研究論文

Scaling of Asymmetric Similarity Matrix with Diagonal Elements

今　泉　　　忠 *
Tadashi IMAIZUMI

Abstract : We find many asymmetric relations from an object to another object of n objects, for example, journal citation data etc. Okada and Imaizumi (1987) proposed a non-metric scaling model and method for analyzing asymmetric similarity matrices. As Okada and Imaizumi's model do not utilize any information of diagonal elements of similarity matrix, some important information will be ignored. A redefined and extended model of Okada and Imaizumi will be proposed to account the diagonal elements and off-diagonal elements of asymmetric similarity matrix. An application to real data set will be shown.

Keywords : Asymmetric Similarity, Diagonal Elements, Radius-Distance Model

Introduction

We find many asymmetric relations from an object to another object of n objects, for example, a trading data between nations, frequency of talking of n persons, journal citation data etc. Tversky(1977) reported the systematic asymmetry in Morse code data. Then we want to explore the complex relationship of n objects in data matrix by some models and methods. Okada and Imaizumi (1987) proposed a non-metric scaling model and method for analyzing asymmetric similarity matrix. This model consists of a common object configuration. In the common object configuration, each object is represented by a point and a circle (sphere, hypersphere) in a Euclidean space. The common object configuration represents pairwise proximity relationships between pairs of objects. Symmetry of data is represented as Euclidean distance between two objects. Asymmetry of data is represented as circles, sphere, and hypersphere etc around the point of object. And this asymmetry is related to the characteristics of each object. Zielman and Heiser's slide-vector model (1996)

* 多摩大学経営情報学部　School of Management and Information Sciences, Tama University

assumes the asymmetric data are induced by sliding of latent dimensions. These models do not utilize any information of diagonal elements of similarity matrix in analyzing data, and are lack of utilizing all information in data. An extended model of Okada and Imaizumi will be proposed to account the diagonal elements and off-diagonal elements of asymmetric similarity matrix. An application to real data set will be shown.

Scaling Model for Asymmetric Similarity Data

Several models and methods have been proposed to analyze asymmetric (dis)similarity matrix. One is the decomposition of similarity matrix, and another is distance-based model. All of these models and methods try to give us the geometric representation of relationship among n objects in similarity matrix. Chino and Okada (1996) reviewed the models and methods of multidimensional scaling for analyzing asymmetric proximity data. Those models were classified into two types, one is a distance based modeling and the other is non-distance based modeling. Constantine and Gower(1978) proposed a decomposition of asymmetric matrix to symmetric matrix and skew-symmetric matrix. Let $\mathbf{S}$ be a asymmetric similarity matrix of size $n \times n$, where n is number of objects we want to analyze. A simple decomposition of this matrix will be

$$\mathbf{S} = \mathbf{A} + \mathbf{B}, \mathbf{A} = (\mathbf{S} + \mathbf{S}')/\mathbf{2}, \mathbf{B} = (\mathbf{S} - \mathbf{S}')/\mathbf{2}. \tag{1}$$

Then the matrix $\mathbf{A}$ is a symmetric matrix, and the matrix $\mathbf{B}$ is a skew symmetric matrix,

$$a_{jk} = a_{kj}, b_{jk} = -b_{kj}$$

Chino(1978) also proposed ASYMSCAL model in which he try to represent $\mathbf{A}$ and $\mathbf{B}$ simultaneous.

$$s_{jk} = a(x_{j1}x_{k1} + x_{j2}x_{k2}) + b(x_{j1}x_{k2} - x_{j2}x_{k1}) + c. \tag{2}$$

This model represents similarity from object j to object k by an inner product between two points j and k, $(x_{j1}x_{k1} + x_{j2}x_{k2})$ and by an outer-product between these two points $(x_{j1}x_{k2} - x_{j2}x_{k1})$. Chino extended ASYMSCAL model to GIPSCAL model. Harshman(1978) proposed DEDICOM model,

$$\mathbf{S} = \mathbf{YAY}', \tag{3}$$

where $\mathbf{A}$ is a asymmetric matrix of $t \times t$. This decomposition is in general form, and Kier(1989) proposed an algorithm to find a $\mathbf{Y}$ and $\mathbf{A}$ from observed $\mathbf{S}$. These approach is based on decomposition of a similarity matrix. And the other approach to analyze an asymmetric similarity matrix will be the distance-based model such as a variants of unfolding model. Let δ_{jk} be dissimilarity from object j to object k,

$$\delta_{jk} \geq 0, \delta_{jj} = 0. \tag{4}$$

And let x_j and y_k be two points in R^t, then An unfolding model assumes

$$\delta_{jk} = \sqrt{\sum_{p=1}^{t}(x_{jp} - y_{kp})^2}, \tag{5}$$

Young(1975) proposed a weighted unfolding model,

$$\delta_{jk} = \left(\sum_{p=1}^{t} w_{ip}(x_{jp} - y_{kp})^2\right)^{1_2}. \tag{6}$$

And Zielman and Heiser (1996) proposed the slide-vector model,

$$\delta_{jk} = \left(\sum_{p=1}^{t}(x_{jp} - x_{kp} + z_p)^2\right)^{1/2},$$

where $\mathbf{z} = [z_p]$ is a so-called 'slide vector'. As $|x_{jp} - x_{kp} + z_p| \neq |x_{kp} - x_{jp} + z_p|$, they represent asymmetry of data. Krumhansl (1978) proposed distance-density model for analyzing similarity matrix.

$$s_{jk} = f^{-1}(\delta_{jk})$$

$$\delta_{jk} = \sqrt{\sum_{p=1}^{t}(x_{jp} - x_{kp})^2} + a\iota_j + b\iota_k,$$

where ι_j is a term of the density of object j, $\iota_j \geq 0$. Okada and Imaizumi(1987) proposed a radius-distance model,

$$m_{jk} = \left(\sum_{p=1}^{t}(x_{jp} - x_{kp})^2\right)^{1/2} - r_j + r_k. \tag{7}$$

The radius $r_j, j = 1, 2, \cdots, n$ is the relative dominance of object j over the other objects. the larger dominant object j is, the smaller the radius of object is. So, the object j is less dominant than the object k, and

$$\delta_{jk} \geq \delta_{kj}$$

As the distance-based model assumes the relation between δ_{jk} and d_{jk} directly, we can capture overall relation among objects by configuration of points. The distance-based models have some advantage than the the decomposition models and methods of asymmetric similarity model in this point. Imaizumi(2010) modified this basic radius-distance model Eq. 7 for the case that the radius r_k changes with the corresponding object j,

$$m_{jk} = d_{jk} - r_j + (1 + a_j + b_k)r_k, b_k \geq 0.$$

This model is a restricted model of

$$m_{jk} = d_{jk} - r_j + c_k.$$

These distance-based models have a weakness on how to handle the diagonal elements of similarity matrix. Rooij and Heiser(2003) proposed a model for analyzing the contingency table(s),

$$s_{jk} = g + r_j + c_k - d_{jk}^2. \tag{8}$$

where g is grand mean, r_j with $r_j > 0$ is row-effect, and c_k with $c_k > 0$ is column-effect. This model analyzes the similarity data as metric data, and the diagonal element s_{jj} of $\mathbf{S} = [s_{jk}]$, is represented by the row-effect and the column-effect. In general, the diagonal elements of similarity matrix will indicate the degree of self-similarity of object itself, and have some important information about the point of brand royalty of that object etc.

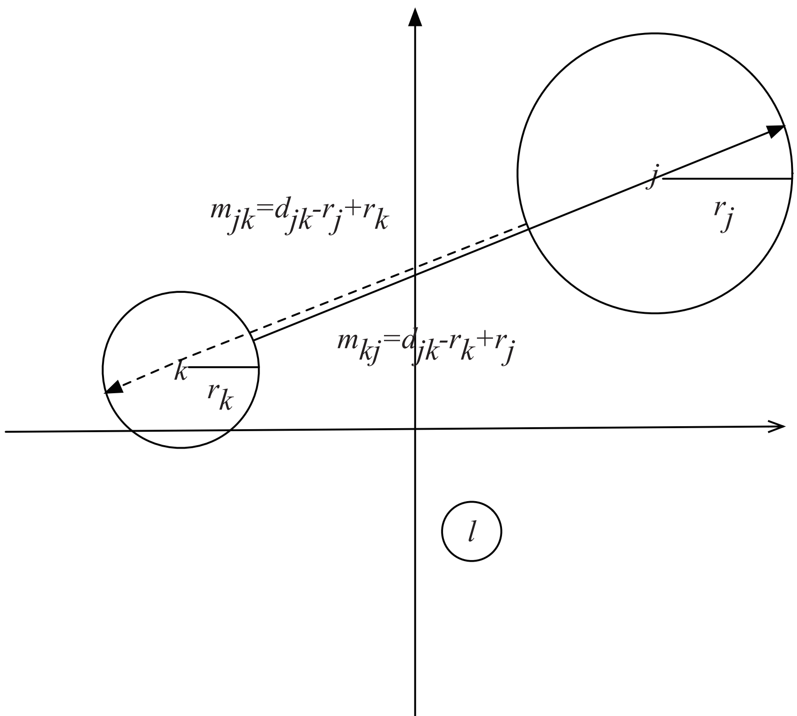

Figure 1. The Radius-Distance Model

Rooij and Heiser's model 8 is an interesting model, but, is appropriate for the data source such as frequency table only. And the self-similarity will be compounded by row-effect and column-effect in their model. And we propose a asymmetric scaling model which handles the diagonal elements of similarity matrix in this paper. Imaizumi(2012) proposed a model for handling these diagonal elements. The proposed model assumed 1) similarity s_{jk} showed the degree of flows of communication channels between two objects, and the asymmetricty of data came from the difference of outflows and inflows of these two objects.

$$s_{jk} = a_j \frac{1}{b_k} e^{-d_{jk}^2 + c_k}$$

So, the log-transformed similarity will be

$$log(s_{jk}) = -\left(d_{jk}^2 - log(a_j) + log(b_k) - c_k \right).$$

This model is similar to that of Roij and Heiser, however, has the difficulty of the assumption on communication channels that the squared distance was assumed to decrease exponentially. This assumption is too strict to analyze the asymmetric similarity by using a distance based model in general. And we need to reconsider this proposed model.

On Distance Based Modeling of the Diagonal Elements

We assume the radius-distance model basically. The difference of similarities between object j and object k, $s_{jk} - s_{kj}$, shows how the object j dominates over the object k. The similarity in the diagonal element will be interpreted how object j dominates over the other $n-1$ objects. In the radius-distance model, the dominance of object j over object k is represented as

$$r_j - r_k.$$

The larger s_{jj} indicates that object j will be an object bound inward from the other $n-1$ objects. We extend the radius-distance model to handle the self-similarity. Let d_j^* be the relative degree of the dominance of object j over the other $n-1$ objects, and be defined by

$$d_j^* = r_j - \frac{1}{n-1} \sum_{k=1, k \neq j}^{n} r_k. \tag{9}$$

The smaller d_j^* indicates that object j more dominant over the other $n-1$ objects. And the similarity will be represented by a modified radius-distance model,

$$m_{jk} = \begin{cases} d_{jk} - r_j + r_k & j \neq k, \\ d_j^* = -\frac{1}{n-1} \sum_{k=1, k \neq j}^{n} r_k + r_j & j = k. \end{cases} \tag{10}$$

Then we want to obtain m_{jk} which is represented as a decreasing function of s_{jk}, especially a monotonic decreasing function of s_{jk}, that is

$$s_{jk} \preceq s_{j'k'} \rightarrow m_{jk} \geq m_{j'k'}. \tag{11}$$

Parameter Estimation

We use a convenient minimizing function to estimate the model parameters $\{x_{jp}\}$ and $\{r_j\}$. We try to find the model parameters which minimizes

$$Loss(\mathbf{X}, \mathbf{r}) = \frac{\sum_{j=1}^{n} \sum_{k=1}^{n} (\hat{m}_{jk} - m_{jk})^2}{\sum_{j=1}^{n} \sum_{k=1}^{n} \hat{m}_{jk}^2}, \tag{12}$$

where $\mathbf{X} = [x_{jp}]$, $\mathbf{r} = [r_j]$. And $\hat{m}_{jk}, j = 1, 2 \cdots, n; k = 1, 2, \cdots, n$ are the monotonic disparities by Kruskal(196) which keeping the order relation among similarities. Figure 2 show the flow chart to seek the model parameters. This computational procedures for fixing the dimensionality consists of two parts mainly. One is to derive the monotonic disparities, and the other is to seek model parameters for fixing $\{\hat{m}_{jk}, j = 1, 2 \cdots, n; k = 1, 2, \cdots, n\}$.
How to derive m_{jk} is one key algorithm in non-metric scaling methods. Let $\{s_{jk}; j = 1, 2, \cdots, n; k = 1, 2, \cdots, n\}$ be a set of similarity from object j to object k, and $\{m_{jk}; j = 1, 2, \cdots, n; k = 1, 2, \cdots, n\}$ be the corresponding estimates. We seek $\{m_{jk}\}$ such that

$$if \quad s_{jk} \preceq s_{j'k'} \quad then \quad m_{jk} \geq m_{j'k'} \tag{13}$$

Unfortunately, the given $\{m_{jk}\}$ did not satisfied the order relation in $\{s_{jk}\}$. Then we need to improve $\{m_{jk}\}$ as they satisfy. We introduce the disparities $\{\hat{m}_{jk}\}$ which satisfy

$$if \quad s_{jk} \preceq s_{j'k'} \quad then \quad \hat{m}_{jk} \geq \hat{m}_{j'k'} \tag{14}$$

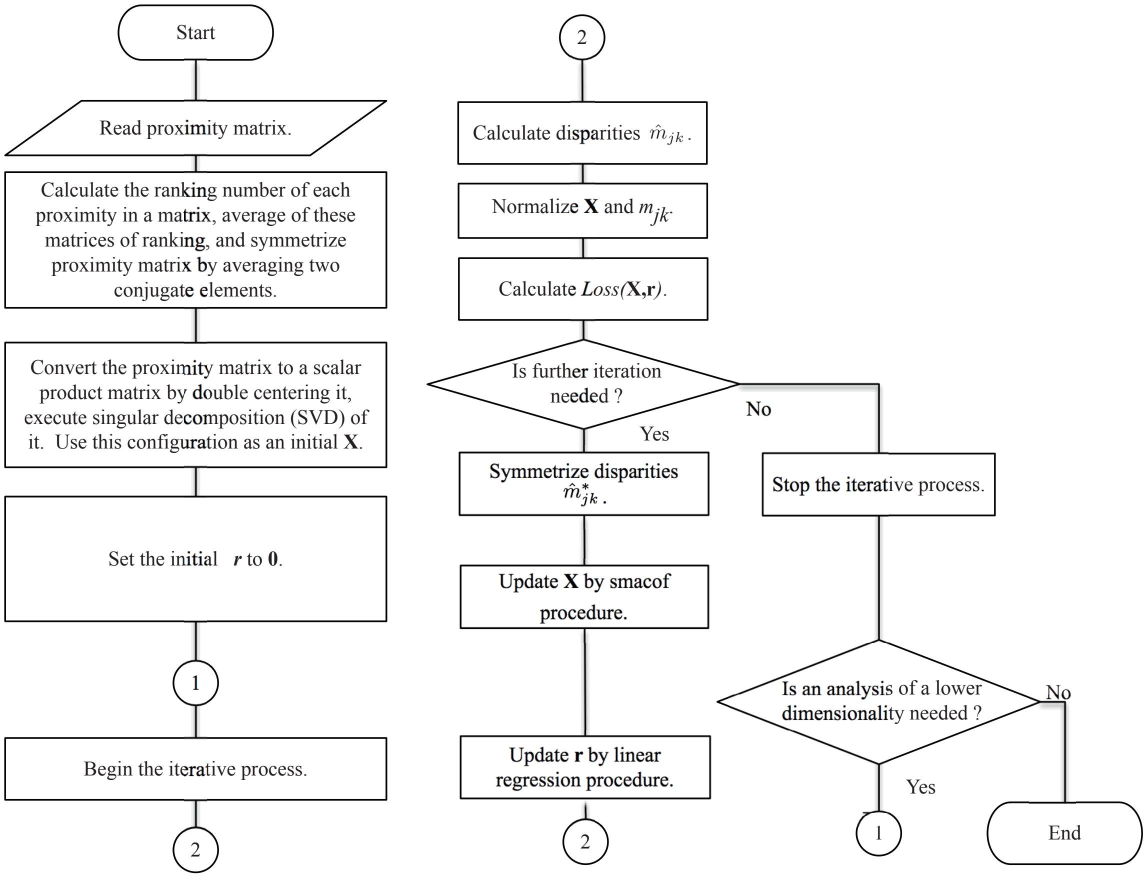

Figure 2. Flow Chart of the proposed computational procedure

And the loss of order relation is defined by

$$S(\hat{m}, m) = \sum_{j=1}^{n}\sum_{k=1}^{n}(\hat{m}_{jk} - m_{jk})^2. \tag{15}$$

How to derive those disparities is the very important. We emploies $\hat{m}_{jk}$ as values given by PAVA Isotonic Regression algorithm(de Leeuw, Hornik and Mair,2009). The original algorithm was given by Kruskal(1964).

The procedure to estimate model parameters is an iterative optimization procedure, and each optimization consists of the two steps. We can compute $\hat{m}_{jk}$ for given $\mathbf{X}$ and $\mathbf{r}$ and $Loss(\mathbf{X}, \mathbf{r})$. We stop the iterative optimization process in the case of $Loss(\mathbf{X}, \mathbf{r})$ being sufficiently small. But, if it was not, we need to improve $\mathbf{X}, \mathbf{r}$. Before applying those inner two steps for improving, $\mathbf{X}$ and $\mathbf{r}$ are normalized such as

$$\sum_{j=1}^{n}\sum_{k=1}^{n}\hat{m}_{jk}^2 = n^2. \tag{16}$$

The first step is to update $\mathbf{X}$ by using smacof procedure (Mair and de Leeuw,2009). And The average of two corresponding two pairs,

$$\hat{m}^*_{jk} = (\hat{m}_{jk} + \hat{m}_{kj})/2,$$

are calculated. Then metric smacof procedure is applied to find $\mathbf{X}$ which minimizes

$$\sum_{j=1}^{n-1} \sum_{k=j+1}^{n} (\hat{m}^*_{jk} - d_{jk})^2.$$

by treating $\{m_{jk}\}$ are observed dissimilarities. Then, the second step is to update $\mathbf{r}$. Let $\mathbf{y}$ be the vectorized deviates of $\hat{m}$ from d_{jk},

$$y_l = \hat{m}_{jk} - d_{jk}, l = 1, 2, \cdots, n^2,$$

where $l = (j-1) \times n + k$. And $\mathbf{P}^* = [p^*_{lh}]$ is a corresponding design matrix of $n^2 \times n$ such

$$p^*_{lj} = -1, p^*_{lk} = 1 \quad \text{if} \quad j \neq k,$$

$$p^*_{lj} = -1, p^*_{lk'} = -\frac{1}{n-1}, k' = 1, 2, \cdots, j-1, j+1, \cdots, n, \qquad \text{if} \quad j = k.$$

Then a linear least squares estimation procedure is applied to update $\mathbf{r}$,

$$\mathbf{y} \approx \mathbf{Pr}$$

Application

We applied the proposed model to Brand Switching Data for colas. Borg and Groenen(2005) analyzed the Brand Switching Data for 15 colas. Table 1 shows names of 15 colas. They applied several asymmetric MDS to the data converted dissimilarity in the sense of the gravity model and compared them.

Table 1

List of 15 colas

Coke decaf	Coke diet decaf	Pepsi diet decaf	Pepsi decaf	Canfield
Coke	Coke classic	Coke diet	Pepsi diet	RC diet
Rite diet	Pepsi	Private label	RC	Wildwood

We applied the proposed model to the same Brand Switching Asymmetric Data for comparison. Figure 3 shows the two dimensional configuration and the joint plot of radius of each object by the proposed model. Figure 4 shows those by the original radius-distance model. Two object configurations look very similar. Colas of decaf buyers change to the colas of same type or the cola of diet type. As Borg and Groenen(2005, Pp.513) pointed out, these figures indicates that more households are changed from Coke decaf to Rite diet or

RC diet. Figure 5 shows scatter plotting of two **r** of 15 colas, the r_j by the original radius-distance model were token on the x-axis of Figure 5, and that by the proposed model were token on the y-axis. Each circle in Figure 5 shows the point of each cola. Each pairs will be plotted on a line, but the slope is greater than 1 and the relative dominance is smaller. Coke decaf is the least dominant object on both models. And it has the largest radius on the proposed model, it will be changed to other colas. Figure 6 shows the ratio of r_j of the proposed model to that of the radius-distance model. Coke classic,Coke diet, Rite diet and Pepsi show the higher ratio in this figure, that is, those object will be relative switchable colas.

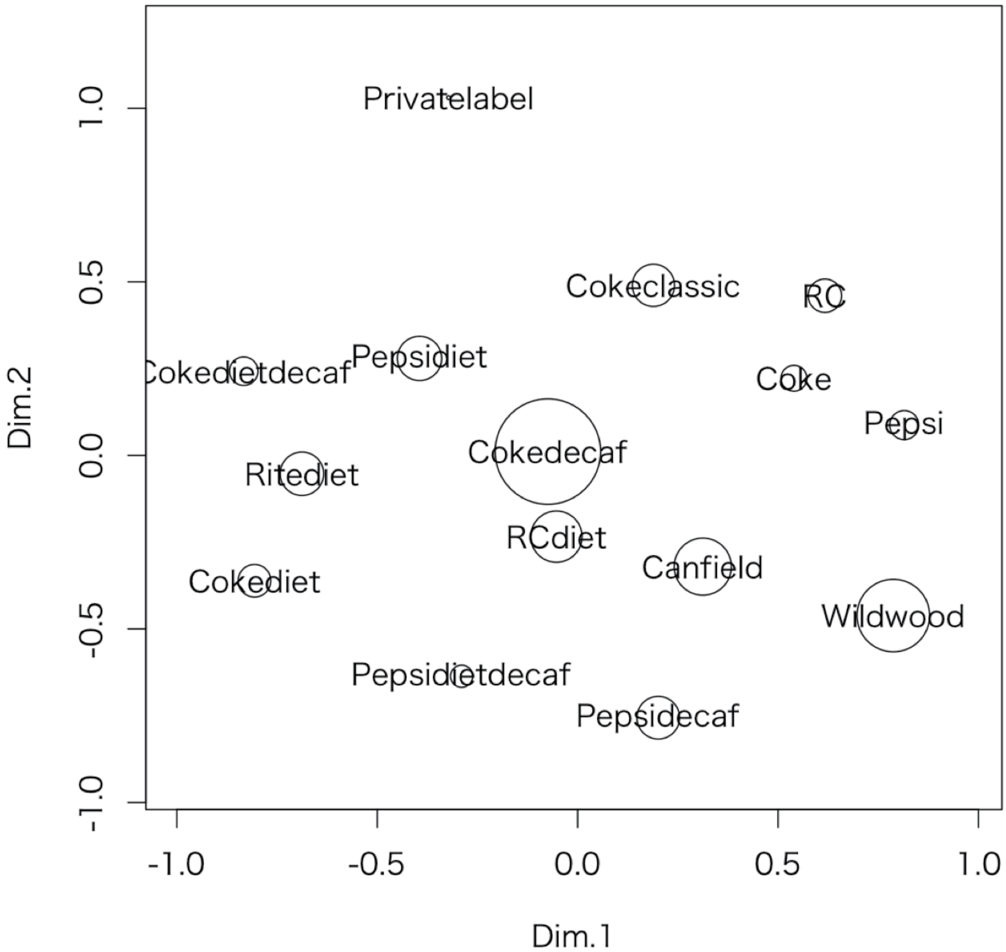

Figure 3. Joint Configuration by the Proposed Model

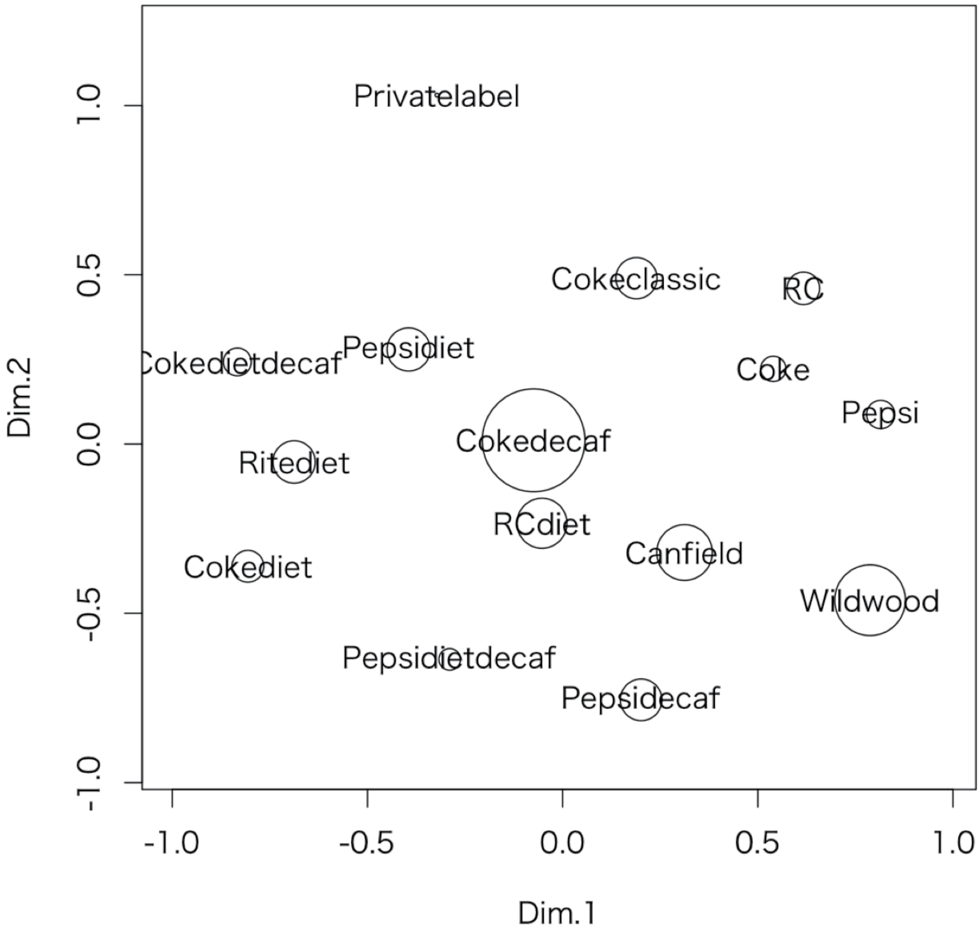

Figure 4. Joint Configuration by the Original Model

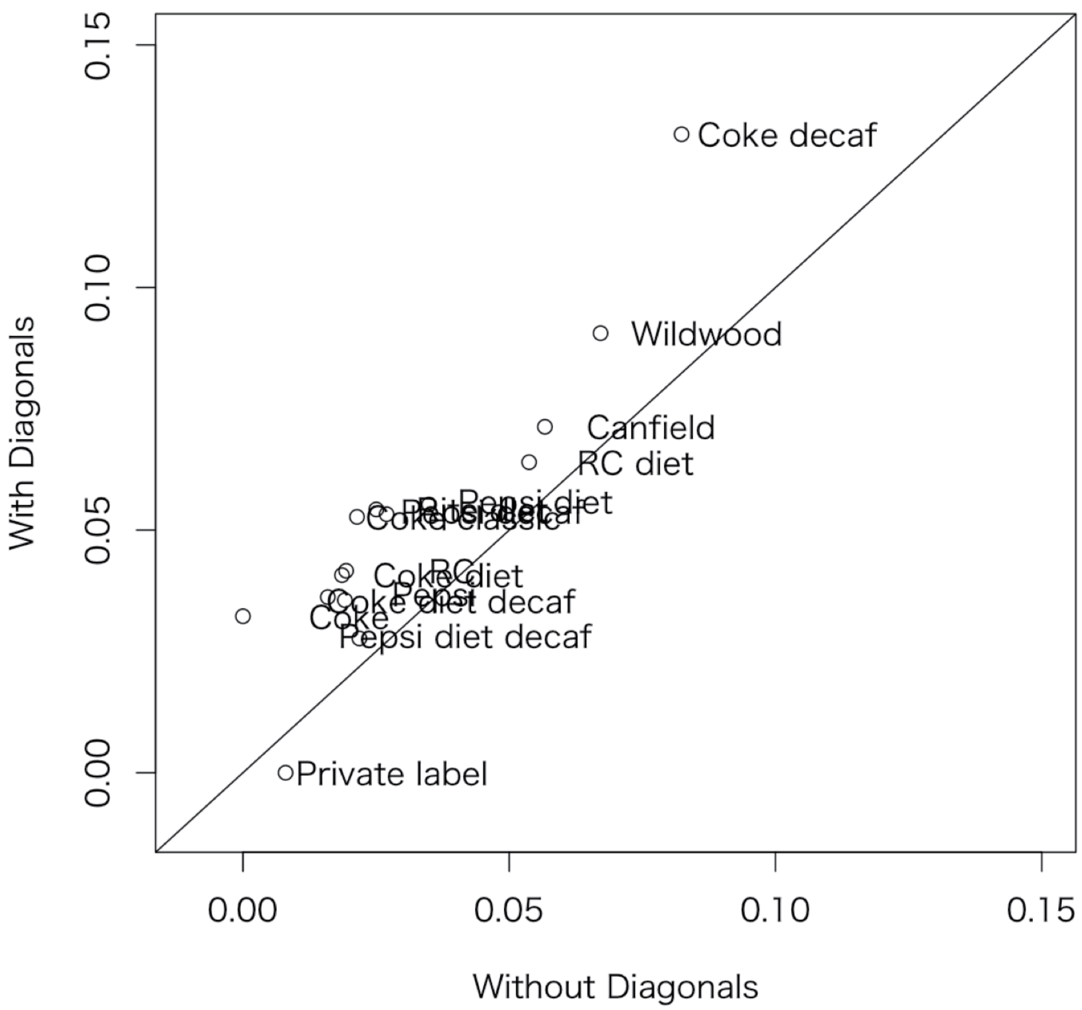

Figure 5. Scatter Plot of Pairs of Radii

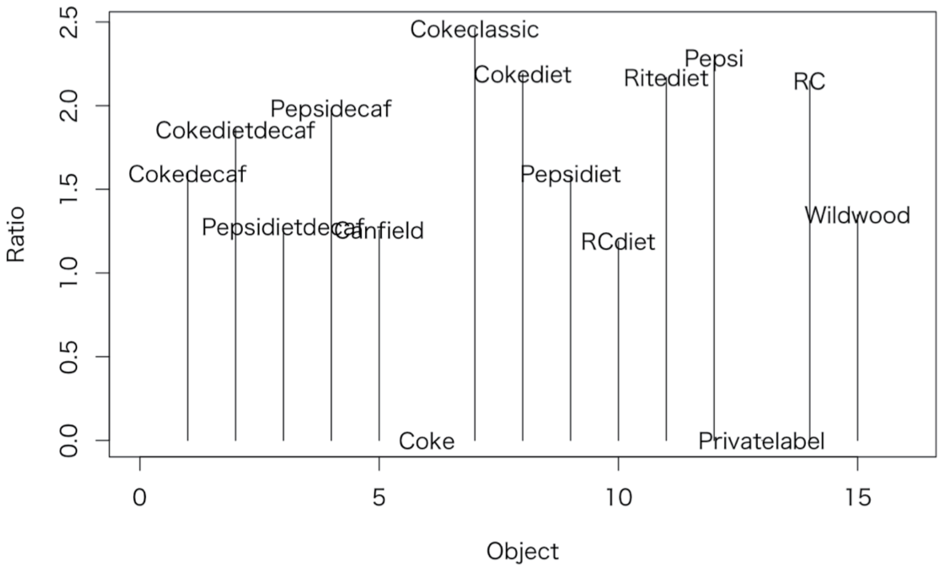

Figure 6. Ratio of Corresponding r_j

Conclusion

We proposed a modified radius-distance model for analyzing asymmetric similarity matrix with diagonal elements. The self-similarity is accounted by the term the relative dominance of each object. In application to the Brand Switching data of 15 colas, the estimated r_j indicates that Coke decaf and Wildwood may have less brand royalty than others. This proposed model will be extended to analyze the two mode three way data from several sources $i, i = 1, 2, \cdots, N$, for example,

$$m_{ijk} = d_{ijk} - u_i r_j + u_i r_k, \tag{17}$$

for $j \neq k$ and

$$m_{ijj} = u_i\left(-\frac{1}{n-1}\sum_{k=1, k\neq j}^{n} r_k + r_j\right), \tag{18}$$

for $j = k$ for each source i.

Acknowledgment

I acknowledge for the anonymous reviewers for their valuable comments.

References

[1] Borg,I., and Groenen, P. J. F: Modern multidimensional scaling. Springer(2005).

[2] Chino, N.,:A graphical technique for representing the asymmetric relationships between N objects. Behaviormetrika, **5**,

[3] Chino, N. and Okada. A.: Asymmetric Multidimensional Scaling and Related Topics. The Japanese Journal of Behaviormetrics,**23**, 130-152(1996). In Japanese.

[4] Constantine, A. G., and Gower, J. C.:Graphic representations of asymmetric matrices. Applied Statisitcs,**27**, 297-304(1978).

[5] de Leeuw, J., Hornik, K., and Mair, P.:Isotone Optimization in R: Pool-Adjacent-Violators Algorithm (PAVA) and Active Set Methods,Journal of Statistical Software, **32**, 1-24(2009).

[6] de Rooij, M., Heiser W.: A distance representation of the quasi-symmetry model and related distance models. In: Yanai, H., Okada, A., Shigemasu, K., Kano, Y., Meul-man, J. (Eds.), New developments in Psychometrics, pp. 487-494. Tokyo: Springer Verlag.(2002).

[7] Harhman, R.: Models for Analysis of Asymmetrical Relationships Among N Objects or Stimuli, Paper presented at the First Joint Meeting of the Psychometric Society and the Mathematical Psycology, McMaster University, Hamilton, Ontatrio, August(1978).

[8] Imaizumi, T.: Asymmetric Multidimensional Scaling for Directed Similarity Matrix, Tama University Journal of Management & Information Sciences,**14**, 63-71(2010). In Japanese.

[9] Imaizumi, T. : Scaling of Asymmetric Similarity Matrix by a Communication Chan-nels Model, Paper presented at Joint meeting of the Japanese Classification Society and the Italian Classification and Data Analysis Group, Villa Orlandi,Anacapri, Italy, September(2012).

[10] Keirs, H.A.L., : An Alternating Least Squares Algorithm for Fitting the Two- and Three-way DEDICOM Model and the IDIOSCAL Model, Psychometrika, **54**, 515-521(1989).

[11] Krumhansl, C. L.: Concerning the Applicability of Geometric Models to Similarity Data: The Interrelationship between Similarity and Spatial Density, **85**, 445-463(1978).

[12] Kruskal, J. B: Nonmetric Multidimensional Scaling: A Numerical Method, Psychometrika, **29**, 115-129(1964).

[13] Mair, P., and de Leeuw, J.: Multidimensional Scaling Using Majorization: SMACOF in R. Journal of Statistical Software, **31**, 1-30 (2009).

[14] Okada, A., and Imaizumi, T.: Geometric models for asymmetric similarity data. Be-haviormetrika, **21**, 81-96 (1987).

[15] Tverksy,A.: Features of similarity. Psychological Review, **84**, 327-352(1977)

[16] Zielman, B., and Heiser, W. J.: Models for asymmetric proximities. Britsh Journal of Mathematical and Statistical Psychology, **49**, 127-147 (1996)