detailed expression of airfoil geometry. Thus, design variables defined by NURBS are transformed into the design variables which are more familiar to aerodynamic engineers as shown in Fig. 9.



DV1: ΔY , DV2: X_{maxU} , DV3: $maxU$, DV4: X_{maxL} , DV5: $MaxL$, DV6: X_{maxTc} , DV7: $maxTc$, DV8: $angTEU$, DV9: $angTEL$, DV10: $angTE$, DV11: $U90F$, DV12: $U90R$, DV13: $U90$

Figure 9. Transformed design variables

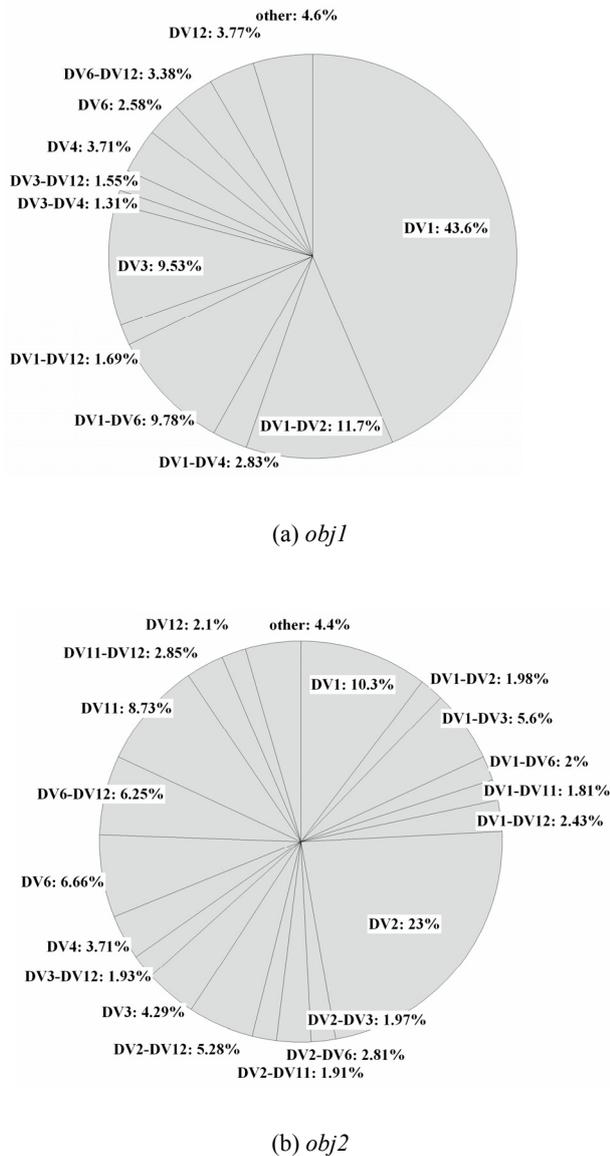
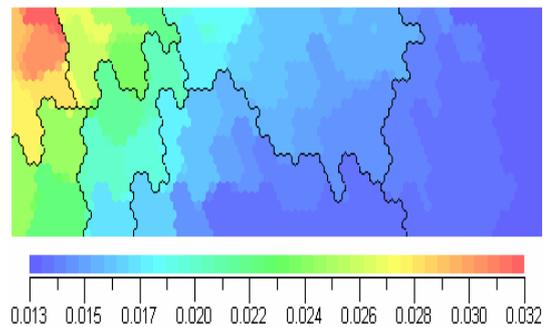


Figure 10. The proportion to the total variance of model

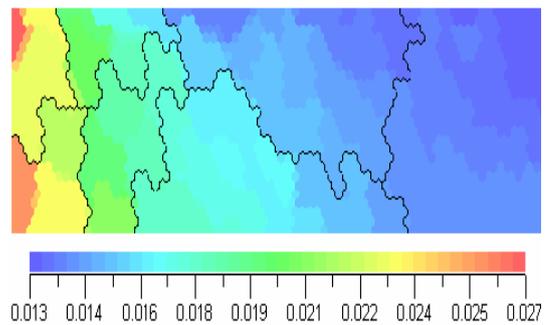
With these design variables, ANOVA is performed again. The results are shown in Fig. 10. According to the results, DV1 (leading-edge radius) gives the largest effect on *obj1* and DV2 (x-coordinate of upper surface crest) gives the largest effect on *obj2*. These findings are consistent with the aerodynamic knowledge. This indicates that ANOVA used for data mining is valid.

4.4.2 SOM

Using 89 sample data, clustering was performed by using SOM. Figures 11(a) and 11(b) show SOM colored by *obj1* and *obj2*, respectively. In Fig.11 (a), the airfoils which have a good *obj1* performance are clustered in right hand side and the airfoil has a bad *obj1* performance are clustered in upper left hand side. In Fig. 11 (b), the airfoils having a good performance are clustered in right hand side and the airfoil having a bad performance are clustered in upper and lower corner of left hand side. Figures 12 (a) and 12 (b) show SOM colored by *dv3* and *dv9* which gives the largest effect on the *obj1* and the *obj2*, respectively, according to the ANOVA. In Fig 12 (a), airfoils which have a large *dv3* value are clustered in upper left side. It is same with that of SOM colored by *obj1*. This means that a large *dv3* value has a relation with a large *obj1* value. In Fig. 12 (b), airfoils which have a small *dv9* value are clustered in the upper and the lower corner of left hand side. It is opposite with that of Fig 11 (b). This means that a small value of *dv9* generates the airfoil which has a good *obj2* performance vice versa. The visualization results of SOM agree with quantitative analysis of the ANOVA.

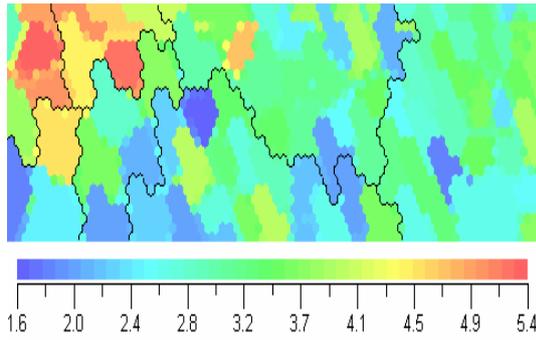


(a) SOM colored by *obj1*

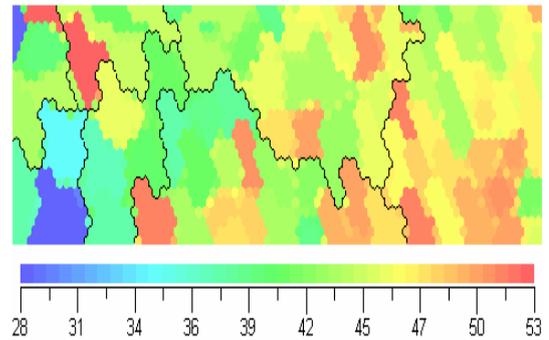


(b) SOM colored by *obj2*

Figure 11. SOM colored by objective functions

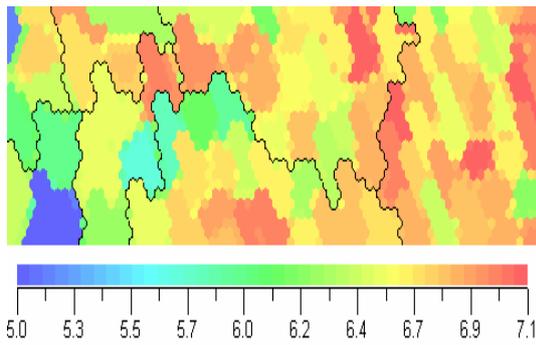


(a) SOM colored by dv3



(b) SOM colored by DV2

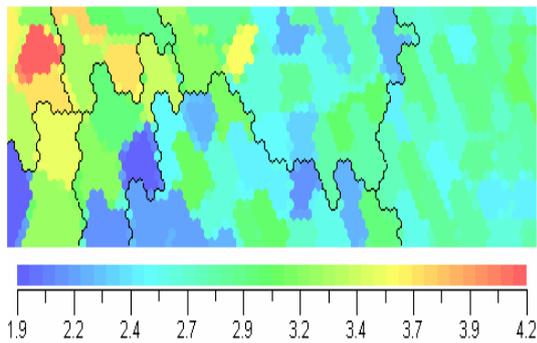
Figure 13. SOM colored by selected design variables



(b) SOM colored by dv9

Figure 12. SOM colored by selected design variables

SOM is also colored by the transformed design variables. Figures 13 (a) and 13 (b) show SOM colored by DV1 and DV2. In Fig. 13 (a), airfoils which have large DV1 values are clustered in upper left side. It is similar to SOM colored by *obj1*. This means that a large DV1 value has a relation with a large *obj1* value. In Fig. 13 (b), airfoils which have a small DV2 value are clustered in the upper and the lower corner of left hand side. It is opposite with that of Fig. 11 (b). This means that a small value of DV2 generates the airfoil which has a good *obj2* performance vice versa. These results also coincide with those of the ANOVA.



(a) SOM colored by DV1

5 Conclusions

In this study, a surrogate model was applied to multi-objective aerodynamic optimization design. For the balanced exploration and exploitation with the surrogate model, objective functions were converted to the expected improvements (EI) and these values were directly used as fitness values in the multi-objective optimization. As an optimizer, genetic algorithm was used here. From the non-dominated solutions about EIs, additional points for the update of the Kriging model were selected.

The present method was applied to a transonic airfoil design. In order to obtain the information about design space, two data mining techniques were applied to design results. One is analysis of variance (ANOVA) and the other is self-organizing map (SOM). ANOVA showed the relation between objective functions and design variable quantitatively and SOM did it qualitatively. For the intuitive understand of aerodynamic performance and the airfoil geometry, the design variables defined by NURBS were transformed into the geometry parameters which are more familiar to aerodynamic engineers. The results of data mining are consistent with the aerodynamic knowledge. These indicate that ANOVA and SOM used for data mining are valid.

Bibliography

- [1] Myers, R. H. and Montgomery, D. C, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, John Wiley & Sons, New York, 1995.
- [2] Vicini, A. and Quagliarella, "Multipoint transonic airfoil design by means of a multiobjective genetic algorithms," AIAA Paper 1997-82.
- [3] Sasaki, D., Obayashi, S and Nakahashi, K., "Navier-Stokes Optimization of Supersonic Wings with Four Objectives Using Evolutionary Algorithm," *Journal of Aircraft*, Vol. 39, 2003, pp621-629.
- [4] Jone, D. R., Schonlau, M. and Welch, W. J, "Efficient Global Optimization of Expensive Black-Box Function,"

Journal of global optimization, Vol. 13, 1998, pp. 455-492.

[5] Jeong, S., Murayama, M. and Yamamoto, K., "Efficient Optimization Design Method Using Kriging Model," *Journal of Aircraft*, Vol. 42, 2005, pp. 412-420.

[6] Knowles, J. and Hughes, E. J., "Multiobjective Optimization on a Budget of 250 Evaluation," *Proceeding of third international conference of EMO 2005*, 2005, pp. 176-190.

[7] Parmee, I. C. and Abraham, J. A. R., "Supporting Implicit Learning via Visualisation of COGA Multi-objective Data," *Proceeding of IEEE Congress on Evolutionary Computation 2004*, 2004, pp. 395-402.

[8] Sack, J., Welch, W. J., Mitchell, T. J. and Wynn, H. P., "Design and analysis of computer experiments (with discussion)," *Statistical Science* 4, 1989, pp. 409-435.

[9] Matthias, S., "Computer Experiments and Global Optimization," Ph.D Dissertation, Statistic and Actuarial Science Dept., University of Waterloo, Waterloo, Ontario, 1997.

[10] Krzysztof, J. C., Witold, P. and Roman, W. S., *Data Mining Methods for Knowledge Discovery*, Kluwer Academic Publisher, 1998.

[11] Eudaptics software gmbh, <http://www.eudaptics.com/somine/index.php?sprache=en>, last access on April 14, 2005.

[12] Lepine, J., Guibault, F., Trepanier, J-Y., and Pepin, "Optimized Nonuniform Rational B-spline Geometrical Representation for Aerodynamic Design of Wings," *AIAA Journal*, Vol. 39, 2001.

[13] Mckay, M. D., Beckman, R. J. and Conover, W. J., "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *Technometric*, Vol. 21, No. 2, 1979, pp. 239-245.