Organization of Ship Design Information Using Cramer's V and Genetic Algorithm

Yusuke Sato[1], Kazuhiro Aoyama[1], Hideaki Ozawa[1], Kenji Takahashi[2], Shingo Maeda[3], Naoki Herai[3].

[1]The University of Tokyo
[2]Imabari Shipbuilding Co., Ltd.
[3]Nihon Shipyard Co., Ltd.

Abstract: In ship design, existing specifications are repeatedly modified in response to shipowner requests. In this process, there are components that are simultaneously changed for design-related reasons, but the relationships among components have been established based on the tacit knowledge of designers owing to a large number of ship components. In this study, we identify specifications that change simultaneously by using Cramer's V. By leveraging the macro-level granularity, the designers can perform efficient checks and create a Design Structure Matrix (DSM). Additionally, the components are grouped by performing DSM clustering, which reflects whether specification changes occur within or outside the department. A case study is conducted using the actual specifications of a merchant's vessel to demonstrate the effectiveness of the proposed method. Consequently,accurate relationships among components are extracted, and the departments that must collaborate with each other are identified.

Keywords: Shipbuilding, specification change, Cramer's V, DSM clustering

1 Introduction

Japan is currently facing a declining birthrate and an aging population, and the manufacturing industry is facing labor shortages owing to the retirement of numerous skilled workers. The skills of such skilled workers must be transferred efficiently to maintain and expand production.

Similarly, shipbuilding industries also face shortages of skilled workers. The shipbuilding industry is based on a build-to-order production system in which shipbuilders use past construction records and standard specifications as a reference for fulfilling the requests of shipowners. The specifications are repeatedly modified to reflect the requests. However, owing to restrictions on ship design, price, and delivery date, other specifications must also be simultaneously altered when specifications are modified per request. Therefore, the process requires considerable effort and time because numerous factors have to be considered.

Understanding the relationships among such specifications can reduce the time and effort required for ship design. However, the relationships among the specification items are not organized because ships are complex systems with many interrelated components, and the design process currently relies on the intuition and experience of skilled personnel. Thus, the current design process relying on tacit knowledge should be discarded when considering the decreasing number of skilled workers.

Prior research on the clarification and organization of design and production processes in manufacturing industries includes the use of a design structure matrix (DSM), which attempts modularization in consideration of design constraints (Kaushik et al., 2019). Additionally, the effectiveness of DSM outside the manufacturing industry has been studied in recent years (Hui et al., 2022). Studies have also attempted to organize design information in a network format instead of a matrix format (Loureiro et al., 2020).

However, in such studies, the application of the suggested methods to products with a large number of components is difficult because the relationships among elements are manually entered, and the analytical results depend on the subjectivity of the person who entered the data. Therefore, the relevance between the specifications of a ship must be extracted automatically and objectively.

There are some works about DSM clustering in ship design specifically. Gui et al. (2021) analyzed the specifications of commercial ships by integrating Bayesian network structure learning and DSM clustering to evaluate the complexity of feature relations at the chapter level for text information on general parts of commercial ship specifications to support the development of standard specifications. The study highlighted the limitations of using only text data. Gunawan et al. (2019) develop an approach for moldularization of engine room using DSM. This study used E-R model (Peter, 1976) for making DSM and genetic algorithm for DSM clustering. These work focused on certain parts or systems but Vessels are organically related to various systems. Thus, in this study, we focus on machinery, electrical, and ship outfitting sections by describing detailed specifications rather than general texts and attempt to analyze detailed data on specifications that are summarized in tabular form instead of text data. It is difficult to check relationships batween whole ship elements so we use a statistical method before making the DSM.

The shipbuilding process is designed by subdivision based on outfitting and systems, which makes specification adjustments across departments more difficult compared to modifications that can be completed within a department. Currently, the departments that should collaborate on a particular specification modification are determined based on experience. Therefore, determining the necessity of cross-departmental collaboration when modularizing components through clustering is important.

Based on the aforementioned context, the main contributions of this study are as follows:

1. By calculating Cramer's V for the content described in the specifications, we automatically and objectively identified candidate specifications that changed simultaneously.
2. We differentiated the costs of specification changes based on whether they crossed departments and modularized the components through DSM clustering, which reflected the cost differences.

2 Theoretical background

Before proposing the method for analysis, we present an overview of Cramer's V and DSM clustering methods, which were used for this study.

2.1 Cramer linkage coefficients

Understanding which specification items require simultaneous modifications to formalize the relationships among ship specification items is necessary.

Cramer's V (Cramer, 1946) is an indicator that quantitatively shows the relationships among categorical data. The indicator is based on an χ^2 test, which is conducted to verify the independence of two variables.

Specifics of the calculation method are as follows. First, a cross-tabulation table is created for two sets of categorical data, and the χ^2 value is calculated. Thereafter, Cramer's linkage coefficient *V* is calculated by substituting *n* for the number of samples in the categorical data and *k* for the smaller of the kinds of two data into equation *(1)*.

$$V = \sqrt{\frac{\chi^2}{n\,(k-1)}} \tag{1}$$

Cramer's V has values between 0 and 1, with larger values indicating a stronger relationship between two sets of data. Conversely, when comparing specifications, the more likely a specification item requires change when another is also changed, the closer the value of Cramer's V to 1.

2.2 DSM clustering

DSM is a matrix representation of the relationships among elements in industrial production, while clustering by rearranging rows and columns makes modularization of parts with the most feedback and presents production procedures with the fewest reworks possible.

Genetic algorithms are the most common clustering methods (Askhøj and Mortensen, 2019), of which is the Idicula-Guitierrez-Theveau Algorithm-plus (IGTA+) (Borjesson and Hölttä-Otto, 2012). The IGTA+ is based on IGTA (Idicula, 1995; Gutierrez Fernandez, 1998; Theveau, 2001), which moves elements from one cluster to another for determining whether the overall change leads to any improvement. In this algorithm, *ClusterBid* determines the fitness of an element to a cluster, based on which the cluster where the element is to be moved is determined. *TotalCost* indicates the appropriateness of cluster division of the entire DSM and is used to determine the suitability of moving an element. A flowchart of IGTA is presented in Figure 1.

The IGTA+ efficiently allocates elements to clusters, calculates *TotalCost*, and reduces the completion time of the algorithm by maintaining a list of untested elements. Borjesson and Hölttä-Otto (2019) have provided a detailed description of the calculation method.

DSM clustering processes a matrix by grouping the parts with high correlation values entered into the matrix in a cluster. The strength and importance of the relationships are reflected by adjusting the entered correlation values. For example, previous studies have focused on the problem of impact area expansion because of the existence of relationships between elements that serve as interfaces between clusters, while other studies on DSM have focused on minimizing the correlation values between modules (Li et al., 2019). The proposed method of penalizing cells that serve as interfaces has made it possible to consider both the modularity and impact range of design changes.

Elsewhere, Engel et al. (2017) focus on the trade-off between reducing production and design costs and flexibly responding to customer requirements through modularisation, and propose a cost-effective modularisation method using genetic algorithms. While the focus of this study is on economics, the purpose of modularization in this study is to visualise information about the coordination between different departments in specification changes. Therefore, the weights in the clustering should reflect the coordination between departments.

By clustering, a systematic understanding of the relationships among numerous specification items of ships can be established. This enables us to determine the extent to which contemplated specification changes may have an impact and to grasp specification items that can be developed independently from ongoing specification changes.

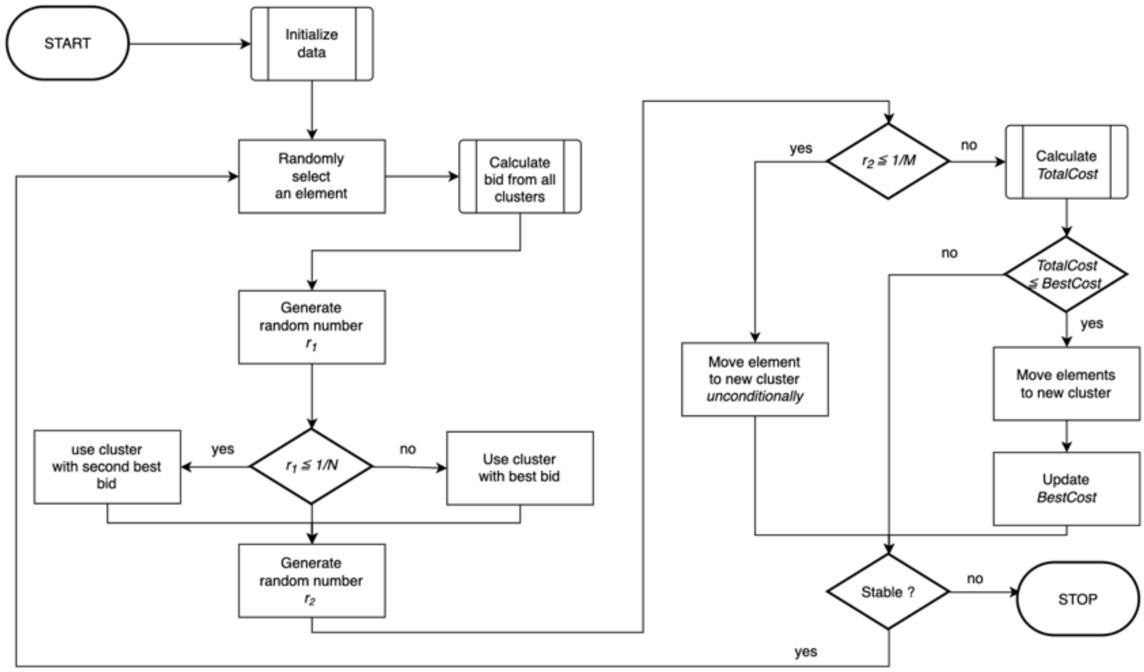


Figure 1. Flowchart of IGTA (Borjesson, F. & Hölttä-Otto, K., 2012)

3 Methods

Using Cramer's V and IGTA+ discussed in section 2, we propose a method for identifying specification items that are simultaneously changed from data of ship specifications and modularize them to reflect inter-departmental coordination.

First, data on specifications are prepared in a tabular form. Then, we calculate Cramer's V for any combination of specification items, check Cramer's V for several pairs of specification items, set an appropriate threshold, and consider specification items with the value of Cramer's V above the threshold to be "relevant," while those without such a coefficient to be "not relevant." The appropriate threshold value varies depending on the analysis and the number of samples; therefore, a flexible threshold value can be set.

The value of Cramer's V becomes high when both variables change together, and determining whether the correlation has occurred coincidently is not possible. Therefore, calculation result checks by designers are more effective than using the results as they are. To efficiently perform such checks, the utilization of ship specifications categorized by systems, components, and parameters (e.g. manufacturer, output) in descending order of magnitude is employed. Specifically, if there are pairs of "relevant" components between two different systems, those systems are considered as "relevant." Subsequently, instead of checking numerous ship specification items individually for the accuracy of the correlation obtained from Cramer's V, the existence of relationships is determined beforehand at a more macro level. Even if the Cramer's V correlation coefficient is high, the relationship between components and parameters belonging to unrelated systems is considered "not relevant." Thereafter, the check proceeds to the level of components and parameters for their correlations. By leveraging the fact that design information has multiple levels of granularity, the effort put in by the designers while checking can be reduced.

Second, for each pair of specification items considered as "relevant," we assign importance to relatedness, reflecting whether the specification change crosses departmental boundaries. Information on ship specifications is divided in descending order into the following categories: systems, components, and attributes (for example, the manufacturer of the MAIN ENGINE in the MAIN PROPULSION UNIT is listed as *x*). As described in Chapter 1, a cost difference between changes in specifications that are completed within a department and those that are performed across departments exists, especially a reduction in design efficiency when the latter occurs frequently. Therefore, systems are regarded as departments in this study. Specification items belonging to different systems are entered with larger values than those belonging to the same outfits or systems when simultaneously changing specification items are expressed in the DSM.

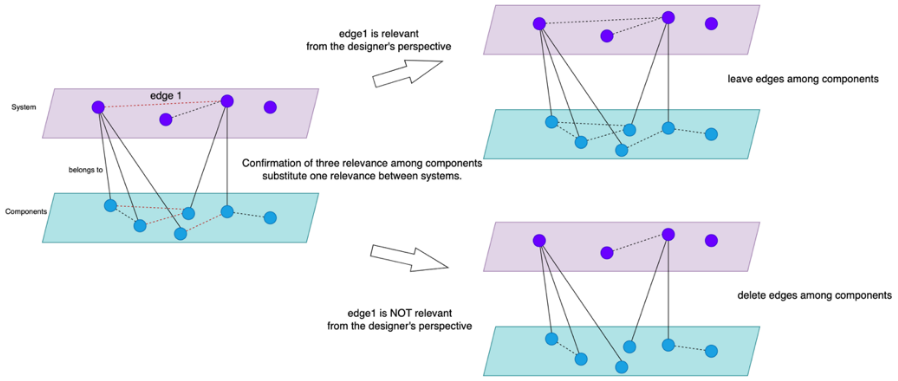


Figure 2. Simplified proposal for checking relationships among elements

Once a DSM reflecting the design cost of the change is created, clustering is performed using IGTA+. While clustering, specification items that differ in the system to which they belong tend to be placed in the same cluster based on the size of the values entered into the DSM. Meanwhile, the assignment of specification items belonging to the same system to the same cluster is relatively reduced; however, even if these specification items are assigned to different clusters, the knowledge of the person in charge can cover them to some extent. This allows related departments to be grouped in advance for specification changes that require cooperation, cannot be completed by a single person, and are assumed to lead systematic specification changes.

4 Results

4.1 Data set

As a case study, we analyzed the specifications of the mechanical, electrical, and hull equipment of 30 vessels in the 63,000 MT D/W TYPE BULK CARRIER or 64,000 MT D/W TYPE BULK CARRIER merchant ship series. The specification data were listed in descending size in tabular form under the categories of outfitting, systems, components, and specifications.

Table 1. Example of specification information

Outfiting	System	Components	Attributes	Ship 1	Ship 2
mecanical	MAIN PROPULSION UNIT	MAIN ENGINE	Maker	A	B
mecanical	MAIN PROPULSION UNIT	MAIN ENGINE	Power	i	j
mecanical	FUEL OIL SERVICE	FUEL OIL TANK	Capacity	x	y

4.2 Calculation of Cramer's V

Before calculating Cramer's V, specification items with only one type of description and those whose independence from other specification items could not be determined were excluded from the analysis. Such data included 30 types of systems, 330 types of components, and 770 types of attributes.

First, Cramer's V was computed at the specification level. A histogram showing the distribution of Cramer's V for 296,065 specification item pairs is shown in Figure 3.

Second, we defined "component-level description" as a description of specifications belonging to the same connected component to calculate the relationship between elements at the component level. For example, if Company A manufactures the main engine, such that the model is *T* and the output power is *x*, then the "description of the main engine" is "*A*, *T*, *x*." Although the simple notation method used commas, specification items for commercial ships were changed together as a set. This notation method was chosen because the number of combinations did not explode. Figure 4 (a) shows the distribution of the number kinds of "component-level descriptions" in the considered notation method, while Figure 4 (b) shows the distribution of the number of combinations of "component-level descriptions" for each pair.

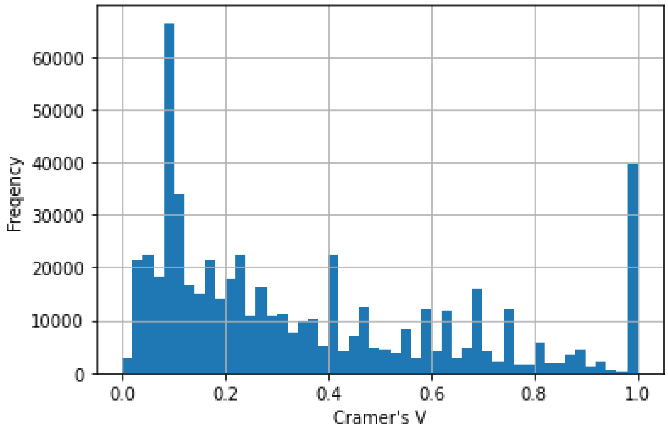
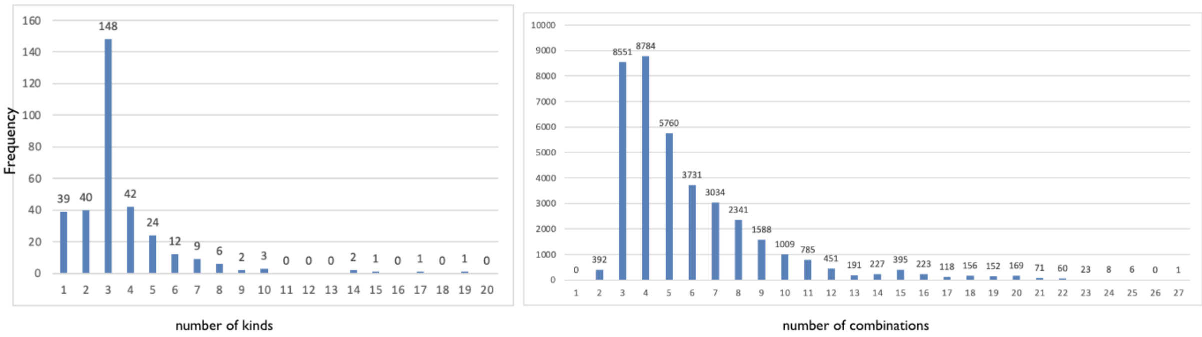


Figure 3. Distribution of Cramer's V at the attributal level



(a)

(b)

Figure 4. Number of types and combinations of "component-level description"

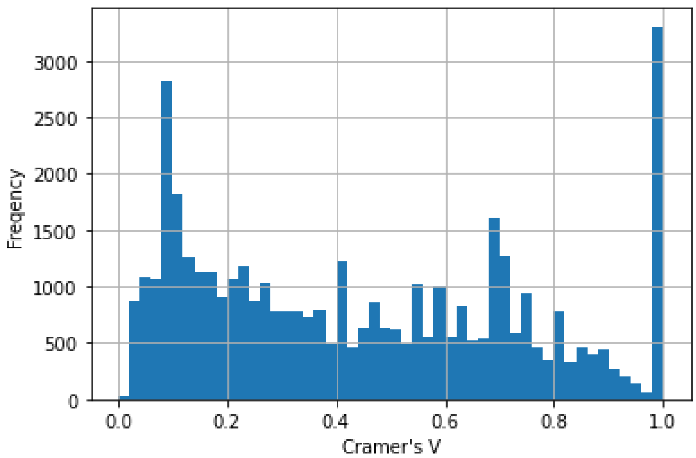


Figure 5. Distribution of Cramer's V at the components level

Additionally, components that have only one type were excluded and Cramer's V was calculated at the component level. A histogram showing the distribution of Cramer’s V for 42,195 component pairs is presented in Figure 5.

Third, the results were reviewed by an experienced shipbuilding design engineer with a set threshold value of 0.96, and Cramer’s V included enough specification items that were simultaneously changed.

While performing the aforementioned checks, a relationship where 30 systems were connected by lines was observed by the designer when Cramer’s V of the components belonging to the systems was referred. Based on the perspective of the designer, the systems that were truly considered to be related were indicated by red lines, as shown in Figure 6.

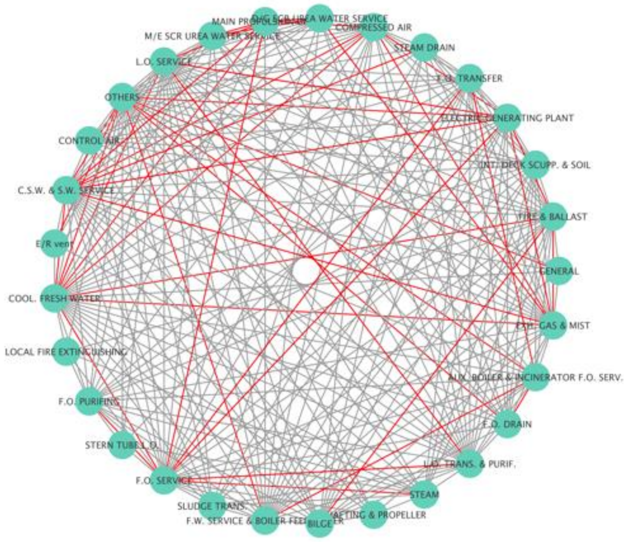


Figure 6. Relationships among ship systems

Table 2. IGTA+ parameter settings

Parameter	Setting
powcc	1.0
powdep	2.0
powbid	0
rand_bid, randaccept	Twice the number of components in DSM
max_size	0.2

In this process, the relationships among components belonging to systems that were deemed unrelated were removed. Finally, the relationships among components within systems or between systems that were deemed related were evaluated by the designer to determine the true existence of the relationships.

4.3 DSM clustering

After identifying pairs of specification items that simultaneously changed at the component or specification level, we classified the importance of the relationships based on their belongingness to the same system or different systems. When the systems to which two related components belonged were different, an input of value 5 was given to the DSM, while when they belong to the same system, a value of 1 was given as input.

IGTA+ was applied to DSM created for clustering. The values of various parameters used in the algorithm are listed in Table 2.

The clustering results of the DSMs at the component level are shown in Figure 7. The clusters were the same in the area circled by the blue line. When the systems to which two components belonged were different, they were represented by red dots, while yellow dots represented the case when the two components belonged to the same system.

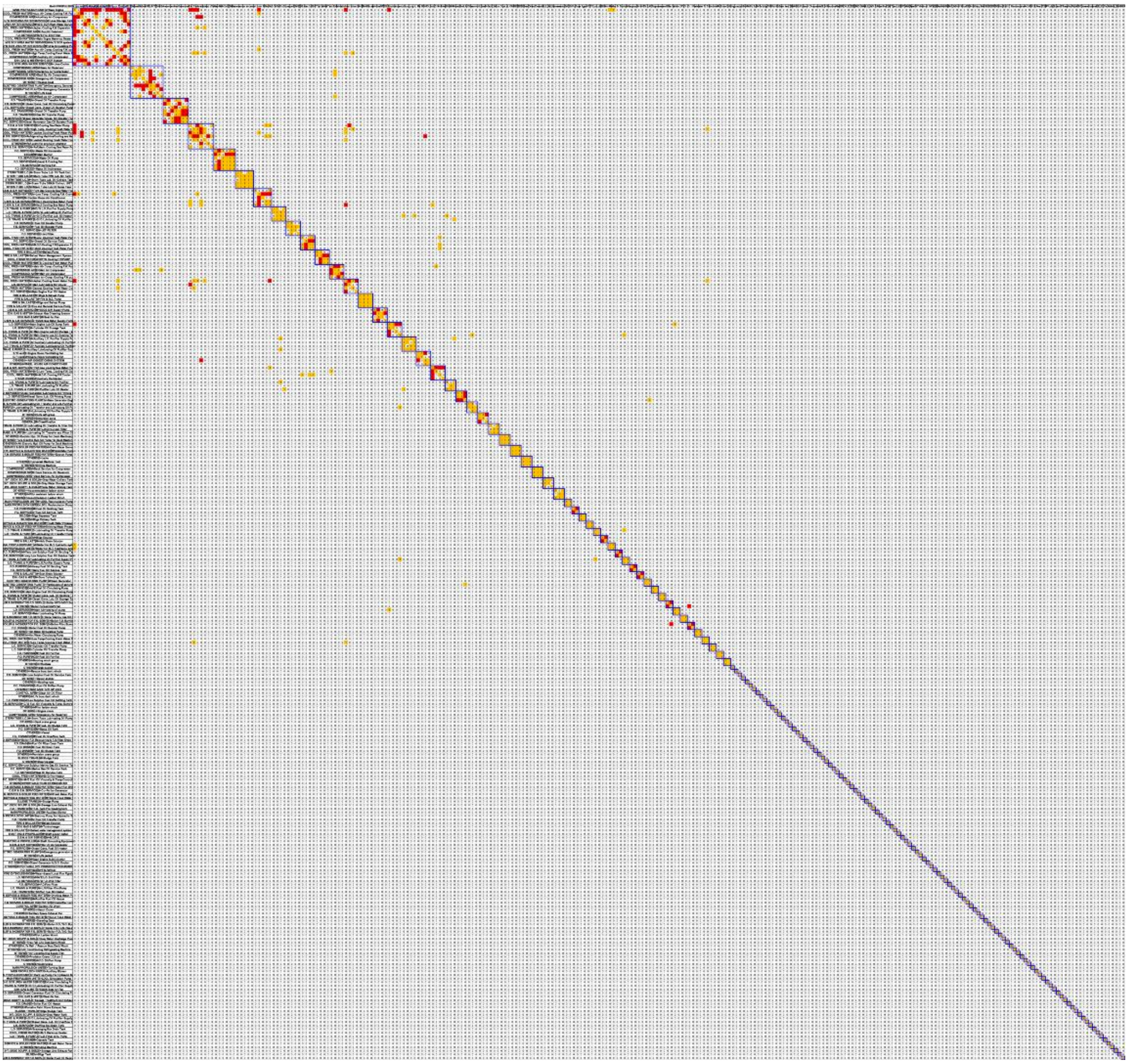


Figure 7. Clustered DSM of components

5 Discussions

This section discusses the extraction accuracy of relationships among specification items using Cramer’s V and modularization through DSM clustering.

5.1 Cramer’s V

Skilled designers reviewed the output results to identify relevant specification items using Cramer’s V. In this study, a value of 0.96 was adopted as the threshold of Cramer’s V. A close examination of the specification items that were considered as “relevant” indicated that the output was somewhat accurate for those pairs of specification items that were expected as output by the skilled designer.

There are also pairs of components that may be relevant from the designer's perspective but have low Cramer's V. Not regarding these pairs of constituent elements as related can be helpful for a precise understanding of specification changes in a particular series of ships.

On the other hand, multiple cases where the Cramer's V was high even for pairs of specification items that did not appear to be related were observed. This was likely due to the fact that the completed specification was created through multiple specification changes. Specifically, some specification items were changed in response to requests from the shipowners, while other parts were changed to meet different requests. However, even when comparing the completed specifications, understanding the order of changes was difficult. Further, all the changes were recognized to occur simultaneously, resulting in a high Cramer's V. To distinguish among multiple specification changes that occurred in response to different requests, information on macro-level concepts such as systems may be helpful.

Since this study was conducted on a single series of vessels, the analysis could be affected by specification changes and fixed specification items specific to the series of vessels. Therefore, if specification changes are assumed based on the general knowledge of ships, a discrepancy between the analytical results and the tacit knowledge of designers can exist. To overcome this problem, the number of target analyses must be increased. Conversely, Cramer's V can be considered to extract features of series vessels that cannot be grasped even by skilled operators.

5.2 Modification of the results from Cramer's V

By using the results of Cramer's V calculations and conducting checks through the designers, we aimed to create an accurate DSM. Since multiple types of components exist in a ship, performing checks at the granularity level of Cramer's V calculations can be cumbersome. Therefore, by conducting initial checks at a macro level, the effort put in by the designers during checks can be reduced. In this study, checking the relevance of 30 lineages has allowed us to correct some of the relevance between about 300 components.When there are different levels of granularity in the specification information, it is effective to choose the appropriate level of granularity for the parts processed opportunistically and the parts that involve human interaction.

Improvements to information accuracy are desired through designer checks; however, Cramer's V can extract the correlation among configuration elements that designers may not be aware of, as aforementioned, and excessive intervention by the designers may prevent them from being reflected in the DSM calculation results.

5.3 Efficiency of DSM clustering

By DSM clustering, we grouped specification items that particularly exhibited strong relationships together. Except for the largest group, the number of components belonging to each group was in single digits. Thus, specification changes tend to occur in a partial manner in ships composed of numerous pieces of equipment. The largest group located in the upper left corner of Figure 7 consisted of specifications related to the main engine, which aligned with the reality that the main engine is the core specification in a ship and has a very large impact range.

Since the relationships among components belonging to different systems were weighed during clustering, some clusters contained multiple systems. Specifically, cases where multiple systems related to fuel or water management belonged to the same cluster were observed. Thus, we identified specification changes that required coordination between departments.

There were 52 associations between components belonging to different clusters, 40 of which were between components belonging to the same lineage. Considering that of the remaining 12 of these, 6 are related to the Main Engine, which is the ship's core component and affects the various components, it can be said that the DSM clustering adequately mapped the related lineages and departments. In addition, the relationship between components belonging to the same system and different clusters can be viewed as a bridge in network analysis, which is important for modularity.

Furthermore, cases where components belonging to the same system were part of multiple clusters were observed. Thus, the existence of several strongly related groups within the same system was established. By obtaining such information in addition to the conventional specifications, designers could gain a more accurate understanding of the impact scope of specification changes. Additionally, since components belonging to the same cluster were functionally related, clustering potentially assisted in optimizing traditional system classification and departmental organization.

IIn addition, focusing on each cluster, some clusters are closely related to each other, while others consist mainly of specific components. For example, in the top left cluster, the Main Engine, and in the second cluster, the Emergency Generator, have many relationships with other components and play a central role. By analysing things like centrality in network analysis, it may be possible to quantify the impact of each component on the overall system.

6 Conclusion and future work

In this study, a DSM was created and clustered based on specification descriptions to understand the tacit knowledge of specification changes in shipbuilding from the perspective of module design. Although previous studies on DSM manually examined the relationships among elements, creating a manual DSM for ships was difficult because of the large number of components. This study identified specification items that were simultaneously changed by using Cramer's V. However, since Cramer's V could not distinguish between multiple specification changes that occurred during the specification document creation stage, the results could contain inaccurate information. Therefore, a check was performed by designers to address the issues. While checking, the information related to macro-level systems was utilized to reduce the workload of the designers. Additionally, considering that specification changes across departments were more expensive than those completed within one department, we proposed DSM creation and clustering using IGTA+. Further, the proposed method was applied to actual specification data and its effectiveness was discussed.

To extract specification information by Cramer's V, the proposed method extracted enough specification items that required simultaneous adjustment, but pairs of specification items with high Cramer's V, unrelated to the design, were also observed. We attempted to increase the number of samples and the types of vessels to be analyzed to improve the extraction accuracy by referring to the information from various perspectives, such as the age of manufacture, shipowner, and text data, in addition to Cramer's V. If specification changes can be accurately extracted based on Cramer's V, the proposed method may find application not only to ships but also to other products with many components and multiple variations.

For DSM clustering, we suggest that by referring to the system to which the specification item belongs (the macro classification of the constituent elements), modularization that reflects the cost of cooperation among departments can be performed for determining the departments that must coordinate in advance. In the future, we plan to conduct trials with multiple values while interviewing the designers because we do not have sufficient evidence on methods to set the values to be entered into the DSM.

References

Askhøj, C. & Mortensen, N. H., 2019. Deciding on the total number of product architectures. Concurrent Engineering, 28(1), 20-31.

Borjesson, F. & Hölttä-Otto, K., 2012. Improved clustering algorithm for design structure matrix. Proceedings of the ASME 2012 International Design Engineering Technical Conferences 87 and Computers and Information in Engineering Conference. 3, 38th Design Automation Conference, Parts A and B. 921-930.

Cramér, 1946. Mathematical Methods of Statistics. Princeton: Princeton University Press, 282.

Engel Avner., Tyson R. Browning, & Yoram Reich., 2017 Designing Products for Adaptability: Insights from Four Industrial Cases. Decision Sciences, 48(5), 875-917.

Gui, C., Zeng, R., Takahashi, K., Herai, N. & Aoyama, K., 2021. Genetic algorithm-based clustering method to formulate standard specifications for merchant ship preliminary design. In International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (Vol. 85376, p. V002T02A050). American Society of Mechanical Engineers.

Gunawan., Yanuar., Farhan Aji Waskita. & Arif Kurniawan., 2020. Modularization of Ship Engine Room Using Design Structure Matrix (DSM) Based on The Genetic Algorithm. Engineering Journal 24.4, 205-216.

Gutierrez Fernandez, C. I., 1998, Integration Analysis of Product Architecture to Support Effective Team Co-Location, Master's Thesis at Massachusetts Institute of Technology, Department of Mechanichal Engineering.

Idicula, J., 1995, Planning for Concurrent Engineering, Gintic Institute Research Report, Singapore.

Li, Z. K., Wang, S. & Yin, W. W., 2019. Determining optimal granularity level of modular product with hierarchical clustering and modularity assessment. Journal of the Brazilian Society of Mechanical Sciences and Engineering 41, 1-14.

Loureiro, G. B., Ferreira, J. C. E. & Messerschmidt, P. H. Z., 2020. Design structure network (DSN): A method to make explicit the product design specification process for mass customization. Research in Engineering Design, 31, 197-220.

Peter Pin-Shan Chen., 1976. The Entity-Relationship Model toward an un-fined View of Data. ACM Transaction on Database Systems, 1(1), 9-36.

Sinha, K., Han, S. Y. & Suh, E. S., 2020. Design structure matrix-based modularization approach for complex systems with multiple design constraints. Systems Engineering, 23(2), 211-220.

Tang, H., Zhong, Q. & Chen, C., 2022. Process optimization for post disaster reconstruction project based on industrial design structure matrix (DSM). Soft Computing, 26(17), 8731-8743.

Thebeau, R. E., 2001. Knowledge management of system interfaces and interactions from product development processes. Doctoral dissertation, Massachusetts Institute of Technology.

Contact: Yusuke Sato, The University of Tokyo, School of Engineering, Room 330, Faculty of Engineering Bldg.3, 7-3-1 Hongo, Bunkyo-ku Tokyo 113-8656, Japan, yusuke_s@m.sys.t.u-tokyo.ac.jp.