Proceedings of the Autumn Conference in 2001
and the 30th Anniversary of the Korean Statistical Society

Exploratory Graphical Tools for Multiresponse Optimization

Dae-Heung Jang[1]

Pukyong National University, Pusan, Republic of Korea, 608-737

The multiresponse optimization problem is more complex than in the single response case. The analysis of multiresponse data requires careful consideration of the mutivariate nature of the data. The scatterplot matrix and the parallel coordinate plot can be useful as exploratory graphical tools for multiresponse optimization.

Introduction

Under single response model, we usually execute the steepest ascent method and the canonical analysis/the ridge analysis as the formal analytic approaches to the first and second-order model analysis, respectively. But, These methods may not work at all for multivariate cases. The multiresponse optimization problem is more complex than in the single response case. The analysis of multiresponse data requires careful consideration of the mutivariate nature of the data. The main difficulty stem from the fact that when response variables are under investigation simultaneously, The meaning of optimum becomes unclear since there is no unique way to order multivariate values of a mutiresponse fuction. Furthmore, optimal condition for one response may be far from optimal or even physically impractical for the other responses. Many researchers studied multiresponse optimization(See Myers and Carter(1973), Derringer and Suich(1980), Khuri and Conlon(1981), Vining and Myers(1990), Del Castillo and Montgomery(1993), Derringer(1994), Luner(1994), Lin and Tu(1995), Copeland and Nelson(1996), Del Castillo(1996), Del Castillo, Montgomery, and McCarville(1996), Vining(1998), and Kim and Lin(1998, 2000).). Recently, Box(1999), Myers(1999), and Carlyle, Montgomery, and Runger(2000) mentioned the multiresponse problems. In multiresponse optimization, it is important to describe the distribution of the mean responses. Heuristically, we can consider superimposing contours of all response variables. This procedure has its limitation in large systems involving several input variables and several response variables though it is simple and straightforward. Hence, we need the tools for describing the distribution of the mean responses in large systems involving several input variables and several response variables. This article

[1]Dr. Jang is a Professor in the Division of Mathematical Sciences.e-mail:dhjang@pknu.ac.kr

is concerned with using the scatterplot matrix and the parallel coordinate plot for describing the distribution of the mean responses in large systems involving several input variables and several response variables.

The scatterplot matrix and the parallel coordinate plot

The scatterplot is an ideal tool for examining and visualizing the relationships between two variables. Scatterplots are especially useful when we are examining the relationship between continuous variables using statistical techniques such as correlation or regression. Another approach to graphing a set of variables is to look at a matrix of all possible pairwise scatterplots of the variables. The scatterplot matrix will produce such a plot. Each variable is plotted with every other variable. Every combination is plotted twice so that each variable appears on both the X and Y axis.

Two dimensional projection may convey the wrong intuition in scatterplot. Wegman(1990) mentioned this fact. Thus, it is desirable to have a simultaneous representation of all coordinates of data vector. We can have this representation through the parallel coordinate plot. Its advantage over other types of statistical graphics is its ability to display multidimensional data in one representation. The parallel coordinate plot was originally proposed and implemented by Inselberg (1985). variations of the device and various applications have been supposed by other statisticians (Gennings, Dawson, Carter, and Myers(1990), Wegman(1990), Miller and Wegman(1991), Bateson and Curtiss(1996), Weber and Desai(1996), Jang and Yang(1996), Becker(1997), Inselberg(1998), Teppola, Mujunen, Minkkinen, Puijola, and Pursiheimo(1998), Chou, Lin, and Yeh(1999), Gröller, Löffelmann, and Wegenkittl(1999), Andrienko and Andrienko(2001), and Falkman(2001)). In the parallel coordinate plot, each observation in a data set is represented as an connected series of line segments which intersect vertical axes, each scaled to a different variable. The observation's line passes through each axis according to its value of the variable of that axis.

Exploratory graphical tools for multiresponse optimization

Let $y_1, y_2, ..., y_r$ be r response variables that depend upon k input variables, $x_1, x_2, ..$ and x_k. Suppose that the response variables can be represented by polynomial regression models in the values of k input variables with a certain interest region R using a design D. Let the i-th predicted response at a point $\mathbf{x} = (x_1, x_2, ..., x_k)$ is $\hat{y}_i(\mathbf{x})$ and the prediction variance of responses at $\mathbf{x}$ divided by the error variance is $V(\mathbf{x})$.

The analytical methods for multiresponse optimization - direct search methods, mathematical optimization algorithms, desirability fuction approach etc. - give the optimal solutions under multiresponse model, but these methods can not provide the distribution of the mean responses as superimposing contours of all response variables in case of $k = 2$ or 3. Therefore, we need the tools for describing the distribution of the mean responses in large systems involving several input variables and several response variables before and/or after we perform the analytical optimization. At this time, the scatterplot matrix and the parallel coordinate plot can be used as tools for describing the distribution of the mean responses in large systems involving several input variables and several response variables.

Before and/or after we perform the analytical optimization, we can use the scatterplot matrix as a exploratory graphical tool for describing the conditional distribution of the mean responses, knowing the interrelationships among resposes, and seeking the optimal solutions heuristically under multiresponse model as following procedure.

1. Get each predicted response fuction $\hat{y}_i, i = 1, 2, ..., r$.
2. Obtain the combinations $(x_1, x_2, ..., x_k, \hat{y}_1, \hat{y}_2, ..., \hat{y}_r, V(\mathbf{x}))$ of input variables and responses which satisfy desirable conditions on the responses using Monte-Carlo simulation.
3. Draw the scatterplot matrix using the combinations obtained in Step 2.
4. Seek the optimal solutions heuristically using brushing in the scatterplot matrix.

Detailed explanation about Step 2 is as following;

(1) Choose the values of input variables $\mathbf{x} = (x_1, x_2, ..., x_k)$ using pseudo random numer generator.

(2) Calculate $\hat{y}_1, \hat{y}_2, ..., \hat{y}_r$ with respect to the the values of input variables which we obtained in (1).

(3) Search if calculated values of response variables $\hat{y}_1, \hat{y}_2, ..., \hat{y}_r$ satisfy desirable conditions on the responses. If satisfied, we obtain a combination $(x_1, x_2, ..., x_k, \hat{y}_1, \hat{y}_2, ..., \hat{y}_r, V(\mathbf{x}))$ of input variables and responses which satisfy desirable conditions on the responses.

(4) Repeat (1) - (4) until we obtain the desired number of combinations.

Brushing technique in the scatterplot matrix was introduced by Becker and Cleveland(1987). The more we have the combinations in Step 2 of upper-mentioned procedure, the better we can get the optimal solutions.

Also, Before and/or after we perform the analytical optimization, we can use the parallel coordinate plot as a exploratory graphical tool for describing the distribution of the mean responses, knowing the interrelationships among resposes, and seeking the optimal solutions heuristically under multiresponse model by using upper-mentioned procedures similarly.

Conclusion

The scatterplot matrix and the parallel coordinate plot can be used as exploratory graphical tools for describing the distribution of the mean responses in large systems involving several input variables and several response variables.

Before and/or after we perform the analytical optimization, we can use the scatterplot matrix and the parallel coordinate plot as exploratory graphical tools for describing the distribution of the mean responses, knowing the interrelationships among resposes, and seeking the optimal solutions heuristically under multiresponse model.

Acknowledgements

This paper was accomplished with Research Fund provided by Korea Research Foundation Support for Faculty Research Abroad(1999).

References

Andrienko, G. and Andrienko, N. (2001). "Exploring Spatial Data with Dominant Attribute Map and Parallel Coordinates". *Computers, Environment and Urban Systems* 25, pp. 5-15.

Bateson, A. and Curtiss, B. (1996). "A Method for Manual Endmember Selection and Special Unmixing". *Remote Sensing of Environment* 55, pp. 229-243.

Becker, O. M. (1997). "Representing Protein and Peptide Structures with Parallel-Coordinates". *Journal of Computational Chemistry* 18, pp. 1893-1902.

Becker, R. A. and Cleveland, W. S. (1997). "Brushing Scatterplots". *Technometrics* 29, pp. 127-142.

Box, G. E. P. (1999). "Statistics as Catalyst to Learning by Scientific Method: Part II-a discussion". *Journal of Quality Technology* 31, pp. 16-29.

Carlyle, W. M.; Montgomery, D. C.; Runger, G. C.(2000). "Optimization Problems and Methods in Quality Control and Improvement". *Journal of Quality Technology* 32, pp. 1-17.

Chou, S.; Lin, S.; and Yeh, C. (1999). "Cluster Identification with Parallel Coordinates". *Pattern Recognition Letters* 20, pp. 565-572.

Copeland, K. A. F. and Nelson, P. R. (1996). "Dual Response Surface Optimization via Direct Function Minimization". *Journal of Quality Technology* 28, pp. 61-70.

Del Castello, E. (1996). "Multiresponse Process Optimization via Constrained Confidence Regions". *Journal of Quality Technology* 28, pp. 61-70.

Del Castello, E. and Montgomery, D. C. (1993). "A Nonlinear Programming Solution to the Dual Response Problem". *Journal of Quality Technology* 25, pp. 199-204.

Del Castello, E.; Montgomery, D. C.; and McCarville, D. R. (1996). "Modified Desirability Functions for Multiple Response Optimization". *Journal of Quality Technology* 28, pp. 337-345.

Derringer, G. (1994). "A Balancing Act: Optimizing a Product's Properties". *Quality Progress* 27, pp. 51-58.

Derringer, G. and Suich, R. (1980). "Simultaneous Optimization of Several Response Variables". *Journal of Quality Technology* 12, pp. 214-219.

Falkman, G. (2001). "Information Visualisation in Clinical Odontology: Multidimentional Analysis and Interactive Data Exploration". *Artificial Intelligencein Medicine* 22, pp. 133-158.

Gennings, C.; Dawson, K. S.; Carter, W. H.; and Myers, R. H. (1990). "Interpreting Plots of Multidimensional Dose-Response Surfaces in a Parallel Coordinate System". *Biometrics* 46, pp. 719-735.

Gröller, E.; Löffelmann, H.; and Wegenkittl, R. (1999). "Visualizations of Dynamical Systems". *Future Generation Computer Systems* 15, pp. 75-86.

Inselberg, A. (1985). "The Plane with Parallel Coordinates". *The Visual Computer* 1, pp. 69-91.

Inselberg, A. (1998). "Visual Data Mining with Parallel Coordinates". *Computational Statistics* 13, pp. 47-63.

Jang, D. and Yang, S. (1996). "The Dynamic Parallel Coordinate Plot and Its Applications". *The Korean Journal of Applied Statistics* 9, pp. 45-52.

Kim, K. and Lin, D. K. J. (1998). "Dual Response Surface Optimization: A Fuzzy Modeling approach". *Journal of Quality Technology* 30, pp. 1-10.

Kim, K. and Lin, D. K. J. (2000). "Simultaneous Optimization of Mechanical Properties of Steel by Maximizing Exponential Desirability Functions". *Applied Statistics* 49, pp. 311-325.

Lin, D. K. J. and Tu, W. (1995). "Dual Response Surface Optimization". *Journal of Quality Technology* 27, pp. 34-39.

Luner, J. (1994). "Achieving Continuous Improvement with the Dual Response Approach: a Demonstration of the Roman Catapult". *Quality Engineering* 6, pp. 691-705.

Khuri, A. I. and Conlon, M. (1981). "Simultaneous Optimization of Multiple Responses Represented by Polynomial Regression Functions". *Technometrics*, 23, pp. 363-375.

Miller, J. J. and Wegman, E. J. (1991). "Construction of Line Densities for Parallel Coordinate Plots". in *Computing and Graphics in Statistics* edited by A. Buja and P. A. Tukey. Springer-Verlag, New York, NY.

Myers, R. H. (1999). "Response Surface Methodolgy - Current Status and Future Directions". *Journal of Quality Technology* 31, pp. 30-44.

Myers, R. H. and Carter, W. H., Jr. (1973). "Response Surface Techniques for Dual Response Systems". *Technometrics* 15, pp. 301-317.

Teppola, P; Mujunen, S; Minkkinen, P.; Puijola, T.; and Pursiheimo, P. (1998). "Principal Component Analysis, Contribution Plots and Feature Weights in The Monitoring of Process Data From A Paper Machine's Wet End". *Chemometrics and Intelligent Laboratory Systems* 44, pp. 307-317.

Vining, G. G. (1998). "Compromise Approach to Multiresponse Optimization". *Journal of Quality Technology* 30, pp. 309-313.

Vining, G. G. and Myers, R. H. (1990). "Combining Taguchi and Response Surface Philosophies: a Dual Response Approach". *Journal of Quality Technology* 22, pp. 38-45.

Weber, C. A. and Desai, A. (1996). "Determination of Paths to Vendor Market Efficiency Using Parallel Coordinates Representation: A Negotiation Tool for Buyers". *European Journal of Operational Research* 90, pp. 142-155.

Wegman, E. J. (1990). "Hyperdimensional Data Analysis using Parallel Coordinates". *Journal of the American Statistical Association* 85, pp. 664-675.

Key Words: *Multiresponse Optimization, Monte-Carlo simulation, The Scatterplot Matrix, The Parallel Coordinate Plot.*