Development of a hypothesis exploration application by Kansei Information approach

Takashi Sonda* and Toshimasa Yamanaka**

* University of Tsukuba, Graduate school of Human Comprehensive Sciences
Japan, sonda@kansei.tsukuba.ac.jp
** University of Tsukuba, Institute of Art & Design
Japan, tyam@geijutsu.tsukuba.ac.jp

Abstract: When a researcher doesn't have clear hypothesis, the hypothesis exploration from different point of views is necessary in order to process the research smoothly. Classification is a popular support tool for hypothesis exploration. When there is a great quantity of objects, for which nature is hard to specify, classifying them is a useful method to organize information easily. For example, classification could be done by similarity or dissimilarity evaluation. By considering such a classification, the researcher can arrange his/her own ideas and then visually get the relations. The result is convenient since it gives hints to consider hypothesis and to clarify one's idea.

We introduce an application that supports hypothesis exploration. The purpose of this study is to examine demerits included in a process of hypothesis exploration and examine functions necessary for the application.

Classification can be done by manually. But Multi Dimensional Scaling (MDS) or cluster analysis can also perform it. If the degree of similarity or dissimilarity judgment is measured using stage evaluation between objects, a similarity (dissimilarity) matrix can be created. Then, this matrix can be used as data for MDS or cluster analysis. This method classifies the objects more objectively than by manually, since manual classification can be greatly influenced by the researcher's subjectively. The scatter diagram or dendrogram obtained by these analyses has also the merit of determining the relation between objects, as well as manual classification does.

However, whether a manual classification either a classification by degree of similarity (dissimilarity) is done, there is no value if the researcher doesn't have insight about which criteria contribute to the classification result. Classification is not the final objective: it is also required to understand the output extractable from the classification.

This point brings us to another issue: once classification is finished, it ends here, and the researcher won't investigate once again on objects' relations. If this investigation can be done with ease, which is the purpose of the application introduced in this paper, the researcher can have the opportunity to reconsider many time the relations by intuitive operation without recreating again the similarity (dissimilarity) matrix and graphs. For example, in the case of divergent thinking

approach, the possibility of multi-viewpoint interpretation, by exploring hypothesis or arranging one's own information, improves the quality of global interpretation. Therefore, we expect our application to contribute hypothesis exploration.

Key words: classification, reconsideration, exploration, MDS, cluster analysis

1. Introduction

To organize ideas gained by brain storming or analyze qualitative data, you must make a hypothesis. If you have no theoretical or empirical grounds, the hypothesis exploration from different point of views is necessary in order to process the research smoothly. Because of making working hypotheses and comparing them is useful to make good hypothesis. Classification is a support tool for hypothesis exploration. When there is a great quantity of objects, for which nature is hard to specify, classifying them is a useful method to organize information easily. For example, classification could be done by similarity or dissimilarity evaluation. By considering such a classification, a researcher can arrange his/her own ideas and then visually get the relations. The result is convenient since it gives hints to consider hypothesis and to clarify one's idea. Although this procedure is useful and frequently-used, it has demerit to disturb the hypothesis exploration from different point of views. And this manual classification can be greatly influenced by the researcher's subjectively. This paper introduces an application that aims the hypothesis exploration. Therefore, firstly we point out demerit of frequently-used procedure. Next, we suggest a solution to improve demerit. Then, a brief analysis of functions necessary for our application shows and realizes these functions by program language.

2. Problematic

2.1 The problem of frequently-used classification

Although how to express ideas is various, usually, one idea is regarded as one item as the expression ideas. Once classification is finished, it ends here, and you may not investigate once again on objects' relations. Because you are satisfied with classification and you may think it's a pain to rethink a relation between items and to do classification again. But, in a context of hypothesis exploration, it is desirable to try various possibilities on hypothesis exploring. Therefore, the environment of multi-viewpoint interpretation, by exploring hypothesis or arranging one's own information will improve the quality of global interpretation and this environment is needed.

And manual classification is greatly influenced by researcher's subjectively since it is difficult for researcher to consider relation between each item minutely. So, researcher considers relation among items roughly and classification is influenced by subjectively.

2.2 Kansei information approach

We try to above problem from Kansei information approach. Although the definition of Kansei is various among researchers, in this paper, we think Kansei has a function to evaluate with value judgment. And we regard Kansei information as a result of evaluation. For example, scales used in questionnaire or Semantic Differential method (SD method) [1] to ask subjects evaluation for objects are one example of Kansei information. One feature of Kansei information is as following. That is, the factors related to evaluation are difficult to be conscious by subjects. Subjects can evaluate objects by using scales, but they can not always be able to

conscious of factors related to evaluation. Although subjects can not explain clear reason why he/she think/feel so, he/she can evaluate. And this evaluation has consistency within him/her. Now, we mention that how to use Kansei information in our application. In this paper, Kansei information is relation among items expressed by similarity (dissimilarity). Similarity (dissimilarity) is used to judge the relation among items and organize items. First, we use Kansei information to improve subjectivity of classification.

2.3 Improvement in subjectivity of classification

Classification can be done by manually. But Multi Dimensional Scaling (MDS) or cluster analysis can also perform it. If the degree of similarity or dissimilarity judgment is measured using stage evaluation between items, a similarity (dissimilarity) matrix can be created. Then, this matrix can be used as data for MDS or cluster analysis. This method classifies the objects more objectively than by manually. If you classify by manually, you find the difficulty to think the relation among all items equally. As the result, you may be able to consider the relation among some parts. If you use MDS or cluster analysis, you don't have to worry above problem. And the scatter diagram or dendrogram obtained by these analyses has also the merit to consider the relation between items, as well as manual classification does. Improvement in subjectivity of classification is led by Kansei information (similarity (dissimilarity) between items). Next we use Kansei information to aim hypothesis exploration from different point of views

2.4 To aim hypothesis exploration from different point of views

In manual classification, it is difficult to explore hypothesis from different point of views. Once classification is finished, researcher may not rethink since he/she may feel it's a pain to rethink a relation between items and to do classification again. So, we aim this problem by our developed application. Concretely, our application aims to make environment that user don't think it's too much of a bother to rethink. If the relation between items changes, then the result of scatter diagram or dendrogram also changes. This change may help person to think from another viewpoint. And if change is conducted many times, it contributes to make working hypothesis and compare them. As a result, arranging ideas and exploring hypothesis may be encouraged. These advantage is led by Kansei information (similarity (dissimilarity) between items). But, it is trouble to return the statistics software to reanalysis to change the relation between items. And returning the statistics software to re-judgment disturbs person's thinking. It is inefficient. Therefore, we make application which can change the relation between items with easy operation and reconstruct the scatter diagram or dendrogram automatically. Next, we examine functions necessary for our application and mention how realizes these functions.

3. Functions of our application

The functions necessary for our application are mainly divided into three points.

a. Making a distance matrix from similarity (dissimilarity) judgment between items
b. Conducting MDS or cluster analysis from a distance matrix and making scatter diagram or dendrogram.
c. Changing the result of the optional part of the distance matrix where user wants to change with easy operation. And then recalculating the result of the changed distance matrix and remaking scatter diagram or dendrogram.

We examine the necessary things to realize each function.

3.1 Making a distance matrix from similarity (dissimilarity) judgment between items

When similarity (dissimilarity) judgment between items is done, it is attended that items to be pairs must be randomly showed. For example, there is five items, A, B, C, D and E. If the order of the judgment is A-B, A-C, A-D, then this order is not good. Because if judgment between items is not randomly, there is a danger that user may refer the former result to conduct the present judgment. To avoid the order effect, it is desirable the order of judgment between items is randomly. Next, we must decide the stage evaluation of similarity (dissimilarity) judgment between items. Although there is no rule that judgment must be certain stage evaluation, we decided to use seven stages. Because we think seven stages reflect the sensitive differences. The judgment is done by alphabet, A to G (1 to 7). If Arabic numerals are used, user is conscious of distance that former judged. So, we use alphabet, not Arabic numerals. If similarity judgment is done, “A” stands for very different, “D” stands for neutral and “G” stands for very close. And if dissimilarity is done, that order becomes the reverse. Next, we mention distance matrix. Distance matrix obtained this application is symmetry matrix. That is the value of A-B and B-A is the same. So, only one pair of the same pairs is needed as the data of distance matrix. For example, there is A-B as pair. If the data of A-B, the value of B-A becomes the value of A-B automatically. And don’t ask user the value of B-A. And although the pairs of diagonal line are not asked, the value becomes 8(similarity matrix) or 0(dissimilarity matrix). We must realize these processing on application.

3.2 Conducting MDS or cluster analysis from a distance matrix and making scatter diagram or dendrogram

Cluster analysis is conducted from distance matrix, but these analyses have many kinds of calculation. So, we must decide to choose which kind of calculation. For example, cluster analysis has many kinds of calculation by choosing the distance between items and how to calculate the distance when clusters are combined. But, we think it is better that calculation way is limited to one than we prepare many kinds of calculation. Because we think user may prefer simple. Therefore, we decided Euclidean distance as distance between item and ward method as calculation method between the distances of clusters when clusters are combined. It is said that ward method often leads meaningful results than other methods [2], so we think this method is proper and decided it. MDS has many kinds of calculation as cluster analysis. But we think simple is better and we decided to use Hayashi's Quantification IV that is regarded as one of MDS. This method is calculated by similarity matrix between items and eigenvectors is gained as the result. To interpret these eigenvectors, researcher can infer the relation between items. Only the first to the second or to the third eigenvectors is used in interpreting. And to draw the scatter diagram using these eigenvectors, researcher can grasp the relation between items. Either analysis is used, to save the user’s trouble, scatter diagram or dendrogram which is gained as the result of the analysis is made automatically by application.

3.3 Changing the result of the distance matrix where user wants to change and recalculating the result of the changed distance matrix

This function is main purpose of application to support hypothesis exploration. User can change the result of the optional part of the distance matrix where user wants to do with easy operation. And then result of the

change of the distance matrix reflects scatter diagram or dendrogram automatically. Application must be able to conduct these functions.

3.4 Program language

We used Visual Basic by Microsoft Corporation to develop our application.

4. Realization by application

We introduce how to realize functions which are mentioned 3.1 to 3.3 are realized by our application. As an example, it is assumed that there are five items.

4.1 Realization of the function making the distance matrix from similarity (dissimilarity) judgment between items

User input each item and then similarity (dissimilarity) judgment between items is asked (See Fig.1 and Fig.2). Because items are five, the number of judgment is five times. Because the parts of diagonal line and the same pairs (e.g. If there is A-B's data, B-A's data is not asked) are excluded (5×5－5)/2= 5). Judgment is conducted by click the button "A" to "G". And although in figure, the results of former judgment are showed in order to explain, actually the former results are hidden in order not to influence the present judgment when user conducts next judgment. And the order of similarity (dissimilarity) judgment is randomly by using random numbers.

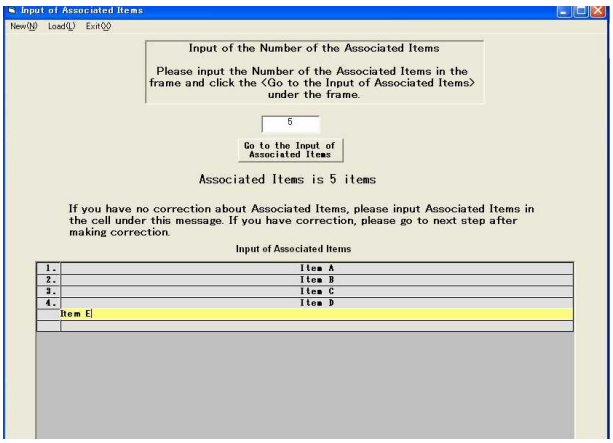


Fig.1 Data Input

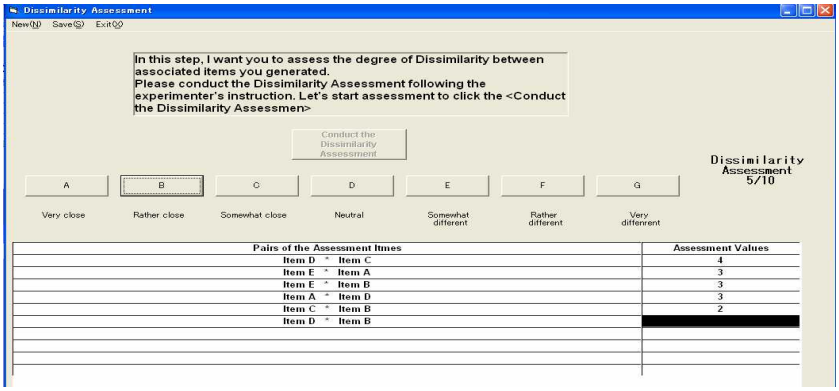


Fig.2 Similarity (dissimilarity) judgment

4.2 Realization of the function conducting MDS or cluster analysis and drawing diagram automatically

Application conducts MDS or cluster analysis from distance matrix as the result of the similarity (dissimilarity) matrix, and draws scatter diagram or dendrogram automatically. If similarity judgment is done, MDS is conducted and dissimilarity judgment is done, cluster analysis is conducted. Drawing in conducting MDS, scatter diagram is drawn by the pair of the first eigenvectors and the second eigenvectors. In this, we show the result of cluster analysis (See Fig.3). To save the user's trouble, application draws dendrogram and shows each item in dedrogram. Scatter diagram as the result of MDS also shows each item.

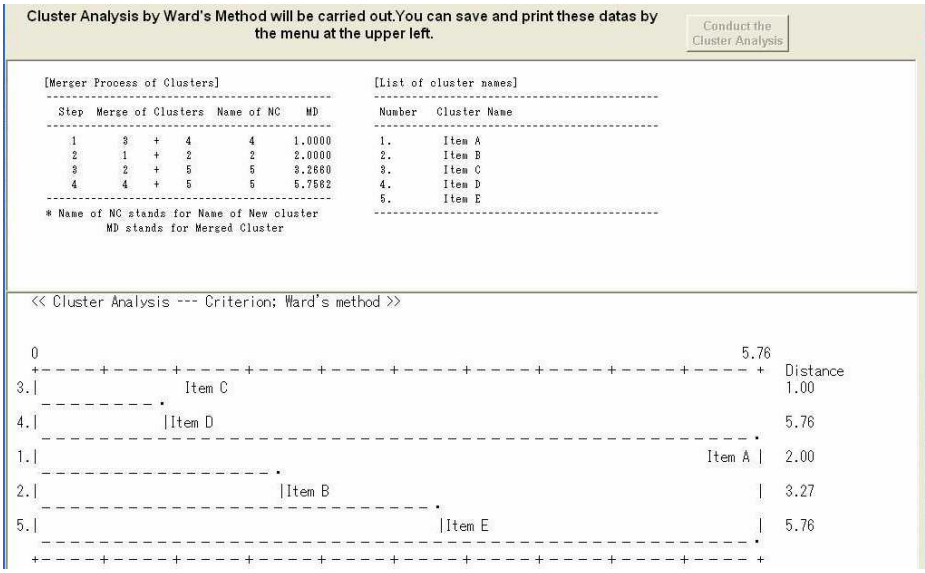
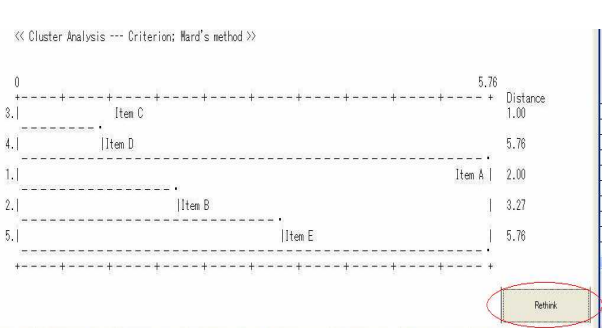


Fig.3 The result of cluster analysis

4.3 Realization of the function changing the result of the distance matrix and recalculating

If ''Rethink" button which is placed under a dendrogram is clicked, the screen which is needed to change the result of the distance matrix is shown (See Fig.4). If user wants to change the similarity (dissimilarity) judgment, user inputs text box on this screen the number of items which user want to do, and click "OK" button next to text box. In this, the relation between 3 and 4(Item C and Item D) is selected as an example.



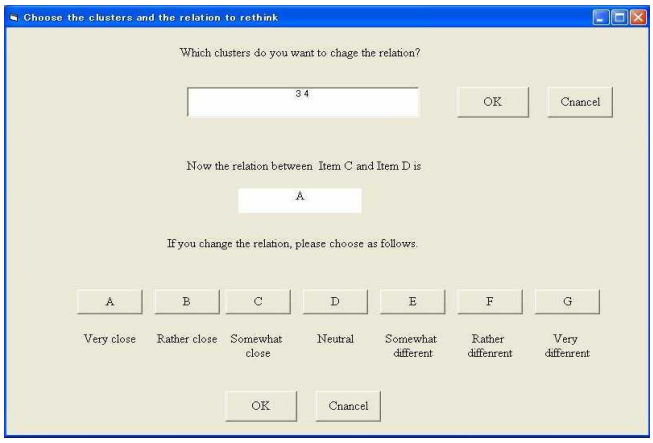


Fig.4 changing the result of the distance matrix

And then, the result of similarity (dissimilarity) judgment between selected items is shown. If user wants to change, to select the button “A” to “G”, user can rethink the judgment. And if “OK” button is clicked, application recalculates MDS or cluster analysis and redraws scatter diagram or dendrogram (See Fig.5).

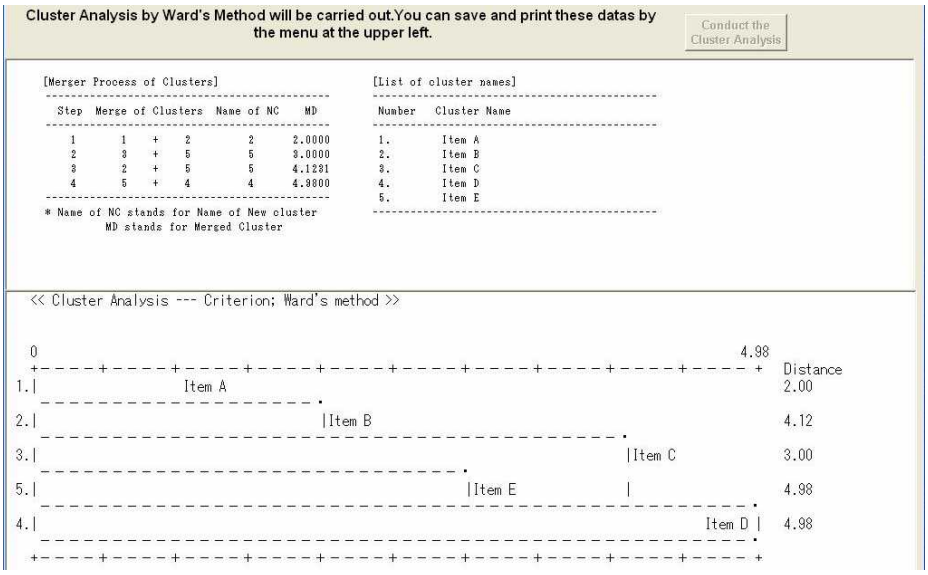


Fig.5 Result of the recalculation

*Since the relation between Item C and Item D is changed, the result of dendrogram is different from Fig.3 and Fig.4.

5. Conclusions

In this study, we point out the demerits that researcher may encounter when he/she explore hypothesis by using classification. That is, to disturb the hypothesis exploration from different point of views and manual classification can be greatly influenced by the researcher’s subjectively. To solve this problem, we tried to develop application. To make this, we examine the functions necessary for application and developed it by Visual Basic. But we have not conducted evaluation experiment. Therefore, we can not write clear opinion if our application is efficient for hypothesis exploring or not. When we will present our study, we will report a result of evaluation experiment.

Acknowledgment

This study supported by 21st Century COE Program, sponsored by the NEXT.

References

[1]Osgood, C.E., 1952, The nature and measurement of meaning, Psychological Bulletin, vol. 49(3), pp.197-237.

[2]Marui, E., 1986, Cluster analysis, In Yanai,H. and Takagi, H.(Eds.), Handbook of multivariate analysis(pp. 224-242), Gendai-Sugakusha, Japan.