Behaviormetrika
Vol.43, No.1, 2016, 1–17

DISCRIMINANT COORDINATES FOR ASYMMETRIC DISSIMILARITY DATA BASED ON RADIUS MODEL

Kensuke Tanioka* and Hiroshi Yadohisa**

Asymmetric multidimensional scaling is a visualization method that reveals relations between objects using their asymmetric dissimilarity data as input. Such visualization is important in areas such as marketing science since it reveals consumer dynamics and brand competition. Large amounts of asymmetric dissimilarity data can be processed by modern information technology; nevertheless, it remains difficult to interpret the asymmetries between many objects. To overcome this problem, we propose a new visualization method called discriminant coordinates for asymmetric dissimilarity data, which interprets the relations between and within a set of object classes, given asymmetric dissimilarity data. The method easily interprets the estimated asymmetries between and within classes, even when the asymmetric dissimilarity data contain noise if some assumptions for asymmetric dissimilarity data is satisfied.

1. Introduction

Asymmetric dissimilarity occurs when the dissimilarity from object i to object j does not necessarily equal that from object j to object i. Asymmetric dissimilarity data appear in various fields such as marketing area. The object relations in asymmetric dissimilarity data can be visualized by asymmetric multidimensional scaling (asymmetric MDS) (Borg and Groenen, 2005; Chino, 2012; Saito and Yadohisa, 2005). The radius model is a well-known asymmetric MDS model (Okada and Imaizumi, 1987) applied to brand switching data. In this application, the brand switching data from i to j are expressed as the purchase frequency of brand j by previous purchasers of brand i. The radius model reveals competitive relations among the brands. It describes the symmetric parts as object coordinates in d-dimensions, and the skew-symmetric parts as the radii of the objects. If the relations between objects are largely asymmetric, object tends to be assigned to the large radii, and the other to the small radii. On the other hand, if the relations have little asymmetry, these objects tend to have almost the same radii.

Large and complex asymmetric dissimilarity data can be processed by modern information technologies. If the number of objects is large, however, it is difficult to interpret asymmetries between objects. In this case, if the information of classes of objects can be given, we can utilize asymmetries between and within classes, instead of asymmetries of objects, which make the interpretation easier, although we need to assume that objects belonging to the same class are similar each other and these

Key Words and Phrases: Asymmetric Dissimilarity Data, Asymmetric Multidimensional Scaling, Clustering, Simultaneous Optimization

* Graduate School of Culture and Information Science, Doshisha University
** Department of Culture and Information Science, Doshisha University
Mail Address: kensuke.t0628@gmail.com
phone: +81 774 65 7657

asymmetric relations are simple. Therefore, it is important to reveal asymmetries between and within classes using the asymmetric dissimilarity data and the classes of objects as inputs if objects belonging to the same class are similar each other and their asymmetries are simple. For example, there exist tea brand switching data. In Japan, consumers may choose among numerous brands and types of tea, such as green tea and oolong. In this case, detecting the asymmetric relations between and within tea brands classes characterized by types is important for interpreting the features of tea, if each class has the common features. If given object classes can be interpreted by asymmetric MDS, such methods may become readily adaptable to the asymmetric dissimilarity data. To achieve this purpose, the following two properties of asymmetric MDS should be satisfied; the resulting classes should be discriminated from each other, and the asymmetries within and between classes should be interpreted easily.

To satisfy the first property, when interpreting class features in asymmetric MDS, objects belonging to the same class should be closely spaced, while those belonging to different classes should be well-separated. One method for visualizing the differences between classes is the discriminant coordinates (DC) method (Seber, 1984), which is applicable to multivariate data rather than dissimilarity data, but which has motivated our approach. Another technique, called cluster difference scaling (CDS) (Heiser & Groenen, 1997) is unsuited to our purpose. CDS is designed only for symmetric dissimilarity data, meaning that asymmetries cannot be interpreted from CDS results. Therefore, no existing method can satisfy the first property.

To satisfy the second property, asymmetries within classes should be simply structured, even at the cost of sacrificing the correctness of the asymmetries to some degree. Succinctly, the asymmetries within classes are described by a few kinds of length of radii, which can be used to classify objects into a small number of clusters within classes based on the lengths of radii. This purpose is known to be achieved by a tandem method of k-means (MacQueen, 1967) and an asymmetric MDS based on the radius model. The tandem method was reported by Kiers et al. (2005) for symmetric MDS.

The tandem method involves of two steps to satisfy the second property. In the first step, the radius model is applied to the asymmetric dissimilarity data, and in the second step, the k-means method is applied to the estimated radii within each class in the second step. From the results of the tandem method, the lengths of radii for objects can be represented by a small number of clusters through different shapes or colors of objects. However, this approach sometimes fails to classify objects based on the lengths of the radii; as an example, the results of tandem clustering are presented in Figure 1. The left and right panels display the true data and the result of tandem clustering with noise, respectively. The true data consist of two classes, each with two kinds of radii lengths. The true asymmetric dissimilarity data are generated from the true data, and noise are added to the dissimilarity data. Next, the tandem method is applied to the asymmetric dissimilarity data. Figure 1 shows that the tandem method sometimes mislabels the estimated object radii if these dissimilarities include noise. Moreover, the objects within each class are spatially dispersed because

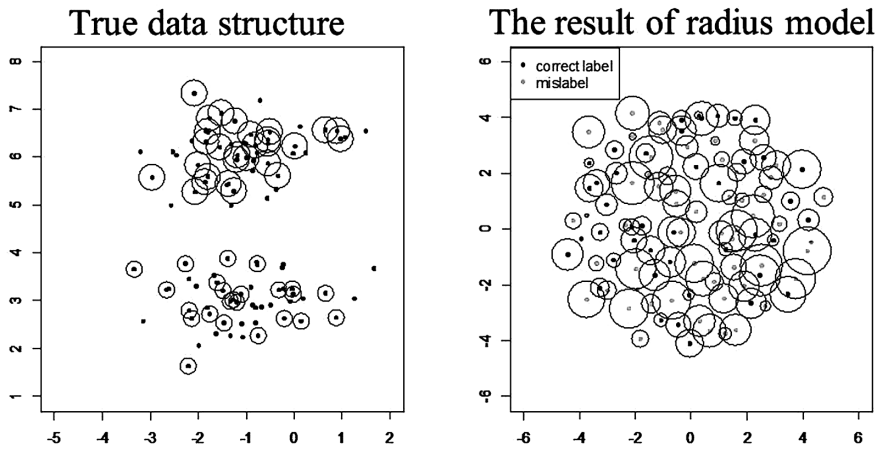

Figure 1: True structure with the number of classes equal to 2 (the left figure) and the result of the tandem clustering with noise (the right figure).

the radius model is unconstrained obscuring the features of the classes. Concretely, when some of the asymmetric dissimilarity data within the same class are noisy, the tandem method delivers poor results, although objects belonging to the same class are assumed to be located close to one another. In this situation, the information about the classes can be used, to recover the true structure.

In this paper, we propose a new technique, called discriminant coordinates for asymmetric dissimilarity data (DCADD) method, which simultaneously discriminates the coordinates of objects based on the object classes to classify the objects based on estimated radii. The proposed method has two advantages. First, the estimated coordinates of objects belonging to the same class are closely located allowing easy interpretation of the features of skew-symmetric parts. In addition, the axis can be easily interpreted using information about the classes. Second, when objects belonging to the same class have a few types of radii lengths, the features of objects having radii within each class becomes simple to interpret. The proposed method can detect the clustering structure of the objects according to radii lengths, even when the asymmetric dissimilarity data are very noisy.

The remainder of this paper is structured as follows. Section 2 introduces the asymmetric dissimilarity data, the proposed model, and the objective function. The properties of the proposed objective function and the estimation process are discussed in Section 3. Finally, the superiority of the proposed method is demonstrated through simulations and real applications in Section 4 and 5, respectively.

2. Model and the objective function of the DCADD method based on the radius model

This section introduces the constrained radius model and defines the proposed objective function. To this end, we define asymmetric dissimilarity data and the

indicator matrix.

The asymmetric dissimilarity data are expressed as follows:

$$\boldsymbol{\Delta} = (\delta_{ij}),\ \delta_{ij} \in \mathbb{R}^{+}\ (i, j = 1, 2, \cdots, n),$$

where $\neg(\delta_{ij} = \delta_{ji})$ and $\delta_{ii} = 0$ for all $i, j = 1, 2, \cdots, n$, n is the number of objects, and $\mathbb{R}^{+}$ is the set of positive real number. If the value of δ_{ij} $(i, j = 1, 2, \cdots, n)$ is large, the difference from i to j is considered to be large. Conversely, if the value of δ_{ij} $(i, j = 1, 2, \cdots, n)$ is small, difference from i to j is considered to be small.

Next, the indicator matrix is defined as follows:

$$\boldsymbol{U} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_k) = (u_{is}),\ u_{is} = \begin{cases} 1 & (\text{object } i \text{ belong to class } s) \\ 0 & (\text{others}) \end{cases} \tag{1}$$
$$(i = 1, 2, \cdots, n; s = 1, 2, \cdots k; k \leq n)$$

where, k is the number of classes.

Given asymmetric dissimilarity data $\boldsymbol{\Delta}$ and an indicator matrix $\boldsymbol{U}$, model of the discriminant coordinates for the asymmetric dissimilarity data are determined by the following radius model approach:

$$\sum_{s=1}^{k}\sum_{t=1}^{k} w_{st}\boldsymbol{\Gamma}_s\boldsymbol{\Delta}\boldsymbol{\Gamma}_t = \boldsymbol{D}(\boldsymbol{X}) - \mathbf{1}(\boldsymbol{\Psi}(\boldsymbol{I}_{k^*} \otimes \boldsymbol{U})\boldsymbol{r}^*)^T + \boldsymbol{\Psi}(\boldsymbol{I}_{k^*} \otimes \boldsymbol{U})\boldsymbol{r}^*\mathbf{1}^T + \boldsymbol{E}, \tag{2}$$

where $\boldsymbol{\Gamma}_s = \mathrm{diag}(\boldsymbol{u}_s)$,

$$w_{st} = \begin{cases} a_s & (s = t) \\ 1 & (\text{others}) \end{cases} \quad \text{subject to } \sum_{s=1}^{k} a_s = 1, \text{ and } a_s \geq 0\ (s = 1, 2, \cdots, k), \tag{3}$$

the Euclidean distance matrix of the estimated coordinates matrix

$$\boldsymbol{D}(\boldsymbol{X}) = (d_{ij}(\boldsymbol{X})),\ \big(d_{ij}(\boldsymbol{X}) = \|\boldsymbol{x}_{(i)} - \boldsymbol{x}_{(j)}\|\big)(i, j = 1, 2, \cdots, n)$$
$$\text{where } \boldsymbol{X} = (\boldsymbol{x}_{(1)}, \boldsymbol{x}_{(2)}, \cdots, \boldsymbol{x}_{(n)})^T, \boldsymbol{x}_{(i)} \in \mathbb{R}^d\ (i = 1, 2, \cdots, n),$$

the indicator matrix of objects for radii

$$\boldsymbol{\Psi} = (\boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \cdots, \boldsymbol{\Psi}_{k^*}), \quad \boldsymbol{\Psi}_\ell = \mathrm{diag}(\psi_{1\ell}, \psi_{2\ell}, \cdots, \psi_{n\ell}),$$
$$\psi_{i\ell} \in \{0, 1\}\ (i = 1, 2, \cdots, n; \ell = 1, 2, \cdots, k^*) \quad \sum_{\ell=1}^{k^*} \psi_{i\ell} = 1\ (i = 1, 2, \cdots, n),$$

where k^* is the number of clusters, esimated radii by clusters

$$\boldsymbol{r}^* = (\boldsymbol{r}_1^{*T}, \boldsymbol{r}_2^{*T}, \cdots, \boldsymbol{r}_{k^*}^{*T})^T, \quad \boldsymbol{r}_\ell^* = (r_{1\ell}^*, r_{2\ell}^*, \cdots, r_{k\ell}^*)\ (\ell = 1, 2, \cdots, k^*),$$
$$r_{s\ell}^* \geq 0\ (s = 1, 2, \cdots, k; \ell = 1, 2, \cdots, k^*),$$

$\mathbf{1} = (1, 1, \cdots, 1)$, $\boldsymbol{I}_{k^*}$ is $k^* \times k^*$ identity matrix, $\otimes$ is kronecker product, and

$\boldsymbol{E} = (e_{ij}), e_{ij} \in \mathbb{R}\ (i, j = 1, 2, \cdots, n)$ is error matrix.

To reduce the internal within-class variation, w_{st} $(s, t = 1, 2, \cdots, k)$ is introduced. Objects belonging to the same class are assumed to have k^* kinds of radii. In the proposed method, the radius model is expressed as follows:

$$\boldsymbol{\Psi}(\boldsymbol{I}_{k^*} \otimes \boldsymbol{U})\boldsymbol{r}^* = \sum_{\ell=1}^{k^*} \boldsymbol{\Psi}_\ell \boldsymbol{U} \boldsymbol{r}_\ell^*, \tag{4}$$

and an elementary description of Eq. (4) is given by:

$$\sum_{s=1}^{k} u_{is} \sum_{\ell=1}^{k^*} \psi_{i\ell} r_{s\ell}^* \ (i = 1, 2, \cdots, n)$$

Therefore, if $\psi_{i\ell} = 1$ and object i belongs to class s, then object i belongs to cluster ℓ of class s :conversely, if $\psi_{i\ell} = 0$ and object i belongs to class s, object i does not belonging to cluster ℓ of class s.

An example of DCADD is presented in Figure 2,

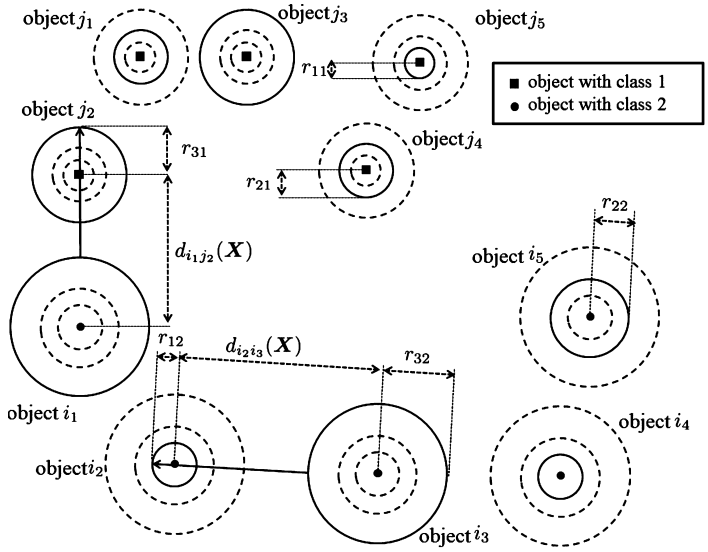

Figure 2: The model of DCADD in case of 2 classes and 3 clusters.

The DCADD result is interpreted in a manner similar that of the radius model, but with an additional consideration: objects belonging to the same class are divisible into k^* clusters, and these radii differ among these clusters.

We next introduce the objective function of the discriminant coordinates of the asymmetric dissimilarity data. This function is based on the radius model and Eq. (2).

$$L(\boldsymbol{W},\boldsymbol{X},\boldsymbol{\Psi},\boldsymbol{r}^*|\ \boldsymbol{\Delta},\boldsymbol{U})$$
$$= \left\| \sum_{s=1}^{k}\sum_{t=1}^{k} w_{st}\boldsymbol{\Gamma}_s\boldsymbol{\Delta}\boldsymbol{\Gamma}_t - \left[\boldsymbol{D}(\boldsymbol{X}) - \mathbf{1}(\boldsymbol{\Psi}(\mathbf{1}_{k^*}\otimes\boldsymbol{U})\boldsymbol{r}^*)^T + \boldsymbol{\Psi}(\mathbf{1}_{k^*}\otimes\boldsymbol{U})\boldsymbol{r}^*\mathbf{1}^T\right]\right\|^2 \longrightarrow \text{Min} \tag{5}$$

subject to $r^*_{s\ell} \geq 0$ $(s = 1, 2, \cdots, k;\ \ell = 1, 2, \cdots, k^*)$, subjection (3), and $\psi_{i\ell} \in \{0, 1\}$ $(i = 1, 2, \cdots, n;\ \ell = 1, 2, \cdots, k^*)$.

The tasks of the objective function are to estimate the weights within the clusters, the object coordinates, the indicator matrix of the radii, and the radius parameters.

3. DCADD algorithm based on the radius model

The DCADD algorithm is based on the alternative least squares criterion (ALS) (Young, et al., 1980). The objective function is first decomposed into its symmetric and skew-symmetric parts, and each parameter in the objective function is then updated. The proposed objective function is decomposed identically to the objective function in the radius model (Borg and Groenen, 2005; Bove and Critchley, 1993).

We put $\boldsymbol{\Delta}^* = \sum_{s=1}^{k}\sum_{t=1}^{k} w_{st}\boldsymbol{\Gamma}_s\boldsymbol{\Delta}\boldsymbol{\Gamma}_t$, and $\boldsymbol{\Delta}^*$ can be decomposed as follows:

$$\boldsymbol{\Delta}^* = \frac{1}{2}(\boldsymbol{\Delta}^* + \boldsymbol{\Delta}^{*T}) + \frac{1}{2}(\boldsymbol{\Delta}^* - \boldsymbol{\Delta}^{*T}). \tag{6}$$

In Eq. (6), $(\boldsymbol{\Delta}^*+\boldsymbol{\Delta}^{*T})/2$ and $(\boldsymbol{\Delta}^*-\boldsymbol{\Delta}^{*T})/2$ denote the symmetric and skew-symmetric part, respectively. From this decomposition, the objective function can be decomposed into symmetric and skew-symmetric parts as follows:

$$L(\boldsymbol{W},\boldsymbol{X},\boldsymbol{\Psi},\boldsymbol{r}^*|\ \boldsymbol{\Delta},\boldsymbol{U})$$
$$= \left\|\boldsymbol{\Delta}^* - \left[\boldsymbol{D}(\boldsymbol{X}) - \mathbf{1}(\boldsymbol{\Psi}(\boldsymbol{I}_{k^*}\otimes\boldsymbol{U})\boldsymbol{r}^*)^T + \boldsymbol{\Psi}(\boldsymbol{I}_{k^*}\otimes\boldsymbol{U})\boldsymbol{r}^*\mathbf{1}^T\right]\right\|^2$$
$$= \left\|\frac{1}{2}(\boldsymbol{\Delta}^*+\boldsymbol{\Delta}^{*T}) + \frac{1}{2}(\boldsymbol{\Delta}^*-\boldsymbol{\Delta}^{*T}) - \left[\boldsymbol{D}(\boldsymbol{X}) - \mathbf{1}(\boldsymbol{\Psi}(\boldsymbol{I}_{k^*}\otimes\boldsymbol{U})\boldsymbol{r}^*)^T + \boldsymbol{\Psi}(\boldsymbol{I}_{k^*}\otimes\boldsymbol{U})\boldsymbol{r}^*\mathbf{1}^T\right]\right\|^2$$
$$= \left\|\frac{1}{2}(\boldsymbol{\Delta}^*+\boldsymbol{\Delta}^{*T}) - \boldsymbol{D}(\boldsymbol{X})\right\|^2 \tag{7}$$
$$+ \left\|\frac{1}{2}(\boldsymbol{\Delta}^*-\boldsymbol{\Delta}^{*T}) - \left[\boldsymbol{\Psi}(\boldsymbol{I}_{k^*}\otimes\boldsymbol{U})\boldsymbol{r}^*\mathbf{1}^T - \mathbf{1}(\boldsymbol{\Psi}(\boldsymbol{I}_{k^*}\otimes\boldsymbol{U})\boldsymbol{r}^*)^T\right]\right\|^2 \tag{8}$$

The terms labeled (7) and (8) in the above expression are the symmetric and skew-symmetric parts of the objective function, respectively. The estimation of $\boldsymbol{X}$ is based solely on the symmetric part (7). Similarly, the estimates of $\boldsymbol{\Psi}$ and $\boldsymbol{r}^*$ are based solely on the skew-symmetric part (8). However, estimates of $\boldsymbol{W}$ are based on Eq. (5).

We now describe the algorithm of the proposed objective function algorithm.

Algorithm

Step 0 Set $\boldsymbol{U}$, k^*, d and a constant value $\varepsilon > 0$ and initialize, $\boldsymbol{X}$, $\boldsymbol{\Psi}$, and $\boldsymbol{r}^*$.
Step 1 Update $\boldsymbol{W}$ based on (5), given $\boldsymbol{X}$, $\boldsymbol{\Psi}$, and $\boldsymbol{r}^*$
Step 2 Update $\boldsymbol{X}$ based on the symmetric part of (7), given $\boldsymbol{W}$
Step 3 Update $\boldsymbol{\Psi}$ based on the skew-symmetric part of (8), given $\boldsymbol{W}$, and $\boldsymbol{r}^*$
Step 4 Update $\boldsymbol{r}^*$ based on the skew-symmetric part of (8), given $\boldsymbol{W}$, and $\boldsymbol{\Psi}$
Step 5 If difference of the values of the objective function in the previous and current iterations is smaller than a threshold value ε, terminate the algorithm: otherwise return to **Step1**

Here, the updated formula corresponding to the parameters of the objective function are derived.

To satisfy constraint (3), the diagonal elements of $\boldsymbol{W} = (w_{st})$ are given by

$$w_{ss} = a_s = \frac{b_s^2}{\sum_{t=1}^{k} b_t^2} \quad (s = 1, 2, \cdots, k).$$

Therefore, the problem of estimating $a_s (s = 1, 2, \cdots, k)$ becomes that of estimating $b_s (s = 1, 2, \cdots, k)$ (Gill et al., 1981). As an estimation method, we adopt the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm. Details of this algorithm are provided in Nocedal and Wright (1999).

Using the decomposition terms of the objective function (7) and (8), the coordinate matrix $\boldsymbol{X}$ is estimated to be equivalent to the ordinal metric MDS weighted by $\boldsymbol{W}$. Therefore, the BFGS is adopted for the estimation of $\boldsymbol{X}$.

Now, let us describe the method of updating $\boldsymbol{\Psi}$ is shown. $\psi_{i\ell}$ $(i = 1, 2, \cdots, n; \ell = 1, 2, \cdots, k^*)$ is estimated to satisfy the following:

$$\psi_{i\ell} = \begin{cases} 1 & (\|\boldsymbol{I}(\boldsymbol{A} + \boldsymbol{1}\boldsymbol{r}_\ell^T\boldsymbol{U}^T)\boldsymbol{T}_i + \boldsymbol{T}_i(\boldsymbol{A} - \boldsymbol{U}\boldsymbol{r}_\ell^*\boldsymbol{1}^T)\boldsymbol{I}\|^2 \\ & \quad \leq \|\boldsymbol{I}(\boldsymbol{A} + \boldsymbol{1}\boldsymbol{r}_{\ell^*}^T\boldsymbol{U}^T)\boldsymbol{T}_i + \boldsymbol{T}_i(\boldsymbol{A} - \boldsymbol{U}\boldsymbol{r}_{\ell^*}^*\boldsymbol{1}^T)\boldsymbol{I}\|^2) \\ 0 & (\text{others}) \end{cases},$$

for $\ell^* = 1, 2, \cdots, k^*$, where $\boldsymbol{T}_i = \mathrm{diag}(t_{1i}, t_{2i}, \cdots, t_{ni})$, $t_{ii} = 1$ $(i = 1, 2, \cdots, n)$ and $t_{ji} = 0$ $(j \neq i; i, j = 1, 2, \cdots, n)$, $\boldsymbol{I}$ is the identity matrix, and

$$\boldsymbol{A} = \frac{1}{2}(\boldsymbol{\Delta}^* - \boldsymbol{\Delta}^{*T}).$$

For the estimation of $\boldsymbol{r}^*$, an constrained numerical solution is used. This is because the value of (8) does not change if the estimate of $\boldsymbol{r}^{**}$ is updated such that $\boldsymbol{r}^* = \boldsymbol{r}^{**} + c\boldsymbol{1}$, where c is some constant value like the radius model (Borg and Groenen, 2005). Therefore, we set $c = -\min r_{s\ell}^*$.

When estimating $\boldsymbol{W}$ and $\boldsymbol{r}$, we adopt BFGS. Whenever these parameters are estimated at each step, the feasible area does not change because all of the updated parameters always satisfy the constraints. The values of the parameters estimated at the previous step are used as initial parameters for $\boldsymbol{W}$, $\boldsymbol{X}$, $\boldsymbol{\Psi}$ and $\boldsymbol{r}^*$. In addition,

when the i th element of diag $\boldsymbol{\Psi}$ is updated, the other elements have no effect on the update. Therefore, if these parameters are updated, the value of the objective function does not increase because parameters at the previous step are included in the feasible areas.

4. Simulation

In this section, the clustering results of the proposed method are evaluated via a simulation study. First, let us present the simulation design.

In this simulation, we introduce four phases. The first phase compares the results of the proposed method with those of tandem clustering. Tandem clustering consists of two steps. In the first step, asymmetric MDS based on the radius model is applied to the asymmetric data. In the second step, k-means is applied to the estimated lengths of the radii within each class to classify the objects into k^* clusters. For the second phase, the number of classes is set to 2, 3, and 5, as implemented in Milligan and Cooper (1988).

Before describing the third phase, let us show how the artificial data are generated, in this simulation. We use the right-hand side of Eq. (2) without $\boldsymbol{E}$, with $d = 2$ as follows:

$$
\begin{aligned}
&X_1 \sim N(-1, 1), \quad X_{s2} \sim N(s-1, 1)\ (s = 1, 2, \cdots, k), \\
&r_{s1} \sim U(0.3, 0.5)\ (s = 1, 2, \cdots, k), \\
&r_{s\ell} = r_{(s-1)\ell} + r^*_{s\ell}, \quad r^*_{s\ell} \sim U(0.3, 0.5)\ (s = 1, 2, \cdots, k; \ell = 2, 3, \cdots, k^*)
\end{aligned}
$$

where X_1 and X_{s2} are random variables selected from a normal distribution of the first and the second dimensions of class s, respectively , the radii r_{s1} and $r^*_{s\ell}$ are uniformly distributed , and elements of $\boldsymbol{\Psi}$ are randomely determined. The objects in the first dimension X_{i1} are concretely assumed to possess no clustering structure, while those belonging to the same cluster, s, in the second dimension X_{s2} are assumed to have a clustering structure. Ninety objects are discriminated into 2, 3, and 5 classes, which end up containing 45, 30, and 18 objects, respectively.

In the third phase, two levels of noise are imposed on the asymmetric dissimilarity data. One level is adding no noise, and the other level is that 75% of the number of dissimilarities within each class are replaced by noise, where noise dissimilarity is $\delta_{ij} \sim U(3, 5)$. The numbers of noises in the 2, 3, and 5 classes are 3037, 2025, and 1215, respectively.

In the fourth phase, three values are set for the number of clusters of radii. Specifically, the numbers are 2, 3, and 5.

Eighteen separate scenarios (3 kinds of classes $\times$ 2 noise levels $\times$ 3 kinds of clusters) are simulated. One-hundred artificial data points are generated for each scenario. The asymmetric clustering result is evaluated by the ARIs (Hubert & Arabie, 1985), which are calculated for all object classes: their mean is used for method evaluation.

The clustering result is displayed as a box plot in Figure 3.

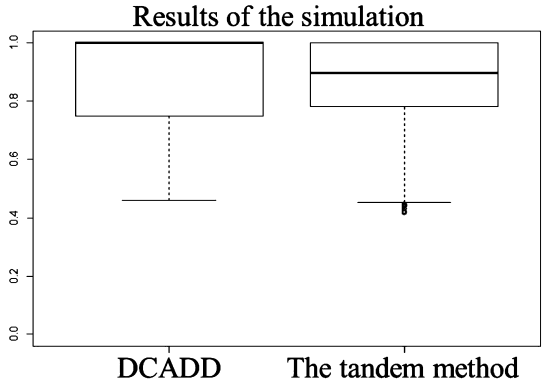

Figure 3: Results of the proposed and tandem methods in the phase 1, and the vertical axis indicates ARI.

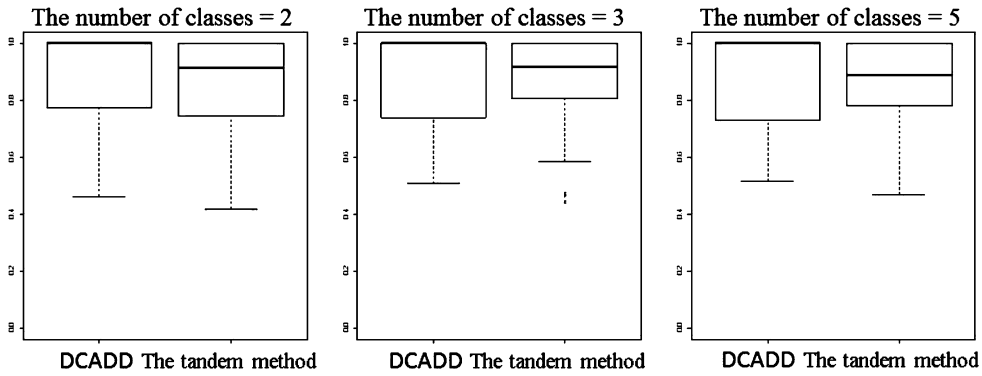

Figure 4: Results of the proposed and tandem methods in the phase 2.

In this figure, the results of the proposed method are compared with those of the tandem method. The median of the ARIs for the proposed method is higher than that for the tandem method, and the internal variations in each case are almost the same. Therefore, the results of the proposed method are superior to those of the tandem method in terms of both representative points and internal variations.

The results of the phase 2 are presented in Figure 4. The results are similar to those of Figure 3. The results of the proposed method are superior to those of the tandem method, irrespective of the number of classes.

The results of phase 3 are presented in Figure 5. The result of the proposed method is superior to that of the tandem method irrespective of the presence of noise. While the results of the proposed method are worse for the case where there is noise than they are for the case of noiselessness, the results of the tandem method are almost the same in either case.

Finally, the results of phase 4 are presented in Figure 6. In the cases where the number of clusters is 2, or 3, the results of the proposed methods are superior to those of the tandem method: however, when the number of clusters is 5, the results of the

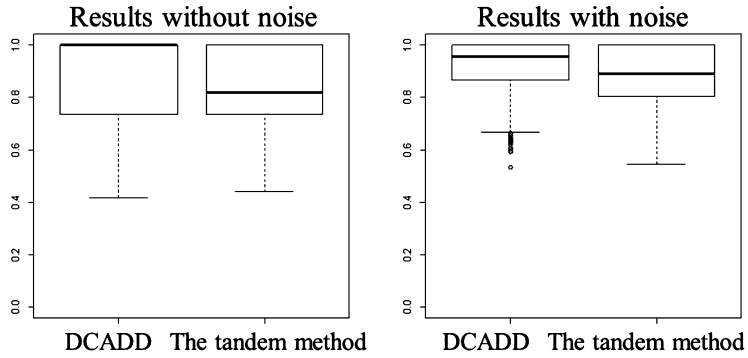

Figure 5: Results of the proposed and tandem methods in the phase 3.

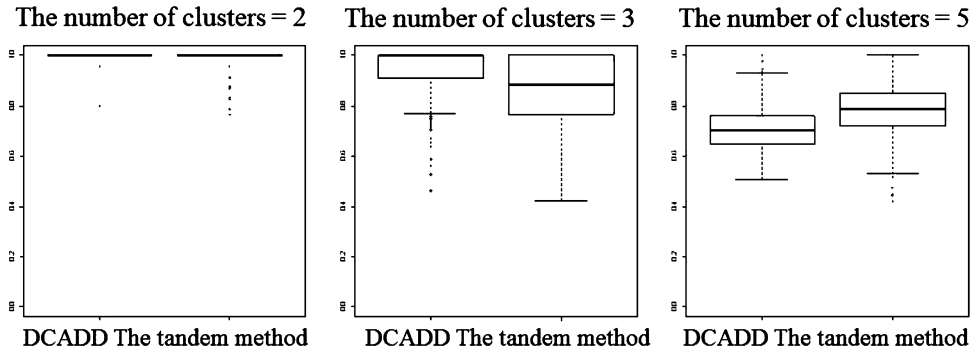

Figure 6: Results of the proposed and tandem methods in the phase 4.

tandem method are superior to those of the proposed method.

5. Application

We now evaluate the proposed method in a real application, namely, the switching data of Japanese tea bottles. These data were generated from scan panel data collected in Tokyo on July 2012 and provided by MACROMILL, INC. The sixteen brands of Japanese tea bottles are listed in Table 1, and the switching data are presented in Table 2. Each entry in Table 2, p_{ij} $(i, j = 1, 2, \cdots, n)$, specifies the frequency of switching from a brand i to another brand j. In this application, "Kinds of tea" in Table 1 are considered as classes.

These object classes and their asymmetries represent the competitive relations among brands and the tendencies of customers to switch to other brands, respectively.

The purpose of this application is to comprehend the asymmetries in the inter- and intra-class relations. The switching data are converted from the asymmetric similarity data to asymmetric dissimilarity data based on three kinds of transformations as follows:

Table 1: Brands and types of Japanese tea bottles (abbreviations shown in parentheses).

Abbreviation	Tea bottle brand	Kinds of tea
1. DG	Daido Green tea	Green tea(G)
2. IG1	Itoen Green tea1	Green tea(G)
3. IG2	Itoen Green tea2	Green tea(G)
4. IG3	Itoen Green tea3	Green tea(G)
5. KaG	Kao Green tea	Green tea(G)
6. SaG	Sangaria Green tea	Green tea(G)
7. SG1	Suntory Green tea1	Green tea(G)
8. SG2	Suntory Green tea2	Green tea(G)
9. 7G	7 & i Green tea	Green tea(G)
10. ABl	Asahi Blended tea	Blended tea(Bl)
11. CBl1	Coca cola Blended tea1	Blended tea(Bl)
12. CBl2	Coca cola Blended tea2	Blended tea(Bl)
13. CBl3	Coca cola Blended tea3	Blended tea(Bl)
14. KaBl	Kao Blended tea	Blended tea(Bl)
15. KBl1	Kirin Blended tea1	Blended tea(Bl)
16. KBl2	Kirin Blended tea2	Blended tea(Bl)

Table 2: Japanese tea bottle switching data.

	1	2	3	4	5	6	7	8
1. DG	32	0	0	0	0	0	0	0
2. IG1	2	283	0	18	2	2	1	40
3. IG2	2	41	177	9	0	2	1	13
4. IG3	0	0	0	18	2	0	3	9
5. KaG	0	0	0	0	58	0	1	3
6. SaG	2	0	0	0	0	24	0	0
7. SG1	0	0	0	0	0	1	14	0
8. SG2	3	0	0	0	0	3	8	246
9. 7G	0	4	5	0	0	0	0	3
10. ABl	0	15	11	5	0	0	2	11
11. CBl1	0	0	0	0	0	0	0	0
12. CBl2	0	0	0	0	0	3	0	0
13. CBl3	2	0	0	0	0	5	0	0
14. KaBl	0	0	0	0	0	0	0	2
15. KBl1	0	0	0	0	0	0	0	0
16. KBl2	0	0	0	0	0	0	0	0
	9	10	11	12	13	14	15	16
1. DG	0	0	0	0	0	0	1	0
2. IG1	0	0	0	5	18	0	17	2
3. IG2	0	0	0	2	10	0	6	0
4. IG3	0	0	1	2	8	0	9	0
5. KaG	0	0	0	1	1	6	0	1
6. SaG	0	0	0	0	0	0	2	0
7. SG1	0	0	0	1	2	0	8	0
8. SG2	0	0	2	2	26	0	21	1
9. 7G	56	0	0	1	1	0	2	0
10. ABl	4	137	2	5	16	2	5	2
11. CBl1	0	0	22	0	0	0	3	1
12. CBl2	0	0	4	114	0	0	13	7
13. CBl3	0	0	7	22	201	0	23	5
14. KaBl	0	0	0	3	2	28	1	2
15. KBl1	0	0	0	0	0	0	192	13
16. KBl2	0	0	0	0	0	0	0	5

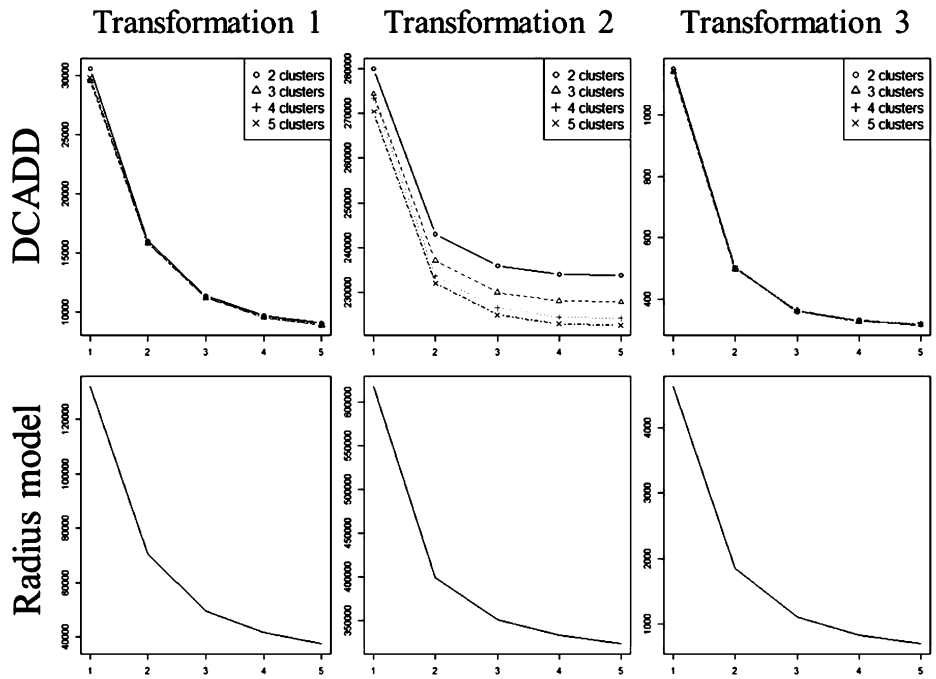

Figure 7: Values of objective function for DCADD and radius model. The horizontal axes indicate the number of dimensions, and the vertical axes indicate values of the objective function.

$$\delta_{ij} = \max_{ij,i\neq j}(p_{ij}) - p_{ij} + o^* \ (i \neq j; i, j = 1, 2, \cdots, n) \tag{9}$$

$$\delta_{ij} = \Big(\frac{o\ m_i\ m_j}{p^*_{ij}}\Big)^{1/2} \ (i, j = 1, 2, \cdots, n), \tag{10}$$

$$\delta_{ij} = (p_{ii} + p_{jj} - 2p_{ij})^{1/2} \ (i, j = 1, 2, \cdots, n), \tag{11}$$

where $p^*_{ij} = p_{ij} + o^*$ ($o^* > 0$ *is a small constant value*) prevents a singularity at $p_{ij} = 0$, m_i and m_j are the sums of the entries in the ith row and jth column of Table 2, respectively, and o is a constant. In the application, we set $o = 1$, and $o^* = 0.1$. Transformation 1 Eq. (9) and 3 Eq. (11) are shown in Cox and Cox (1994), and transformation 2 Eq. (10) is called the as gravity model (Tobler and Wineburg, 1971).

In the application, when transformation 1 (9) is applied, self-similarities are not considered. On the other hand, transformations 2 (10) and 3 (11) can consider the self-similarities. In previous studies for applying asymmetric MDS to brand-switching data, the gravity model was used (Borg and Groenen, 2005; Groenen and Heiser, 1996; Zielman and Heiser, 1993).

To determine the number of dimensions, and the number of clusters, see Figure 7. In this application, the candidates of the numbers of dimensions are 1, 2, 3, 4, and 5 and those of clusters are 2, 3, 4, and 5. Therefore, there are 20 pair-wise combinations of the numbers of dimensions and clusters for the DCADD method. For the case of

the radius model, there are 5 candidates because there is no need to determine the number of clusters. These parameters are determined using changes in the values of the objective functions. When the value of the objective function for each candidate is calculated, the DCADD or asymmetric MDS method is applied to the asymmetric dissimilarity data 50 times for various initial values, and the results corresponding to the minimum value are selected from among them.

As shown in Figure 7, the number of dimensions is set to 2 for each result. For transformations 1 (9), 2 (10) and 3 (11), the numbers of clusters are set to 3, 4 and 3, respectively.

First, the results of the radius model and the DCADD method corresponding to transformation 1 (9) are shown.

As shown in Figure 8, for the result of the radius model, the asymmetric relation between the brands can be easily interpreted and these brands belonging to the same class are located closely each other. However, the length of the radii is different depending on the degree of asymmetry and it is difficult to interpret these asymmetric relations as a whole, especially when the number of objects is large.

From Figure 9, it is easy to interpret the asymmetric relation within and between the classes because the brands within each class have only three kinds of radii. Showing the same tendency as radius model result, KBl1 has a relatively large radius and ABl, IG1 and IG2 have relatively small radii. Additionally, it is easy to interpret the asymmetric relations within each class and between different classes because brands belonging to the same class are located closely each other.

Next, we show the results of transformation 2 (10) for the radius model and the proposed method in Figures 10, and 11, respectively. For the results of the radius model, the asymmetries are difficult to be discerned and the structures of the classes and asymmetries cannot be interpreted because the length of each radii gets large by estimation, although brand belonging to the same class are located closely. On the other hand, for the results of the DCADD method, it is easy to interpret the asymmetries within and between classes. For the results of the radius model, it is difficult to interpret the relation between brands located near the $(0, 0)$ although dissimilarities within class are small. On the other hand, we can easily to understand the feature within and between classes since brands belonging to the different classes are discriminated each other. The tendency is not the same as that of transformation 1 (9) because transformation 1(9) does not consider the self dissimilarities.

Finally, the results of the radius model and the DCADD method corresponding to transformation 3 (11) are shown in Figures 12 and 13. From the results of transformation 3 (11), the asymmetries cannot be represented in low-dimensions. This is because the values of self-similarities tend to be larger than those of non-self-similarities.

As a result of the applications, if it is difficult to interpret the asymmetries within and between classes, DCADD can be useful in some cases. In fact, when transformation 2 (10) is used to consider the self similarities, the DCADD method provides us with a mean of interpreting the asymmetric relations easily in this application. The lengths of radii for the features of the DCADD method tend to be smaller than those

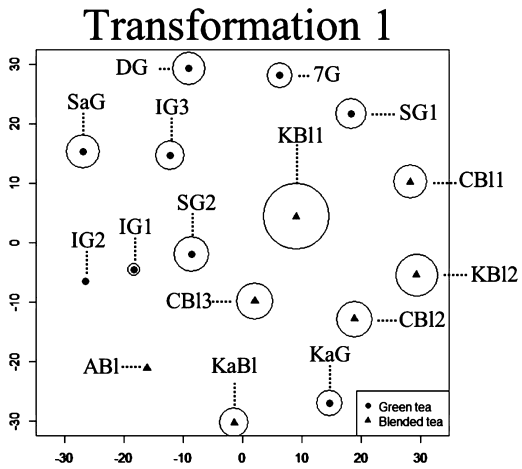

Figure 8: The result of the radius model with transformation 1 (9).

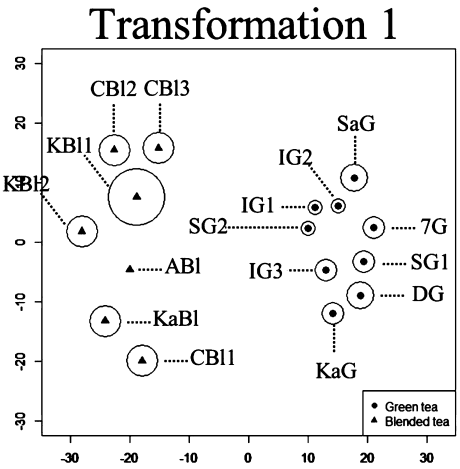

Figure 9: The result of the proposed method with transformation 1 (9).

for the radius model. This is because the asymmetric dissimilarities within classes become small through the estimation of w_{ss} $(s = 1, 2, \cdots, k)$. However, when DCADD is applied to asymmetric dissimilarity data, we have to check whether the assumption is satisfied or not. In the application, from the both results by the radius model and DCADD, it can be confirmed that brands belonging to the same class are located closely. Therefore, we consider the data as the one including class structure.

6. Conclusion

In this paper, we proposed the DCADD method to reveal the asymmetry within and between object classes, given asymmetric dissimilarity data. In the case of noisy asymmetric dissimilarity data, the proposed method detects asymmetries in the clus-

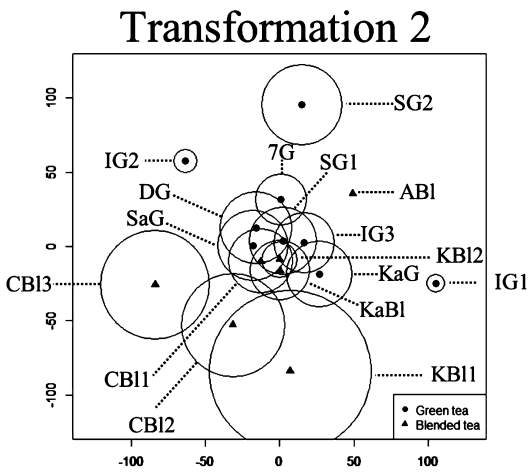

Figure 10: The result of the radius model with transformation 2 (10).

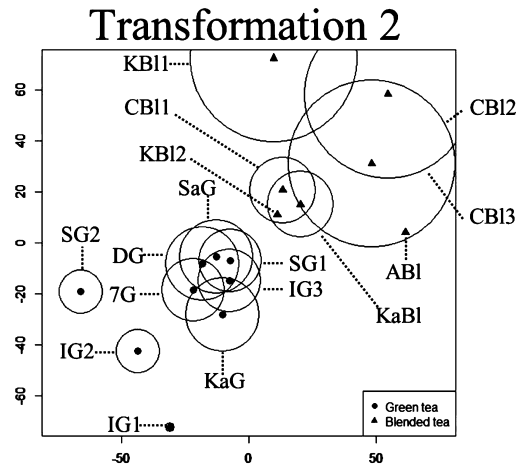

Figure 11: The result of the proposed method with transformation 2 (10).

tering structure that cannot be revealed by the tandem method combining the radius model with k-means clustering. When DCADD parameters are estimated, there exist a lot of local optimal solutions. Therefore, DCADD does not guarantee global optimal solutions. In practice, DCADD must be applied to the data with various kinds of initial values.

In the future work, we must solve the remaining three problems. The first problem is that the result of the DCADD method does not hold if the number of clusters are very large. The second problem is the means of checking the assumptions of DCADD. There are two assumptions. First, the objects within the same class are cohesive, and second, that lengths of radii within the same class can be represented by k^* kinds of radii. In the application, we compared the results of DCADD to those of radius model and checked the appropriateness. However, it needs a criteria to check whether

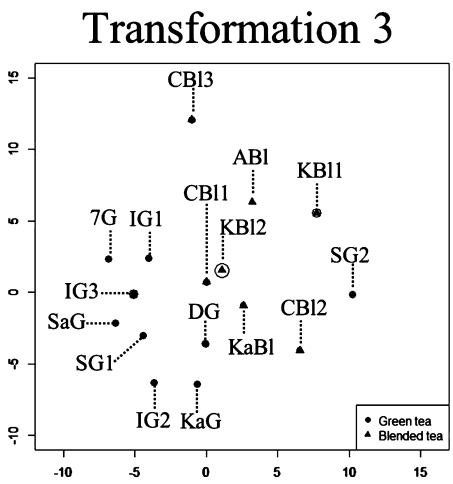

Figure 12: The result of the radius model with transformation 3 (11).

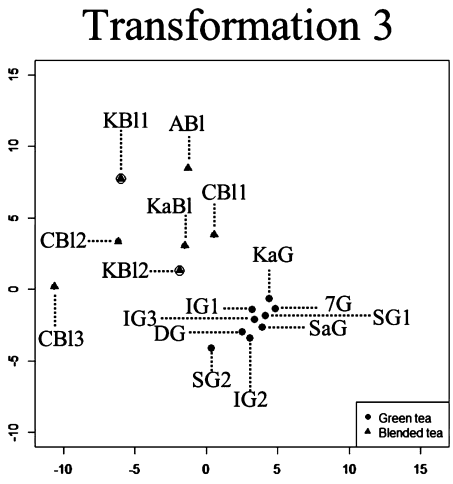

Figure 13: The result of the proposed method with transformation 3 (11).

the assumptions is satisfied or not.

Finally, the original discriminant coordinates is proposed based on discriminant analysis. Therefore, DCADD should consider the notion of discriminant analysis such that variations of objects in low- dimensions within each class are minimized and those between classes are maximized. It can be possible to propose method asymmetric MDS with penalty term of discriminant analysis for the coordinates of objects. However, it is difficult to determine the tuning parameters for the term related to discriminant analysis and the results depend on the tuning parameters. Therefore, in this paper, we adopt different approach. To overcome this problem, how to determine the tuning parameters should be carefully considered.

Acknowledgements

We would like to express our gratitude to MACROMILL, INC for providing QPR scan panel data. In addition, we would like to express our deepest appreciation to the two reviewers and the editor for their helpful comments.

References

Borg, I. & Groenen, P. (2005). *Modern multidimensional scaling: theory and application*, New York: Springer.

Bove, G. & Critchley, F. (1993). Metric multidimensional scaling for asymmetric proximities when the asymmetry is one-dimensional. In R. Steyer, K. F. Wender, & K. F. Widamann (Eds.). *Pcychometric Society in Trier* (pp.55–60), Stuttgart: Gustav Fischer Verlag.

Chino, N. (2012). A brief survey of asymmetric mds and some open problems. *Behaviormetrika*, **39**, 127–165.

Cox, T. F. & Cox, M. A. A. (1994). *Multidimensional Scaling*, London: Chapman & Hall.

Gill, P. E., Murray, W. & Wright, M. H. (1981). *Practical optimization*, London: Academic Press.

Groenen, P. J. F. & Heiser (1996). The tunneling method for global optimization in multidimensional scaling, *Psychometrika*, **61**, 529–550.

Heiser, W. J. & Groenen, P. (1997). Cluster differences scaling with within-cluster loss component and a fuzzy succesive approximation strategy to avoid local minima. *Psychometrika*, **62**, 529–550.

Hubert, L. & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, **2**, 193–218.

Kiers, H. A. L., Vicari, D. & Vichi, M. (2005). Simultaneous classification and multidimensional scaling with external information. *Psychometrika*, **70**, 433–460.

MacQueen, J. B. (1967). Some methods for classification and Analysis of Multivariate Observations. *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability* (pp.281–297), Berkeley, University of California Press.

Milligan, G. W. & Cooper, M. C. (1988). A study of standardization of variables in cluster analysis. *Journal of Classification*, **5**, 181–204.

Nocedal, J. & Wright, S. (1999). *Numerical optimizations. Springer series in operation research*, New York: Springer.

Okada, A. & Imaizumi, T. (1987). Geometric models for asymmetric similarity data. *Behaviormetrika*, **21**, 81–96.

Saito, T. & Yadohisa, H. (2005). *Data analysis of asymmetric structures: advanced approaches in computational Statistics*, New York: Marcel Dekker.

Seber, G. A. F. (1984). *Multivariate observations*, New York: Wily.

Tobler, W. & Wineburg, S. (1971). A cappadocian speculation. *Nature*, **231**, 39–41.

Young, F. W. & de Leeuw, J. & Talane, Y. (1980). *Quantifying qualitative data*, In E. D. Lantermann & H. Feger (Eds.), Similarity and choice. Bern: Huber. *Nature*, **231**, 39–41.

Zielman, B. & Heiser, W. J. (1993). Analysis of asymmetric analysis by slide-vector. *Psychometrika*, **58**, 101–114.

(Received January 20 2014, Revised June 15 2015)