EVOLUTIONARY AND DETERMINISTIC METHODS FOR DESIGN, OPTIMIZATION AND CONTROL
C. Poloni, D. Quagliarella, J. Périaux, N. Gauger and K. Giannakoglou (Eds.)

EFFICACY OF ROUGH SET THEORY AS DECISION MAKING – INVESTIGATION INTO FORMULATED OPTIMIZATION PROBLEM

Kazuhisa Chiba
Department of Mechanical Systems Engineering
Hokkaido Institute of Technology
7-15-4-1, Maeda, Teine, Sapporo 006-8585, Japan
Email: thousandleaf@gmail.com
Web page: http://www.geocities.jp/thousandleaf_k/index_e.html

Abstract. Data mining methods are used in decision making to select compromise solution from the results of multi-objective optimization. In the present study, the efficacy of the rough set theory is investigated applying the method to a known multi-objective optimization problem. The results indicate that the rough set theory might address local design knowledge although it is difficult to directly drive automated decision.

Key words: Multi-objective Optimization, Decision Making, Data Mining, Rough Set Theory.

1 INTRODUCTION

The results produced by evolutionary multi-objective optimizations are not individual optimal solution but rather an entire set of optimal solutions. That is, the result of a multi-objective optimization is not the end of the process while from a practical point of view the designers need a conclusive shape (or few candidates of a conclusive shape). Thus a informatics procedure approach is proposed which is constructed as the combination between optimization and data mining as decision making procedure for an innovative design. The essence of this approach is the comprehension of design space. Designers can have a possibility of breakthrough innovation when they understand what are the characteristics of a present design and what prevents ideal performances to be reached . This efficacy does not depend on the upstream and downstream of design process. The effectiveness of functional analysis of variance (ANOVA) and self-organizing map (SOM) as data mining technique is adequately confirmed in the past studies but other techniques should be investigated in order to comprehend better and more intuitively design space. The objective of this study is to investigate the efficacy of the rough set theory (RST) [1] as data mining tool also on the basis of previous experience [2]. In the present study, RST would be applied to the simple optimization result with the formulated objective functions, and then the efficacy of RST is investigated to determine whether it is useful or not to understand the design space.

2 OPTIMIZATION PROBLEM

2.1 Design Variables

Diameter D [in.] and height H [in.] are used as two design variables so that a can is simply described as cylinder.

2.2 Objective Functions

The following two objective functions are defined by using volume and surface area of a can in the present problem. There is a tradeoff between them.

1. profit per a can p_c [-]:

$$p_c = 1.7 \times \frac{\pi D^2}{4} H - 0.02 \times \left(\frac{\pi D^2}{4} H \right)^2 - 0.1 \times \pi D \left(H + \frac{D}{2} \right) \tag{1}$$

2. profit per unit volume p_o [-]:

$$p_o = \frac{p_c}{\frac{\pi D^2}{4} H} \tag{2}$$

2.3 Constraints

The following three constraints regarding diameter, volume, and aspect ratio of can are set. They are considered to include commercial can shape and to be also a practical can shape.

$$\begin{aligned} &1.5 \leq D \text{ [in.]} \leq 3.5 \\ &9.0 \leq \frac{\pi D^2}{4} H \text{ [in.}^3\text{]} \leq 27.0 \\ &1.3 \leq \frac{H}{D} \text{ [-]} \leq 3.0 \end{aligned} \tag{3}$$

3 ROUGH SET THEORY

RST is a new approach to address the issue of vagueness [1,3] which has become a topic of interest for many researchers in computer science. It is an important method for decision support systems and data mining tools. In fact, it is a new mathematical approach to analyze data.

The basic idea behind RST is to construct approximations of sets using the binary relation R_A. The indiscernibility sets $R_A(x)$ form basic building blocks from which subsets $X \subseteq U$ can be assembled. If X cannot be defined in a crisp manner using attributes A, we can circumscribe them through lower and upper approximations $\underline{A}X$ and $\overline{A}X$. The lower approximation consists of those objects that certainly belong to X whereas the upper approximation consists of the objects that possibly

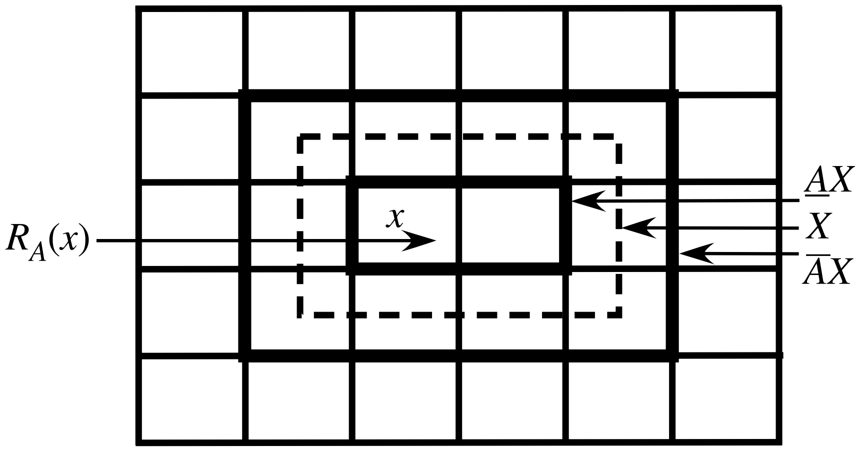

Figure 1: Illustrated idea of RST.

belong to X. The boundary region is defined as the difference between the upper and the lower approximation, and consists of the objects that we cannot decisively assign as being either members or non-members of X. The outside region is defined as the complement of the upper approximation, and consists of the objects that are definite non-members. An RST is any subset $X \subseteq U$ defined through its lower and upper approximations. Figure 1 shows these ideas graphically. In the data mining by RST, as minimum combination of attributes is found by using reduction to determine the decision attribute, decision rules are extracted. RST fundamentally implements classification, discretization, reduct generation, and filtering one after another, and then rules are generated. Since several methods are employed regarding discretization, reduct generation, and filtering, each result would be compared. The typical methods are summarized as follows.

- Classification
  - k-means method [4]
  - Self-organizing map
- Discretization
  - Boolean reasoning algorithm [5]
  - Entropy/Minimum description length (MDL) algorithm [6]
  - Equal frequency binning [7]
  - Naive algorithm [8]
- Reduct generation
  - Genetic algorithm (GA) [9]
  - Johnson's algorithm [10]
  - Holte's 1R rule [11]
  - Exhaustive calculation
- Filtering
  - Cost filtering
  - Performance filtering

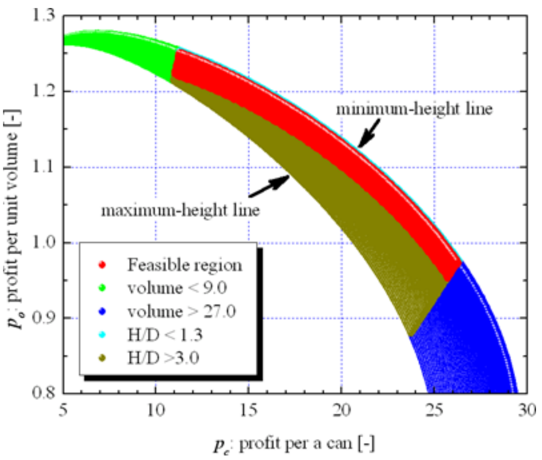

Figure 2: All solutions in objective-function space.

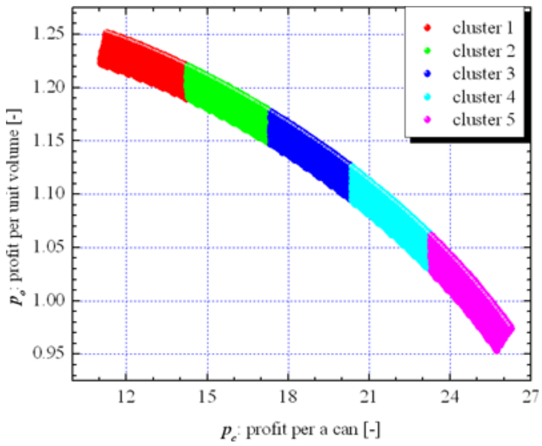

Figure 3: All solutions in feasible region divided into five clusters by k-means method.

The clustering is performed by using k-means method as classification. Discretization, reduct, and filtering are also implemented by using ROSETTA [12].

4 DATA-MINING RESULTS

Figure 2 shows the view of the objective-function space. The feasible region is narrow because of the constraints needed to build a feasible shape of the can is considered. Since the candidates of compromise solution are selected from feasible region, the individuals in feasible region are the subject of data mining. Figure 3 shows the feasible region divided into five clusters by k-means method with silhouettes [4]. The dividing number of cluster is set by considering the balance between the number of cluster and sihouette width. RST is carried out for each cluster, thus, the rules are obtained for each cluster. It is notable that Fig. 3 shows that five clusters form a line. That is, there is not much difference in the design space between each neighboring cluster. Therefore, it is expected that there is not much difference among rules generated for each neighboring cluster.

For discretization, five methods are used as boolean reasoning algorithm [5], entropy/minimum description length (MDL) algorithm [6], equal frequency binning [7], naive algorithm [8], and semi-naive algorithm. And also, GA [9], Johnson's algorithm [10], and Holte's 1R rule [11], are adopted for reduct computation. Note that the results acquired by Johnson's algorithm for reduct strictly correspond to those by GA. Therefore, the results obtained by Johnson's algorithm are omitted.

The rules generated by RST are generally more than one. At least one rule is generated when there is one individual number with a support rule (support number). When the generated rule supports only one individual the rule does not have global relevance. Rules supporting many individuals should be used to identify rules with global applicability. Therefore, the rule with the highest support number is selected in the present study.

The obtained rules are summarized in Tables 1 to 10. Since the obtained rules are represented by using the range of design variables, rule absolutely narrows the design-variable space. The individuals in the closed design-variable space can be also plotted in the objective-function space for the visualization of rule. Figures 4 to 13 show the plotted objective-function space for each rule. The acquired rules do not directly have physical meaning because of machine learning. An observer should find physical implication (if it exists) from the objective-function space. In this study, a simple optimization problem on two dimensions is considered, and physical implications can be intuitively comprehended by the observation of the objective-function space.

The results obtained by any algorithms do not extract the characteristic individuals in each cluster. These figures show the following four facts: 1) There is no unification among the results. There is no resemblance between the rules generated for each neighboring cluster. The rules generated by same algorithm do not give the commonness for each cluster. Although the Pareto solutions are extracted in some clusters (Fig. 8 cluster5, Fig. 9 cluster5, and Fig. 11 cluster5), the rules for the other clusters generated by the same algorithm do not address the Pareto solutions. 2) There is no assurance to generate the rule which supports the assigned cluster only, or rather, when there is the affinity among each cluster, the similar rule might be generated for different clusters (for example Table 2). As the affinity depends on the algorithms, the selection of the algorithms for discretization and reduct is difficult. 3) Design space is absolutely narrowed by rule due to the discretization of design-variable space. Therefore, globalization is lost. The narrowness of the range of design variable means the decrease of the support number of individuals for rule. 4) Since discretization is independently implemented into each design variable, the generated rule has limitation regarding space expression.

These facts indicate the following two key points: 1) The index of RST to determine an useful rule is not necessarily the support number of data. In fact, the rules having the second and third value of the support number also give the above tendency (the applicable number of the design variables for all generated rules is not employed because of a small number of all generated rules due to a small number of the design variables). 2) RST is not the best method to find out the general rule in cluster. In other words, the rule can represent the local information in a specific part of the design space. ANOVA and SOM, which are commonly used data mining techniques, cannot find the local information because these techniques seek general design information. ANOVA and SOM find out global design knowledge, and RST find out local design knowledge. When the specific character is found by RST the method can draw decisions but the decision might be not sufficiently robust with respect to design variables.

The three problems related to the use of RST as data mining technique are summarized as follows; 1) the obtained results cannot be intuitively employed like as ANOVA and SOM, 2) the understanding of the physical implication of the rules might be difficult, and 3) the selection of the key rules out of the many generated in a complex environment might be difficult.

5 CONCLUSION

The efficacy of the rough set theory has been investigated using a known optimization problem. The support number was used to select an individual rule from all rules generated by the rough set theory. Consequently, the selected data addresses without a common principle among divided clusters even when different methods

Table 1: Summary of the obtained rules by RST with discretization using equal frequency binning and reduct using GA for each cluster.

Cluster 1	$dv1 < 2.07$	$\wedge$	$dv2 < 4.02$
Cluster 2	$2.07 \leq dv1 < 2.33$	$\wedge$	$dv2 < 4.02$
Cluster 3	$2.07 \leq dv1 < 2.33$	$\wedge$	$4.02 \leq dv2 < 4.84$
Cluster 4	$2.07 \leq dv1 < 2.33$	$\wedge$	$4.84 \leq dv2$
Cluster 5	$2.33 \leq dv1$	$\wedge$	$4.84 \leq dv2$

Table 2: Summary of the obtained rules by RST with discretization using equal frequency binning and reduct using Holte's 1R rule for each cluster.

Cluster 1	$dv1 < 2.07$	$\wedge$	$dv2 < 4.02$
Cluster 2	$dv1 < 2.07$	$\wedge$	$dv2 < 4.02$
Cluster 3	$2.07 \leq dv1 < 2.33$	$\wedge$	$4.84 \leq dv2$
Cluster 4	$2.33 \leq dv1$	$\wedge$	$4.84 \leq dv2$
Cluster 5	$2.33 \leq dv1$	$\wedge$	$4.84 \leq dv2$

Table 3: Summary of the obtained rules by RST with discretization using boolean reasoning algorithm and reduct using GA for each cluster.

Cluster 1	$dv1 < 1.99$	$\wedge$	$dv2 < 3.80$
Cluster 2	$dv1 < 1.99$	$\wedge$	$4.42 \leq dv2 < 5.20$
Cluster 3	$1.99 \leq dv1 < 2.19$	$\wedge$	$4.42 \leq dv2 < 5.20$
Cluster 4	$2.19 \leq dv1 < 2.41$	$\wedge$	$4.42 \leq dv2 < 5.20$
Cluster 5	$2.19 \leq dv1 < 2.41$	$\wedge$	$5.20 \leq dv2$

Table 4: Summary of the obtained rules by RST with discretization using boolean reasoning algorithm and reduct using Holte's 1R rule for each cluster.

Cluster 1	$dv1 < 1.99$	$\wedge$	$dv2 < 3.80$
Cluster 2	$dv1 < 1.99$	$\wedge$	$dv2 < 3.80$
Cluster 3	$1.99 \leq dv1 < 2.19$	$\wedge$	$4.42 \leq dv2 < 5.20$
Cluster 4	$2.41 \leq dv1$	$\wedge$	$5.20 \leq dv2$
Cluster 5	$2.41 \leq dv1$	$\wedge$	$5.20 \leq dv2$

Table 5: Summary of the obtained rules by RST with discretization using entropy/MDL algorithm and reduct using GA for each cluster.

Cluster 1	$dv1 < 2.05$	$\wedge$	$dv2 < 3.60$
Cluster 2	$dv1 < 2.05$	$\wedge$	$4.46 \leq dv2 < 4.68$
Cluster 3	$2.43 \leq dv1 < 2.45$	$\wedge$	$dv2 < 3.60$
Cluster 4	$2.29 \leq dv1 < 2.45$	$\wedge$	$4.46 \leq dv2 < 4.68$
Cluster 5	$2.77 \leq dv1$		

Table 6: Summary of the obtained rules by RST with discretization using entropy/MDL algorithm and reduct using Holte's 1R rule for each cluster.

Cluster 1	$dv1 < 2.05$	
Cluster 2	$dv1 < 2.05$	
Cluster 3	$dv1 < 2.05$	
Cluster 4		$4.46 \leq dv2 < 4.68$
Cluster 5	$2.77 \leq dv1$	

Table 7: Summary of the obtained rules by RST with discretization using naive algorithm and reduct using GA for each cluster.

Cluster 1	$dv1 < 1.73$
Cluster 2	
Cluster 3	
Cluster 4	
Cluster 5	

Table 8: Summary of the obtained rules by RST with discretization using naive algorithm and reduct using Holte's 1R rule for each cluster.

Cluster 1	$dv1 < 1.73$
Cluster 2	$1.85 \leq dv1 < 1.87$
Cluster 3	$1.99 \leq dv1 < 2.01$
Cluster 4	$2.13 \leq dv1 < 2.15$
Cluster 5	$2.77 \leq dv1$

Table 9: Summary of the obtained rules by RST with discretization using semi-naive algorithm and reduct using GA for each cluster.

Cluster 1	$dv1 < 1.85$	$\wedge$	$3.84 \leq dv2 < 5.34$
Cluster 2	$1.85 \leq dv1 < 1.97$	$\wedge$	$3.84 \leq dv2 < 5.34$
Cluster 3	$2.11 \leq dv1 < 2.25$	$\wedge$	$3.84 \leq dv2 < 5.34$
Cluster 4	$2.25 \leq dv1$	$\wedge$	$3.84 \leq dv2 < 5.34$
Cluster 5	$2.25 \leq dv1$	$\wedge$	$3.84 \leq dv2 < 5.34$

Table 10: Summary of the obtained rules by RST with discretization using semi-naive algorithm and reduct using Holte's 1R rule for each cluster.

Cluster 1		$3.84 \leq dv2 < 5.34$
Cluster 2		$3.84 \leq dv2 < 5.34$
Cluster 3		$3.84 \leq dv2 < 5.34$
Cluster 4		$3.84 \leq dv2 < 5.34$
Cluster 5	$2.25 \leq dv1$	

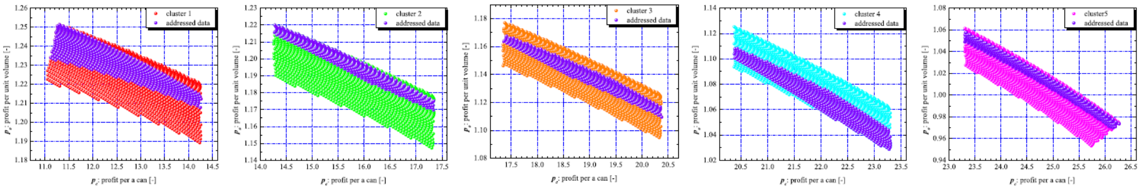

Figure 4: Addressed data by RST with discretization using equal frequency binning and reduct using GA for each cluster.

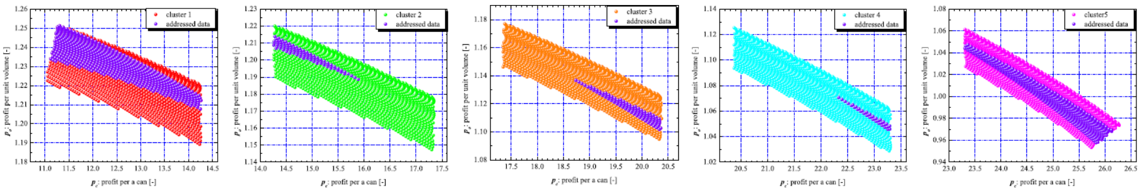

Figure 5: Addressed data by RST with discretization using equal frequency binning and reduct using Holte's 1R rules for each cluster.

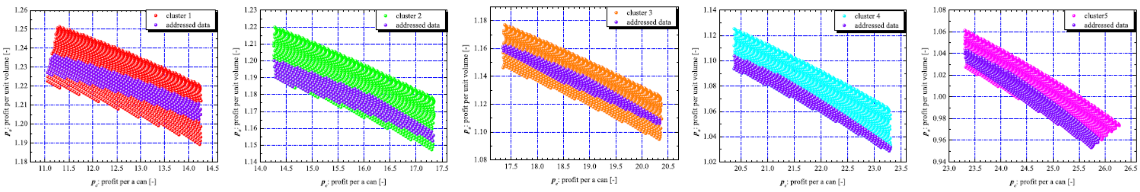

Figure 6: Addressed data by RST with discretization using boolean reasoning algorithm and reduct using GA for each cluster.

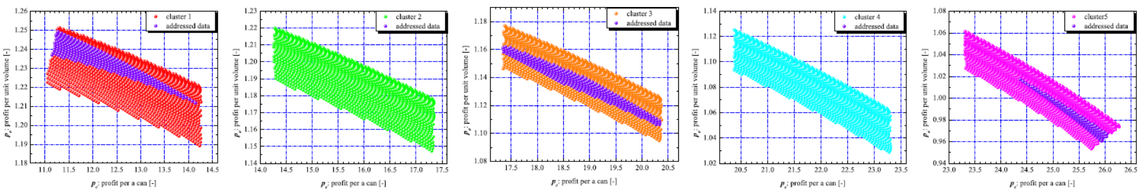

Figure 7: Addressed data by RST with discretization using boolean reasoning algorithm and reduct using Holte's 1R rules for each cluster.

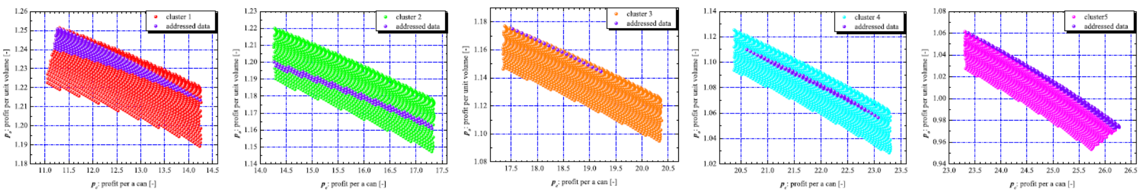

Figure 8: Addressed data by RST with discretization using entropy/MDL algorithm and reduct using GA for each cluster.

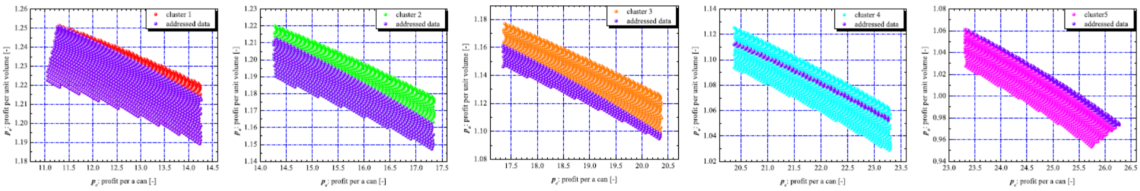

Figure 9: Addressed data by RST with discretization using entropy/MDL algorithm and reduct using Holte's 1R rule for each cluster.

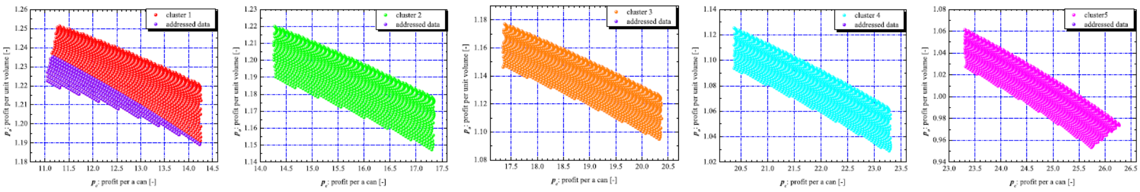

Figure 10: Addressed data by RST with discretization using naive algorithm and reduct using GA for each cluster.

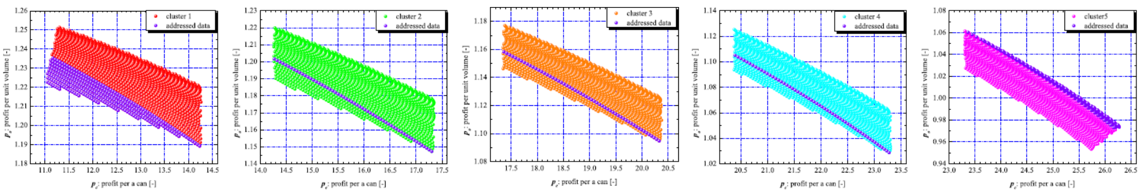

Figure 11: Addressed data by RST with discretization using naive algorithm and reduct using Holte's 1R rule for each cluster.

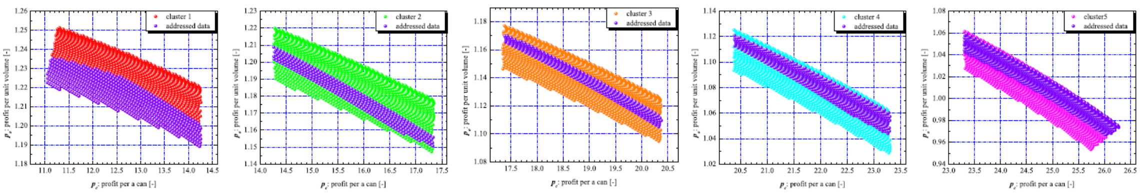

Figure 12: Addressed data by RST with discretization using semi-naive algorithm and reduct using GA for each cluster.

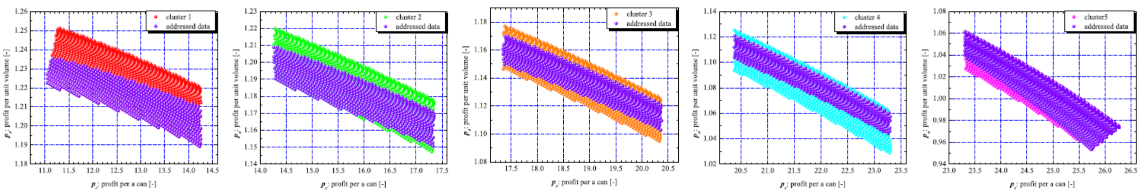

Figure 13: Addressed data by RST with discretization using semi-naive algorithm and reduct using Holte's 1R rule for each cluster.

are used for discretization and reduct computation. The principal rule should not be selected by using the support number because the rough set theory is not the method to find out global knowledge. The physical implication of the rule generated by the rough set theory should be comprehended. The peculiarity of rules as essence should be also selected. These problems need to be solved so that the rough set theory is directly intuitively used as decision maker.

ACKNOWLEDGMENT

This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Young Scientists (B) 22700155, 2010.

REFERENCES

[1] Kusiak, A., "Rough Set Theory: A Data Mining Tool for Semiconductor Manufacturing," *IEEE Transactions on Electronics Packaging Manufacturing*, Vol. 24, No. 1, 2001, pp. 44–50.

[2] Chiba, K. and Obayashi, S., "Knowledge Discovery in Aerodynamic Design Space for Flyback-Booster Wing Using Data Mining," *Journal of Spacecraft and Rockets*, Vol. 45, No. 5, 2008, pp. 975–987.

[3] Pawlak, Z., "Rough Sets," *International Journal of Computer and Information Science*, Vol. 11, No. 5, 1982, pp. 341–356.

[4] Rousseeuw, P. J., "Silhouettes: A Graphical Aid to the Interpretation and Valication of Cluster Analysis," *Journal of Computational and Applied Mathematics*, Vol. 20, No. 1, 1987, pp. 53–65.

[5] Nguyen, H. S. and Skowron, A., "Quantization of Real-Valued Attributes," *Proceedings of the 2nd International Joint Conference on Information Sciences*, Wrightsville Beach, NC, 1995, pp. 34–37.

[6] Dougherty, J., Kohavi, R., and Sahami, M., "Supervised and Unsupervised Discretization of Continuous Features," *Proceedings of the 12th International Conference on Machine Learning*, Morgan Kaufmann, 1995, pp. 194–202.

[7] Marzuki, Z. and Ahmad, F., "Data Mining Discretization Methods and Performances," *Proceedings of the International Conference on Electrical Engineering and Informatics*, Institut Teknologi Bandung, Indonesia, 2007, pp. 535–537.

[8] Knuth, D. E., J. H. Morris, J., and Pratt, V. R., "Fast Pattern Matching in Strings," *Journal on Computing*, Vol. 6, No. 2, 1977, pp. 323–350.

[9] Vinterbo, S. and Øhrn, A., "Minimal Approximate Hitting Sets and Rule Templates," *International Journal of Approximate Reasoning*, Vol. 25, No. 2, 2000, pp. 123–143.

[10] Johnson, D. S., "Approximation Algorithms for Combinatorial Problems," *Journal of Computer and System Sciences*, Vol. 9, 1974, pp. 256–278.

[11] Holte, R. C., "Very Simple Classification Rules Perform Well on Most Commonly Used Datasets," *Machine Learning*, Vol. 11, No. 1, 1993, pp. 63–91.

[12] Øhrn, A., *Discernibility and Rough Sets in Medicine: Tools and Applications*, Ph.D. dissertation, Department of Computer and Information Science, Norwegian University of Science and Technology, 1999.