Sakyo & Ukyo: Visualization of Clustered Matrix Data Applying Dual Hierarchical Data Visualization Technique

Haruho Tachibana Takayuki Itoh
Graduate School of Humanitics and Sciences, Ochanomizu University
{haruho, itot}@itolab.is.ocha.ac.jp

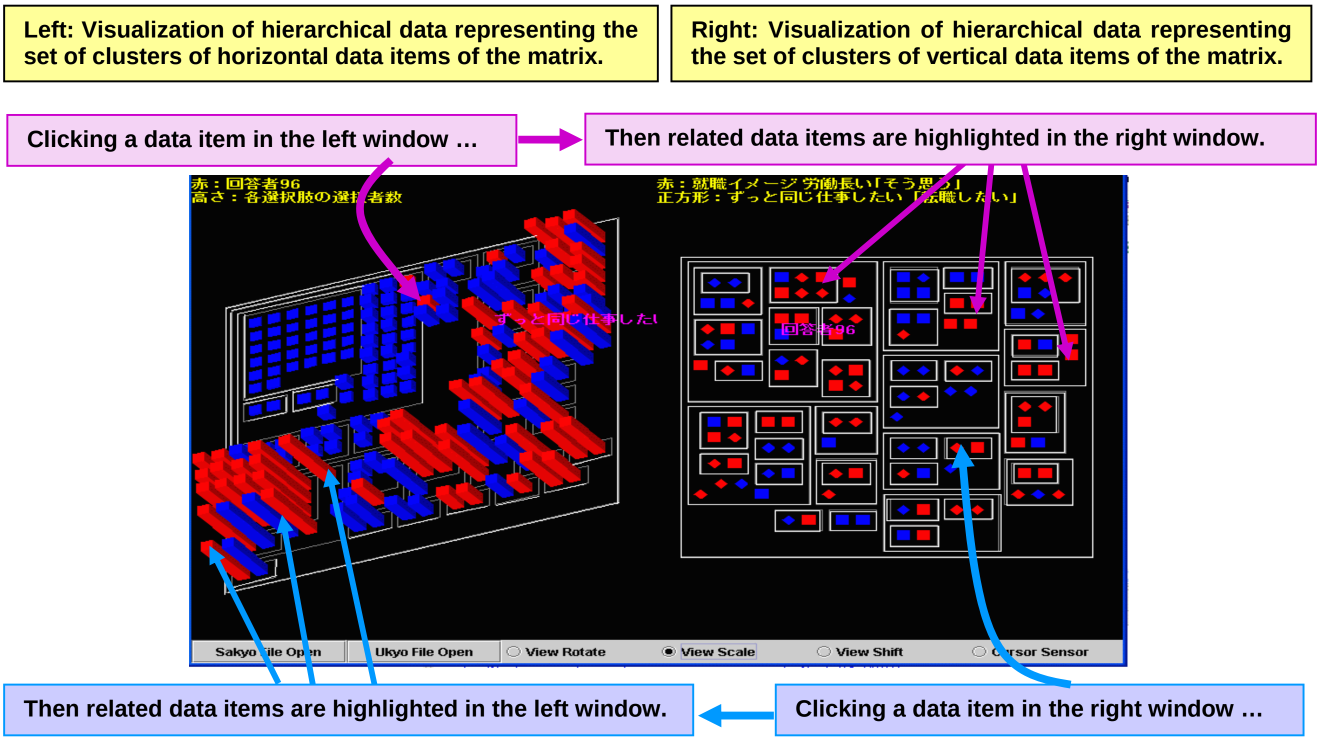

Figure 1. Overview of the presented technique. The technique first generates clusters of horizontal and vertical data items of matrix data. It then visualizes the two sets of clusters applying dual hierarchical data visualization technique. The dual visualization can be interacted each other.

Abstract

Clustered matrix data is very popular data structure, and therefore various visualization techniques for such data have been already presented. However, some of the data in our daily life or work are very sparse, and therefore the usage of display spaces is not always reasonable while using existing table or matrix data visualization techniques. Many past visualization studies have attempted to save the display space usage by converting the table or matrix data into hierarchical or graph data.

This report presents a visualization technique matrix data applying dual hierarchical data visualization technique. The presented technique first applies hierarchical clustering for both horizontal and vertical data items of the matrix data. It then visualizes the two sets of clusters applying dual hierarchical data visualization components. We apply HeiankyoView, a large-scale hierarchical data visualization technique, for the visualization of the two sets of clusters. This report introduces an example applying the matrix data of a questionnaire result.

Keywords: Hierarchical data, HeiankyoView, Clustered Matrix Data, Rectangle Packing, Questionnaire.

1. Introduction

Matrix data is a very common data in our daily life, and therefore matrix data visualization technique is an active research topic. Recent typical popular matrix data is microarray data, which collect expression values of genes. Often they are clustered according to relativity among genes or microarrays. Microarray visualization technique is also an active research topic [Sar04].

In many cases matrix data in our daily life is very sparse, and it is not always reasonable if we represent such sparse data using matrix-oriented visualization techniques. For example, if we have n×m matrix and values of 90% of the data items are zero, we may feel the matrix-oriented representation looks redundant. Many past visualization studies have attempted to save the display space usage by converting such sparse matrix data into node data such as tree or graph. It can drastically save the display space usage in many cases, because it represents only n or m data items. Matrix-oriented and node-oriented visualization techniques have their own bottlenecks, and some technical papers compare the readability of visualization results between the techniques [Gho04].

This paper presents a new node-oriented visualization technique for such sparse matrix data. Figure 1 shows an example of visualization result using the presented technique, and Figure 2 denotes the processing flow of the presented technique. The technique first generates clusters of horizontal and vertical data items of the matrix data. It then generates two hierarchical data from the clusters, and visualizes those using dual hierarchical data visualization technique. We applied "HeiankyoView", a large-scale hierarchical data visualization technique, for the dual visualization of the two hierarchical data. In our implementation, the visualization components can interact with the other component, so that users can explore horizontal and vertical data items alternately.

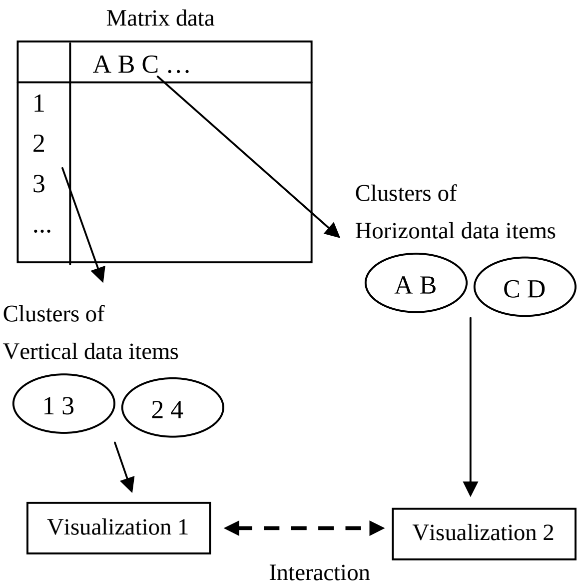

Figure 2. Processing flow of the presented technique.

The paper introduces an example of visualization result using the technique. We collected a questionnaire data and converted the data into a matrix data. We discovered some interesting trend in the questionnaire data using the presented visualization technique.

2. HeiankyoView: A Large-scale Hierarchical Data Visualization Technique

The presented technique applies a hierarchical data visualization technique "HeiankyoView". Figure 3 is an example of the visualization by HeiankyoView, which represents leaf-nodes as black square icons, and branch-nodes as rectangular borders enclosing the icons.

The visualization technique places thousands of leaf-nodes into one display space while satisfying the following conditions:

- It never overlaps the leaf-nodes and branch-nodes in a single hierarchy of other nodes,
- It attempts to minimize the display area requirement and draws all leaf-nodes by equally shaped and sized icons.

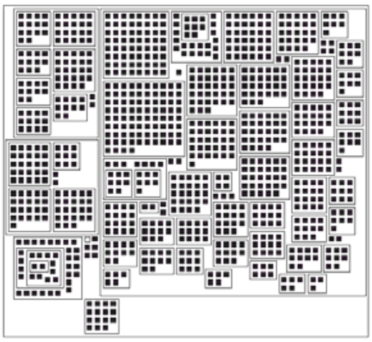

Figure 3. Example of hierarchical data visualization using a rectangle packing algorithm.

This style is suitable to equally visualize thousands of leaf-nodes of hierarchical data in one display space.

The technique first packs icons, and then encloses them in rectangular borders. Similarly, it packs a set of rectangles that belong to higher levels, and generates the larger rectangles that enclose them. Repeating the process from the lowest level toward the highest level, the technique places all of the data onto the layout area. The packing algorithm for icons and rectangles is the key technology for the visualization technique. A triangular-mesh-based rectangle packing algorithm has been first presented in [Yam03][Ito04] for the hierarchical data visualization, and then an improved rectangle packing algorithm is applied to HeiankyoView, as described in [Ito06a][Ito06b].

3. Sakyo & Ukyo: A Matrix Data Visualization Technique Applying Dual HeiankyoView

The paper presents a new technique to visualize matrix data by applying dual HeiankyoView. It applies HeiankyoView as "Visualization 1" and "Visualization 2" in Figure 2.

This section describes detailed implementation of the technique, called "Sakyo & Ukyo". The technique supposes to display two HeiankyoView in one display, looking like left- and right- side of the map of Heiankyo, an ancient palace in Kyoto-city. That is why we call the new technique "Sakyo & Ukyo".

3.1 Clustering of horizontal and vertical data items

Let us describe the data items of matrix data as follows: horizontal data items h_1 to h_m, vertical data items v_1 to v_n, and values a_{11} to a_{nm}, as shown in Figure 4. Here the technique treats horizontal data items $H=(h_1, \dots, h_m)$ as n-dimensional vectors, such as $h_i = (a_{1i}, \dots, a_{ni})$. Similarly, the technique treats vertical data items $V=(v_1, \dots, v_n)$ as m-dimensional vectors, such as $v_j = (a_{j1}, \dots, a_{jm})$.

The technique generates clusters of horizontal data items by the following procedure. It first calculates Euclidian distances between every possible pairs of data items. Let the distance between h_i and h_j as d_{ij}, and the maximum distance of the all distances as D_{max}. Here the technique calculates the relativity value r_{ij} for d_i and d_j by the following equation:

$$r_{ij} = 1.0 - \frac{d_{ij}}{D_{\max}}$$

Here 0 is the minimum value of r_{ij}, and 1 is the maximum value of r_{ij}, that denotes d_i and d_j are entirely same vectors.

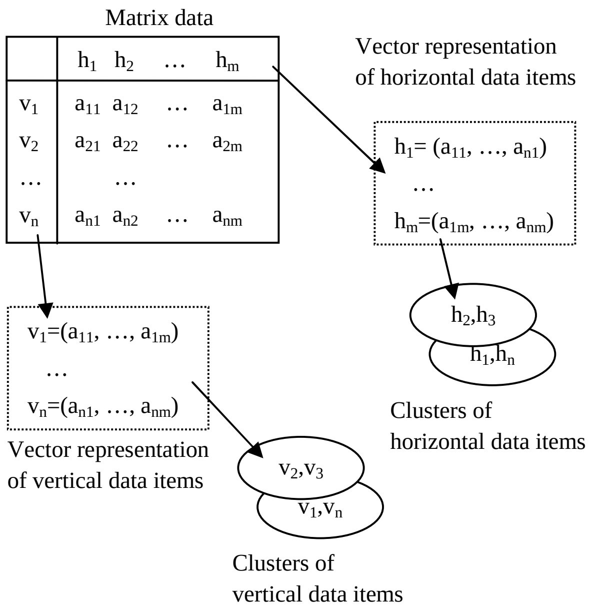

Figure 4. Clustering of horizontal and vertical data items of matrix data.

The technique then generates clusters of horizontal data items, applying hierarchical clustering method, where non-hierarchical clustering methods (i.e. Self organizing map, k-means method) can be also applied. Our implementation simply generates combination of data items by iteratively coupling data items or groups of data items according to their relativity values, and then generates nested clusters by grouping the data items according user-defined threshold values.

Figure 5 denotes the processing flow of hierarchical clustering and illustration of the nested clusters visualized by HeiankyoView. The clustering process couples data items or groups of data items, where the couples have the largest r_{ij} values at each coupling process. After generating a tree of couples as shown in Figure 5, our implementation applies several threshold values to generate nested clusters. S_1 and S_2 in Figure 5 are the examples of threshold values. Our implementation first generates smaller clusters that consist of data items whose r_{ij} values are higher than S_1. It then generates larger clusters that consist of data items and smaller clusters whose r_{ij} values are higher than S_2. Illustration in Figure 5 denotes that the process generates two-level clusters by applying two-level threshold values.

Above processes are also applied to vertical data items as well as horizontal data items. Finally the technique generates two sets of nested clusters, and they can be treated as two hierarchical data.

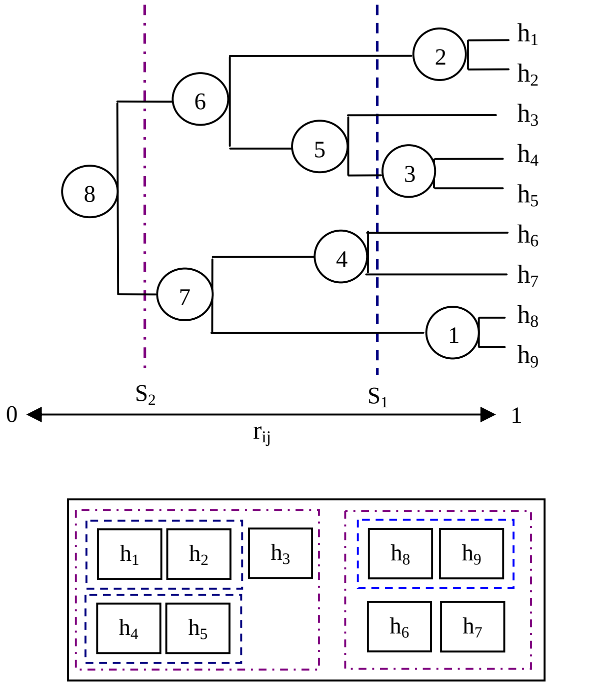

Figure 5. (Upper) Hierarchical clustering. (Lower) Illustration of the clusters visualized by HeiankyoView.

3.2 Visualization of clusters by HeiankyoView

The technique then visualizes the two hierarchical data using two HeiankyoView. Let HeiankyoView for vertical data items as Sakyo, and HeiankyoView for horizontal data items as Ukyo. Therefore, Sakyo visualizes n data items v_1 to v_n, and Ukyo visualizes m data items h_1 to h_m. They represent the data items as three dimensional bar charts, where colors, shapes, and heights of the bars vary based on application-oriented customization.

3.3 Interaction between two HeiankyoView

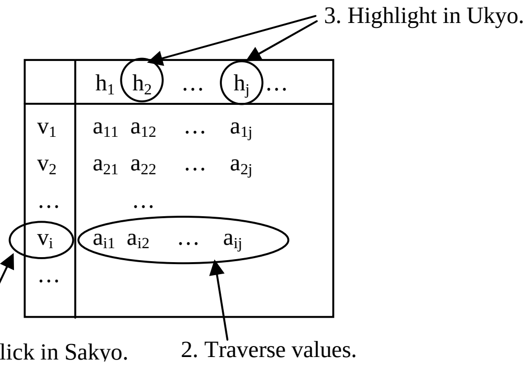

Figure 6. Extraction of horizontal data items clicking a vertical data item.

Sakyo and Ukyo have capability to interact each other, so that users can interactively explore the data items. When users click a data item in Sakyo, then the technique highlights data items in Ukyo, which are related to the clicked data item in Sakyo. Similarly, when users click a data item in Ukyo, then the technique highlights data items in Sakyo, which are related to the clicked data item in Ukyo. When a user clicks v_i in Sakyo, the technique traverses values a_{i1} to a_{im}, and then extracts a horizontal data item h_j, if the value a_{ij} satisfies user-defined conditions. It finally highlights the set of extracted horizontal data items in Ukyo. The above processing flow is shown in Figure 6.

Similarly, when a user clicks h_j in Ukyo, the technique extracts the set of vertical data items, and highlights them in Sakyo.

4. Experiments with Questionnaire Data

This section shows an example of our experiments using the technique. We applied questionnaire data as a sample matrix data. Our experiment supposed that the questionnaire included n choices, and answers were collected from m answerers. As shown in Figure 7, our questionnaire supposed all questions were to be answered by selecting one or more items. We created a matrix data from the collection of answers, which included m horizontal data items and n vertical data items, and values were “1” if an answerer selected the choice, and “0” if he/she did not select it.

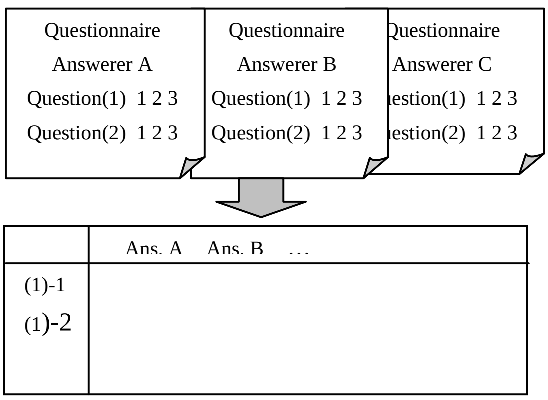

Figure 7. Questionnaire and matrix data.

We prepared a questionnaire about job interview. Answerers were 118 female university students, belonging to department of information sciences, whose grades were B1 to M2. The questionnaire included 31 questions and 179 choices. Therefore the matrix data created from the collection of the answer includes 118 horizontal data items and 179 vertical data items. Figure 1 shows an example of visualization result using the questionnaire data.

As described in Section 3.2, visual attributes of nodes (i.e. shapes, colors, and heights) can be controlled according to application-oriented semantics in our implementation. In this section we show Sakyo and Ukyo customized for questionnaire visualization. In our experiments Ukyo denoted 118 answerers, and Sakyo denoted 179 choices. Since Sakyo represented clusters of choices, highly related choices were closely displayed in Sakyo. Similarly, since Ukyo represented clusters of answerers, highly related answerers were closely displayed in Ukyo.

We modified interaction mechanism between Sakyo and Ukyo to customize for questionnaire data visualization, as follows:

- Heights of icons in Sakyo denote the number of answerers that selected the corresponding choices.
- When a cursor points one of the icons of choices, Sakyo indicates the meaning of the corresponding choice.
- When a user clicks an icon of choice in Sakyo, Ukyo then represents icons of answerers who selected the choice as red, and other icons as blue. When a user clicks another icon of choice in Sakyo, Ukyo then represents icons of answerers who selected the choice as rhombus, and other icons as square.
- When a cursor points to one of the icons of answerers, Ukyo indicates the name of the corresponding answerer.
- When a user clicks an icon of answerer in Ukyo, Sakyo then represents icons of choices which the answerer selected as red, and other icons as blue.

Figure 8 shows visualization result of Ukyo. In the figure, shapes of icons denote the answer for the question “Do you respect suggestions of your parents for your job interviews?”, where square icons denote answerers selected “yes”, and rhombus icons denote answerers selected “no”. In this figure colors of icons denote the answer for the question “Do you live with your parents?”, where red icons denote answerers selected “yes”, and blue icons denote answerers selected “no”. The figure denotes that rhombus icons got closer in the display, so it seems that the first question highly contributed to the clustering result. Similarly, blue icons got closer, so did red icons. It also seems that the second question highly contributed to the clustering result. However, the figure denotes that blue and red rhombus icons are mixed in small figures. Before this experiment we imagined that the two questions are highly related and many students living with their parents have more respects to suggestions of their parents for their job interviews. However, the visualization result seems that the two questions are not highly related. It was an unexpected discovery for us.

Figure 9 shows a visualization result of Sakyo. In the figure heights of icons denote the number of answerers that selected the corresponding choices. Red icons denote choices that an answerer selected, and blue icons denoted choices that she did not select. Here we made an attention to a small cluster (denoted by a blue circle in Figure 9) that contains two choices selected by the answerer. The two choices were as follows:

- “Yes” for the question “Did you have part time jobs with the conscious to your job interview”

- “Yes” for the question “Do you feel that your job interview is fun?”

We imagined that students who had part time jobs with the conscious to their job interviews have already used to think of their job interviews, and finally the students felt their job interviews fun. We did not think of such relationship between the two questions, so it was also an unexpected discovery for us.

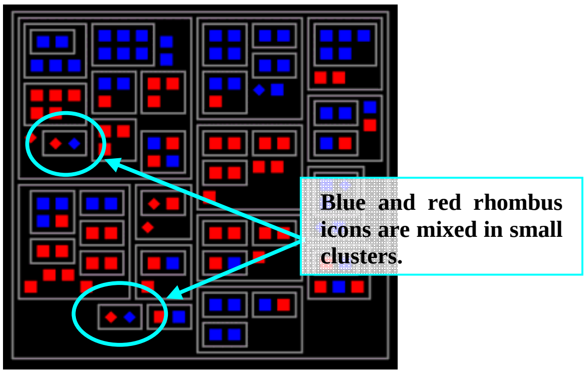

Figure 8. Result (Ukyo). Ukyo visualizes relationships between job interview activity of answerers and their parents.

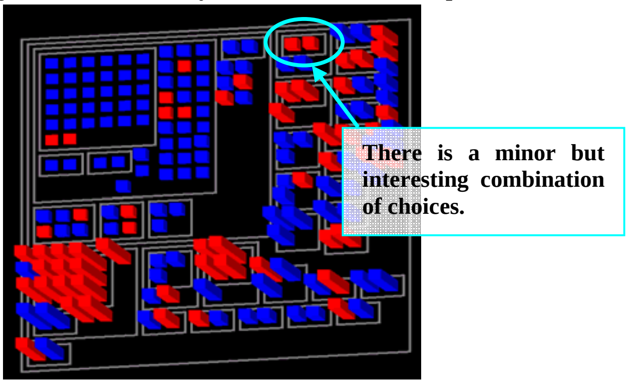

Figure 9. Result (Sakyo). Sakyo visualizes relationships between part time jobs of answerers and their impressions of job interview activity.

5. Related Work

5.1 Matrix data visualization

Visualization of matrix data is an active research topic. Some excellent works for visualization of tables, matrices, and spreadsheets, focused on exploration and navigation of the data [Rao94][Rao95][Chi99].

Matrix data visualization can be often applied as an alternative of graph or network data visualization [Bec95]. This approach often improves occlusion problems of graph/network visualization techniques. However, display usage of the approach is not very efficient if the input data is not dense. Readability the visualization techniques has been compared in [Gho04].

Microarray data visualization is a new, active research topic, which matrix data visualization techniques are often applied [Eis98] [Seo02]. Again, display usage of the approach is not very efficient if the input data is not dense. Many of this approach use horizontal and vertical axes of the matrix to assign microarrays and genes, and apply hierarchical clustering to reorder the microarrays and genes. Sakyo & Ukyo can be a good alternative of matrix-based microarray data visualization techniques.

5.2 Hierarchical data visualization

5.2.1 Nest-based techniques

HeiankyoView represents hierarchy of input data by nested rectangles. This representation is suitable for representing all lower data items in one display.

Against HeiankyoView places data items and nested metaphors in two-dimensional spaces, some three-dimensional nest-based visualization techniques have been also presented, including Information Cube [Rek93] and H-BLOB [Spr00]. These techniques require capability and skill of 3D graphics, and adjustment of semi-transparency may be difficult for large-scale data visualization.

5.2.2 Space-filling techniques

Space-filling visualization technique is the other approach to represent all lower-level data items in one display space as well as HeiankyoView. The technique subdivides display spaces to represent each portion of hierarchical data. TreeMaps [Joh91] recursively subdivides the display spaces into rectangular regions to form nested bar charts. HeiankyoView is analogous to TreeMaps because both techniques subdivide display spaces into rectangular area.

Variation improved TreeMaps have been recently proposed. Squarified Treemap [Bru00] subdivides display spaces into rectangles as much as square. Ordered Treemap [Shn01] subdivides display spaces and assigns the subregions in the predefined order of leaf-nodes. Quantum Treemap [Bed02] applies modified Squarified or Ordered Treemap so that it can represent leaf-nodes that are equally shaped and sized. Target of the Quantum Treemap is very similar to HeiankyoView, and actually Quantum Treemap can be an alternative of HeiankyoView for the purpose of this paper. Experiments described in [Ito04] discusses trade-offs between Quantum Treemap and HeiankyoView.

5.2.3 Other hierarchical data visualization techniques

Tree visualization techniques are other well-known hierarchical data visualization techniques. Hyperbolic Tree [Lam96], Cone Tree [Car95], and Fractal Views [Koi95] provides navigation and exploration capabilities for large-scale hierarchical data. We did not apply such approach because our target is all-in-one visualization of lower-level items of hierarchical data, rather than interactive navigation and exploration of the data starting from the top-level.

6. Conclusion and Future Works

This paper presented a new technique for visualizing matrix data, applying dual hierarchical data visualization technique. The technique first clusters horizontal and vertical data items of the matrix data, and then visualizes the two sets of clusters using HeiankyoView. The two HeiankyoView can be interacted each other, so that users can interactively explore the data items of the matrix data.

We call the name of the technique “Sakyo & Ukyo”, since Heiankyo is an ancient palace whose blocks are always orthogonally aligns, and “Sakyo” and “Ukyo” are right- and left-sides of Heiankyo.

The following are our future works on this technique:

- Experiments with larger questionnaire data.
- Experiments of various data in addition to questionnaire, such as matrix of documents and keywords.
- Usability tests.
- Optimal configuration of clustering process for the presented visualization technique.

Acknowledgements

We appreciate Dr. Michael de Hoon in Colombia University, for providing and advising clustering software. We also appreciate Prof. Michiko Mifune, Prof. Keiko Hatae, Prof. Nobuko Nagase, Prof. Midori Kasai, and Ms. Junko Kozono, in Ochanomizu University, for advising about questionnaire productions. We appreciate 118 students in Ochanomizu University for answering our questionnaire. This work has been partially supported by Japan Society of the Promotion of Science under Grant-in-Aid for Scientific Research (C) No. 18500074.

References

[Bec95] Becker R. A. Eick S. G., Wilks A. R., Visualizing Network Data, IEEE Transactions on Visualization and Computer Graphics, Vol. 1, No.1, pp. 16-28, 1995.

[Bed02] Bederson B., Schneiderman B., Ordered and Quantum Treemaps: Making Effective Use of 2D Space to Display Hierarchies, ACM Transactions on Graphics, Vol. 21, No. 4, pp. 833-854, 2002.

[Bru00] Bruls D.M., Huizing K., Wijk J. J., Squarified Treemaps, Proceedings of Data Visualization 2000, pp. 33-42, 2000.

[Car95] Carriere J., Kazman R.., Research Paper: Interacting with Huge Hierarchies beyond Cone Trees, IEEE Information Visualization 95, pp.74-81, 1995.

[Chi99] Chi E. H., A Framework for Information Visualization Spreadsheets, Ph.D. Thesis, University of Minnesota, 1999.

[Eis98] Eisen M., Spellman P., Brown P., Boststein D., Cluster Analysis and Display of Genome-wide Expression Patterns, Vol. 95, Issue 25, pp. 14963-14968, 1998.

[Gho04] Ghoniem M., Fekete J., Castagiloia P., A Comparison of the Readability of Graphs Using Node-Link and Matrix-Based Representations, IEEE Information Visualization 2004, pp. 17-24, 2004.

[Ito04] Itoh T., Yamaguchi Y., Ikehata Y., Kajinaga Y., Hierarchical Data Visualization Using a Fast Rectangle-Packing Algorithm, IEEE Transactions on Visualization and Computer Graphics, Vol. 10, No. 3, pp. 302-313, 2004.

[Ito06a] Itoh T., Takakura H., Sawada A., Koyamada K., Hierarchical Visualization of Network Intrusion Detection Data in the IP Address Space, IEEE Computer Graphics and Applications, Vol. 36, No. 2, 2006.

[Ito06b] Itoh T., Yamashita F., Visualization of Multi-dimensional Data of Bioactive Chemicals Using a Hierarchical Data Visualization Technique "HeiankyoView", Asia Pacific Symposium on Information Visualization (APVIS) 2006, pp. 23-29, 2006.

[Joh91] Johnson B., et al., Tree-Maps: A Space Filling Approach to the Visualization of Hierarchical Information Space, IEEE Visualization '91, pp. 275-282, 1991.

[Koi95] Koike H., Fractal Views: A Fractal-Based Method for Controlling Information Display, ACM Transactions on Information Systems, Vol. 13, No. 3, pp.305-323, 1995.

[Lam96] Lamping J., Rao R., The Hyperbolic Browser: A Focus+context Technique for Visualizing Large Hierarchies, Journal of Visual Languages and Computing, Vol. 7, No. 1, pp. 33-55, 1996.

[Rao94] Rao R., Card S. K.., The Table Lens: Merging graphical and symbolic representations in an interactive focus + context visualization for tabular information, Computing Systems (CHI'94), pp. 318-322, 1994.

[Rao95] Rao, R. and Card, S. K., Exploring large tables with the table lens, in Proceedings of ACM Conference on Human Factors in Computing Systems (CHI'95), pp. 403-404, 1995.

[Rek93] Rekimoto J., The Information Cube: Using Transparency in 3D Information Visualization, Third Annual Workshop on Information Technologies & Systems, pp.125-132, 1993.

[Sar04] An Evaluation of Microarray Visualization Tools for Biological Insight, IEEE Information Visualization 2004, pp. 1-8, 2004.

[Seo02]Seo J., Shneiderman B., Interactively Exploring Hierarchical Clustering Results, it IEEE Computer, Vol. 35, pp. 80-86, 2002.

[Shn01] Shneiderman B., Wattenberg M., Ordered Treemap Layouts, IEEE Information Visualization Symposium 2001, pp. 73-78, 2001.

[Spr00] Sprenger T. C., Brunella R., Gross M. H., H-BLOB: A Hierarchical Visual Clustering Method Using Implicit Surfaces, IEEE Visualization 2000, pp. 61-68, 2000.

[Yam03] Y. Yamaguchi, T. Itoh, Visualization of Distributed Processes Using "Data Jewelry Box" Algorithm, CG International 2003, pp. 162-169, 2003.