Literature Survey: The Reorderable Matrix

Marco Garcia, Weinan Huang, Christin Seifert & Wolfgang Wallisch

May 5, 2010

Abstract

The reorderable matrix is a widely used data processing technology with long history, which aims at providing better insights into data. Jacuqes Bertin mentioned it as a necessary step of graphic information analysis. The short history, basic idea, and general features are introduced at the beginning of this survey. Construction and reconstruction are two main steps of matrix permutation. Construction step builds a matrix, and reconstruction step rearrange it. Manual approaches, interactive approaches, and algorithmic approaches are used to carry out construction and reconstruction. In this review, we give an example "township" to show how matrix permutation works. The application of reorderable come from diverse sources, we also discuss the application domains after we introduce the basic idea of reorderable matrix. Usability of reorderable matrix is discussed through three examples.

Contents

Chapter 1

Introduction

Reorderable matrix(or called matrices permutation, matrices reordering) is a widely used data processing technology with long history, which aims at providing better insights into data. From the age that matrices have not been put forward as an independent mathematical term, people began to do the things similar to matrices reordering. Maybe it is a human instinct to change the order of columns or rows of a matrix when they can not figure out some rational things to do. The only difference is that the object they operated was not "matrix", but a concrete version of it: "table". The oldest example relating to matrices reordering can be traced back to *Organon* collection of the works by Aristotle [10]. In [10], a short history of research/application of matrices reordering has been summarized as figure 1.1. It is interesting to find that reorderable matrix was seldom discussed as an independent topic, but derived from other application background, such as anthropology, sociology, cartography, etc. One of the most famous research and application was done by Jacques Bertin [3, 2], this is why we always refer reorderable matrix as "Bertin reordering matrix" now. Bertin [3, 2] described a graphical data analysis strategy, and took matrices reordering as a necessary step of the whole process (so-called Bertin matrices). This is the start point of this survey work. This is to say, we focus on the matrices reordering, especially from the aspect of information visualization.

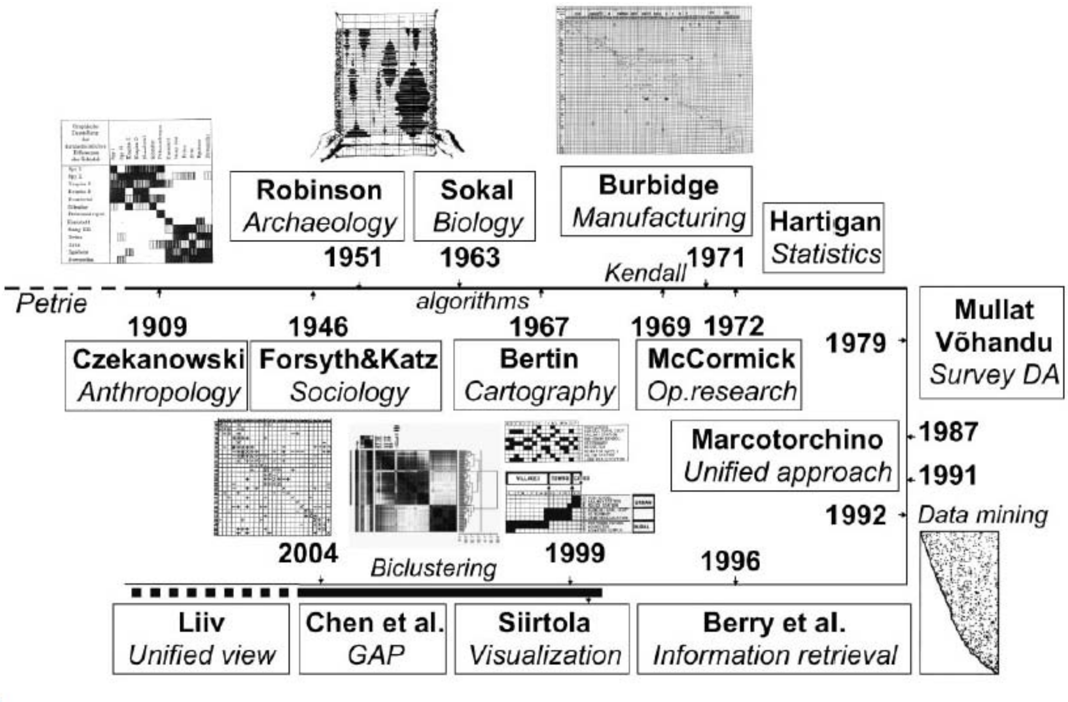


Figure 1.1: Short history of reorderable matrix

As a visualized data mining technology, matrices reordering can be used in a lot of situations. One of the possible reason is that matrix reordering handles matrix directly, and matrix is the most basic concept in modern mathematical (it's not surprising that the most popular mathematical software is named as "Matrices Laboratory"), we use matrix to describe so many things.

Until now, we have not given a clear definition on matrices reordering in this paper. From the abstract aspect, matrices reordering is a serial of operations that change the location of rows and columns of a matrix, without losing any of them. During the process, no information will be lost, Bertin described it as "simplifying without destroying". This is an important advantage of matrices reordering, compared with other data mining method. Some researchers consider that reordering matrices do lose some information [17], but maybe they had a confusion of matrices reordering and matrix analysis [3].

We will not do permutation on matrices without purpose. The general purpose of reorderable matrix is to make some useful information visible. The most attractive feature of matrix permutation is: it is hard to imagine which kind of information can be showed by the reordered matrix before you done it. Bertin's township example [3, 18] is the popular example to explain the effect of reorderable matrix, which will be cited in chapter 2. In this example, there are several towns with some facilities listed in a table. After permutation, you can find that it is obvious that the towns can be grouped into 3 types. But, there is a dilemma: if we do not know which kind of matrices we want, we can not produce them effectively, however, if we do not permute the matrices effectively, we do not know what information it will give. Fortunately, during the research we found that the useful matrix always has some common attributes, one of most important attributes is that they are always ordered by some sequence, for example, the elements has similar values should be placed close. Seriation is another widely used concept associated with matrix reordering, which can be considered as a special case of matrix reordering. "Seriation is descriptive analytic technique, the purpose of which is to arrange comparable units in a single dimension such that the position of each unit reflects its similarity to other units" [10]. A much more general way to describe the situation is that they are clustered (seriation is a kind of clustering). Base on this clustering principle, we define an objective (or an objective function, and it is also called purity function [5]) to measure the effect of clustering, the goal of matrices reordering can be described as pursuing the maximum (or minimum) clustering effect. Effects of different ordering principle are discussed in [6].

There are two necessary steps for matrices permutation: first, construct a matrix; second, reconstruct the matrix. The first step is to decide how to describe the target data set with a matrix, and the second step is about how to reorder it and make information visible. The method to handle reorderable matrix has been changing. In the age of Bertin did their work, computer is not a usual tool. Therefore the manual method is discussed most. At this moment, automatic methods based on computer algorithm are developed, for example, Sugiyama's algorithm.

As mentioned above, reorderable matrix is a kind of data processing technology, so it can be used in nearly all the data mining work. The diverse application will be introduced in chapter 4.

But, no method is perfect. Some limits exist in reorderable matrix. The most significant one is the limitation of dimensions of data it can process [3, 2]. At the most common scenario, an element should have 3-dimensions: the column number, the row number, and the value. Another noticed disadvantage of it comes from automatic approach, matrix permutation is considered as a NP-complete problem, it means that no one can guarantee the optimal solution in a finite time.

Some developments have been done with reorderable matrix to overcome the dimension limitation. Clustering heat map [22] and combining parallel coordinates with the reorderable matrix [17] are typical examples. And, more efficient heuristic algorithm is introduced into this field, in order to get better permutation result.

This survey is organised as following: First we explain the concept of the matrix reordering on a small example in (chapter 2). Chapter 3 then introduces methods for construction and reconstruction of the reorderable matrix. In chapter 4 application domains of the reorderable matrix are discussed. An overview on usability studies is presented in chapter 5. Finally, chapter 6 discusses a serial of further questions on reorderable matrix.

Chapter 2

Examples

In the following section we will show an example on how reordering the rows and columns of a matrix can help analyze data and find correlations between variables within a dataset.

The dataset used for the example can be seen on figure 2.1 by using the 2D Sort presented on the last chapter it is possible to influence the end result of the matrix by choosing a variable to be analyzed. In our example we decided to analyze the locals variable and thus decided to put this row on the top of the matrix prior to the application of the 2D Sort. This can be seen on figure 2.2.

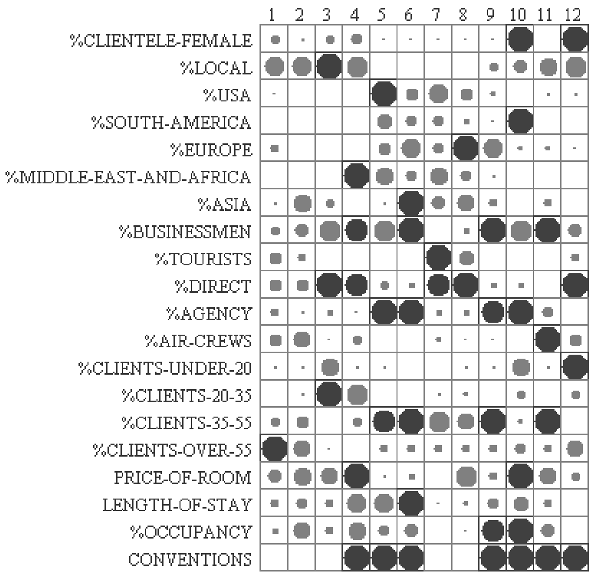


Figure 2.1: An example of analysing data with reorderable matrixes

The results of the automatic sort can be seen on figure 2.3. After the application of the algorithm it is now easy to deduct the following:

- locals clients stying at this hotel are usually under 35 years.
- locals pay more for the room.
- they book their room directly

Another example (Irish Referendum data) is discussed in detail in [5].

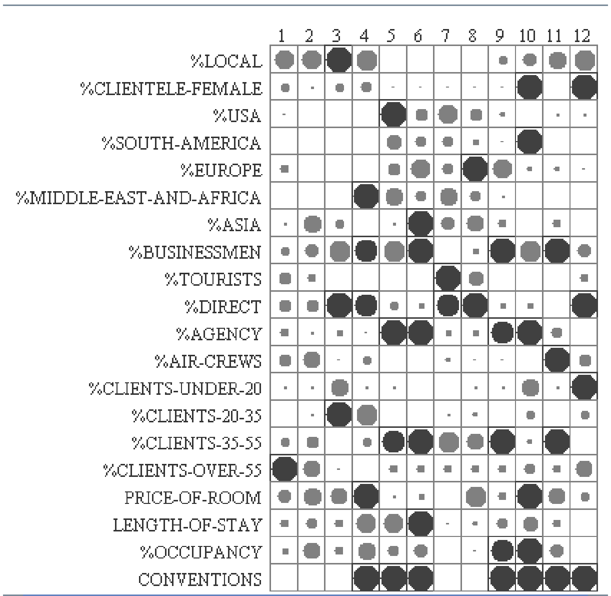


Figure 2.2: Selecting the variable to be analyzed

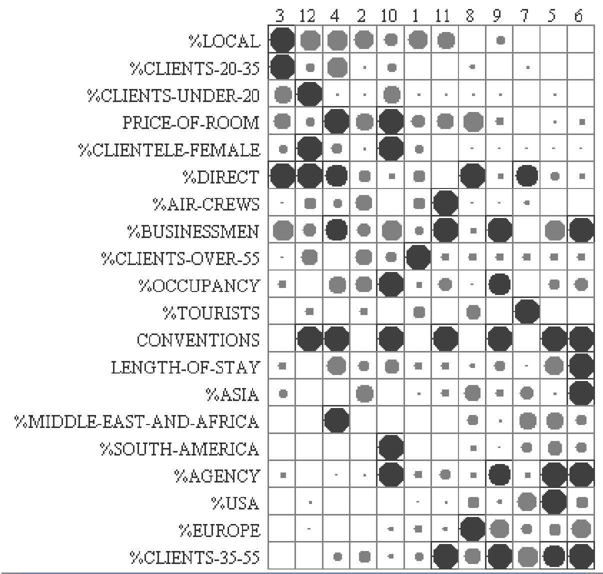


Figure 2.3: An example of analysing data with reorderable matrixes

Chapter 3

Construction and Reconstruction

This section gives an overview of existing approaches for constructing and reconstructing the reorderable matrix. Bertin[3] uses the term *construction* for the initial generation of a reorderable matrix from a data table, *reconstruction* for the permutation of rows and columns of the constructed matrix. Methods for construction and reconstruction can be divided into three categories: (i) manual approaches without the use of computer technology, (ii) interactive approaches where users manually re-order the matrix with the help of information visualisation tools, and (iii) fully automatic algorithmic approaches.

3.1 Manual Approaches

Initially, the permutation of the matrix was performed with paper and pencil, the matrix had to be redrawn after each permutation. The process was documented by taking photocopies or photographs of the reordering steps. Bertin developed permutation equipment which made the redrawing unnecessary. The basis of the devices were small elements, marked with eleven visual steps and special signs for missing data (see figure 3.1c). The first devices, Domino 1 (120x140 elements), and Domino 2(600x100 elements) were combined with photography to document the steps. Domino 3 (see figures 3.1a and 3.1b) was portable and could be used with standard (in 1970s) photocopy devices.

3.2 Interactive Approaches

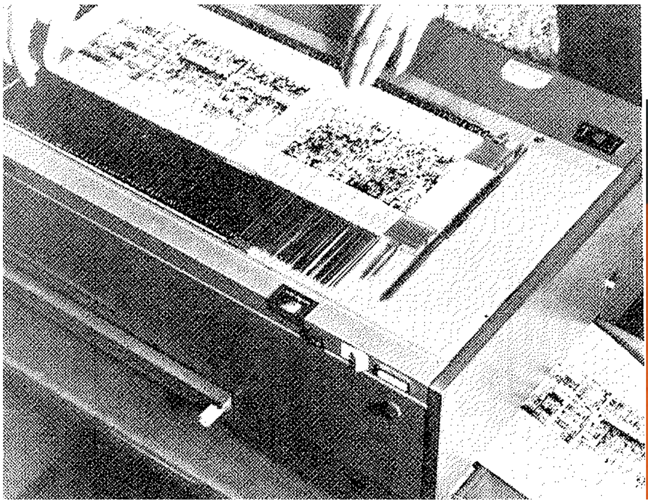
Interactive approaches to matrix reordering [16, 18] are visualisations of the reorderable matrix combined with appropriate interaction strategies. An example user interface is shown in figure 3.2. Figure 5.3 on page 16 shows the same user interface annotated with implemented user interactions for the usability studies.

In [16] the user could manipulate the matrix using four operations:

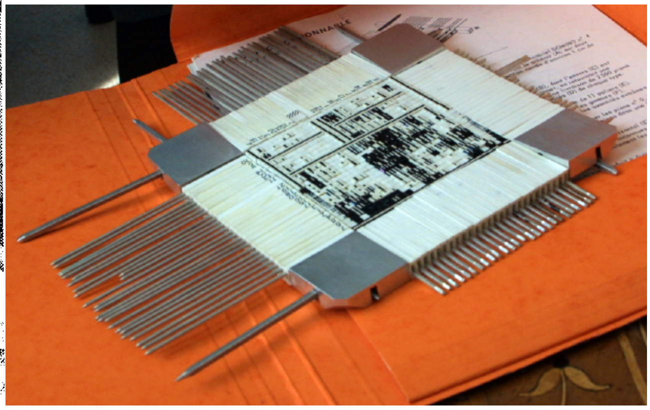
- sorting of rows or columns
- dragging of rows or columns

In [18] the possible manipulations were extended to the following operations:

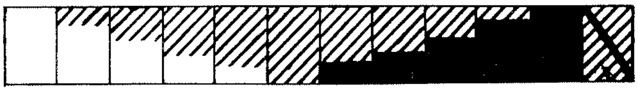
- automatic arrangement of rows or columns (includes sorting of rows or columns)
- moving of rows, columns or areas(includes dragging of rows or columns)
- locking of neighbouring rows or columns
- threading along a row or column (sorting the matrix along this row or column) according to a given characteristic



(a) Domino 3 and copier



(b) Domino 3



(c) visual elements

Figure 3.1: Elements and devices for matrix construction and reconstruction, taken from [3], last taken from [7].

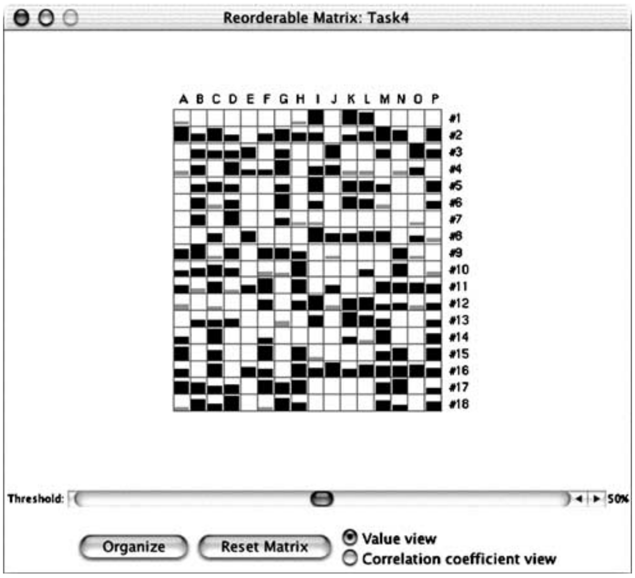


Figure 3.2: The user interface for the reorderable matrix, taken from [18].

3.3 Algorithmic Solutions

Algorithmic solutions for Bertin's reoderable matrix have been proposed by Mäkinen and Siirtola in [11, 12]. Henry [7] (pp. 81) include an extensive overview of reordering techniques that can be applied to matrices. s Matrix reordering is related to problems which are known to be NP-complete [11], e.g. graph bandwidth minimisation and matrix denomination. Graph bandwidth minimisation aims at permutating rows and columns of a matrix such that the resulting matrix that all non-zero elements are positioned near the main diagonal - in a band as thin as possible. Because of the NP-completeness of the reordering problems two heuristic approaches were proposed: an 2D sort method and Sugiyama's algorithm. In the following both approaches will be discussed and applied to the famous township example. Figure 3.3 shows the township dataset as a reorderable matrix, figure 3.4 show one optimal reordering that reveals the three clusters *cities*, *towns* and *villages*. Table 3.1 shows the three groups contained in the data and the corresponding columns.

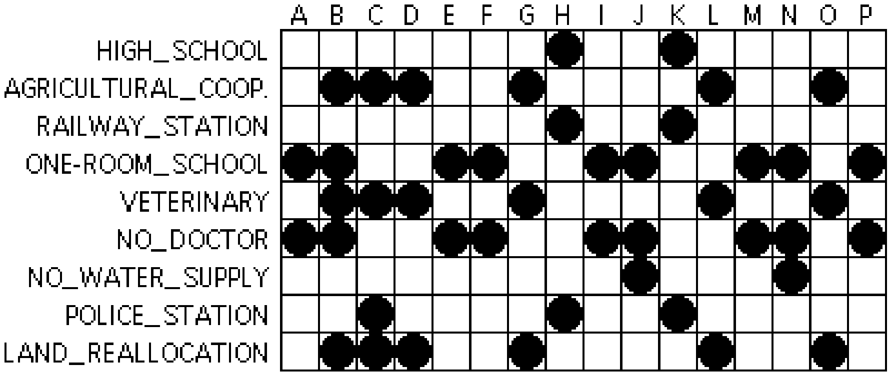


Figure 3.3: Reordering matrix of the township example, taken from [11]

group	**rows**
cities	K, H (first two rows in fig 3.4)
towns	C, D, G, L, O, B (third to eighth row in fig 3.4)
villages	A, I, E, M, F, P, J, N (ninth two last in fig 3.4)

Table 3.1: Data groups in the township example

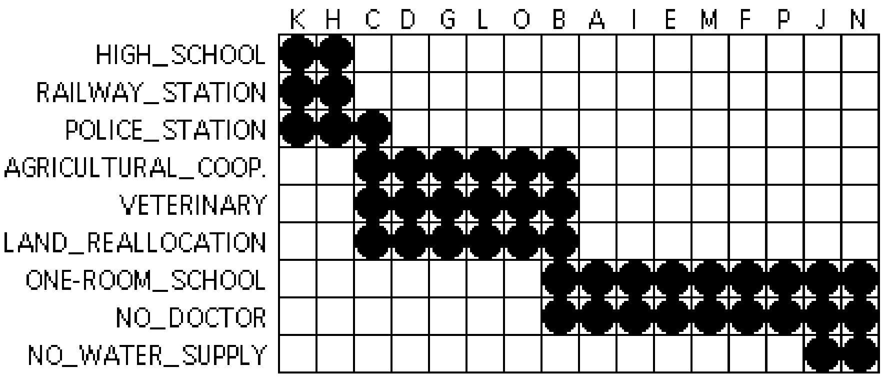


Figure 3.4: An optimal solution for the township example, taken from [11]

3.3.1 2D Sort

The assumption behind 2D sort is that interesting patterns are revealed if black areas are built on the top left and the bottom right of the matrix. The algorithms consists of the following steps:

1. calculate weithed row sums, weight = column position of the cell
2. arrange matrix ascending according to this row sums (sort rows)
3. calculate weighted column sums, weight = row position of the cell
4. arrange matrix ascending according to this column sums (sort columns)
5. repeat step 1.-4. until no reordering occurs

As can be seen in figure 3.5 the 2D sort algorithm reorders the cells similar to the optimal solution of figure 3.5, but does not find an optimal solution.

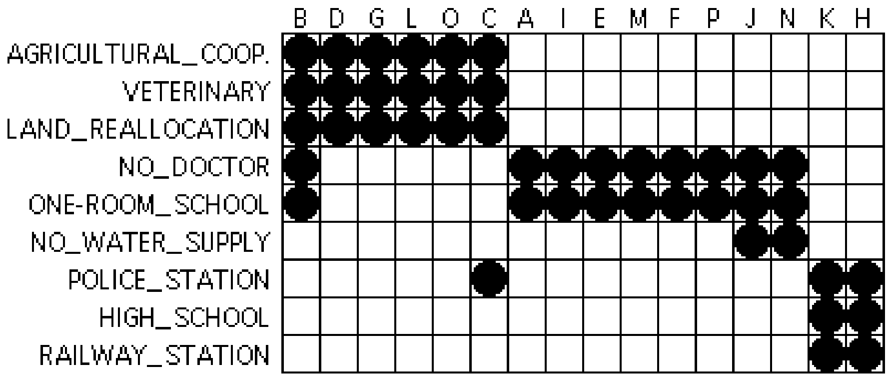


Figure 3.5: 2D Sort applied to the township example, taken from [11]

3.3.2 Sugiyama's Algorihm

A binary reordering matrix can be seen as an adjacency matrix of a bipartite graph. The rows of the matrix represent elements of one partition and the columns represent the elements of the other partition. A cell of the matrix contains the value 1 if an edge exists between the vertices of the corresponding row and column. The optimal layout of bipartite graphs, i.e. a layout with a minimum of edge crossings, is known to be NP-complete. An heuristic approach for the optimal layout of bipartite graphs is Suyiyama's algorithm. The heuristic repeatedly orders the vertices of the two sets according to the averages of their adjacent vertices in the opposite sets. The applied heuristic tends to generate "black areas" on the top left and on the bottom right. For details of the algorithm see [20]. To apply Sugiyama's algorithm to real-valued matrices, the simplest solution is to set a threshold; values below the threshold are mapped to 0 (no edge exists) and values above the threshold are mapped to 1 (edge exist). Figure 3.6 shows the result of Sugiyama's algorithm applied to the township matrix. The ordering of the rows and columns is close to the optimal solution (figure 3.4), only three columns need to be reordered further.

3.3.3 Comparison of the Algorithms

Table 3.2 gives an overview of the results obtained for the different algorithms. Sugyiama's algorithm reveals exactly the same groups as the manual solution. The 2D sort is nearly optimal, only two towns (C, B) and two villages (J, N) are not as clearly identifyable as in the manual solution.

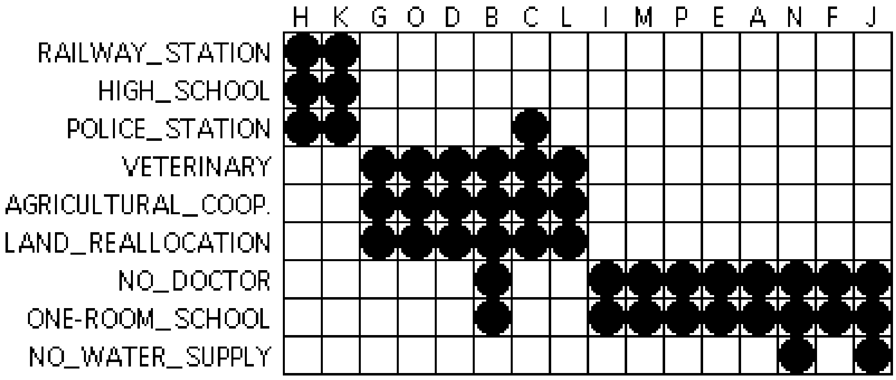


Figure 3.6: 2D Sort applied to the township example, taken from [11]

algorithm	cities	towns	villages	problematic
manual/optimal	K, H	C, D, G, L, O, B	A, I, E, M, F, P, J, N	-
2D sort	K, H	-, D, G, L, O, -	A, I, E, M, F, P, -, -	C, B, J, N
Sugiyama	K, H	C, D, G, L, O, B	A, I, E, M, F, P, J, N	-

Table 3.2: Comparison of reordering result quality for 2D sort and Sugyiama's algorithms in the township example

3.4 Systems including the Matrix Reordering

Apart from the application domains summarised in section 4, the concept of matrix reordering is used for example in the following information visualisation systems (the citation in brackets indicates who classified the system as an extension or adaption of the reordering matrix):

- Table Lens [13] (classified by [18])
- Voyager [15] (classified by [18])
- FOCUS/InfoZoom [19] (classified by [7])
- CHART [1] (classified by [7])
- MatrixExplorer [8] (classified by [7])
- Reorderable Matrix [18] (classified by [7])
- Clustering Heatmap (classified by [22])

Chapter 4

Application Domains

The application domains in which matrix reordering is used are very extensive, matrix reordering is currently used in fields that vary from Archeology to Bioinformatics Liiv [10] wrote a historical overview on the matrix reordering methods and the application domains in which these methods are used. The following chapter summarizes the most relevant application domains mentioned in the review. Intrested readers may want to read to original review for a more extensive description.

4.1 Archaeology and Anthropology

According to the author [10] the first systematic method for seriation was developed by the English Egyptologist W. M. Flinders Petrie, who called it sequence dating. Even though Petrie did not use any classical mathematical notations, his observations and methods are considered to be the first to clearly formulate the idea of sequencing objects on the basis of their incidence or abundance.

Petrie examined about 900 graves and assigned them sequence dates using mainly the characteristics of the found pottery. He was able to seriate the pottery chronologically by nearly looking at the characteristics of the handles. Petries work influenced influenced several prominent American anthropologists and archaeologists like George Andrew Reisner, Alfred Vincent Kidder, Alfred Louis Kroeber,Nelson, Leslie Spier, James A. Ford who applied, popularized, and further developed the methodology to better suit the practical needs or relative dating.

Brainerd and Robinson [14] proposed a desired final form for the matrix: the highest values in the matrix should be along the diagonal and monotonically decrease when moving away from the diagonal. This matrix is widely known as *Robinson Matrix* or *R-Matrix*.

Robinson's approach was criticized for not taking into account the differences in the size of the collections. Dempsey and Baumhoff proposed a method to cope with such problem. The method is called occurance seriation. They argue that types that occur with low frequency may be among the best time-indicators and the presence of single specimens of certain types may be crucial in establishing chronologies.

Bertin himself also made a study about archeology. He examined 59 Merovingian artifacts, described according to 26 characteristics [4].

4.2 Sociology and Sociometry

Forsyth and Katz [14] were the rst to introduce an approach of rearranging the rows and columns of the sociomatrix for a better presentation of the results of sociometric tests. There seems to be neither an obvious nor an implicit inuence of previous works with rearranging the matrices, and the motivation for method development seems to descend directly from Morenos work on sociograms. Forsyth and Katz credited the sociogram as clearly advantageous over verbal descriptions and relationship listings, but confusing to the reader, especially if the number of subjects is large. Katz also argued that the sociometric art has simply progressed to the point where pictorial representation of relationships is not enough and quantications of the data should be sought. It was hoped that the sociomatrix and the development of methods for analyzing the matrices would ll that gap. Sociogram drawing was a manual process and there were still decades until automatic graph drawing algorithms started to emerge in computer science, and be used across disciplines.

Matrix reordering techniques have been used to study social relationships in social networks, it is easy to find clusters (cliques) using sociomatrixes. Sociograms and sociomatrixes both offer certain advantages and seem to supplement each other. Some reserchers however seem to find sociomatrixes inferior to sociograms since even pair relations are hard to find. But the strength of the sociomatrixes seems to be in the finding of more complex structures, such as triangles, chain reactions and stars.

4.3 Psycology and Psychometrics

Hubert [9] adapted seriation algorithms that were developed for archeology in the psychology field. He performed analysis on both, one mode and two mode matrices. Hubert used a parallelogram analysis for finding patterns within a martix for paired- comparison data.

4.4 Ecology

In Ecology, seriation was often considered to be the best practice to perform clustering whithout explicitly distinguishing between the two. It is in this field that seriation is more widely used than classical clustering which seems to be the standard in many other fields. It is probably for this reason that packages that were developed for ecological studies are amogst the most mature. Kulczynski studied plant associations using the matrix coding and visualization approach developed by Czekanowski.

4.5 Biology

Seriation methods in biology have similar methodological roots to those that are used in the discipline of ecology. The paradigm of data analysis using the reordering of rows and columns was introduced to the community of biologists by the famous monograph of Numerical Taxonomy by Sokal and Sneath *Sneath1973*, which created a lot of controversy for the strong statements and criticism against the traditional way of creating taxonomies in biology. Sokal and Sneath *Sneath1973* (p. 176) introduced matrix reordering techniques, using the name differential shading of the similarity matrix, and referred to the result of the seriation procedure as a cleaned up diagram. It was thought that Robisons procedure could find a optimum structure in the system.

Chapter 5

Usability Studies

There are some usability studies to determine whether working with Bertin's reorderable matrix is useful and maybe better than using the traditional way of pen and paper. Several people are included in such a usability study to solve a special task to check the subjective user satisfaction, the efficiency and simplicity of the tested scenario, and how it eventually could be used in everyday work. Following we will describe three different usability studies concerning Bertin's reorderable matrix.

5.1 Study: Bertin's Hotel Example

The "Hotel Example" [16] is the example used from Bertin himself to illustrate the functionality of the reorderable matrix. The data contains different values about the clientele of a fictitious hotel including age, nationality, gender, reason and length of stay. This kind of data would be useful for hotel managements to plan their offers and price policy. The same example is used as an example for using the reorderable matrix in chapter 3 of this document.

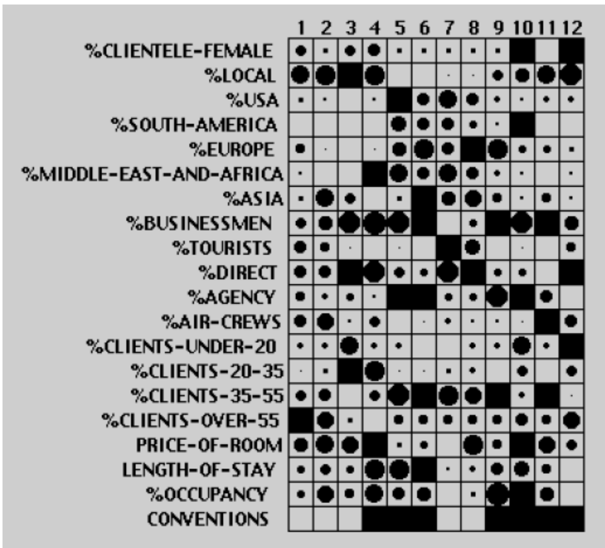


Figure 5.1: Bertin's Hotel Example

The task for the participants of this usability study was to find as many as possible correlations in the data within 20 minutes. The matrix could be manipulated in various ways: rows and columns could be sorted in ascending or descending order by simply clicking the row or column to be sorted outside of the matrix, rows and columns could be dragged into a new position, and already finished moves could be undone or redone by clicking on the appropriate button. It was also possible to drag a row and a column simultaneously.

Eleven people participated in the study, 5 of which were women and 6 were men. All of them where fluent computer users, most of them working in research about user interfaces and usability, but none of them had any idea of Bertin's book or Bertin's work at all. They also weren't familiar with the domain of hotel management.

As for the results of this empirical study, most of the users found the reorderable matrix to be interesting and exciting and they liked to work with it. They just mentioned that the program didn't indicate how to sort rows or columns for they had to find out for themselves to click outside the border to sort rows or columns. They also would have liked to have some grouping so that after sorting rows or columns groups like nationalities or age groups would stand out because of colouring or something like that. They also weren't so happy with the feature to move a row and a column simultaneously except one of the participants, who used this possibility mainly.

The participants could be classified into four major groups concerning the two most used operations. 4 of the participants preferred to sort the matrix along one specific row and then move other rows to look for correlations, which is similar to Bertin's preferred strategy. 3 of the users did more sorting operations than moving operations. The third category which was done also by 3 of the users was to mainly sort either rows or columns, and the last one participant did no sorting at all but only moving along the rows and columns.

Performance and correctness could also be augmented by automating the move operations after a sort operation using an algorithm like Sugiyama's Algorithm, because if similar rows or columns would lie just next to each other, the probability for the user to overlook such a correlation would be minimised. The question is, why the reorderable matrix visualisation isn't used more these days, because implementation isn't very complicated, and it would be a nice feature completing static visualisations like the ones in Microsoft Excel or anything comparable.

5.2 Study: Clustering Example

The second usability example we present in our paper is an example to find clusters in matrix-like ordered data. There is no specific domain data, the data is merely some fictitious random data to be sorted and scanned for clusters to reveal the pure functionality of reorderingmatrices.

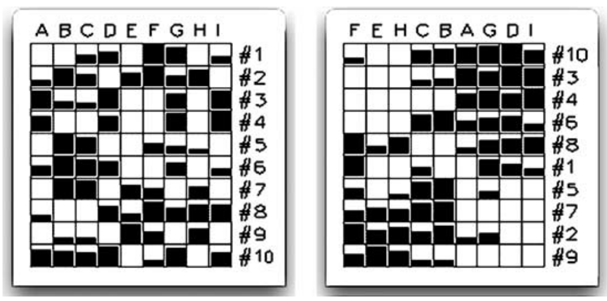


Figure 5.2: The Clustering Example - left: initial state - right: one of the goal settings

The task for the participants was to find groups of similar cases in a given reorderable matrix, so that the cases had roughly the same values in most of their characteristics [18]. The operations in the matrix were nearly the same like in the usability study mentioned before - clicking at the end of a row or column sorted the row/column ascending towards the position of the mouse, a dragging operation inside the matrix moves the specific row/column. However, there were no "undo" or "redo" buttons, but there was an "organise" button to sort the matrix automatically and a "reset" button to reset the matrix to its initial arrangement. For organising the matrix automatically there was also a slider to define a threshold for the process organisation.

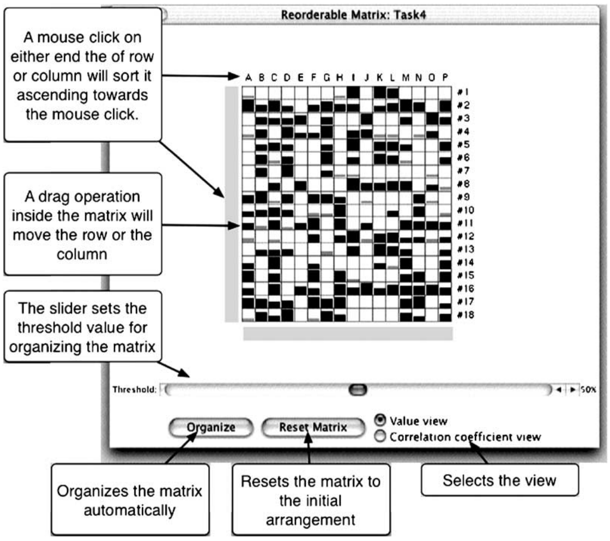


Figure 5.3: The Clustering Example - the User Interface

Twelve subjects participated in this usability study, seven of them where men and five were women. All of them where either students or employees of the research unit where the initiators of the study came from, and they had rather good experience in the use of computers, but they had so far no idea of specialised visualisation tools. The median age of the users was 26 years.

There were three conditions to be met:

- Paper and pencil. The subjects received some paper with a picture of the reorderable matrix and a pen and had to find clusters by only using pen and paper.
- Computer program with manual operations. There was a computer program implementation of the reorderable matrix like described before.
- Computer program with automatic ordering. Additionally to the manual operations of the computer program implementation there was an automatic method to sort the matrix using the Barycenter heuristic.

The participants were randomly divided into two groups, where both groups had to do the task with pen and paper, but one group had to do the manual only part, and the other group had to do the automatic part.

As a result, the pen and paper solution took the most time and was also the most inaccurate. The traditional solution (manual ordering) was better both in performance and in correctness, and the automatic solution was best in speed and correctness.

The subjective satisfaction of the users was quite low, although they considered the technique of reordering matrices quite better than the ”antique” solution using pen and paper. It is easier to find similar columns with the matrix application than with pencil and paper because of the correlation view, and it is more useful to show matrix cells as graphics than numbers.

5.3 Study: Ecological Interface for Process Control Health Monitoring

In the last usability study we examine in our paper we look at the difference between two different views of the same data [21]. The first view of the data examined is a control chart with parameters on the horizontal axis, values on the vertical axis and tools resp. machines encoded by means of icons (see figure 5.4).

The second view (see figure 5.5) is a redesigned matrix of the data, where the tools are on the vertical axis, the parameters are still on the horizontal axis, but the values are encoded as the thickness of the data points

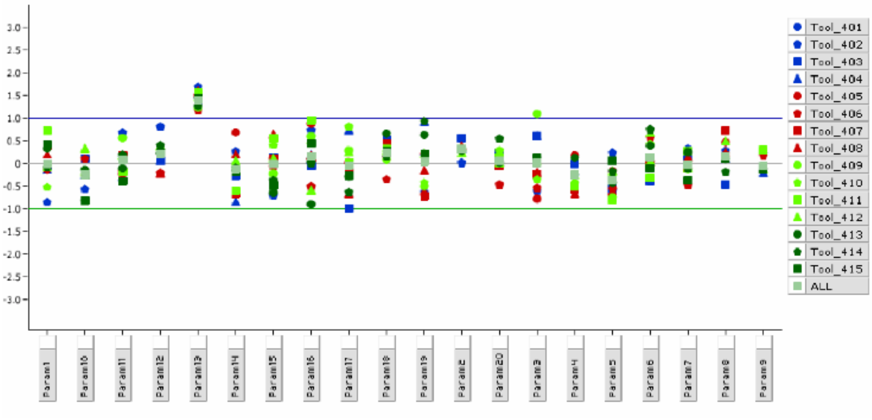


Figure 5.4: Ecological Interface for Process Control Health Monitoring

in the matrix. The goal of this experiment was to determine whether the redesign of the view would have an impact on the usability and readability of the data.

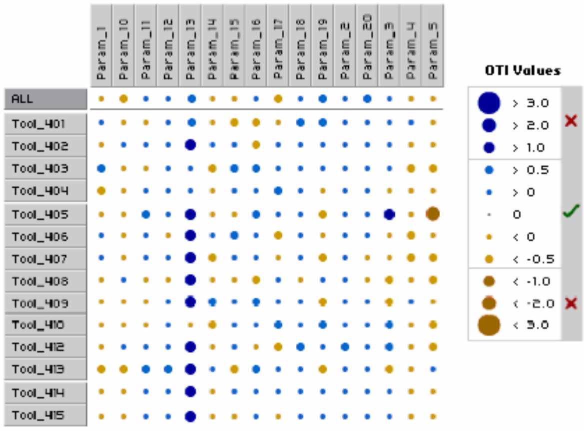


Figure 5.5: Ecological Interface for Process Control Health Monitoring - Redesigned matrix

The task of this study was to find outliers in tools and/or parameters and/or values using both the original chart and the redesigned chart showing the reorderable matrix.

The study was done by 20 participants, where 14 were males and 6 were females. Their age range spanned from 22 to 40 years, half of them were postgraduate students from the computer science department of Trinity College in Dublin, the other half were industry employees. They all were experienced computer users, but had no knowledge whatsoever of the domain of the process control health monitoring.

For the results of this usability study the most of the participants (about 16 out of 20) considered the redesigned chart to be better for finding outliers in tools, parameters and values of the chart. Also the efficiency and accuracy is much better when using the redesigned chart. So the conclusion holds, that the chart type is strongly responsible for improvements in the performance, and the reorderable matrix is a proper and good tool to analyse sets of data up to a certain amount.

5.4 Summary

In all the usability studies mentioned before the users were very satisfied by using the reorderable matrix. The comparison to the traditional method using pen and paper was also successful, because using the reorderable matrix clearly was a lot faster and more reliable. Surely, using a reorderable matrix with also an automatic sorting algorithm was rated better because of the performance and the correctness of found correlations and/or clusters. Also the comparison with other views of the same data showed that the reorderable matrix brought improvement in the readability of the data. Handling of the programs used for the usability studies was mostly intuitive and no problem for the users being mostly fluent computer users. The question remaining is why Bertin's reorderable matrix isn't used much more in everyday work, mainly in technical sectors, although it apparently would simplify most of the work to be done.

Chapter 6

Discussion

Matrix permutation is still a "mysterious" method until now. Several points on it deserve discussion.

From the aspect of application, the usability is still an open question. As mentioned in section 5, we pointed out that although the usability of this method has been proven in some experimental researches, this method is not a popular analytical method. We think this due to practical usability issues. This problem can be divided into two sub-problems: The first problem is: Does the matrix reordering give some predictable useful result? The second problem is: Is this method easy to use? The second problem also includes scalability issues, i.e. to which extend is the method applicable to large data sets. We do have some positive examples to support affirmative answers to the two problems, but no proof from the aspect of theory.

Effective algorithms for matrix permutation, performed either manually or automatic, are another important problem. The complexity of the permutation algorithm is the core of this problem. This problem of effective algorithms also can be divided into two sub-problems: first, how to define the measurement of permutation effects (i.e. how to rate a specific permutation), and second, how to find a better permutation. For the first problem, a dilemma was mentioned in chapter 1 of this survey. To find an attribute which relates to a better permutation effects is not only a technology but also an art. As we do not know what to expect from the data it is hard to define a target function for optimisation a priori. The second problem, the optimal ordering, is considered as an NP-complete problem, and the proposed heuristics are not guaranteed to find a globally optimal solution. The traditional principles, such as seriation or clustering, work well but not well enough, because they do not consider the matrix layout. For instance, similarity measures for clustering are either defined for a fixed ordering of attributes (constructed feature vectors remain unchanged during the process) or do not consider any ordering at all (e.g., Euclidean distance on feature vectors treats all attributes equally).

The extension of matrix permutation is worth noticing. Apart from the pending problems mentioned before, the dimension limitation of matrix permutation is obvious. The most popular method to make up this disadvantage is to combine reorderable matrix with other methods. For instance, the combination of the reordering matrix and node-link diagrams (e.g., in the MatrixExplorer [8]) seems promising.

In summary, reorderable matrix is an attractive method, but it has not been investigated systematically. This method will become a complete-defined data mining method or not, is still pending.

Bibliography

[1] William H. Benson and Bernard Kitous. Interactive analysis and display of tabular data. In *SIGGRAPH '77: Proceedings of the 4th annual conference on Computer graphics and interactive techniques*, pages 48–53, New York, NY, USA, 1977. ACM.

[2] J. Bertin. *Graphics and graphic information processing*, chapter 1D, 2D, 3D, pages 62–65. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1999.

[3] Jacques Bertin. *Graphics and Graphic Information Processing*. Walter de Gruyter & Co., Berlin, New York, 1981. (Originally La graphique et le traitemente graphique de linformation, 1967, translated in English by William J. Berg and Paul Scott).

[4] Jacques Bertin. Matrix theory of graphics. *Information Design Journal*, 10:5–19, 2001.

[5] A. de Falguerolles, F. Fredrich, and G. Sawitzki. A tribute to J. Bertin's graphical data analysis. In W. Bandilla and F. Faulbaum, editors, *SoftStat 97, Advances in Statistical Softwar*, pages 11–20, Stuttgart, 1997. Lucius and Lucius.

[6] Michael Friendly and Ernest Kwan. Effect ordering for data displays. *Computational Statistics & Data Analysis*, 43(4):509 – 539, 2003. Data Visualization.

[7] Nathalie Henry. *Exploring Social Networks with Matrix-based Representations*. PhD thesis, University South Paris and University of Sydney, July 2008.

[8] Nathalie Henry and Jean-Daniel Fekete. Matrixexplorer: a dual-representation system to explore social networks. *IEEE Transactions on Visualization and Computer Graphics*, 12:677–684, 2006.

[9] L. J. Hubert and R. G. Golledge. Matrix reorganization and dynamic programming: Applications to paired comparisons and unidimensional seriation. *Psychometrika*, 46(4):429–441, 1981.

[10] Innar Liiv. Seriation and matrix reordering methods: An historical overview. *Wiley Periodicals*, 3(2), 2010.

[11] Erkki Mäkinen and Harri Siirtola. Reordering the reorderable matrix as an algorithmic problem. In *Diagrams '00: Proceedings of the First International Conference on Theory and Application of Diagrams*, pages 453–467, London, UK, 2000. Springer-Verlag.

[12] Erkki Mäkinen and Harri Siirtola. The barycenter heuristic and the reorderable matrix. *Informatica (Slovenia)*, 29(3):357–364, 2005.

[13] Ramana Rao and Stuart K. Card. The table lens: Merging graphical and symbolic representations in an interactive focus+context visualization for tabular information. In *Proc. CHI'94*, pages 318–322, Boston, Massachusetts, USA, April 1994. ACM.

[14] W. S. Robinson. *A method for chronologically ordering archaeological deposits*, volume 16. Am Antiq, 1951.

[15] Gunther Sawitzki. Extensible statistical software: On a voyage to oberon. *Journal of Computational and Graphical Statistics*, 5(3):263–283, 1996.

[16] Harri Siirtola. Interaction with the reorderable matrix. *Information Visualisation, International Conference on*, 0:272, 1999.

[17] Harri Siirtola. Combining parallel coordinates with the reorderable matrix. In *CMV '03: Proceedings of the conference on Coordinated and Multiple Views In Exploratory Visualization*, page 63, Washington, DC, USA, 2003. IEEE Computer Society.

[18] Harri Siirtola and Erkki Mäkinen. Constructing and reconstructing the reorderable matrix. *Information Visualization*, 4(1):32–48, 2005.

[19] Michael Spenke, Christian Beilken, and Thomas Berlage. Focus: the interactive table for product comparison and selection. In *UIST '96: Proceedings of the 9th annual ACM symposium on User interface software and technology*, pages 41–50, New York, NY, USA, 1996. ACM.

[20] Kozo Sugiyama, Shojiro Tagawa, and Mitsuhiko Toda. Methods for visual understanding of hierarchical system structures. *IEEE Trans. Systems, Man and Cybernetics*, 11(2):109–125, February 1981.

[21] Connor Upton, Gavin Doherty, and Myra ORegan. Usability evaluation of an ecological interface for process control health monitoring. Technical Report TCD-CS-2007-32, Department of Computer Science, Trinity College Dublin, 2007.

[22] Leland Wilkinson and Michael Friendly. The History of the Cluster Heat Map. *The American Statistician*, 63(2):179–184, 2009.