2014 IEEE Congress on Evolutionary Computation (CEC)
July 6-11, 2014, Beijing, China

A Proposal on Analysis Support System Based on Association Rule Analysis for Non-dominated Solutions

Shinya Watanabe
College of Information and Systems,
Muroran Institute of Technology
27-1, Mizumoto-cho, Muroran 050-8585, JAPAN
Email: sin@csse.muroran-it.ac.jp

Yuta Chiba
Accenture Japan Ltd
Email: hartmann1661@gmail.com

Masahiro Kanazaki
Faculty of System Design,
Tokyo Metropolitan University
Email: kana@sd.tmu.ac.jp

***Abstract*—This paper presents a new analysis support system for analyzing non-dominated solutions (NDSs) derived by evolutionary multi-criterion optimization (EMO). The main features of the proposed system are to use association rule analysis and to perform a multi-granularity analysis based on a hierarchical tree of NDSs. The proposed system applies association rule analysis to the whole NDSs and derives association rules related to NDSs. And a hierarchical tree is created through our original association rule grouping that guarantees to keep at least one common features. Each node of a hierarchical tree corresponds to one group consisting of association rules and is fixed in position according to inclusion relations between nodes. Since each node has some kinds of common features, the designer can analyze each node with previous knowledge of these common features.**

To investigate the characteristics and effectiveness of the proposed system, the proposed system is applied to the concept design problem of hybrid rocket engine (HRE) which has two objectives and six variable parameters. HRE separately stores two different types of thrust propellant unlike in the case of usual other rockets and the concept design problem of HRE has been provided by JAXA. The results of this application provided possible to analyze the trends and specifics contained in NDSs in an organized way unlike analysis approaches targeted at the whole NDSs.

I. Introduction

Recently, a new problem in the field of evolutionary multi-criterion optimization (EMO) emerges along with the progress of EMO algorithms. It is how to analyze non-dominated solutions (NDSs) with many design variables and many objectives. NDSs can be regarded as a high potential subspace of being selected in whole search space and have the feature of being non-inferior to other solutions. Therefore, an analysis of NDSs and a feature extraction of NDSs are very important for getting useful information of design[1], [2]. Deb pointed out the usefulness of extracting a feature of problem by analyzing NDSs and called this as "Innovization", which means that a new innovation can be brought to designers by analyzing NDSs[1].

There have been proposed some previous approaches related analysis of NDSs. These approaches used various visualization and mining methods, such as self-organizing map (SOM), multiple discriminant analysis (MDA), and kernel dimensionality reduction (KDR). However, most of these results or visualization maps could only extract the overall trends in NDSs and were not easy to analyze and understand.

Therefore, we proposed a new analysis support system using association rules. The proposed system applies logical analysis to the whole NDSs and the whole NDSs are divided into some subsets according to the obtained association rules. This can be viewed as a hierarchical tree in which each node is a subgroup of NDSs. The main feature of the proposed system is to perform a multi-granular analysis based on a hierarchical tree of NDSs. Since every subset in a hierarchical tree have at least one common character, user can analyze a subsets with a prior-knowledge.

To investigate the characteristics and effectiveness of the proposed system, we apply the proposed system to the concept design problem of hybrid rocket engine(HRE)[3], which is presented by JAXA. Although HRE has various objectives, such as thrust force, curb weight, attainment and etc., we define this as two objectives and six variable parameters problem in this paper.

II. Related study

There are many researches about multiple criteria decision making (MCDM) and multiple criteria decision aid (MCDA)[4]. Also, some researches related to an analysis of non-dominated solution (NDS) have been proposed, though the number of them is limited.

In this session, we introduce previous studies for analyzing NDSs from two different viewpoints; visualization and data analysis.

A. Visualization

Obayashi et al. proposed an approach for visualization and analysis of NDSs using self organizing map(SOM)[2]. SOM can perform a nonlinear map for many dimensional data and user can estimate a trend of solutions through a cluster analysis of the obtained 2-dimension map. However, it is very difficult to interpret the visualization map when the number of solutions is large.

Yamashiro et al. proposed an approach based on fuzzy c-means (FCM) and fuzzy multiple discriminate analysis

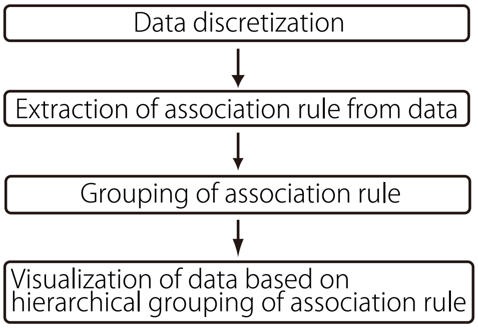


Fig. 1. The flow of the proposed system.

(FMDA)[5]. In their approach, FCM makes a cluster of NDSs and FMDA reduces the dimension of NDSs for visualization. Their approach is very similar with our proposed one in term of analysis based on group (cluster). But, thier approach didn't consider a various granularity analysis or a feature extraction mechanism. Also, FMDA in thier method doesn't support a nonlinear map because FMDA is basically based on linear discriminate.

B. Data analysis

Even though user can extract a general trend of NDSs by visualizing NDSs, it is hard to understand details of relationships between objectives and variables or local distinctive features in NDSs. Generally, analysis of variance (ANOVA) or correlation analysis is used for analyzing relationships between objectives and variables, but these methods are not necessarily to reveal an association inherent in problem especially when variable dependency is strong.

Yamashiro[5] and Watanabe[6] presented analysis methods that divide a whole NDSs into sub clusters and perform an analysis based on clustering. These methods firstly divide a whole NDSs into some subsets according to user's interest and then presents the features information of cluster obtained by an analysis based on an intergroup comparison.

On the other hand, Obayashi reported the utility of logical analysis using rough set for NDSs analysis. Rough set is one of logical analysis methods and is able to induce rules from an inconsistent data. While statistical analysis needs user to judge the results subjectively in order to extract features contained in data, logical analysis presents obviously features in the form of association rule. Therefore, we think logical analysis is more suitable to analyze NDSs objectively.

III. PROPOSED SYSTEM

In this work, we proposed a new analysis support system for analyzing non-dominated solutions (NDSs) derived by evolutionary multi-criterion optimization (EMO). The main features of the proposed system are a recursive grouping using association rules and a multi-granular analysis. The results of the recursive grouping can be viewed as a hierarchical tree in which each node is a subgroup of NDSs. The proposed system performs a multi-granular analysis based on a hierarchical tree of NDSs and it is possible to extract features of NDSs from a macro to a micro view point.

A. The procedure of the system

The procedure of the proposed system is shown in Fig. 1. Firstly, the proposed system discretizes continuous data for applying to logical analysis, and then association rules are derived from the discretized data through logical analysis. As the user tends to interest in objective function values rather than variable values, we only extracted the association rules concerning objective function values at antecedent or consequent part.

And the association rules are gradually integrated into subsets according to the degree of coincidence at antecedent or consequent part. Finally, the proposed system constructs a structured hierarchy tree by grouping these subsets based on inclusive relation.

It is important that the proposed system creats subsets of NDSs not based on a similarity between NDSs, but an explicit feature inherent in each NDS.

The details of a grouping method for association rules and a way of creating a structured hierarchy tree are shown below.

B. Grouping method for association rules

A hierarchization tree can be easily created through grouping similar association rules and a multi-granular analysis based on the hierarchization tree enable user analyze in response to user's demand.

Our system tried to integrate rules into subsets not based on mere similarity, but in accordance with coincidence of contents at antecedents and consequents of rules. And a structured hierarchy tree is easily built on the basis of inclusive relationship between subsets.

Actually, all elements that contained in all extracted association rules are selected and every combination of these elements is calculated without overlaps. And then the subsets are created for gathering association rules that have the same elements as these combination.

Therefore, all integrated subsets must have some kind of common features. In other word, every node of a hierarchization tree have at least one common character and this node's character becomes a very useful clue when user choices an interested node.

The following is the detail procedure of this grouping method.

Step 0: Discretizing every variable and objective values in NDSs.
Step 1: Extracting the association rules from discretized NDSs by appling logical analysis, which satisfy minimum confidence factor (CF_{min}) and minimum support factor (SF_{min}) and have at least one elements concerning objective function at antecedent or consequent part.
Step 2: Sorting out every discretized variable and objective values contained within extracted association rules. And calculating all combinations of these discretized values without overlaps.
Step 3: Searching an extracted association rules having the same elements at antecedent or consequent as these combination in Step 2 and creating subsets.

TABLE I. THE OBJECTIVES OF HYBRID ROCKET ENGINE DESIGN PROBLEM.

Initial total working weight $[kg]$	Minimize M_{tot}
Highest attainment altitude $[km]$	Maximize $H_{\max}$

TABLE II. THE DESIGN VARIABLES OF HYBRID ROCKET ENGINE DESIGN PROBLEM

Design variable	Range of design
Oxidant flow rate $[kg/sec]$	$1.0 \leq \dot{m}_{\text{oxi}} \leq 30.0$
Length of fuel room$[m]$	$1.0 \leq L_{\text{fuel}} \leq 10.0$
Initial port radius $[mm]$	$1.0 \leq r_{\text{port}} \leq 200.0$
Combustion time$[sec]$	$15.0 \leq t_{\text{burn}} \leq 35.0$
Fuel room pressure $[MPa]$	$3.0 \leq P_{\text{ch}} \leq 4.0$
Open area ratio [-]	$5.0 \leq \varepsilon \leq 7.0$

Step 4: Building a hierarchization tree on the basis of inclusive relationship between the features of subsets in Step 3.

The concept of the above procedure is shown in Fig. 2. Fig. 2. These figures assume that ten association rules with five discretized values (discretized variable value A1, A2, B4 and discretized objective value C2, D1) could be extracted in Step1. A1 means first category of discretized variable A and B4 also means forth category of discretized variable B.

Firstly, all possible combinations of five discretized elements (A1, A2, B4, C2 and D1) were calculated in Step 2. Step 3 tried to detect association rules having the same elements at antecedent or consequent as combinations in Step 2 and make subsets based on the elements of these association rules. Lastly, these subsets derived in Step 3 were hierarchized based on inclusion relations between combinations of elements (A1, A2, B4, C2 and D1) and a hierarchization tree having these subsets as nodes were built.

Through grouping association rules and hierarchizing groups(subsets) described above, a hierarchization tree which every node(subset) keep at least one common characters can be created. As every node has explicit feature, the proposed system can provide not only a multi-granular analysis, but also facilitate users selecting node based on their interest.

Also, a structured hierarchy tree enable to find distinctive trends from vertical and transverse hierarchized relations. For example, we can expect what kind of impact the differences between features of nodes have on results through a transverse comparisons with parallel nodes having the same parents, and we can reveal which components consists of the target node through a vertical comparisons with parent-child relationship nodes (parent is target node).

IV. THE DESIGN PROBLEM OF HYBRID ROCKET ENGINE

In this paper, we used the concept design problem of hybrid rocket engine (HRE)[3] which has been provided by JAXA.

HRE separately stores two different types of thrust propellant unlike usual rockets. The aim of optimizing this problem is to get a design knowledge through analyzing NDSs of this problem. Since this problem should treat a wide variety of elements and there are strong dependence relationships between parameters, it is difficult to optimize this multiobjective problem. Therefore, we formulated the concept

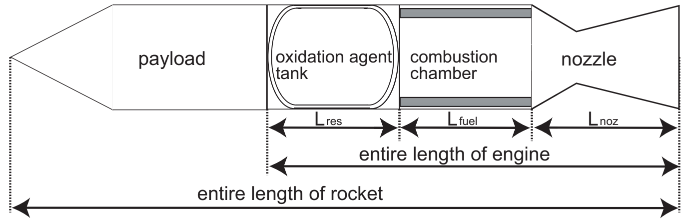


Fig. 3. The concept of HRE rocket.

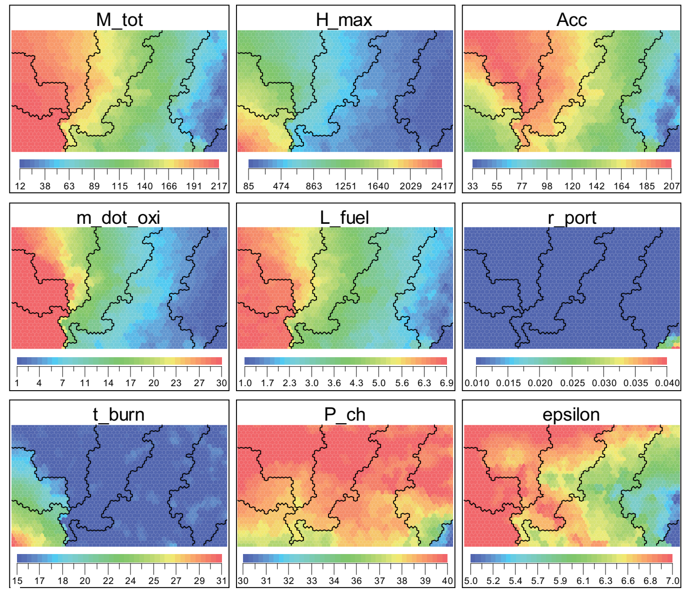


Fig. 4. The SOM results of the whole NDSs.

design problem of HRE as a two-objectives and six-design variables optimization problem. These objectives and variables are liseted in Table I and reftb:dv, respectively. The concept figure of HRE is shown in Fig. 3. In here, we omitted the setting of HRE here because of space limitations(the details setting of HRE can be found in reference [3]).

V. NUMERICAL EXPERIMENT

To investigate the effectivity of the proposed analysis system, the concept design problem of Hybrid rocket engine (HRE) having two objectives and six design variables. In this experiment, we used 504 non-dominated solutions (NDSs) obtained by the method of this reference[7] and applied these NDSs to the proposed system.

In this section, we discussed two different cases: a conventional analysis approach using SOM for the whole NDSs and the proposed analysis system using hierarchization node, in order to investigate the effectiveness of the proposed one.

Also note that SOM was used as a visualization technique for high dimensional data in this experiment, but our proposed system don't depend on any specific visualization technique. The proposed system only tries to translate the whole NDSs into structured hierarchy tree.

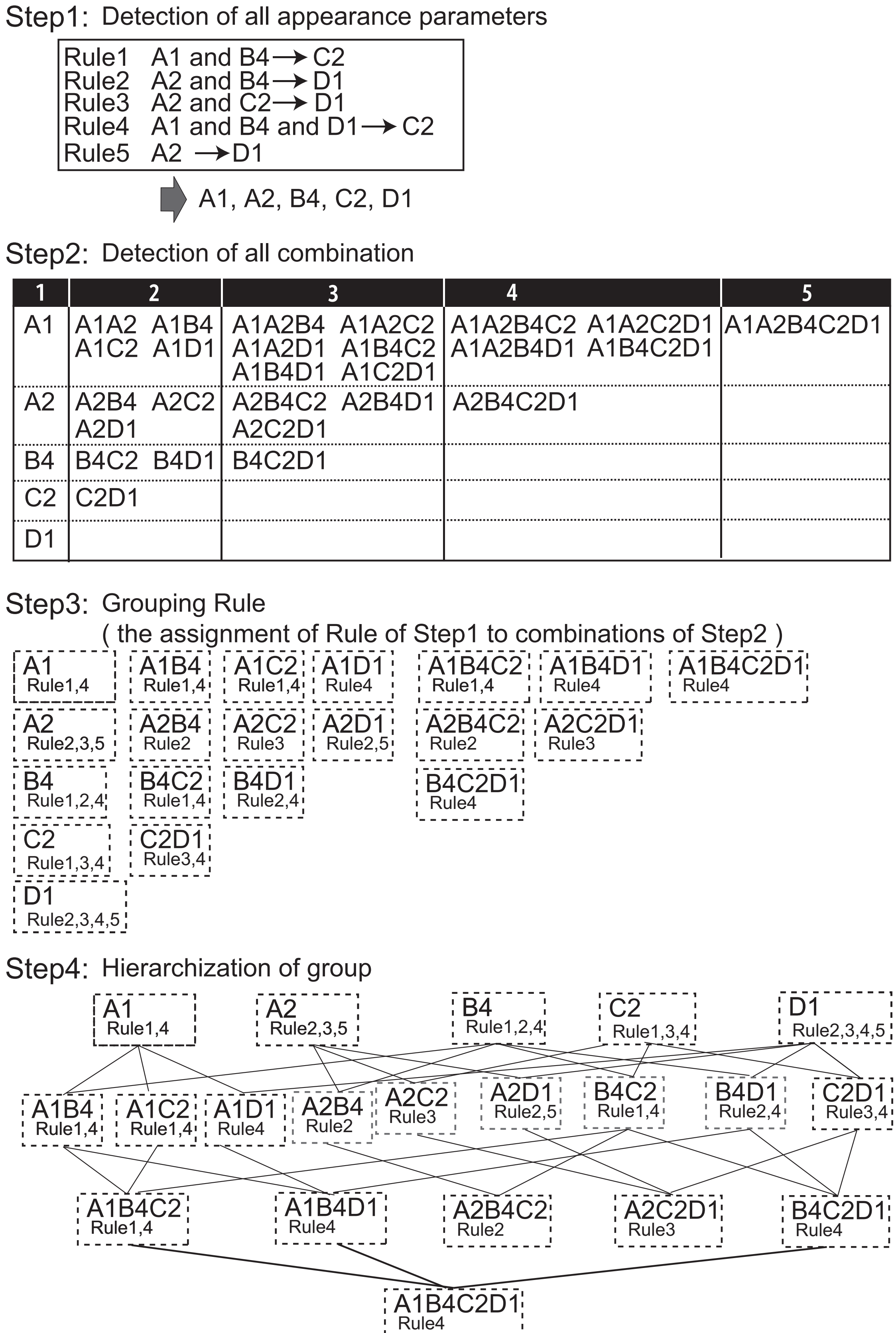


Fig. 2. The concept of grouping rules.

TABLE III. THE CORRELATION COEFFICIENT OF EACH COMPONENT IN THE WHOLE NDSS.

M_tot	0.86	0.82	0.94	0.98	-0.18	0.5	0.21	0.84
	H_max	0.42	0.94	0.92	-0.09	0.83	-0.04	0.68
		Acc	0.65	0.74	-0.22	-0.07	0.43	0.74
			m_dot_oxi	0.99	-0.11	0.6	0.11	0.75
				L_fuel	-0.16	0.57	0.17	0.81
					r_port	-0.04	-0.53	-0.25
						t_burn	-0.26	0.39
							P_ch	0.36
								epsilon

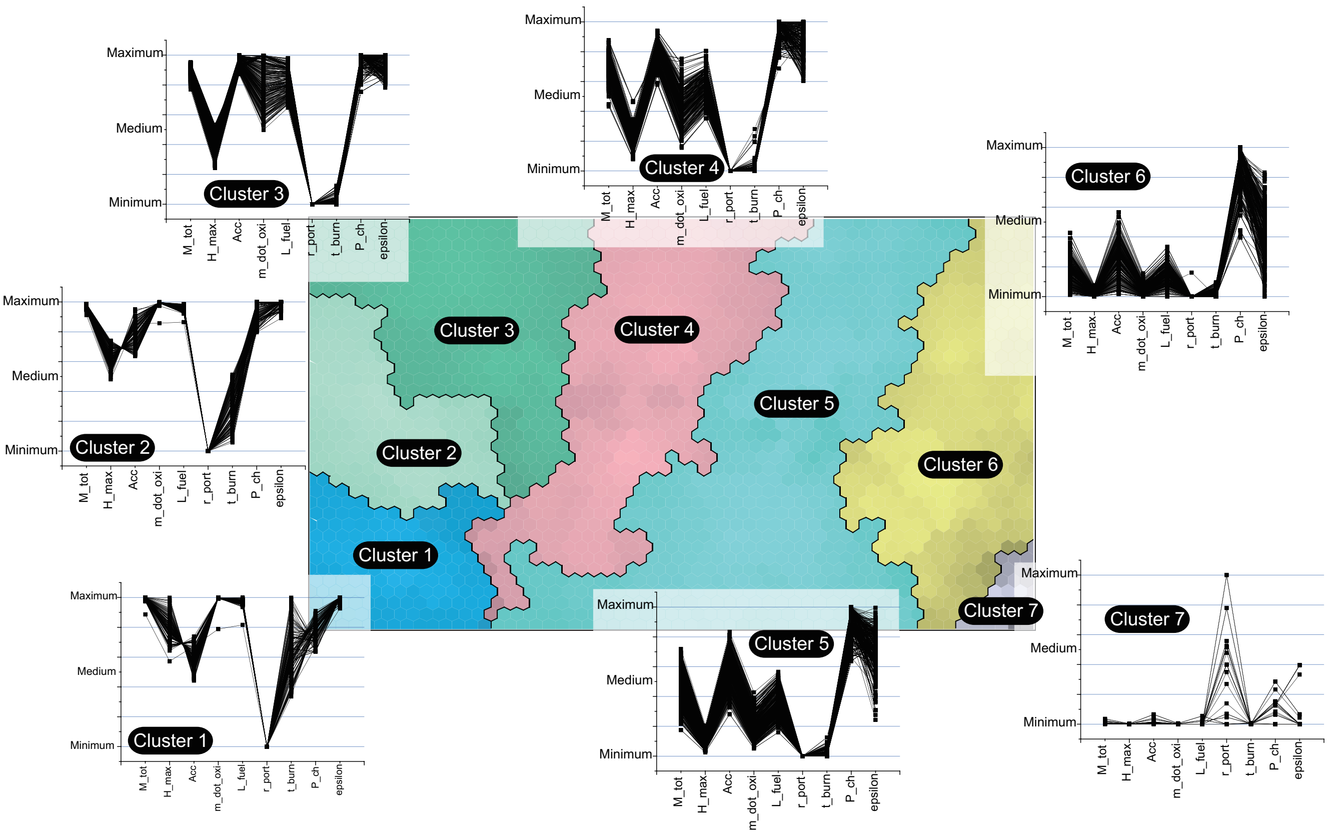


Fig. 5. The distributions of each component in each SOM cluster.

A. Analysis of a conventional approach

We used SOM analysis approach as a conventional approach intended for the whole NDSs. Fig. 4 represented the result of applying SOM to all 502 NDSs[1].

In this figure, there are nine values in total: two objectives, six design variables and acceleration. And minimum values of these were indicated in deep blue, mediums were in green and maximums were in red. Demarcation lines in these maps were the results of cluster analysis, and we could find that the whole NDSs was divided into seven clusters in total.

In order to analyze the characteristics of seven clusters derived by Fig. 4, the distribution maps of these variables in each cluster were shown in Fig. 5.

And we indicated a correlation table Table III between objectives and design variables as a quantitative evaluation.

From Fig. 4 and Table III, correlationships and distributions of these objectives and variables in NDSs could be detected. Specifically, initial total working weight(M_tot), highest attainment altitude (H_max), acceleration(Acc)and oxidant flow rate (m_dot_oxi) had a strong correlation, while initial port radius (r_port) and combustion time (t_burn) values in a number of NDSs were very low. Through these two results targeted at the whole NDSs, we could understand the whole picture of NDSs while it was impossible to grasp partially trends and features of NDSs.

On the other hand, we could find that there were some nodes having different features in NDSs from Fig. 5. For example, regarding the value of M_tot, cluster 1 and 3 have relatively high value while cluster 4 and 5 are around middle, and cluster 6 and 7 are relatively low. From this result, cluster analysis seems to be good for partial analysis of NDSs in comparison with an analysis targeted at the whole NDSs. However,

[1]In this experiments, we used Viscovery® SOMine 4.0 as SOM application and SOM parameters were set according to this application's recommended setting.

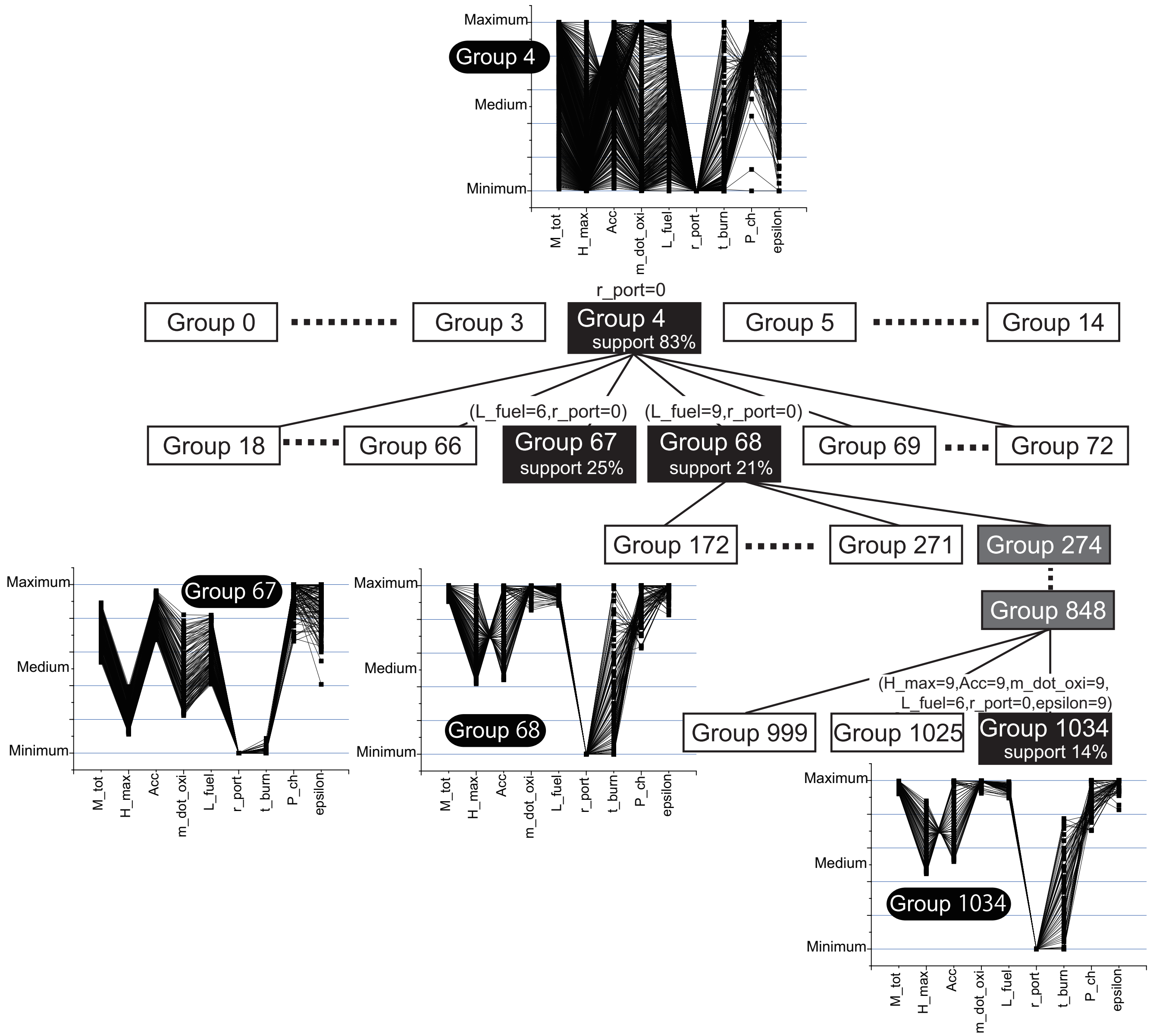


Fig. 6. The hierarchical relationship of node 4, 68 and 1034.

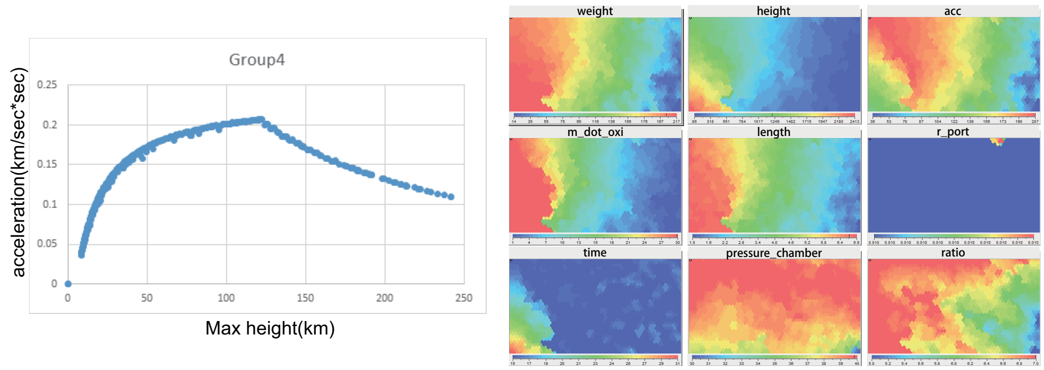
cluster analysis based on proximity between solutions is very difficult to grasp features and trends of high interest to user (or decision maker) because clustering is performed without reference to user's mind. Of course, if clustering is designed with user's preference, user's interested trends of NDSs may be perceived but any other important information cannot be obtained in this case. Important thing is that cluster analysis is effective only if user's interest is limited and very clear, and is hard to extract details information inherent in NDSs.

B. Analysis of the proposed system

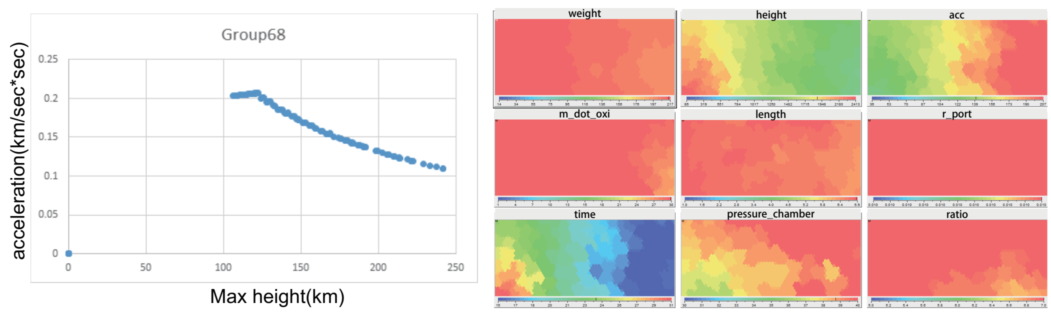
In the proposed system, all continuous attributes of NDSs were discretized into ten interval because logical analysis cannot treat continuous value directly. And Weka[8] based on apriori algorithm was used as extraction method of association rules. In here, the extraction of association rules was performed under the condition that minimum support rating and minimum confidence rating were over 0.1 and 0.9, respectively.

As a results, 1171 association rules were obtained and a hierarchical tree of NDSs having 1111 different nodes were created by the grouping method of section III-B. Since these nodes must have at least one common character, we can easily select an interested node and analyze a trend of NDSs belonging this node.

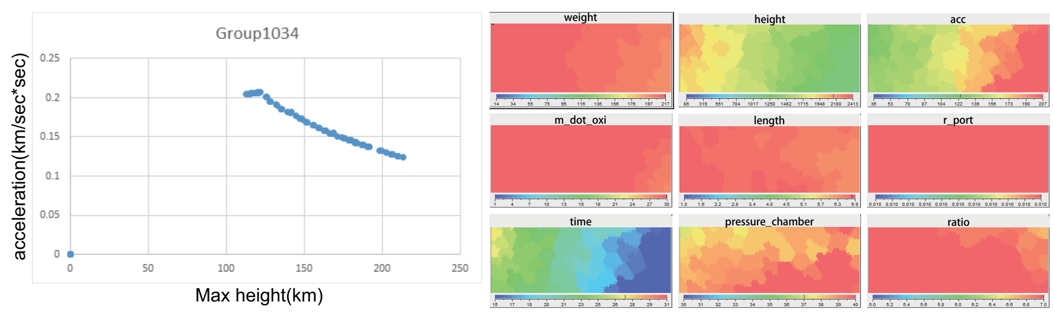
The preliminary demand from HRE designer was to reveal the feature regarding acceleration of NDSs in which highest attainment altitude is over 100km. With that in mind, we focused the results of three nodes (node 4, 68 and 1034) that were hierarchical relationship. While node 4 was the highest support rating node in all, node 68 was located on lower layer nodes of node 4, and node 1034 was also one of further lower layer nodes. The hierarchical relationship of three nodes were



(a) Top layer group (group 4)



(b) Medium layer group (group 68)



(c) Bottom layer group (group 1034)

Fig. 7. The distribution and SOM results of three nodes.

shown in Fig. 6.

1) Top layer node (node 4): Node 4 consisted of 960 association rules that have a following common feature and this node was the largest node that 83 % of all NDSs belonged to.

> Initial port radius is dLevel_0[2].

As the result of node 4, self-organizing maps (SOM) results and the relationship between acceleration and highest attainment altitude were shown in Fig. 7(a). In Fig. 7, minimum value was indicated in deep blue, medium was in green and maximum was in red.

From Fig. 7(a), the map of initial port radius were painted all blue that means minimum and the value of this was identical with a common feature of node 4. This fact was proof that data set of node 4 falled under a common feature. We found that combustion time (time) was relatively low, but fuel room pressure (p_ch) and open area ratio (ratio) were relatively high. On the other hand, initial total working weight(m_tot), highest attainment altitude (h_max), length of fuel room(l_fuel) and oxidant flow rate (m_oxi) seemed not to be correlated with initial port radius. From these result maps, we could confirm correlations between objectives and design variables from a broader perspective.

2) Medium layer node (node 68): Node 68 had 246 association rules that had following common features and 21% of all NDSs belonged to this node. As can be seen from common features of node 68, this node were located directly below node 4.

> Length of fuel room is dLevel_9 and initial port radius is dLevel_0.

[2]dLevel means discretization level. dLevel_0 corresponds to the minimum discretized interval while dLevel_9 corresponds to the maximum.

The SOM visualization results and the relationship between acceleration and highest attainment altitude of NDSs belonging node 68 were shown in Fig. 7(b).

Fig. 7(b) indicated that highest attainment altitude of all NDSs in this node were over 100km, but acceleration values were relatively variable. In comparison with Fig. 7(a) and (b), we found that each values of node 68 were totally homogenized as compared to those of node 4. This fact implied that these homogenized values of each variable are necessary to achieve high value of highest attainment altitude. Also, we could detect that combustion time (time) is non-correlated to highest attainment altitude.

3) Bottom layer node (node 1034): Node 1034 was on bottom layer which are below node 4 and node 68, and was situated at the terminal of hierarchy. This node consisted of only one association rule that 14% of all NDSs belonged to. The association rule of this node was below.

Initial total working weight is dLevel_9, oxidant flow rate is dLevel_9, open area ratio is dLevel_9,length of fuel room is dLevel_9 and initial port radius is dLevel_0.

The SOM results of NDSs belonging to node 1034 were shown in Fig. 7(c) in the same way as above examples.

From the comparison between Fig. 7(b) and (c), we could confirm that the trends of node 68 and 1034 are very similar. As content rates of node 68 and node 1034 differed only 7%, we could expect that data sets belonging these nodes were almost same and these trends of node 68 and 1034 were also similar. Although difference between common features of node 68 and 1034 were fairly big, these solutions of node 68 and 1034 were very similar. From this fact, we could assume that common feature " length of fuel room is dLevel_9" of node 68 and 1034 has strong influence with many variables including highest attainment altitude.

Since higher layer nodes contain more association rules and NDSs, these nodes are suitable to estimate general trend which is applicable to many NDSs. On the other hand, lower layer nodes that consist of small data set are better suited for analyzing a feature of NDSs in a restricted range.

As shown above, our proposed system could extract the detail relationships between objectives and design variables by comparing different layer level nodes which were inclusive relations. Although we didn't refer in this paper, it is possible to clarify a distinctive feature of NDSs through comparison between different nodes which are exclusive relationship.

VI. Conclusion

In this paper, we proposed a new analysis support system for analyzing non-dominated solutions (NDSs). Since the proposed system is enable a multi-granular analysis using a hierarchical tree of NDSs, the designer can extract not only a general trend of the whole NDSs but also specific features of restricted NDSs. The proposed system applies logical analysis to the whole NDSs and the whole NDSs are divided into some subsets according to the obtained association rules. And a recursive grouping using association rules is performed in order to create a hierarchical tree in which each node is a subgroup of NDSs. As a hierarchical tree is linked to the granularity of data, a multi-granular analysis is possible in this system.

Through applying the proposed system to the concept design problem of hybrid rocket engine (HRE), we could confirm that our system worked well for creating a hierarchical tree based on grouping association rules and it was possible to extract the detail relationships between objectives and design variables by comparing different nodes which are inclusion relationship but on different layer level.

As our future works, we will promote a development of an efficient discretization for logical analysis and investigate the use of other logical analysis methods, such as rough set and concept lattice.

Acknowledgment

This research was funded in part by a Grant-in-Aid for JSPS fellows (Scientific Research (C) (General), No. 23500268, Representative: Shinya Watanabe).

References

[1] K.Deb, *Innovization: Innovative Solution Principles Using Multiobjective Optimization*. Springer-Verlag New York, Inc., 2012.

[2] S. Obayashi and D. Sasaki, "Visualization and data mining of pareto solutions using self-organizing map," in *Second International Conference on Evolutionary Multi-Criterion Optimization (EMO 2003)*, 2003, pp. 796–809.

[3] Y.Kosugi, A.Oyama, K.Fuji, and M.Kanazaki, "Conceptual design optimization of hybrid rocket engine(in japanese)," in *Proceedings of Space Transportation Symposium*, 2010, pp. STCP–2009–75.

[4] D. Cvetković and I. C. Parmee, "Designer's preferences and multi–objective preliminary design processes," in *Proceedings of the Fourth International Conference on Adaptive Computing in Design and Manufacture (ACDM'2000)*, 2000, pp. 249–260.

[5] D.Yamashiro, T.Yoshikawa, and T.Furuhashi, "Support to select satisfying solutions using visualization method in multi-objective optimization problem(approaches to optimization problems)," *journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, vol. 20, no. 6, pp. 850–859, 2008.

[6] S.Watanabe and R.Minato, "Development of a design support system that can efficiently utilize non-dominated solutions," in *Joint 4th International Conference on Soft Computing and Intelligent Systems and. 9th. International Symposium on advanced Intelligent Systems(SCIS & ISIS 2008)*, 2008, pp. FR–A4–3.

[7] S.Watanabe and N.Yokouchi, "A new local search method with the guarantee of local pareto optimality," in *Proceedings of the 6th International Conference on Soft Computing and Intelligent Systems & 13th International Symposium on Advanced Intelligent Systems (SCIS & ISIS 2012)*, 2012, pp. CDROM–paper (F3–43–2).

[8] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems)*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2005.