Comparison of Matrix Reordering Algorithms Based on Monotone Systems

Grete Lind and Rein Kuusik

Tallinn University of Technology, Tallinn, Estonia
{grete.lind,rein.kuusik1}@taltech.ee

Abstract. The aim of this paper is to shed light on matrix reordering methods: namely scale of conformity, minus technique, plus technique and mixed technique. All of them are based on the theory of monotone systems. Presented methods are applicable both to NxN (entity-to-entity) and NxM (entity-to-attribute) data tables. All the methods can use larger set of discrete values (not only binary ones). Rows and columns are reordered separately. The result does not depend on the initial order of rows and columns. We compare the results of these methods through stress measure (both in the von Neumann neighborhood and in the Moore neighborhood) using binarized representations of well-known data sets.

Keywords: Matrix Reordering, Monotone Systems, Conformity, Stress.

1 Introduction

Seriation is an exploratory data analysis technique to reorder objects into a sequence along a one-dimensional continuum so that it best reveals regularity and patterning among the whole series [1]. Seriation is often called *matrix reordering*, when applied to two-way datasets [2]. Two-way one-mode data table means entity-to-entity (NxN) data table and two-way two-mode is entity-to-attribute (MxN) data table [1].

Matrix reordering has been used in many different fields, from archaeology to operations research. A thorough historical overview is given by Liiv [1].

In this paper, we introduce little known matrix reordering methods that are based on the theory of monotone systems (MS) created by Mullat [3–5]. These methods – scale of conformity, minus technique, plus technique – have been created by Leo Võhandu at Tallinn University of Technology, department of Informatics [6–8]. We propose a novel MS-based method by Rein Kuusik called mixed technique [8].

According to Liiv [1] the main future goal for seriation is to make it ubiquitously usable, reordering the matrices should be a common practice for everybody inspecting any data table. Recently Liiv and Võhandu [2] experiment with asymmetric one-mode two-way (NxN) data tables. They propose that "seriation methods can be applied to analyze asymmetric one-mode two-way datasets as if

 Published in Francisco J. Valverde-Albacete, Martin Trnecka (Eds.): Proceedings of the 15th International Conference on Concept Lattices and Their Applications, CLA 2020, pp. 281–286, 2020.

they were two-mode two-way datasets while continuing to keep the information about entities actually belonging to one class".

The paper is organized as follows. In the following subsections, we introduce stress measures and describe MS-based reordering methods. Section 2 is dedicated to the mixed technique. Experiments are presented in section 3 and conclusion in section 4.

1.1 Stress

Stress is a dissimilarity measure, it compares the values in a matrix with their neighbors. For an $n \times m$ matrix X, the local stress measure for element x_{ij} is defined for two types of neighborhood [9]:

1. in the Moore neighborhood–a square-shaped neighborhood comprising (at most) eight adjacent entries:

$$s_{ij} = \Sigma_{k=max(1,i-1)}^{min(n,i+1)} \Sigma_{l=max(1,j-1)}^{min(m,j+1)} (x_{ij} - x_{kl})^2 \ , \tag{1}$$

2. in the von Neumann neighborhood–a diamond-shape neighborhood comprising (at most) four adjacent entries:

$$s_{ij} = \Sigma_{k=max(1,i-1)}^{min(n,i+1)} (x_{ij} - x_{kj})^2 + \Sigma_{l=max(1,j-1)}^{min(m,j+1)} (x_{ij} - x_{il})^2 \ . \tag{2}$$

A global stress measure for the whole matrix is the sum of the local stresses (in either neighborhood) for all entries of the table:

$$STRESS = \Sigma_{i=1}^{n} \Sigma_{j=1}^{m} s_{ij} \ . \tag{3}$$

In case of binary data, the local stress s_{ij} is just the count of the neighbors that have different values.

1.2 Reordering Methods Based on Monotone Systems

Here we briefly introduce three methods for reordering data tables: scale of conformity, minus technique and plus technique [6–8]. Because of limited space we cannot present their algorithms and examples.

All the methods reorder rows and columns separately. Thus, the same algorithm can be applied for both rows and columns, using transposed table for one of them. Their result does not depend on the initial order of rows/columns.

The scale of conformity [6] reorders the data by object's typicality using the conformity measure that is the sum of all attribute-value frequencies (of the row).

In case of iterative techniques [7], to wit minus technique, plus technique, rows/columns are removed from table one-by-one, after each removal the weights of the remaining rows/columns are recomputed. Iterative techniques use conformity as a weight function. Actually different weight functions (e.g. influence [6]) can be used.

Algorithm: Mixed technique for matrix rows

S1. Calculate frequencies $FT(t, j)$ for every attribute's values $t = 1, 2, \ldots, Kj$ in columns j, where $j = 1, \ldots, M$
S2. For every row $i = 1, 2, \ldots, N$ find the sums (weights) $W(i) = \Sigma\, FT(t, j)$, $j = 1, \ldots, M$
S3. Find $R = min\, W(i)$; remember i
S4. Eliminate row i from the matrix
S5. If there are yet rows in the matrix then goto S6 else goto S12
S6. Nullify the frequency table FT, weights $Wprev$
S7. Increase frequencies of the eliminated row i elements by one: $FT(t, j) = FT(t, j) + 1$
S8. For every row $i = 1, 2, \ldots, N$ find the sums (weights) $W(i) = \Sigma\, FT(t, j)$, $j = 1, \ldots, M$
S9. Find $R = max\, W(i) - Wprev(i)$; remember i
S10. Eliminate row i from the matrix
S11. If there are yet rows in the matrix then $Wprev = W$; goto S7 else goto S12
S12. Reorder matrix rows in the order of elimination
S13. End

These reordering methods allow to process non-binary data and zeros are treated the same way as other values. Using a different weight function, we can treat zeros differently.

The data table will be reordered in order to better visualize the data. In case of conformity scale and plus technique, the most homogeneous group forms in the upper-left corner and the most atypical in the lower-right corner. In case of minus technique–on the contrary.

While these techniques help finding homogeneous groups, we miss a method offering possibly smooth changes. In the next section we will propose such a method.

2 Mixed Technique

Here we present a novel iterative method for matrix reordering, called mixed technique [8]. It is aimed to create a gradual way of changes, starting from the first object/attribute. The user chooses from which object/attribute to start. If the user cannot decide, then it can be chosen by minimal weight (as in minus technique). Further the closest object/attribute to the just eliminated one (by the number of coincidences) is chosen for removal. The weight is not based on the frequencies of (diminishing) initial table (as in case of plus technique and minus technique), but on the (growing) table of already removed objects/attributes.

Assume that we have an $N \times M$ data matrix X. Every element X_{ij}, $i = 1, \ldots, N$, $j = 1, \ldots, M$, has a discrete value from an interval $[1, K]$.

Algorithm (see above) starts like the minus technique (S1..S5) and after the first iteration, it continues similarly to the plus technique (S6..S13).

Table 1. (a) Initial data matrix, (b) order of elimination of objects (6 iterations), (c) order of elimination of attributes (5 iterations), (d) reordered data matrix

(a)

	A1	***A2***	***A3***	***A4***	***A5***
O1	1	2	2	2	2
O2	2	1	2	1	1
O3	2	1	2	1	1
O4	1	1	2	1	2
O5	2	2	1	2	1
O6	2	1	1	1	1

(b)

	It1	*It2*	*It3*	*It4*	*It5*	*It6*
O1	0					
O2	0	1	4			
O3	0	1	4	9		
O4	0	3				
O5	0	2	2	4	6	9
O6	0	0	2	6	10	

(c)

	A1	***A2***	***A3***	***A4***	***A5***
It1	0	0	0	0	0
It2		2	2	2	0
It3			4	8	4
It4			6		8
It5			10		

(d)

	A1	***A2***	***A4***	***A5***	***A3***	***W(Oi)***
O1	1	0	0	0	0	*0*
O4	1	1	1	0	0	*3*
O2	0	1	1	1	0	*4*
O3	0	1	1	1	0	*9*
O6	0	1	1	1	1	*10*
O5	0	0	0	1	1	*9*
W(Aj)	*0*	*2*	*8*	*8*	*10*	

In order to demonstrate the mixed technique we use data given in Table 1 (a). The conformity i.e. weight of $O1$ is 2+2+4+2+2=12 ($count(A1 = 1) + count(A2 = 2) + count(A3 = 2) + count(A4 = 2) + count(A5 = 2)$). $O1$ has the smallest weight among objects, therefore it is selected first. After that starts the iterative part of the algorithm.

Table 1 (b) shows the weights of objects (rows) during 6 iterations and the order of elimination of rows, while Table 1 (c) shows the same for columns (during 5 iterations). Reordered data table is presented in Table 1 (d). Weights of rows $W(Oi)$ and weights of columns $W(Aj)$ present the weights at the moment of elimination of a row/column.

Compared to the previous techniques, mixed technique gives different information. By maximizing the similarity between consecutive rows/columns, it reveals an "evolutionary" way of changing/developing the initial object/attribute.

3 Experiments

We compare the results of four introduced algorithms by two stress measures [9] (see section 1.1) using 20 different data sets.

We use binarized representations of 20 data sets from UCI Machine Learning Repository[1], the list of data sets, ordered by size, is given in Table 2.

Table 3 shows global stress values in both neighborhoods for all the considered data sets for all 4 reordering algorithms based on monotone systems: conformity scale (conf), plus technique (plus), minus technique (min) and mixed technique (mixed).

[1] `http://archive.ics.uci.edu/ml/`

Table 2. Data sets from UCI ML repository

Data Set	Size	Data Set	Size
1 adult-stretch	20*10	11 balance-scale	625*23
2 lenses	24*12	12 tic-tac-toe	958*29
3 shuttle-landing	15*23	13 car	1728*21
4 post-operative	90*25	14 flare2	1066*32
5 hayes-roth	132*18	15 soybean-large	307*133
6 zoo	101*28	16 dermatology	366*130
7 audiology	26*110	17 breast-cancer	699*110
8 servo	167*19	18 chess	3196*75
9 spect-test	187*23	19 nursery	12960*31
10 house-votes-84	435*18	20 mushroom	8124*119

The smaller the stress value, the better is the reordering. In each row, the smallest values for both neighborhoods are shown in bold. In all cases the mixed technique achieves the best result and conformity scale has the worst result. In most cases, minus technique gives better result than plus technique (16 data sets out of 20 by both measures). This complies with observation by Liiv ([10], p.61): "*the "minus" technique outperformed the "plus" technique on the average with all three measures*", although used data sets and evaluation measures are different.

4 Conclusion

In this paper, we have introduced three matrix reordering methods based on monotone systems and proposed a new one, called mixed technique. We have evaluated all four methods using stress measures. We have found that the novel mixed technique performs better than the other three algorithms. Future challenges include handling big data and dealing with incremental update.

Acknowledgements. The authors would like to thank prof. Sadok Ben Yahia for helpful comments and suggestions.

References

1. Liiv, I.: Seriation and matrix reordering methods: An historical overview. Statistical analysis and data mining 3(2), 70–91 (2010)
2. Liiv, I., Vohandu, L.: Seriation and Matrix Reordering Methods for Asymmetric One-Mode Two-Way Datasets. In: Imaizumi, T., Nakayama, A., Yokoyama, S. (eds.), Advanced Studies in Behaviormetrics and Data Science: Essays in Honor of Akinori Okada, Behaviormetrics: Quantitative Approaches to Human Behavior 5, https://doi.org/10.1007/978-981-15-2700-5_10 (2020)

Table 3. Comparison of reordering algorithms by stress in the von Neumann neighborhood and in the Moore neighborhood.

	Stress in the von Neumann neighborhood				Stress in the Moore neighborhood			
	conf	**plus**	**min**	**mixed**	**conf**	**plus**	**min**	**mixed**
1	404	248	248	**200**	824	516	516	**468**
2	518	386	388	**334**	1054	834	818	**770**
3	338	306	286	**248**	684	644	598	**526**
4	2602	2370	2322	**1720**	5116	4952	4848	**4162**
5	3526	2302	2198	**1830**	6642	5410	5172	**4596**
6	2480	2044	1824	**1164**	5716	5008	4484	**2964**
7	1268	1128	1120	**908**	2636	2384	2376	**2050**
8	3804	2980	2624	**2362**	7272	6392	6050	**5634**
9	5462	5026	4658	**3234**	11408	10752	10078	**7338**
10	11712	7152	7074	**5080**	27204	15868	15534	**12612**
11	15266	14346	13958	**11532**	34766	32590	32348	**29088**
12	38300	35544	34262	**28388**	79808	77764	76332	**69814**
13	46776	38280	38280	**35708**	112804	94208	94208	**92986**
14	19664	18016	18012	**14238**	51728	47568	47896	**40206**
15	27782	24982	23728	**15434**	60224	56060	54472	**38058**
16	37078	34990	32290	**24164**	76834	73466	69046	**55414**
17	25886	25594	24790	**18804**	56068	55816	55020	**44712**
18	241918	229222	212218	**114190**	487386	469402	452762	**308070**
19	471404	369884	383684	**335936**	1162272	952904	977356	**920826**
20	647100	543784	451436	**298416**	1408528	1282768	1132470	**824468**

3. Mullat, J.E.: On the Maximum Principle for some Set Functions. Proceedings of the Tallinn Technical University, 313, 37–44 (1971). `http://www.datalaundering.com/download/modular.pdf`, last accessed 2016/02
4. Mullat, J.E.: Extremal Subsystems of Monotonic Systems. I. Automation and Remote Control, 37(5), 758–766 (1976). `www.datalaundering.com/download/extrem01.pdf`, last accessed 2016/02
5. Mullat, J.E.: Extremal Subsystems of Monotonic Systems. II. Automation and Remote Control, 37(8), 1286–1294, (1977). `www.datalaundering.com/downlaods/extrem02.pdf`, last accessed 2016/02
6. Vyhandu, L.: Rapid data analysis methods. Transactions of Tallinn Technical University, 464, 21–39 (1979). (in Russian)
7. Vyhandu, L: Fast Methods in Exploratory Data Analysis. Transactions of Tallinn Technical University, 705, 3–13 (1989)
8. Võhandu, L., Kuusik, R., Lind, G.: Monotone systems and their applications. Tallinn: Tallinn University of Technology, Department of Software Science (2018). (in Estonian) `issuu.com/monotsys/docs/monotoonsed_s_steemid`, last accessed 2019/07
9. Hahsler, M., Hornik, K., Buchta, C.: Getting things in order: An introduction to the R package seriation, `https://rdrr.io/cran/seriation/`, last accessed 2019/06/27
10. Liiv, I.: Pattern Discovery Using Seriation and Matrix Reordering: A Unified View, Extensions and an Application to Inventory Management. Ph.D. Thesis, Tallinn University of Technology (2008)